



ACTES DES JOURNÉES D'INFORMATIQUE MUSICALE 2023

organisées par le Centre de Recherches Informatique et
Création Musicale (CICM), Laboratoire Musidanse,
Université Paris 8

Maison des Sciences de l'Homme Paris Nord
24 - 26 mai 2023

Mot d'accueil

Le CICM – Centre de Recherches Informatique et Création Musicale, équipe du Laboratoire MUSIDANSE, est heureux d'organiser les Journées d'Informatique Musicale 2023 à la Maison des Sciences de l'Homme Paris Nord et de vous y accueillir, 10 ans après les JIM 2013 à l'université Paris 8. L'esprit de ces JIM est celui d'un festival, conjuguant conférences invitées, présentations d'articles, démonstrations, workshops, concerts, performances. Ce festival est le reflet de notre communauté, dans une interdisciplinarité qui a autant besoin de ses chercheurs que de ses compositeurs, autant besoin de l'informatique musicale que de la *computer music*. Nos deux conférenciers invités, Florence Levé et Curtis Roads, témoignent de cette diversité si nécessaire et qui nous constitue.

Cette édition, du 24 au 26 mai 2023, cultive des thèmes importants et habituels (comme la spatialisation sonore, les représentations musicales, les outils pour la composition et l'improvisation, l'analyse-synthèse) et met particulièrement l'accent sur les quatre thématiques suivantes, dans le sillage d'ISEA 2023 et du projet ANR BBDMI :

- **Faire de la musique avec l'électricité du corps et du cerveau** : dans le sillage du projet ANR BBDMI, les musiciens s'approprient les techniques de captation électromyographique (EMG) et électroencéphalographique (EEG) du signal électrique du corps. États de l'art, expérimentations technologiques et musicales sont au rendez-vous.
- **Open-source, open hardware, petits dispositifs embarqués** : en informatique musicale comme ailleurs, l'accès ouvert aux connaissances s'étend au matériel. Open Software, Open Hardware ont gagné leurs lettres de noblesse aussi bien du côté des instruments modulaires que du côté des petits dispositifs embarqués. Bienvenue aux communautés des *makers*.
- **La symbiose, nouveaux modèles d'interactions** : la thématique de la symbiose est mise à l'honneur, dans le sillage d'ISEA 2023 à Paris dont la MSH Paris Nord est partenaire. Une définition possible : la symbiose désigne la co-existence d'organismes à relations variables. Comment coexiste-t-on avec nos machines ? A l'heure de l'Intelligence artificielle au cœur de nos systèmes, peut-on repenser des modèles d'interaction pour faire musique en nous inspirant du vivant ?
- **Hommage au compositeur-chercheur Horacio Vaggione, à l'occasion de ses 80 ans** : figure de référence de la recherche musicale, de la musique électroacoustique, de l'informatique et de la création musicale, son style singulier et inimitable a traversé les époques, et inspiré grand nombre de jeunes musiciens dans le monde entier. Horacio Vaggione est à l'origine de la création du CICM à l'université de Paris 8. Nous lui rendrons hommage lors d'un concert carte-blanche le jeudi 25 mai.

Les JIM 2023 sont co-organisés par l'AFIM, le CICM/MUSIDANSE de l'université de Paris 8, avec l'aide de la MSH Paris Nord, l'ANR BBDMI, et le Ministère de la Culture et de la Communication. Merci à ces institutions de soutenir et contribuer au financement de ces journées.

Le présent volume comporte l'ensemble des articles soumis et acceptés par les relecteurs du Comité Scientifique, que nous remercions pour leur concours. Un second volume comporte les notes de programme des concerts, performances et démonstrations artistiques, en miroir des présentations scientifiques.

Nous vous souhaitons de belles Journées d'Informatique Musicale 2023, riches des nombreux échanges qui leur donneront vie !

Alain Bonardi, pour le Comité d'Organisation des JIM 2023.

Articles scientifiques

Session 1 « Spatialisation »

- *Développements récents d'outils de spatialisation sonore du GRIS : CONTROLGRIS et SPATGRIS*, Robert Normandeau (Université de Montréal, CIRMMT), Nicola Giannini (Université de Montréal, CIRMMT), Gaël Lane Lépine (GRIS), pages 2-8.
- *Sound Trajectory en pratique live*, Maxence Mercier (Université Côte d'Azur), pages 9-21.
- *Émergences : de la poésie au prototype*, Paul Goutmann (CICM MUSIDANSE Université Paris 8, EUR ArTeC), pages 22-29.
- *Approche de la synthèse par front d'onde (WFS) utilisant une antenne à faible densité*, Raymi Morales (CICM MUSIDANSE Université Paris 8), Anne Sèdes (CICM MUSIDANSE Université Paris 8), pages 30-33

Session 2A « Faire de la musique avec l'électricité du corps et du cerveau »

- *L'électromyographie pour la performance et la composition : trois études de cas dans le cadre du projet BBDMI*, David Fierro (CICM MUSIDANSE Université Paris 8), Alain Bonardi (CICM MUSIDANSE Université Paris 8), Guilherme Carvalho (RIRRA 21, Université Paul Valéry), Anne Sèdes (CICM MUSIDANSE Université Paris 8), pages 35-41.
- *Mycelium Garden*, Diane Schuh (CICM MUSIDANSE Université Paris 8, EDESTA), David Fierro (CICM MUSIDANSE Université Paris 8), pages 42-48.
- *M.I.X.C.S. Music, Internet and Cognitive Sciences*, Pierre-Henri Vulliard (SCRIME / Arts Sciences Information Sarl), pages 49-52.
- *Identification par EEG et utilisation dans cadre artistique des états de concentration : vers une sonification de l'attention ?* Armand Bazin (Université Gustave Eiffel), Ulysse Del Ghingaro (MUSIDANSE / C.et.I. Université Paris 8), pages 53-60.

Session 3 « Représentations musicales »

- *Analyse et essai de reconstruction de structures arborescentes des grilles de jazz*, Patrice Thibaud (Université de Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL), Mathieu Giraud (Université de Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL), pages 62-67.
- *A constraint formalization of Fux's counterpoint*, Damien Sprockeels (Université catholique de Louvain), Thibault Wafflard (Université catholique de Louvain), Peter Van Roy (Université catholique de Louvain), Karim Haddad (IRCAM), pages 68-77.
- *Une approche électroacoustique des pavages rythmiques à partir des canons d'entrées régulières*, Matias Rosales (Université de Strasbourg, HEAR), pages 78-85.
- *LiveScaler : contrôler en live l'harmonie d'un morceau de musique électronique*, Alice Rixte (Université de Bordeaux), pages 86-95.
- *Vers une architecture logicielle pour l'enseignement hybride de l'harmonie*, Anaïs Binet (Université Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800), Myriam Desainte-Catherine (Université Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800), pages 96-104.
- *Étude comparative des articulations du jeu au piano, au métapiano et au midifileperformer au moyen des S-Langages*, Myriam Desainte-Catherine (Université Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800), Jean Haury (SCRIME, Inria), Bernard Serpette (SCRIME, Inria), Sylviane Schwer (USPN, CNRS, LIPN UMR 7030), pages 105-115.

Session 4 « Computer Music »

- *Introducing Computer Music in Secondary School*, Christopher Dobrian (University of California, Irvine), pages 117-122.
- *L'empreinte de John Chowning dans la construction de l'Ircam : de sa rencontre avec Pierre Boulez en 1973 à la création de Stria en 1977*, François-Xavier Féron (STMS – CNRS, Ircam, Sorbonne Université), Laura Zattra (Conservatoires de Rovigo, Milan, Vicenza), pages 123-132.
- *Recherche et acoustique musicales à Marseille entre 1968 et 1989*, Baptiste Bacot (STMS – CNRS, Ircam, Sorbonne Université), Vincent Tiffon (Université Aix-Marseille, CNRS, PRISM), François-Xavier Féron (STMS – CNRS, Ircam, Sorbonne Université), pages 133-142.

Session 5 « Symbioses »

- *L'art de l'attention IV : Symbioses*, Diane Schuh (CICM MUSIDANSE Université Paris 8, EDESTA), Tania Cortés (Universidad Politécnica de Valencia), Sebastián Sarasa Molina (CRD de Pantin, CRD de Noisy), pages 144-152.
- *Objets connectés au service de la performance intermédia, exemple de dispositif interactif pour la co-composition*, Laurent Di Biase (CICM MUSIDANSE Université Paris 8), pages 153-160.
- *Headphonics : Mobile Music on the Metro*, Steve Jones (Teesside University, Royaume-Uni), pages 161-170.

Session 6A « Outils pour la composition et l'improvisation »

- *Synthèse concaténative en réalité virtuelle*, Jonathan Bell (Prism-CNRS, XR2C2-CTEL), pages 172-182.
- *Simone : un instrument distribué pour l'étude des interactions improvisées collectives*, Aliénor Golvet (STMS – CNRS, Ircam, Sorbonne Université), Benjamin Matuszewski (STMS – CNRS, Ircam, Sorbonne Université), Frédéric Bevilacqua (STMS – CNRS, Ircam, Sorbonne Université), pages 183-190.
- *Une station de travail audio-numérique pour la plateforme WEB*, Antoine Vidal-Mazuy (Université Côte d'Azur, CNRS, INRIA), Michel Buffa (Université Côte d'Azur, CNRS, INRIA), pages 191-199.

Session 6B « Démonstrations du groupe de travail AFIM Musiques en Réseau »

- *Allégories de l'audio en réseau*, Peter Sinclair (École Supérieure d'Art d'Aix-en-Provence, Locus Sonus Locus Vitae), pages 201-204.
- *New Atlantis, métavers du sonore*, Ludmila Postel (École Supérieure d'Art d'Aix-en-Provence, Locus Sonus Locus Vitae), Jonathan Tanant (développeur indépendant), page 205.
- *Art & Games in Metaverse*, Frédéric Curien (École européenne supérieure de l'image Angoulême-Poitiers), Hervé Jolly (École européenne supérieure de l'image Angoulême-Poitiers), Axel Buendia (École nationale des jeux et médias interactifs numériques), pages 206-209.
- *HERMES V2 : création musicale en réseau*, João Svidzinski (CICM MUSIDANSE Université Paris 8, Ircam), Robin Montferme (Université Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800), Joseph Larralde (Université Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800), pages 210-214.
- *Quacktrip en octophonie : étude de cas dans HERMES V2*, João Svidzinski (CICM MUSIDANSE Université Paris 8, Ircam), Miller Puckette (Department of Music, University of California San Diego), page 215.
- *Networked Music Performance in PATCHXR*, Jonathan Bell (PRISM-AMU), Océane Wandji (PRISM-AMU), Shiho Yamada (XR2C2-UCA), pages 216-217.

Session 7 « Analyse-synthèse »

- *A Musicological Pipeline for Singing Voice Style Analysis with Neural Voice Processing and Alignment*, Yann Teytaut (STMS – CNRS, Ircam, Sorbonne Université), Antoine Petit (Passages XX-XXI), Céline Chabot-Canet (Passages XX-XXI), Axel Roebel (STMS – CNRS, Ircam, Sorbonne Université), pages 219-228.
- *Analyse interdisciplinaire de la synthèse instrumentale : étude de cas avec Partiels de Grisey*, Charles de Paiva Santana (Université Aix-Marseille, CNRS, PRISM), Vincent Tiffon (Université Aix-Marseille, CNRS, PRISM), Mylène Gioffredo (Université Aix-Marseille, CNRS, PRISM et InCIAM), Micael Antunes (Université Aix-Marseille, CNRS, PRISM et NICS/IA, Université de Campinas), Javier Elípe Gimeno (Université Aix-Marseille, CNRS, PRISM), Samuel Poirot (Université Aix-Marseille, CNRS, PRISM), Sølvi Ystad (Université Aix-Marseille, CNRS, PRISM), pages 229-238.

Articles en relation au programme artistique des JIM 2023

- *Hermes v1 : création audiovisuelle en réalité virtuelle*, João Svidzinski (CICM MUSIDANSE Université Paris 8, Ircam), page 240.
- *BOATNOA : an audiovisual reflection on a modern city*, Oscar Bahamonde (Universitat Politècnica de València), Tania Cortés (Universitat Politècnica de València), Gabriela Yáñez (Sound Designer indépendante), pages 241-243.

Actes des Journées d'Informatique Musicale 2023
24-26 mai 2023
Maison des Sciences de l'Homme Paris Nord

SESSION « SPATIALISATION »

DÉVELOPPEMENTS RÉCENTS D'OUTILS DE SPATIALISATION SONORE DU GRIS : CONTROLGRIS ET SPATGRIS

Robert Normandeau
Université de Montréal, CIRMMT
robert.normandeau@umontreal.ca

Nicola Giannini
Université de Montréal, CIRMMT
nicola.giannini@umontreal.ca

Gaël Lane Lépine
GRIS
glanelepine@gmail.com

RÉSUMÉ

Cet article présente les développements récents relatifs aux outils de spatialisation du son, conçus par le Groupe de Recherche en Immersion Spatiale (GRIS) – ControlGRIS et SpatGRIS – ainsi que les projets de développement à venir. On y expose le contexte et les motivations qui ont présidé à la création d'un couple plugiciel/logiciel ainsi que les plus récents ajouts qui leur ont été apportés. Nous présentons un nouvel algorithme maison, le MBAP (Matrix Based Amplitude Panning) sur lequel le mode CUBE est construit et qui permet de concevoir pratiquement toutes les formes de dispositifs de haut-parleur imaginables. Nous y faisons la présentation d'une proposition de format standard d'échange d'œuvres en format multipiste qui permettra une meilleure diffusion des musiques entre les individus et les organismes. Depuis 2009, avec un financement prévu au moins jusqu'en 2025, ce projet de recherche-crédation poursuit l'objectif d'offrir une solution libre de droits, gratuite, simple et efficace pour une spatialisation multidirectionnelle du son, intégrée tout au long du processus de composition, tant pour la création musicale, que pour les arts et les installations sonores, ou encore le cinéma indépendant.

1. INTRODUCTION

Sur le plan acoustique, l'immersion sonore est comprise comme la capacité d'un environnement ou d'un dispositif technique à diffuser du son dans plusieurs directions autour d'un auditoire [11]. Toutefois, l'écoute est un phénomène complexe et il serait réducteur d'associer la sensation subjective d'immersion à la simple capacité de diffusion multidirectionnelle d'un dispositif – en la considérant indépendamment des autres paramètres simultanément composables du son, tel que le contenu spectral et l'intensité, ou encore, de l'intentionnalité d'écoute de l'individu [3]. Généralement, les outils de spatialisation audio numériques représentent l'espace virtuel de diffusion en trois dimensions (3D), selon des coordonnées sphériques ou cartésiennes. Mais la

perception du son dans l'espace ne dépend pas uniquement de la direction relative d'une source sonore virtuelle, par rapport à une position centrale, tel que fréquemment représentée dans la plupart des environnements de spatialisation actuellement disponibles. Elle dépend aussi de la nature des sons, de leur complexité, de leur contexte de perception, du sens qu'on peut leur inférer, ainsi qu'une foule d'autres facteurs que les compositeurs et créateurs sonores aiment pouvoir manipuler. Conséquemment, il devrait être possible de composer la position relative des sources sonores virtuelles, ainsi que leur mouvement dans l'espace de diffusion, simultanément aux autres propriétés du son évoluant dans le temps, tel que l'intensité et le timbre [7].

2. CONTEXTE ET MOTIVATIONS

Depuis 2009, le Groupe de Recherche en Immersion Spatiale (GRIS) poursuit le développement d'outils de spatialisation audio numériques dans l'optique d'offrir aux compositeurs un contrôle indépendant sur la localisation et les trajectoires des sources sonores dans l'espace de diffusion, tout en permettant une organisation simultanée de leurs mouvements en relation avec d'autres composantes du son. Le projet du GRIS s'appuie sur deux prémisses : [1] de nos jours, la majorité des compositeurs de musique électroacoustique emploient un environnement de travail audio numérique (DAW) comme outil principal de composition ; [2] la composition spatiale doit coïncider avec la composition temporelle du son, et ce, durant tout le processus de création plutôt qu'à une étape ultérieure de la production.

Historiquement, la spatialisation sonore fait partie intégrante de la musique électroacoustique¹. De nombreux dispositifs de spatialisation d'objets sonores [5] ont vu le jour et sont de plus en plus utilisés en dehors des centres de recherche universitaires – par l'industrie du cinéma, les planétariums, les dômes audiovisuels et la réalité virtuelle (RV), notamment. Mais, certains de ces outils peuvent être coûteux, conçus principalement comme outils de mixage et non de création, ou encore, pour une application liée à des

¹ Notamment, le potentiomètre d'espace, employé avec un ensemble de quatre haut-parleurs, au début des années 50, par Pierre Schaeffer et Jacques Poullin [5, p.26].

domaines surtout visuels, tels que le cinéma, la réalité virtuelle (RV) ou les jeux vidéo. De plus, certains logiciels propriétaires se fondent sur un seul algorithme de spatialisation, dont le principe de fonctionnement est relativement contraignant et parfois obscur. Du côté des outils libres de droits, ceux-ci offrent majoritairement une spatialisation uniquement fondée sur l'ambisonie d'ordre élevé (HOA) [4] ou sur le Vector Base Amplitude Panning (VBAP) et n'incluent pas d'autres algorithmes qui pourraient mieux s'appliquer dans certains contextes.

L'un des principaux objectifs du projet de développement présenté dans cet article est d'offrir aux artistes sonores et aux compositeurs un outil de spatialisation libre de droits, simple, ouvert et efficace, favorisant l'expérimentation dans la diffusion spatialisée du son, et ce, avec un minimum de contraintes sur le plan des ressources informatiques (CPU) tout en offrant une précision spatiale optimale. L'article présente les derniers progrès et les prochaines étapes de développement – effectués en collaboration avec les compositeurs-testeurs du GRIS² – pour les outils de spatialisation audio numériques ControlGRIS/SpatGRIS.

3. SPATGRIS

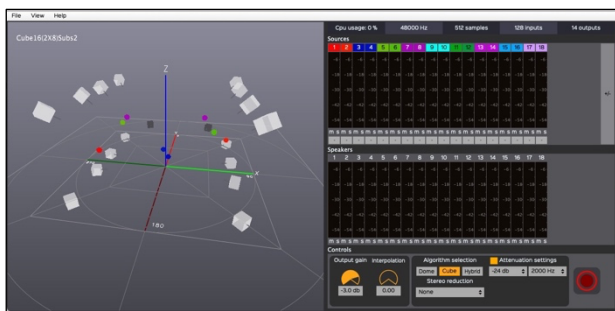


Figure 1. SpatGRIS en mode CUBE.

3.1. Présentation

SpatGRIS offre un « service » de spatialisation à des logiciels « clients » pouvant transmettre du signal audio numérique. Cette approche permet aux utilisateurs d'y connecter simultanément et dynamiquement plusieurs outils de création sonore, que ce soit pour un usage en temps réel ou différé.

3.2. Architecture

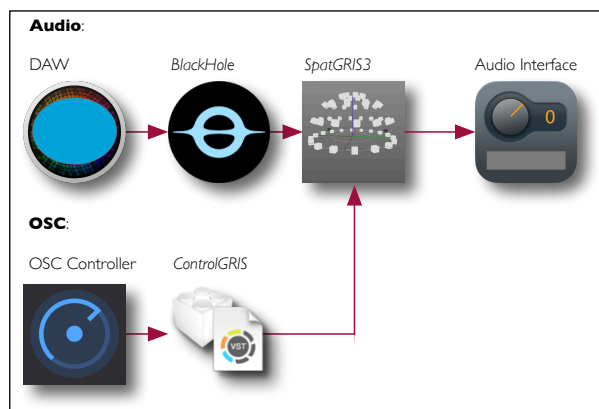


Figure 2. Architecture de connexion.

SpatGRIS reçoit les signaux audio sortants de chaque client, selon leur nombre de canaux respectifs [Figure 2]. Celui-ci peut recevoir jusqu'à 128 entrées et diffuser jusqu'à 128 sorties, grâce au plugiciel³ HAL BlackHole⁴, installé avec le logiciel. En parallèle, SpatGRIS reçoit des coordonnées spatiales (métadonnées), transmises au format Open Sound Control (OSC) par un contrôleur externe, notamment le plugiciel ControlGRIS (section 4). En combinant les signaux entrants et leurs métadonnées correspondantes, le logiciel s'occupe alors de redistribuer chaque signal en fonction de la configuration de haut-parleurs sélectionnée pour la diffusion. Par définition, ce type de système – combinant des signaux audios avec des métadonnées – présente une spatialisation basée sur des objets sonores [12], ce qui permet de composer facilement un espace sonore virtuel, en abstraction du dispositif de diffusion actuel. En somme, l'architecture de SpatGRIS permet : une intégration facile des outils audio numériques privilégiés ; d'adapter toute diffusion à une variété de configurations de haut-parleurs ; et d'employer différents algorithmes de spatialisation, selon les ressources disponibles et les besoins des créateurs .

3.3. Algorithmes de spatialisation

La version actuelle de SpatGRIS⁵ offre quatre modes de spatialisation différents [8] :

DOMÉ : construit sur le VBAP (Vector-Based Amplitude Panning) [10] qui spatialise les sources sonores sur la surface d'un dôme de haut-parleurs préalablement configuré ;

CUBE : construit sur le MBAP (Matrix-Based Amplitude Panning), un algorithme relativement récent

² L'article présente un point de vue de compositeur et non de programmeur.

³ Le terme plugiciel, formé à partir des mots PLUs et loGICIEL, constitue un « plus » en augmentant les performances du logiciel principal. Ce terme a été proposé par l'Office québécois de la langue française en juin 1996 (2013) [9].

⁴ Depuis la version 3 et sur Mac seulement. Originellement, le plugiciel HAL utilisé était *JackRouter*, et il est toujours de mise sur Windows et Linux.

⁵ La version 3.2.7 au moment d'écrire ces lignes.

dont le principe de fonctionnement se trouve décrit à la section 3.3.2 ;

BINAURAL : pour une écoute simulant la spatialisation aux écouteurs. Ce mode exploite une banque de 16 réponses impulsionnelles (RI) de type HRTF⁶;

STEREO : pour l'écoute sur un système audio à deux haut-parleurs.

3.3.1. Dome

Le DOME est basé sur l'algorithme VBAP décrit par Ville Pulkki [10] en 2001. Celui-ci étant déjà connu et très répandu, il n'est pas nécessaire d'élaborer davantage ici.

3.3.2. Cube

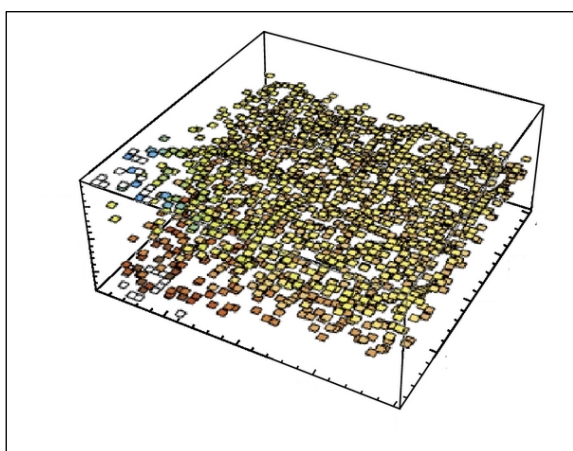


Figure 3. Matrix Based Amplitude Panning (MBAP).

Le CUBE est une représentation virtuelle d'un espace vide à l'intérieur duquel on peut concevoir n'importe quelle disposition de haut-parleurs. Il a été développé afin de permettre l'utilisation de dispositifs asymétriques, irréguliers, inédits comme on en trouve dans les acousmoniums, les installations sonores audiovisuelles ou tout autre dispositif particulier. Le dispositif de haut-parleurs est fabriqué à l'intérieur du CUBE [Figure 1], et permet le passage du son à l'intérieur ou à l'extérieur de celui-ci (contrairement au VBAP, prisonnier de la surface du dôme). Le CUBE est basé sur le MBAP [Figure 3], un algorithme de spatialisation relativement nouveau, développé et implémenté au sein des outils du GRIS par Gaël Lane Lépine, programmeur en titre du groupe. Il se distingue du VBAP par son cadre de référence spatiale non concentrique. C'est-à-dire que la position des sources sonores n'est pas représentée en fonction d'une position centrale à l'intérieur d'une sphère, communément appelée *sweetspot*, mais plutôt en

⁶ Head-Related Transfer Function.

⁷ Selon l'algorithme VBAP, la distance d'une source par rapport au point central est toujours constante. Ainsi, la valeur de distance d'une

fonction des dimensions et des frontières de l'espace de diffusion [12, p. 253-254].

En ce sens, l'algorithme de spatialisation MBAP consiste en la superposition de « matrices » tridimensionnelles (matrix) de haut-parleurs librement positionnés dans l'espace de diffusion. L'algorithme précalcule des matrices de volume en fonction de la position des haut-parleurs dans l'espace 3D. Ces matrices de volume sont ensuite lues, avec interpolation, selon la position de la source. Le gain de la source pour chacun des haut-parleurs sera pondéré en fonction de la distance de la source.

Le MBAP permet de représenter et spatialiser les sources virtuelles non seulement sur la surface du dispositif des haut-parleurs, comme c'est le cas avec le VBAP, mais également à l'intérieur et à l'extérieur de l'environnement d'écoute délimité par les haut-parleurs. Cette possibilité requiert toutefois que l'interface du logiciel – qui transmet les coordonnées spatiales des sources virtuelles au SpatGRIS – permette non seulement un contrôle en 2D (azimut et distance⁷), mais aussi en 3D (X, Y et Z). C'est à cet effet que nous avons créé ControlGRIS.

3.3.3. Hybride

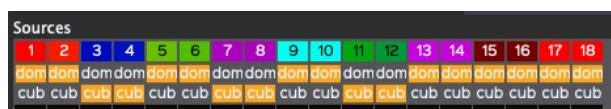


Figure 4. Le mode HYBRIDE.

Le mode HYBRIDE est une combinaison de l'algorithme DOME, pour ce qui est de la configuration du dispositif de haut-parleurs, avec le comportement de l'algorithme CUBE pour certaines sources [Figure 4]. Il est ainsi possible d'avoir en même temps, sur un dôme de haut-parleurs, des sources qui se positionnent sur la surface du dôme et d'autres qui peuvent le traverser ou s'en éloigner.

3.4. Global Sound Diffusion

Contrairement au VBAP, qui gère les niveaux de chacun des haut-parleurs afin de conserver une intensité constante lors des déplacements, le MBAB n'effectue aucune correction de cet ordre, sauf une correction panoramique de base, à peu près semblable à celle utilisée en stéréo. Or ici la distance physique réelle entre les haut-parleurs, qui n'est jamais prise en compte par SpatGRIS, joue un rôle important dans la fluidité des mouvements. Afin d'aider les artistes à bien ajuster les niveaux d'atténuation entre les haut-parleurs, nous avons ajouté un réglage supplémentaire dans la fenêtre Speaker Setup Edition qui se nomme *Global Sound Diffusion*. Celui-ci permet d'ajuster la courbe d'atténuation entre chaque

source représentée en 2D (cercle) sur l'interface du logiciel devient l'angle zénital dans l'environnement 3D (hémisphère).

haut-parleur. De presque rien — valeur à 0.00, passage discret d'un haut-parleur à l'autre — à la valeur maximum — valeur à 1.00, qui met à contribution un grand nombre de haut-parleurs afin qu'aucune perte de signal ne soit perceptible dans le déplacement. L'ajustement minimum correspond à des localisations très précises alors que le maximum correspond à des situations plus immersives.

3.5. Cube et Sphère au complet

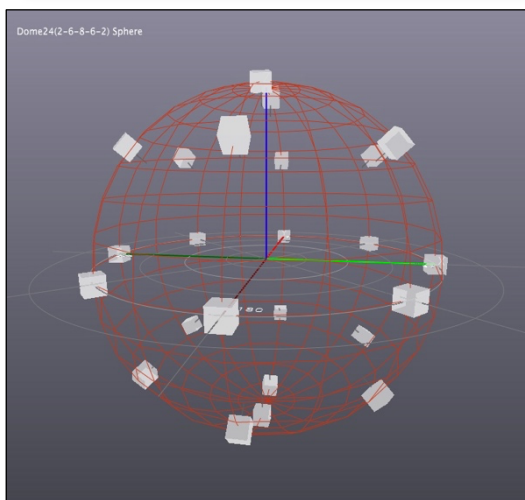
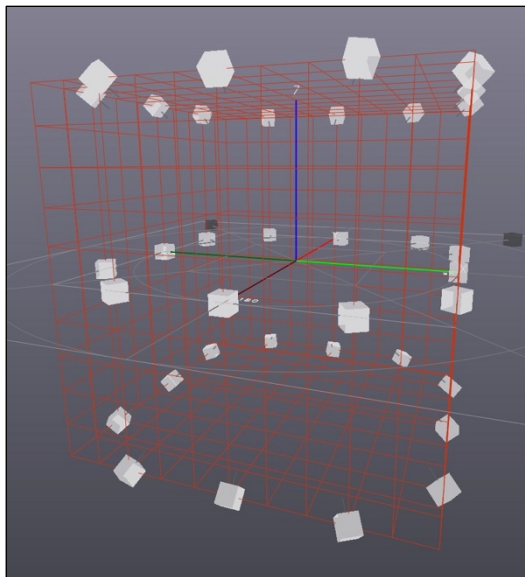


Figure 5. Cube et Sphère peuvent être utilisés au complet.

Si on possède un dispositif qui permet de placer des haut-parleurs sous le plancher [Figure 5], il est possible de spatialiser le son en tenant compte de ceux-ci.

4. CONTROLGRIS

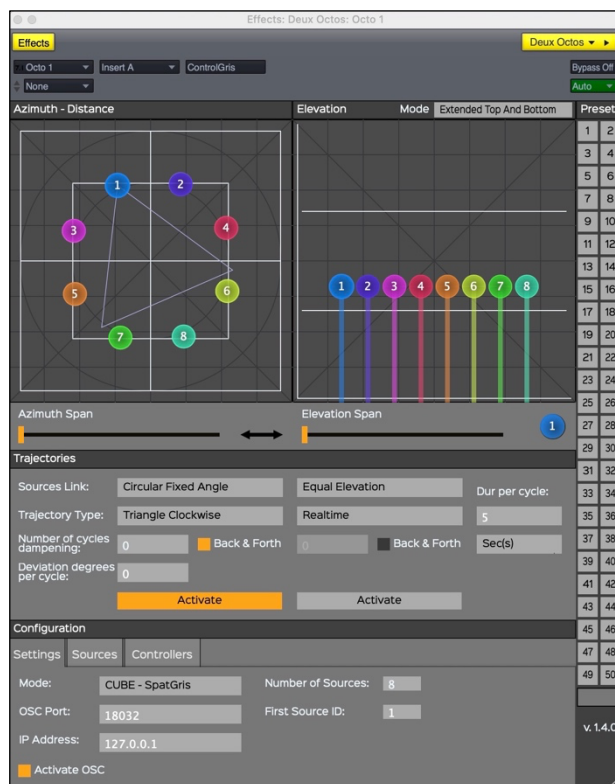


Figure 6. Les réglages de ControlGRIS en mode CUBE.

ControlGRIS est conçu pour transmettre des métadonnées de spatialisation en provenance des pistes d'un séquenceur vers SpatGRIS, ce dernier avait été développé initialement sur la base d'une spatialisation à distance (radius) constante, par rapport à un point central d'un dispositif concentrique (*sweetspot*) comme un dôme de haut-parleurs. En effet, l'interface graphique 2D de l'ancien plugiciel (SpatGRIS1) ne permettait d'assigner le positionnement et le mouvement d'une source que sur la surface d'une sphère. L'interface du ControlGRIS en mode CUBE en contrepartie [Figure 6], permet non seulement de modifier l'élévation des sources à l'intérieur du dispositif, mais également de les positionner au-delà des frontières établies par la configuration des haut-parleurs⁸.

Le couple DAW ou générateur OSC/ControlGRIS simplifie le processus en faisant converger toute spatialisation vers SpatGRIS. Cela rend l'outil d'autant plus utile aux utilisateurs de DAW qui n'offrent que des pistes mono ou stéréo, outre la sortie maître. De plus, le couple ControlGRIS/SpatGRIS peut être contrôlé par plusieurs interfaces physiques externes [voir Figure 1], incluant le joystick, l'iPad, le Leap Motion et, plus récemment, le Lemur ou encore par des logiciels qui gèrent l'OSC, comme Max, PureData ou SuperCollider.

⁸ Ceci n'est vrai que pour le mode CUBE ; le mode DOME impose tout positionnement des sources sur la surface d'une sphère.

D'autres interfaces de contrôle pourront éventuellement s'ajouter à cette liste dans le futur.

Enfin, l'interface du ControlGRIS permet d'y dessiner des trajectoires graphiques, qui peuvent ensuite être parcourues par les sources sonores virtuelles à des vitesses variables, contrôlables durant l'enregistrement de celles-ci sous forme d'automation dans le DAW. Cette possibilité facilite et accélère grandement l'écriture et l'arrangement spatial d'un certain nombre de trajectoires, et ce, de manière très intuitive.

5. LE LECTEUR (PLAYER)

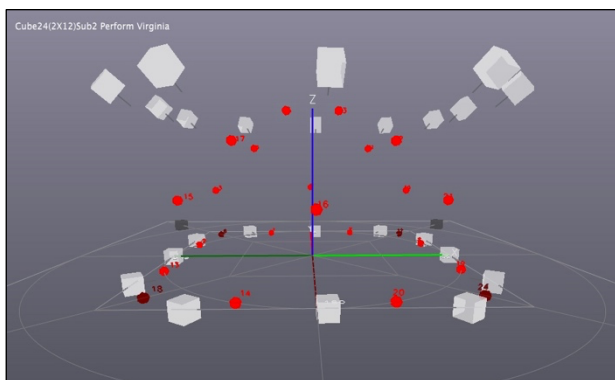


Figure 7. Reproduction d'une musique composée en mode DOME sur un dispositif CUBE.

L'ajout d'un lecteur multipiste universel, destiné à la circulation des œuvres de musique spatiale conçues pour un nombre indéterminé de haut-parleurs, nous est apparu nécessaire pratiquement au début du développement de SpatGRIS, mais ce n'est que récemment que celui-ci s'est concrétisé. En effet, un des problèmes soulevés par la création d'œuvres immersives est qu'elles sont souvent destinées à un dispositif spécifique, rendant leur circulation très difficile. Nous avons donc fait la proposition au ICMC18⁹ d'incorporer un enregistreur/lecteur au format 16.2 (48 kHz, 24 bits) qui permettrait aux compositeurs d'enregistrer leur musique et aux auditeurs de l'écouter sous ce format ou en mode binaural [8], grâce au lecteur multipiste qui les reproduirait [Figure 7].

À l'heure actuelle, contrairement au cinéma, par exemple, qui a uniformisé sa distribution à travers des standards¹⁰, la circulation des musiques autres que stéréophoniques est très limitée. Néanmoins, avec le concept du dôme de haut-parleurs, employé autant par le VBAP que par le HOA, une certaine forme de standardisation a pu être envisagée. En effet, ces modes de spatialisation permettent la portabilité d'une œuvre déjà spatialisée vers n'importe quelle autre configuration en forme de dôme. Dans ce contexte, la composition d'une musique spatiale est alors indépendante de toute configuration de haut-parleurs. Si cela s'applique actuellement aux situations de concert, ce n'est pas

⁹ International Computer Music Conference, tenue à Daegu en Corée du sud en août 2018.

toujours le cas dans un contexte d'écoute individuelle, pour la diffusion sur le web, ou encore lors d'un appel d'œuvre dans le cadre d'un concours. C'est pourquoi nous avons ajouté au SpatGRIS un lecteur multicanal entièrement autonome, qui ne nécessite pas l'usage d'un DAW. Il est désormais possible d'y lire une session audio multipiste préalablement enregistrée directement, dont les positions des sources sont déterminées par l'enregistrement du dispositif originel sur lequel l'œuvre a été composée.

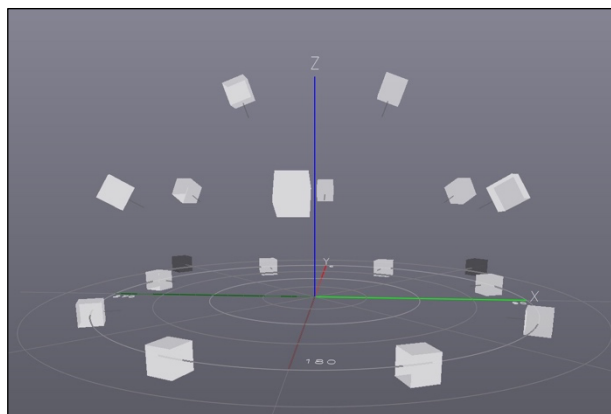


Figure 8. Default Speaker Setup 8-6-2 plus 2 subs.

Le Player peut jouer jusqu'à 128 canaux, comme SpatGRIS dans son ensemble. Nous avons décidé de proposer le format 16.2 comme format multicanal standard – avec une disposition à trois niveaux : 8-6-2 plus deux haut-parleurs de graves [Figure 8]. Cela correspond au format par défaut de SpatGRIS car celui-ci nous a démontré, au fil des ans, être une configuration minimale adéquate pour la reproduction de musiques immersives multicanal. De plus, ce format étant destiné à une écoute individuelle, il constitue un bon compromis sur le plan de la taille des fichiers (environ 2,3 Go, pour une durée de 15 minutes). Ce format aurait été impensable il y a 15, voire 10 ans, mais maintenant les réseaux de distribution sont suffisamment performants pour une telle distribution.

On peut noter en terminant que le Player n'est pas limité au format par défaut et qu'il peut s'adapter à n'importe quel format d'enregistrement et de reproduction.

6. DÉVELOPPEMENTS FUTURS

6.1. Trajectoires concrètes

À ce jour, SpatGRIS comporte un système basé sur le contrôle direct de trajectoires linéaires – cercles, pendules, spirales et autres mouvements aléatoires – dont les paramètres ne sont modifiables que par l'utilisateur, sans toutefois permettre de tenir compte de la nature des

¹⁰ Les Dolby Surround™ en format 5.1, 6.1 ou 7.1 et ATMOS (jusqu'à 64 canaux).

sons. Ces trajectoires demeurent donc abstraites du contenu sonore. Or, notre perspective de développement va plutôt dans le sens de trajectoires concrètes (au sens de musique « concrète ») où celles-ci pourraient être directement associées à des propriétés du son, notamment grâce à l'analyse de signal. Cela implique que ControlGRIS (ou un autre plugiciel à programmer) puisse intégrer des données issues de l'analyse du signal de la piste audio, afin de moduler certains paramètres de spatialisation envoyés vers SpatGRIS sous forme de métadonnées (OSC).

6.2. Analyse de signal

En intégrant des outils d'analyse du signal de type descripteurs audio basés sur la perception – par exemple les Vamp Plugins, développés à la Queen Mary University of London (UK) [2] – à ceux du GRIS, il deviendrait possible d'exploiter une certaine quantité de données complexes pour moduler certains paramètres de spatialisation. À l'instar de ControlGRIS, ces plugiciels ne transmettent pas de signal audio en temps réel. Ils extraient plutôt des données d'un signal en entrée afin de les étiqueter par composante du son. Par exemple, en analysant le signal audio, transmis par un client connecté au SpatGRIS, le compositeur pourra alors choisir d'associer le centroïde et l'étendue spectrale du signal reçu à la redistribution spatiale de celui-ci dans l'environnement de diffusion. Si le centroïde se situe dans l'aigu, le signal serait spatialisé vers le haut de l'environnement de diffusion. À l'inverse, si le centroïde est plus grave, le signal serait distribué vers le bas. Dans un autre cas, la position du centroïde spectral pourrait plutôt servir à moduler la vitesse de déplacement de certaines trajectoires sonores. Finalement, les artistes auraient la liberté de redéfinir toute relation signal/spatialisation et d'en modifier les paramètres de redistribution, afin de créer des situations musicales inédites. Une telle approche – qui fait d'ailleurs écho à une technique de spatialisation du timbre présentée par Normandeau [07] – n'est qu'un exemple très simple parmi une foule d'autres applications rendues possibles par l'intégration prochaine des outils d'analyse de signal à nos outils de spatialisation audionumériques.

6.3. Ambisonie d'ordre élevé (HOA)

Enfin, nous prévoyons également intégrer au SpatGRIS un décodeur ambisonique d'ordre élevé (HOA¹¹), afin d'offrir une solution adaptée à l'univers sonore des jeux vidéo et, surtout, de la réalité virtuelle qui utilisent en abondance ce format. Bien que la qualité spatiale et sonore des rendus en HOA soit relativement comparable à celle du VBAP [6], le HOA a l'avantage de facilement permettre le décodage de contenus audio vers différents formats et plateformes de diffusion. Dans l'optique où nous voulons que nos outils favorisent la composition de l'espace simultanément au contenu

sonore – incluant les situations de musique en direct –, la flexibilité et la polyvalence de ceux-ci représentent un critère important de leur développement. C'est pourquoi nous souhaitons éventuellement ajouter le HOA à la liste des modes de spatialisation actuellement offerts par SpatGRIS.

6.4. Amélioration du mode binaural

Grâce au lecteur multipiste et au format de transmission 16.2, SpatGRIS est en mesure d'associer chaque piste à un haut-parleur virtuel correspondant – représentant une réponse impulsionnelle (RI) de type HRTF – dans l'espace de diffusion binaurale. En transmettant les œuvres de cette façon plutôt qu'au format binaural – déjà réduit à partir d'une banque de RI (HRTF) génériques dont l'efficacité peut considérablement varier d'une personne à l'autre – cela permettrait éventuellement aux auditeurs de choisir une banque de RI (HRTF) adaptée à leur écoute.

6.5. Écoute en ligne

Parallèlement, les fournisseurs de contenu audio en ligne (web) pourraient utiliser SpatGRIS en amont de leur chaîne de diffusion et ainsi ne transmettre que le contenu binaural d'un fichier audio 16.2, réduisant ainsi la taille de la bande passante requise. Encore là, l'application de RI (HRTF) personnalisées sur l'écoute en ligne deviendrait également possible.

7. CONCLUSION

Pour résumer, à travers le développement du couple ControlGRIS / SpatGRIS, nous souhaitons contribuer au progrès des pratiques artistiques employant l'espace comme terrain de jeu expérimental pour la création : en facilitant le contrôle et l'écriture (automatisations) simultanés du son et de l'espace ; en mettant en place un système d'analyse de signal permettant d'associer directement des propriétés sonores à des trajectoires ; en favorisant la transmission et la diffusion des œuvres spatialisées ; et en intégrant de manière continue de nouveaux modes de spatialisation et outils de contrôle.

8. REMERCIEMENTS

Ce projet de recherche-crédation est soutenu par : le Centre interdisciplinaire de recherche en musique, médias et technologie (CIRMMT) ; le Conseil de Recherches en Sciences Humaines (CRSH) du Canada, avec le numéro de subvention 435-2021-0230; le Fonds de Recherche du Québec - Société et Culture (FRQSC), avec le numéro de subvention 2020-RC1-269290 ; la Fondation Canadienne pour l'Innovation, avec le numéro de subvention 33558 ; et la faculté de musique de l'Université de Montréal. Enfin nous remercions David Ledoux à qui ce texte doit quelques paragraphes.

¹¹ High Order Ambisonics.

9. RÉFÉRENCES

- [1] Ivica Ico BUKVIC. « Introducing D4 : An Interactive 3D Audio Rapid Prototyping and Transportable Rendering Environment Using High Density Loudspeaker Arrays ». Dans : *Proceedings of the International Computer Music Conference 2016* (2016). URL : <https://pod/dod-idx/introducing-d4-an-interactive-3d-audio-rapid-prototyping.pdf?c=icmc;idno=bbp2372.2016.096;format=pdf>.
- [2] Queen Mary University of London CENTRE FOR DIGITAL MUSIC. *The Vamp audio analysis plugin system*. online. URL : <https://www.vamp-plugins.org>.
- [3] David LEDOUX et Robert NORMANDEAU. « An Immersive Approach to 3D-Spatialized Music Composition. Tools and Pilot Survey ». Dans : *AM'18 : Proceedings of the Audio Mostly 2018 on Sound in Immersion and Emotion*. dir. de Stuart CUNNINGHAM et Richard PICKING. Wrexham, United Kingdom : ACM, 2018.
- [4] David G. MALHAM. « Higher Order Ambisonic systems for the spatialisation of sound ». Dans : *Proceedings of the 1999 International Computer Music Conference (ICMC 1999)*. Pékin, Chine, oct.1999.
- Peter MANNING. *Electronic and Computer Music*. New York : Oxford University Press, 2013.
- [6] G. MARENTAKIS, F. ZOTTER et M. FRANK. « Vector-Base and Ambisonic Amplitude Panning : A Comparison Using Pop, Classical, and Contemporary Spatial Music ». Dans : *Acta acustica united with Acustica. the journal of the European Acoustics Association (EAA); international journal on acoustics* 100.5 (oct. 2014), p. 945-955.
- [7] Robert NORMANDEAU. « Timbre Spatialisation : The medium is the space ». Dans : *Organised Sound* (déc. 2009), p. 277-285.
- [8] Robert NORMANDEAU et al. « SpatGRIS/ServerGRIS, Creative tools for 2D and 3D sound spatialization ». Dans : *Proceedings of the 2018 international computer music conference*. Daegu, Korea : The International Computer Music Association, 2018, p. 291-297.
- [9] OQLF. *Banque de terminologie du Québec. Plugiciel*. Français. Office québécois de la langue française. 2013.
- [10] Ville PULKKI. « Virtual Source Positioning Using Vector Base Amplitude Panning ». Dans : *Journal of the Audio Engineering Society* 45 (1997), p. 456-466.
- [11] Agnieszka ROGINSKA et Paul GELUSO, dir. *Immersive Sound : The Art and Science of Binaural and Multi-Channel Audio*. New York et Londres : Routledge, 2018.
- [12] Nicolas TSINGOS. « Object-Based Audio ». Dans : *Immersive Sound : The Art and Science of Binaural and Multi-Channel Audio*. dir. d'Agnieszka ROGINSKA et Paul GELUSO. New York et Londres : Routledge, 2018, p. 244-275.

SOUND TRAJECTORY EN PRATIQUE LIVE

Maxence Mercier

Université Côte d'Azur, CTELA¹, XR2C2², France

maxence.mercier@univ-cotedazur.fr

RÉSUMÉ

Depuis 2017, le logiciel *Sound Trajectory*³ propose des outils visant à simplifier les processus de création sonore immersive. Sa modularité est adaptée au modelage et à l'articulation en temps réel des principaux éléments de la spatialisation sonore.

A l'occasion des JIM 2023, le code source de ce logiciel sera mis à disposition sous licence GPL⁴ afin de permettre une utilisation de ses fonctionnalités dans l'environnement Max et garantir une certaine pérennité de ces concepts au-delà du programme original. Cet article présente le logiciel et trois cas pratiques de production scénique qui documenteront la publication du code source. Le premier cas décrit l'usage de l'échantillonneur de *Sound Trajectory* en combinaison avec un processeur audio externe. Le deuxième cas envisage la recréation virtuelle du *Cybernéphone*, un des premiers dispositifs de diffusion pour la musique électroacoustique. Le troisième cas expose l'utilisation live du logiciel lors d'une performance solo pour piano, synthétiseur et électronique. Cette étude bénéficie d'un soutien du programme ANR-15-IDEX-01 dans le cadre d'un contrat doctoral en création/recherche à l'université Côte d'Azur.

1. INTRODUCTION

L'espace a longtemps été considéré comme un cadre de la représentation musicale plus que comme un paramètre de jeu musical. Si d'un point de vue historique on retrouve des considérations pour la spatialisation sonore déjà dans les chœurs antiphonaires [1], l'essor des musiques électroacoustiques dans la deuxième moitié du XX^e siècle ont rapidement tiré parti des innovations de l'électronique pour jouer par l'intermédiaire des haut-parleurs sur la localisation du son et la cinétique de son déplacement. Depuis, chercheurs et compositeurs ont proposé une grande diversité d'approches techniques et esthétiques. Avec la montée en puissance de l'informatique, tout créateur sonore a désormais accès à une offre abondante de solutions de spatialisation. Cependant des besoins de documentations, de formations et de vulgarisations sont nécessaires pour démocratiser réellement l'utilisation effective de la spatialisation sonore dans les productions artistiques. L'ouverture du code source de *Sound Trajectory* vise à approfondir un

travail de documentation des processus d'écriture spatiale qu'il permet. Ce logiciel s'articule autour d'un éditeur de trajectoires tridimensionnelles permettant de spatialiser tout type de source sonore, de créer des masses spatiales, de synthétiser des espaces acoustiques virtuels réalistes ou abstraits. La problématique de cet outil est d'offrir le maximum d'expressivité avec le minimum de gestes, paramètres et commandes. Il apporte des solutions face à la complexité empirique des processus de création sonore immersive en offrant une interface dédiée à une écriture de la spatialisation utilisable en studio, sur scène, lors de la conception d'installations et pour la scénographie sonore en réalité virtuelle.

La vulgarisation récente du mixage objet comme technique indépendante de la configuration des haut-parleurs permet de réaliser des compositions spatiales pour de nombreux domaines d'applications. Les outils tels que les *panners* intégrés au *DAW* sont généralement simples d'emploi et nécessitent des connaissances minimales [2]. Mais lorsque la composition spatiale exige une approche plus élaborée, les méthodes d'organisation de la production d'un projet manquent de références.

La gestion des nombreux paramètres dans les systèmes audio 3d exige des méthodes rigoureuses pour gérer efficacement l'intégralité de l'écriture des mouvements de spatialisation. De plus, chaque domaine d'application a ses propres besoins, qu'il s'agisse du cinéma, du théâtre, de performances musicales en direct ou encore d'installations sonores interactives. Ces contextes partagent de nombreux outils, mais nécessitent souvent des interfaces adaptées à des contrôleurs spécifiques. La recherche sur l'interaction entre ces contrôles et les tâches audio immersives est cruciale pour développer des méthodes intuitives et fiables de gestion de scènes sonores en 3d, qui seront de plus en plus courantes dans le futur [3].

Le choix des workflows de création influence directement l'esthétique et la perception d'une œuvre spatialisée [4]. L'approche semi-modulaire proposée au sein de *Sound Trajectory* permet d'établir des stratégies de composition et de mixage flexibles tout en offrant des possibilités d'interconnexion aux solutions de spatialisation existantes.

1 Centre Transdisciplinaire en Épistémologie de la Littérature et des Arts vivants : <https://ctel.univ-cotedazur.fr/>

2 Centre de référence IDEX en réalité étendue : <https://xr2c2.univ-cotedazur.fr>

3 *Sound Trajectory* : <https://www.tripinlab.com>

4 <https://www.gnu.org/licenses/gpl-3.0.html>

2. HISTORIQUE

Le projet *Sound Trajectory* (ST) s'inscrit dans la lignée du logiciel Holo-Edit⁵, il en prolonge les principes conceptualisés dès la fin des années 90 dans les travaux de Laurent Pottier [5]. Il partage également les objectifs du logiciel *Spat Gris*⁶ du Groupe de recherche en immersion spatiale à Montréal piloté par Robert Normandeau, visant à faciliter l'accès à des outils d'écriture spatiale robustes, simples et efficaces à destination de compositeurs et créateurs sonores employant l'espace comme paramètre expressif d'écriture [6]. *Sound Trajectory* propose néanmoins des logiques différentes car la gestion du temps est opérée par un seul contrôle continu définissant la position d'une source sonore sur une trajectoire. Pour l'utilisateur, un seul paramètre exprimé en pourcentage discrétise la gestion des triplets de coordonnées polaires ou cartésiennes couramment employés pour positionner une source sonore dans un espace tridimensionnel.

En 2014, j'ai élaboré les fondations du dispositif du Synesthésium⁷, un dôme géodésique équipé de 24 haut-parleurs destiné à une diffusion ambisonique 3d. Ce dispositif de création et diffusion était destiné à accueillir le Cauchemar merveilleux⁸, une œuvre sonore composée par Léonore Mercier sur un recueil de poésies d'Arthur H [7]. Léonore Mercier avait besoin d'un outil lui permettant de s'approprier seule les possibilités de mixage dans un espace 3d sans que cela n'implique un apprentissage trop complexe. Sur la base des développements réalisés pour mon opéra multimédia *Iquisme* [8], utilisant les matrices et les objets visuels en OpenGL⁹ de la bibliothèque Jitter¹⁰, j'ai conçu un éditeur visuel permettant de positionner sur des trajectoires les sources sonores d'un mixage 3d. Une fois qu'une trajectoire était dessinée dans l'espace, le seul paramètre à automatiser depuis un DAW était une variable exprimée en pourcentage définissant la position de la source sonore sur cette trajectoire. Le logiciel Iannix¹¹ aurait pu permettre cet usage, mais l'implémentation des traitements audio dans une même application était plus commode à la gestion du projet [9]. Une application standalone Max fut alors réalisée en intégrant la bibliothèque Spat de l'Ircam [10].

Cet outil permettait le mixage de la scène spatiale avec une écriture intuitive, "à l'oreille", dynamique et flexible. Avec le plug-in Tosca [11] chargé dans des bus Pro-Tools¹², il était alors possible d'automatiser la position et les éventuels mouvements de la source sur la trajectoire. La compositrice a pu réaliser en trois semaines un répertoire de plus d'une heure de musique entièrement composé en mixage objet, avec des pièces comprenant près d'une centaine de pistes sonores. Tout le processus d'écriture et de diffusion a été assuré par le même outil.

Ce projet est documenté par plusieurs vidéos lors de présentations publiques couvrant la création du Cauchemar merveilleux.

Le logiciel *Sound Trajectory* fut ensuite commercialisé en janvier 2018, il est la synthèse des outils d'écriture développés pour le projet du *Synesthésium*.

Sound Trajectory est aujourd'hui utilisé par une communauté d'un millier de créateurs sonores de styles et d'horizons divers : musique, cinéma, installations, jeux vidéo.

Le logiciel s'articule autour d'un éditeur de trajectoires tridimensionnelles permettant de spatialiser toute source sonore, de créer des masses spatiales, de synthétiser des espaces acoustiques virtuels vraisemblables ou à l'inverse totalement abstraits.

Le moteur audio de *Sound Trajectory* utilise la bibliothèque ambisonique de l'ICST développée par Jan Schacher, et Philippe Kocher [12].

Les derniers développements du logiciel explorent les possibilités de mixage en temps-réel par le biais de son interface modulaire offrant des processus d'automatisation en interaction avec des suiveurs d'enveloppes, par intensités ou analyses de spectres ainsi que par des contrôleurs OSC [13] et MIDI.

Le logiciel propose des sessions dynamiques, chacune pouvant embarquer un nombre illimité de scènes, ces dernières servant à sauvegarder et rappeler tous les paramètres édités. Les protocoles MIDI et OSC sont intégrés en entrée et sortie afin de contrôler *Sound Trajectory* depuis un autre logiciel et permettre le contrôle d'autres solutions de spatialisation, logiciels (Spat Révolution¹³, Panoramix [14], Dolby Atmos Panner etc.) ou matériels comme HoloPhonix¹⁴, L-ISA¹⁵ parmi les plus courants. Une suite de *devices* Max4Live pour Ableton Live¹⁶ est également disponible avec une intégration plus poussée des capacités d'automatisation du système.

3. FONCTIONNALITÉS

L'objectif du logiciel est d'appréhender l'expressivité de la spatialisation avec le minimum de gestes. Sa modularité est adaptée au modelage et à l'articulation en temps réel des composantes de la synthèse spatiale pour tout type de dispositif.

Son interface permet de naviguer dans un espace 3d pour y dessiner des trajectoires. Les points de chaque trajectoire sont modifiables dans la visualisation 3d ou dans un tableau de valeurs exprimées en coordonnées cartésiennes ou polaires. Les trajectoires peuvent être transformées par différents procédés de rotation, changement d'échelle, translation, symétrie et inversion. L'éditeur offre différents modes de visualisation et angles

5 Holo-Edit : <https://gmem.org/holophon>

6 Spat Gris : <http://gris.musique.umontreal.ca/>

7 Le Synesthésium : <https://www.synesthesium.com>

8 Le cauchemar Merveilleux, reportage fondation Hermès, 2015, <https://www.youtube.com/watch?v=v1hIFdBnls0>

9 <https://www.opengl.org/>

10 <https://cycling74.com/products/jitter>

11 Iannix : <https://www.iannix.org/fr/>

12 Pro-Tools <https://www.avid.com/fr/pro-tools>

13 Spat Révolution : <https://www.flux.audio/project/spat-revolution/>

14 HoloPhonix : <https://holophonix.xyz/>

15 L-ISA : <https://l-isa.l-acoustics.com/>

16 Ableton Live : <https://www.ableton.com/fr/live/>

de vue activables depuis un bandeau d'icônes représentant les fonctions disponibles. La figure 1 présente l'éditeur d'espace sonore en point de vue libre, on distingue le chutier des différentes pistes et la courbe d'atténuation des sources par la distance.

Le logiciel gère jusqu'à 64 trajectoires, associées à des sources audio, mono comme multipistes. Les sources multipistes supportent elles-mêmes jusqu'à 16 canaux.

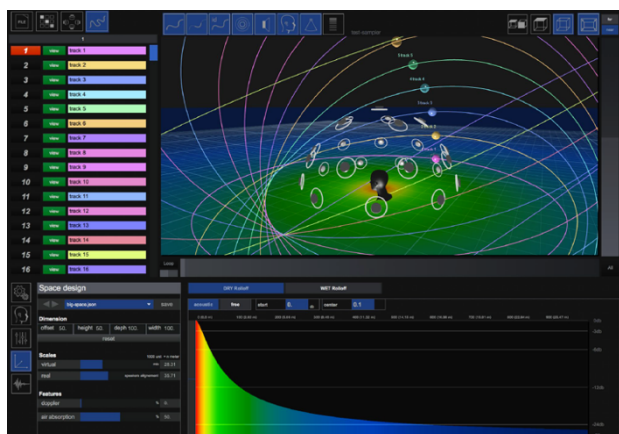


Figure 1 interface principale de Sound Trajectory

3.1. Écriture de trajectoire

Les trajectoires sont la fonctionnalité emblématique de *Sound Trajectory* et la majorité des autres fonctions du programme en sont dépendantes.

La trajectoire invite à faire se mouvoir les sons, mais elle peut se contenter de n'être qu'un guide voire simplement un repère visuel. Elle est représentée sous la forme d'un trait, l'élément le plus simple après le point, essentiel au dessin, à la notation et l'écriture. Au sein de *Sound Trajectory*, la trajectoire est un support d'écriture et repère de déploiement du son dans l'espace. La trajectoire est une virtualité reproductible et appréhendable indépendamment des contraintes imposées par les dispositifs de diffusion. Ainsi, elle peut traverser les époques en bénéficiant des avancées technologiques liées aux questions d'algorithmes d'encodage et de décodage des sources sonores.

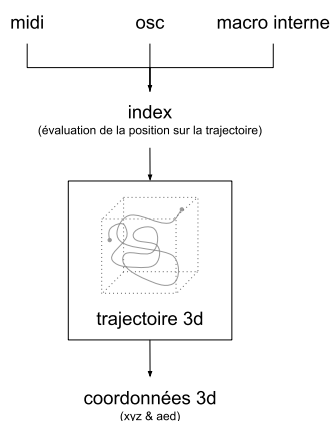


Figure 2 Évaluations des coordonnées d'une trajectoire

Les facteurs influençant la perception du son en mouvement sont nombreux, la représentation simplifiée d'une trajectoire invite à expérimenter les modalités d'expression du son dans l'espace. Ce potentiel ne se heurte pas à des limites techniques.

3.2. Space designer

Le module *Space designer* définit les paramètres de réverbération et d'amplitude par rapport à l'auditeur (*listener*). Le module représente les dimensions et les propriétés acoustiques de la scène virtuelle.

3.2.1. Doppler & atténuation du son par la distance

La gestion de l'effet *doppler* et de l'atténuation du son par la distance, couplés à la réverbération, permettent de jouer sur les rapports physiques de l'environnement virtuel. Au bas de la figure 1, on visualise la courbe d'atténuation du volume par la distance calculée entre le *listener* et la source. Elle est paramétrable suivant des caractéristiques acoustiques vraisemblables ou abstraites. Cet outil est particulièrement utile pour contrôler les sensations d'enveloppement du son par rapport à la gestion du point chaud de l'environnement (*sweet spot*). En dehors des conventions acoustiques, ce paramètre a un fort pouvoir créatif pour simuler la propagation du son dans des environnements non aériens : solide, liquide ou gazeux.

3.2.2. Réverbération

La réverbération à convolution de ST est basée sur l'algorithme *convolve~* de la bibliothèque *HISSTool* [15], Le logiciel est livré avec une collection d'une quarantaine d'impulsions et l'utilisateur peut charger ses propres impulsions. La réverbération est calculée sur un bus auxiliaire. Le traitement est appliqué sur un encodage HOA décodé en double octophonie (une couronne pour le haut et une autre pour le bas) puis recodé en HOA.

3.3. Échantillonneur

L'échantillonneur permet de déclencher la lecture de sons sur des trajectoires avec une précision proche de l'échantillon audio. Il résout les problématiques de latence d'un système OSC dans le cadre d'interaction rapide en temps-réel.

3.4. Motion Listener

Il s'agit d'un principe élémentaire utilisé dans les moteurs audio 3d de jeux vidéo permettant de déplacer la position d'écoute virtuelle. Cette fonctionnalité calcule la position de toutes les sources sonores par rapport à la position du *listener*. C'est une des premières fonctions implémentées dans *Sound Trajectory*. Les options de visualisation permettent d'adapter la vue 3d à la manière de la vue subjective des jeux vidéo. Ses coordonnées sont modifiables manuellement ou définies par une trajectoire.

Le module est entièrement accessible par OSC et peut donc être contrôlé par des capteurs de *headtracking*.

3.5. Mapping interactif

Cette page est destinée à la programmation semi-modulaire d'interactions pouvant contrôler les principales fonctions du logiciel. La variété des modules à base de descripteurs audio, MIDI et OSC ouvre la perspective de scénarios interactifs entièrement dynamiques et relativement simples d'usage. Des triggers et des contrôles continus sont à affecter à une action (séquenceur, boucle, compteur et autres fonctions), elle-même affectée à un paramètre spatial de trajectoires, de sources ou du *listener*.

Les modules *Analyser* et *Events* se combinent pour explorer de multiples couples d'interactions spatiales. Les modules d'analyse produisent des signaux sous forme d'impulsion ou de continuum. Le signal d'impulsion déclenche une action. Il est activé par le franchissement d'un seuil d'amplitude audio, une note, un contrôleur MIDI ou un événement OSC.

La description de chaque module dépasse le cadre de cet article, mais quelques possibilités peuvent être évoquées comme : les modules de marche aléatoire qui vont contraindre une source à bouger sur une trajectoire ; le module de morphing entre deux modèles de trajectoire ; une collection de différents types de séquenceurs activés selon des conditions logiques ; quelques opérateurs permettant de boucler le système sur lui-même pour des processus auto-générateurs par rétroactions.

3.6. Type de pistes audio

Différents types de pistes audio sont disponibles pour couvrir des fonctions particulières.

Les modèles *Ambisonic mix* et *Dbap mix* sont dédiés au mixage d'entrées audio acceptant une à seize sources sonores. L'ordre HOA est défini par le réglage de la session. L'algorithme DBAP est quant à lui dépendant de la configuration des haut-parleurs [16].

Le modèle *Ambisonics HOA input* permet de mixer une piste au format Ambix du 1er au 11e ordre.

Le modèle *direct output* est destiné à la conception de consoles pour l'interprétation des musiques électroacoustiques. Des fonctionnalités de pitchshift, délais et filtrages sont implémentées en suivant les paradigmes du Gmebaphone conceptualisé par Christian Clozier pour créer des registres de voix de diffusion différenciés par bande de fréquences, phases et transpositions de quelques commas.

3.7. Multi-délai

Le multi-délai offre une palette de traitements sonores qui ont été par ailleurs largement étudiés au travers des applications de la bibliothèque Hoa¹⁷ du CICM [17]. Leur

expressivité est documentée au travers d'une abondante littérature de Kendall [18], Vaggione [19], Sèdes [20] à Goutmann [21] formalisant les possibilités de la décorrélation microtemporelle ou bien encore la modulation spatiale en anneau. Le multi-délai vise les mêmes objectifs par le biais d'une interface et d'un fonctionnement adapté au sein du logiciel *Sound Trajectory*.

Ce traitement permet des variations drastiques des champs sonores en modelant de manière souple les morphologies de distribution du son dans l'espace. L'interface du traitement expose les paramètres essentiels du contrôle d'un espace paramétrique plus vaste. Les paramètres accessibles à l'utilisateur sont optimisés pour accéder rapidement à une palette de textures spatiales adaptées à la composition en situation multicanal avec une grande variété d'anamorphoses temporelles et spectrales possibles.

La polyphonie des voix du multi-délai permet d'obtenir des effets classiques comme les retards multiples ou les échos, la réverbération et le déphasage, ou des effets de filtrage complexes lorsque les retards sont très proches les uns des autres. Les sources peuvent être déplacées le long d'une trajectoire de référence de manière reproductible.

Le feedback, paramètre de réinjection du son retardé à l'entrée de la ligne à retard, associé à un harmonisateur, permet de créer des transpositions continues du signal. Pour un signal non transposé, le feedback est un paramètre intéressant pour générer un signal complexe proche des réflexions d'un algorithme de réverbération.

Le multi-délai combine au sein de son interface, le contrôle des lignes de retard, la transposition, la prolifération par boucle de feedback et un contrôle de la distribution spatiale polyphonique. Chaque voix de retard est contrôlée indépendamment. Les paramètres variables sont : le temps de retard, le volume, la rétroaction (volume de réinjection) et la hauteur de transposition. Un filtre appliqué à la boucle de feedback peut colorer et atténuer le signal réinjecté à chaque itération dans la bande de fréquence filtrée.

3.8. Vidéo

Pour les projets audiovisuels, un lecteur et un écran virtuel intégrés dans l'environnement 3d du logiciel permettent de visualiser une source vidéo au format standard ou immersif. Elle peut être synchronisée via MTC (Midi Time Control) et provenir de sources externes.

3.9. Moteur audio

Sound Trajectory gère un nombre illimité d'enceintes. Un éditeur de positionnement des haut-parleurs permet de disposer chaque enceinte dans l'espace réel de diffusion. Toutes les compensations de volume et de temps de délais sont automatiquement calculées pour

¹⁷ HoaLibrary : <http://hoalibrary.mshparisnord.fr/>

optimiser la réponse du système et limiter les problèmes de phases. Il est ainsi possible d'adapter le type de mixage aux configurations de diffusion sonore : sphère ambisonique, full dome, planétarium, salle de concert, théâtre, studio, auditorium, cinéma, etc. Le moteur de rendu audio possède deux bus de sorties. Le premier est conforme au dispositif défini dans l'éditeur de configuration des haut-parleurs. Le second est configurable au format ambiX du 1er au 11e ordre suivant les spécifications de l'utilisateur.

3.10. Workflows

Un routage précis des signaux audio et de contrôle OSC ou MIDI sont un préalable à tout projet d'écriture de la spatialisation. La figure 3 montre un workflow simple où le DAW envoie les sources à spatialiser dans *Sound Trajectory*. Le son spatialisé est envoyé dans le système de diffusion par les sorties physiques de la carte son. Le son peut retourner au DAW pour être enregistré depuis les sorties principales ou le bus ambiX.

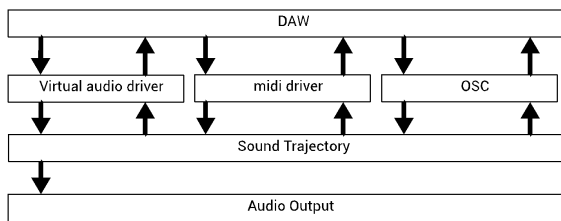


Figure 3 Workflow avec un séquenceur

3.11. Contrôle de moteur audio externe

Le moteur audio de *Sound Trajectory* peut être désactivé si l'utilisateur souhaite utiliser un processeur audio externe. Dans ce cas, seule l'automation des paramètres OSC est assurée par le logiciel ST. Le logiciel a été conçu dès son origine pour se situer à l'interface des solutions utilisées pour le studio, la scène et les besoins d'installations non standards. La figure 4 illustre le diagramme d'un workflow avec un séquenceur et un processeur audio externe.

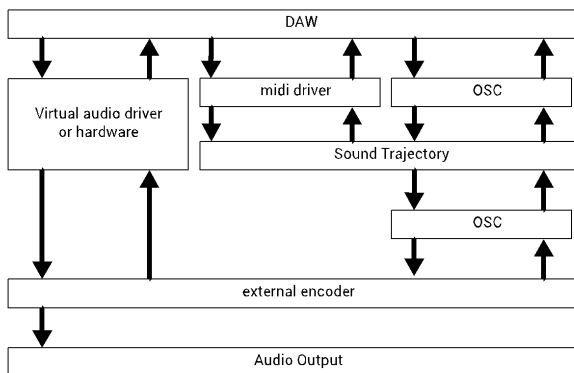


Figure 4 Workflow avec un séquenceur et un processeur audio externe

Après la présentation du logiciel, cet article expose trois cas pratiques où les héritages du passé se télescopent avec les dernières innovations en matière de mixage et de performance sur dispositif immersif virtuel ou réel.

4. CAS PRATIQUE N°1 : LA VIE

Sorti en février 2023, l'album *La Vie* d'Arthur H et Léonore Mercier a donné lieu à une série de concerts spatialisés à Radio France lors de l'HyperWeek Festival du 19 au 22 janvier 2023. En collaboration avec le département innovation de la Maison de la Radio et de la Musique, un nouveau dispositif utilisant le code source de *Sound Trajectory* fut réalisé pour permettre à Léonore Mercier d'interpréter en direct une partie de la spatialisation du projet musical.

4.1. Dispositif

Cette adaptation consistait à utiliser le moteur de trajectoire et l'échantillonneur de *Sound Trajectory* pour alimenter un processeur L-ISA de L-Acoustics. Le processeur recevait les voies de polyphonies de l'échantillonneur de *Sound Trajectory* avec les informations OSC de spatialisation associées, rafraîchies toutes les 8 ms en moyenne (120Hz).

La performance live consistait à repenser l'album de manière immersive, où le public est considéré comme un personnage de la narration.

L'album a été préalablement remixé au format 12.0 par Hervé Déjardin et Léonore Mercier depuis les sessions de studio originales. Toute la matière des arrangements fut réorchestrée spatialement pour être redéployée en concert sur une couronne de 12 points de diffusion assurée par un des trios d'enceintes Syva produits par L-Acoustics. L'encodage des signaux de diffusion en VBAP [22] était réalisé en temps réel par un processeur L-ISA par lequel convergeait toutes les sources sonores : micro voix, échantillonneur, bande-son multipistes. La figure 5 présente le routage audio et OSC du dispositif performatif.

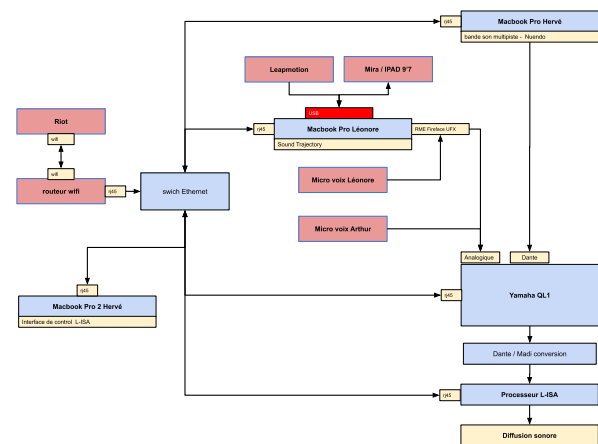


Figure 5 Routage audio et OSC du dispositif performatif

Léonore Mercier réinterprète les sons de l'album qu'elle a soustrait au mixage et spatialise la voix d'Arthur Higelin. Hervé Dujardin est à leur côté sur la scène, pour veiller à la spatialisation du mixage. La scène est placée au centre de la salle de concert. Les pistes de l'album sont lues depuis une station Nuendo¹⁸.

4.2. Les Capteurs

Le dispositif performatif de Léonore emploie des contrôleurs directement connectés à *Sound Trajectory*. Il s'agit de contrôles gestuels gyroscopiques et optiques permettant une grande liberté de geste.

4.2.1. Capteur Riot

Le *Riot*¹⁹ développé par Emmanuel Fletty pour l'Ircam, fournit à haute résolution des données 3d de mouvement. Il embarque plusieurs capteurs intégrés au sein d'une même puce électronique : gyroscopes, accéléromètres, magnétomètres. Léger et sans fil, il est adapté pour s'épingler sur un costume, se glisser dans un bracelet, etc. La version de Léonore renferme le capteur dans un boîtier pour être tenu facilement en main comme une petite télécommande.

Le *Riot* est connecté en WIFI et transmet des données par protocole OSC. Les données de mouvement sont traitées directement au sein de Max avec l'aide de la bibliothèque Mubu [23].

4.2.2. Le Leapmotion

Le *Leapmotion*²⁰ est une caméra infrarouge stéréo détectant la position des mains en 3 dimensions dans le champ visuel de la caméra. La caméra est échantillonnée à 120 hertz avec une latence très faible, permettant un contrôle gestuel fin et précis. L'objet permettant d'utiliser le *Leapmotion* dans Max a été développé par Jules François lorsqu'il était doctorant au sein de l'équipe de l'Ircam "Interaction son musique mouvement" qui produit la bibliothèque Mubu [24].

4.3. Les modes de jeux de l'échantillonneur

Une interface pour Mira²¹ sur iPad représentée en figure 6, permet de piloter le comportement de l'échantillonneur en déclenchant la lecture des sons selon 3 modes de jeux :

4.3.1. Échantillonneur en mode Tracks

Le premier mode « tracks » joue les sons de l'échantillonneur déclenchés depuis les pads de l'interface Mira. Les sons sont spatialisés en suivant une trajectoire définie dans l'écran de visualisation 3d de *Sound Trajectory*. La particularité de l'échantillonneur de *Sound Trajectory* est de pouvoir lire les échantillons en les spatialisant sur des trajectoires. La position du son sur

la trajectoire est automatiquement définie selon la position de la tête de lecture de l'échantillon.

4.3.2. Échantillonneur en mode Leap

Le deuxième est déclenché également par les pads de l'interface Mira, mais la spatialisation est assurée par la position de la paume de la main au-dessus du capteur *Leapmotion*. Cette position est retranscrite en données AED (azimut, élévation, distance) envoyées au processeur audio par message OSC.

4.3.3. Échantillonneur en mode Riot

Le troisième mode est déclenché et contrôlé par le mouvement du Riot. Les sons à jouer sont alors sélectionnés sur l'interface Mira, puis lorsqu'un changement brusque est détecté sur l'accéléromètre du *Riot*, le programme déclenche l'échantillon à jouer. Les boussoles et gyroscopes permettent de spatialiser en temps réel la position des échantillons dans l'espace.

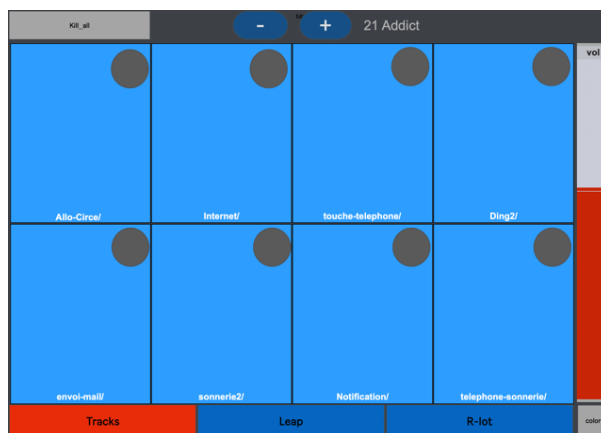


Figure 6 Interface de sélection des échantillons et modes de déclenchements

4.4. Dispositif de diffusion

Le format 12.0 est apprécié par Hervé Déjardin pour sa capacité à reproduire entre chaque enceinte un multiple de la distance angulaire idéale pour une paire d'enceintes stéréo, si bien que les images fantômes en sont amoindries, à défaut de ne pas posséder de canaux d'élévation.

Malgré les multiples entrées permises par le processeur L-ISA, la configuration de la performance ne permettait d'allouer que 6 objets L-ISA à l'échantillonneur de *Sound Trajectory*. Les capacités polyphoniques de cet échantillonneur ont néanmoins permis de jouer dans l'espace une multitude de sons de manière dynamique. La seule limite était de restreindre la polyphonie de l'échantillonneur à 6 voix simultanées. Les signaux audios étaient acheminés en analogique depuis une carte son RME Fireface vers une console

¹⁸ <https://www.steinberg.net/nuendo/>

¹⁹ RIOT par Emmanuel Fletty : <https://ismm.ircam.fr/riot/>

²⁰ Leapmotion : <https://www.ultraleap.com/product/leap-motion-controller/>

²¹ Mira : <https://cycling74.com/products/mira>

Yamaha QL1. Les signaux OSC de spatialisation étaient quant à eux interprétés par le processeur L-ISA. La configuration était ainsi la plus simple possible au regard des exigences techniques.

4.5. Contraintes & logiques

Si la production est expérimentale et que les premières représentations exigent la présence de développeurs, ingénieurs ou RIM, il est impératif que ceux-ci disposent, sur le plateau scénique, d'une plage de temps incompressible et respectée, dans un calme relatif et sans sollicitation extérieure.

Même si les systèmes sont parfaitement fonctionnels, ils peuvent s'avérer complexes à être initialisés pour la scène. Dans notre cas, les tâches étaient simples, mais imposaient aux personnes en charge de pouvoir être bien concentrées. Moi-même et Emmanuel Fletty, le concepteur des *Riot*, avions à calibrer les différents capteurs vis-à-vis de l'orientation de la scène avec le dispositif de diffusion sonore. Les programmes utilisant les capteurs nécessitent un point de référence pour différencier le devant du derrière, le côté droit du gauche, le haut du bas. Le *Leapmotion* se base sur ce que ses caméras capturent, un offset azimutal suffit à caler le dispositif. Les *Riots* quant à eux se réfèrent au magnétisme terrestre et il leur faut donc compenser leurs réponses selon l'orientation magnétique de la scène. Face à un problème mineur de conversion de valeur (-360. 360 à 0. et 1.) exigé par le processeur L-ISA, nous avons dû attendre la pause déjeuner de toute l'équipe pour résoudre en cinq minutes au calme, nos problèmes de quadrant polaire, inversant la linéarité des coordonnées suivant les axes polaires nord/sud.

4.6. Intérêt visuel

Le public apprécie généralement le rapport synesthésique d'une gestuelle bien visible contrôlant la spatialisation sonore. Le lien de causalité offre une lecture évidente des intentions d'interprétation spatiale de l'artiste. Les capteurs WIFI lui permettent de se déplacer sur scène sans être attachée à un écran. Le *Leapmotion* autorise des mouvements lents et doux extrêmement précis. Alors que l'essentiel de la musique était fixé sur les bandes, Léonore pouvait exploiter les systèmes de diffusion dans un rapport d'interprétation en adaptant son jeu de manière organique, c'est-à-dire avec la réponse acoustique de la salle, la réception du public et la réaction de son partenaire. Ce même dispositif avait été préalablement utilisé en 2016 pour le spectacle *Le cauchemar merveilleux* de Léonore Mercier et Arthur H.

Bien que le Thérémine ait plus d'un siècle d'existence, les modes de jeux électroniques dans l'éther suscitent toujours l'intérêt et la curiosité du public, étonné par l'efficacité des gestes.

5. CAS PRATIQUE N°2 : CYBERNEPHONE X

Le *Gmebaphone* est le premier instrument dédié à l'interprétation/diffusion en concert des musiques électroacoustiques. Conçu par Christian Clozier, l'instrument fut inauguré le 13 juin 1973 pour les concerts du festival Synthèse à Bourges. François Bayle présent au concert inaugural présentera huit mois plus tard l'Acousmonium [25, 26 & 27].

Ma collaboration avec Christian Clozier vise à réaliser une nouvelle version open source du *Cybernéphone* utilisable sur dispositif réel et en réalité virtuelle. Il permettra de reproduire les concepts du projet initial avec ses traitements sonores et ses logiques matricielles de distribution des canaux de diffusion.

Natasha Barret avec le Virtualmonium tout comme Christine Webster et al. ont chacun démontré leur intérêt pour virtualiser un dispositif de haut-parleurs, à la fois pour l'étude de l'interprétation en concert, la recherche et la création de nouvelles œuvres également au sein d'environnement virtuel [28 & 29]. Dans la même démarche, la virtualisation du *Cybernéphone* permettra de remettre à l'étude les particularités de ce dispositif précurseur, avec notamment les filtrages opérés sur les groupes de haut-parleurs ainsi que les interfaces de commandes tactiles et visuelles dédiées à l'interprétation.

5.1. Présentation du Gmebaphone

Les concepts musicaux qui définissent l'instrument dès ses débuts proposaient une approche de mise en relief musical et sonore par un contrôle des timbres, du temps et de l'espace. Les modalités d'interprétation permettaient d'exprimer dans la salle de concert, les potentialités expressives d'œuvres électroacoustiques distribuées sur des réseaux de haut-parleurs discrétisés. En 36 ans d'activité, les versions successives du processeur audio et de son pupitre de commande ont évolué de circuits analogiques à un système entièrement numérique. Rebaptisé *Cybernéphone* à la fin des années 90, ce dernier tirait parti de la toute récente couche MSP de Max [30].

C'est un instrument constitué d'un ensemble hiérarchisé de systèmes, accès et opérateurs, doté de mémoires, séquenceurs d'automatisation, de règles et modes de jeux combinatoires fondant une rhétorique expressive de l'interprétation électroacoustique.

Depuis 1973, le dispositif connaîtra 9 développements techniques et théoriques importants au fil de 7 versions instrumentales réalisées. Certaines consoles sont visibles à la BNF et au musée de la Musique de la Philharmonie de Paris.

5.2. Fondamentaux organologiques :

Le principe initial de l'instrument est de diviser les timbres des voix stéréophoniques en 6 registres de fréquences, répartis en bande d'octaves, de l'extrême grave à l'extrême aigu. Ces bandes de fréquence sont ensuite distribuées aux haut-parleurs adéquats répartis et

orientés dans l'espace du concert. Selon son concepteur, le projet consistait en quelque sorte à dé-mixer les polyphonies sonores réalisées en studio.

Le *Cybernéphone* repose sur la superposition de plusieurs réseaux de haut-parleurs illustrés en figure 7, les uns filtrés par bande de fréquences, les autres pleine bande utilisés comme référence conventionnelle.

5.2.1. Les V

Les réseaux dénommés « V » sont filtrés. Ils opèrent une mise en perspective déterminée par le timbre des éléments sonores. La distribution spatiale et spectrale répartit les voix d'un signal stéréo sur 6 registres fréquentiels : d'extrêmes graves aux suraigus.

Le signal stéréo en entrée se subdivise en une multitude de « points virtuels » qui créent de la densité et de la présence. Cette répartition permet des couplages stéréophoniques artificiels entre chaque haut-parleur et espace de timbre. Elle permet également de créer des diagonales entre les haut-parleurs de spectre voisin. À cela s'ajoutait l'illusion psychoacoustique de distance propre aux graves et de proximité propre aux aigus.

5.2.2. Les références

Les réseaux « références » sont des couples large bande, diffusant l'intégralité du spectre sans filtrage, mais dotés de possibilités de réverbération, de délai et de transposition. Ces couples de haut-parleurs non traités permettent d'édifier une mise en espace conventionnelle, à partir d'une pièce stéréo, multipiste ou des musiques mixtes avec instrumentiste. En complément aux réseaux de haut-parleurs filtrés, les haut-parleurs pleine bande étaient disposés en position octophonique permettant des plans classiques mettant en valeur par opposition les effets de relief et profondeur générés par les V.

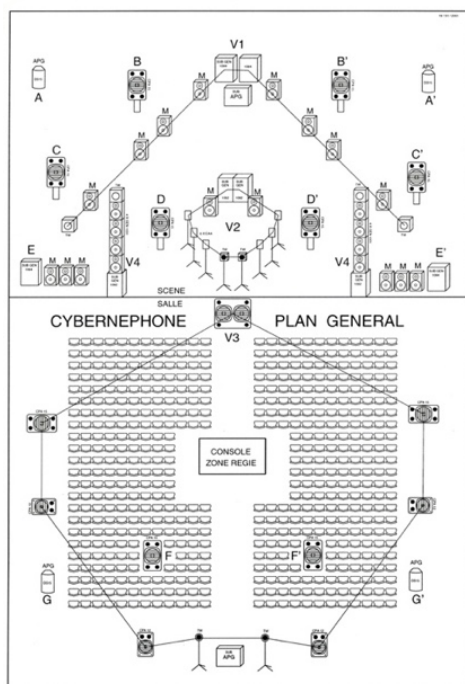


Figure 7 Dispositifs scéniques du *Cybernéphone* en 1999

5.3. Virtualisation

Ces travaux visent à réactiver des stratégies de spatialisation efficaces qui répondent encore aux problématiques esthétiques de la spatialisation pour les nouveaux supports de diffusion immersifs.

Les solutions de virtualisation offrent une alternative aux problématiques d'accès au dispositif de spatialisation sonore qu'il soit en studio ou sur scène. L'accès à ce type de dispositif n'est bien souvent possible que très peu de temps avant le concert pour une courte répétition ne permettant pas d'appréhender sereinement les capacités expressives des dispositifs de diffusion. Ces simulations sont donc utiles pour renforcer la pratique de l'interprétation électroacoustique et contribuer dans une plus large mesure aux connaissances pratiques que l'on retrouvera au-delà de son cadre initial (cinéma, théâtre, installation).

5.4. Techniques

La majorité des traitements et fonctionnalités de matricage nécessaire pour reproduire un *Cybernéphone* préexiste dans *Sound Trajectory*. Recréer cet instrument sous forme virtuelle consiste à dessiner de nouvelles interfaces fonctionnelles et ergonomiques respectant les concepts originaux du dispositif. D'autre part la navigation à 6 degrés de liberté (6dof) couplée à la modélisation de la radiation spectrale des enceintes offrira une navigation dans l'espace de la salle de concert virtualisé [31]. La sortie audio pourra être rendue vers tout type de dispositif de l'écoute binaurale au Dôme ambisonique en passant par des dispositifs non conventionnels. De plus le logiciel inclut des sorties au format standardisé : stéréo, 5.1, 7.1.4 et octophonique, permettant l'usage du logiciel non seulement comme un outil d'interprétation, mais également comme outil de mixage multi-sortie.

5.5. Perspectives

Ce projet est envisagé dans le cadre de concerts réels et dans des ateliers pédagogiques de la Cité de la Musique qui présente dans sa collection la console de diffusion de la deuxième version du *Gmebaphone* (1975).

La réalisation de ce dispositif sera finalisée au troisième trimestre 2023. Il sera utilisé dans le cadre de concerts et d'ateliers d'interprétation et de découverte de la musique électroacoustique sous une forme logicielle. Il accompagnera la documentation du code source de *Sound Trajectory* et sera distribué gratuitement sous forme logicielle.

Les singularités du *Cybernéphone* pourront ainsi être comparées au projet de virtualisation des acousmoniums existants. Ces efforts de reproduction se complètent pour offrir une approche élargie de la mise en espace des plans sonores ainsi qu'une meilleure compréhension des processus créatifs que ces solutions ont offert à la diffusion de la musique électroacoustique.

6. CAS PRATIQUE N°3 : SOLO POUR PIANO, SYNTHÈSE ET DISPOSITIF IMMERSIF

J'utilise *Sound Trajectory* sur scène pour spatialiser mes instruments et leurs traitements sonores. Ma performance s'articule autour d'un piano Yamaha CP70B, celui-ci est muni de micros piezo sous chaque corde, qui permettent d'amplifier l'instrument aussi simplement qu'une guitare électrique. Le piano est accompagné par deux synthétiseurs polyphoniques analogiques MIDI. La figure 8 présente la disposition des instruments.

Les différents modules de *Sound Trajectory* s'articulent de manière organique avec mon instrumentation électronique. De nombreux processus génératifs et automatiques sont ainsi couplés avec les fonctionnalités de ST pour assurer un mixage semi-supervisé lors d'interprétations et d'improvisations sur différents claviers et interfaces gestuelles (Contrôleur MIDI, capteur vidéo *Leapmotion*). Les traitements sonores sont spatialisés individuellement comme sources sonores à part entière.

Au sein du même environnement (Max), les processus de feedback adaptés à la spatialisation, par contrôle paramétrique ou signal audio, permettent de sublimer l'étendue des propriétés sonores des traitements. (granulation, échantillonneurs, gel, SuperVP scrub [32], multi-délais, time lag accumulator²², reverse, harmoniser, modulation, distorsion, délais, réverbération.)

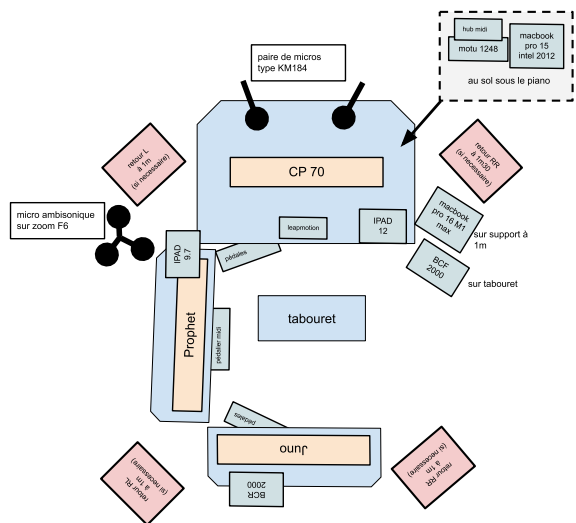


Figure 8 Dispositif instrumental

6.1. Matriçage

La réelle difficulté d'un mixage objet réside dans la capacité à gérer une multitude de sources sonores en mouvement. Lorsque celles-ci sont placées à des positions fixes, on tire tout de même partie des capacités

des moteurs de rendu audio 3d en employant des approches traditionnelles de panoramisation et de distribution multicanal. C'est ainsi qu'il est plus commode et rapide d'utiliser des bus multipistes plus ou moins standardisés qui permettront de poser dans l'espace virtuel d'un mixage objet, un espace préconçu (au format stéréo, quadri, 5.1 7.1, octophonique, 7.1.4.) Ces espaces préconçus pourront être re-interprétés en les redispasant dans l'espace. Bien que limités, ils offrent des possibilités d'éditions simples dans les DAW professionnelles comme Pro-Tools et Nuendo.

6.2. Les sources statiques

La figure 9 expose les affectations communes de canaux entre les logiciels Max, Live et Nuendo alimentant *Sound Trajectory* avec des bus octophoniques. La définition spatiale est faible, mais le recours à ce format permet de lire facilement, depuis différents environnements, du matériel sonore possédant un espace interne pré-spatialisé. Suivant la capacité des logiciels, d'autres formats peuvent être acheminés directement à *Sound Trajectory* comme un mix complet en 7.1.4. Du côté de ST, tous les outils sont disponibles pour réinterpréter ces formats de manière à pouvoir les adapter à tout autre format de diffusion. Même si ces standards ne sont pas optimaux, ils ont le grand avantage d'être disponibles dans les studios dolby Atmos avec les outils d'édition adéquats.

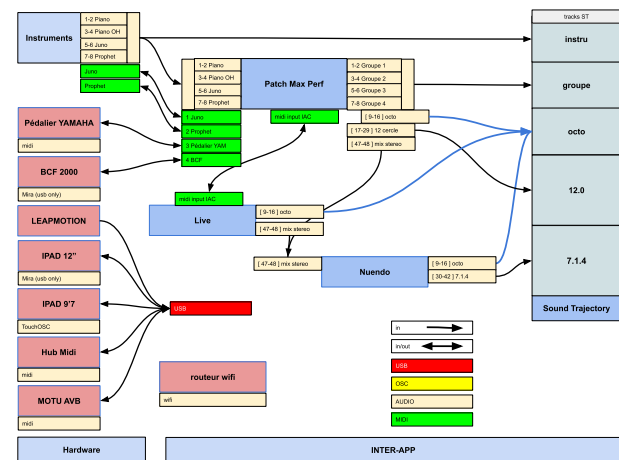


Figure 9 Routages des canaux audio, MIDI et OSC

6.3. Les Sources dynamiques

De manière générale, toutes les autres sources sont susceptibles d'être animées dans l'espace. S'il ne s'agit pas de déplacement azimutal ou d'élévation, cela pourra être la multiplication d'une même source sonore à des points différents, ce qui lui donnera de la présence, accentuera la sensation de largeur ou au contraire aplatira tout relief selon les proportions utilisées. Pour cela chaque piste de *Sound Trajectory* peut être configurée

22 Time lag accumulator : <https://musicstudios.calarts.edu/max-msp/time-lag-accumulator/>

d'une à seize sources sonores. Des doublons de pistes peuvent être créés et activés suivant la conduite du projet. La figure 10 montre un exemple typique d'entrée stéréo étendue au format quadripophonique, ainsi la sensation de présence de la source réelle sera bien plus prononcée, car démultipliée sur deux canaux supplémentaires. En inversant les canaux avant et arrière, le croisement des voix éclate très simplement une stéréo dans l'espace.

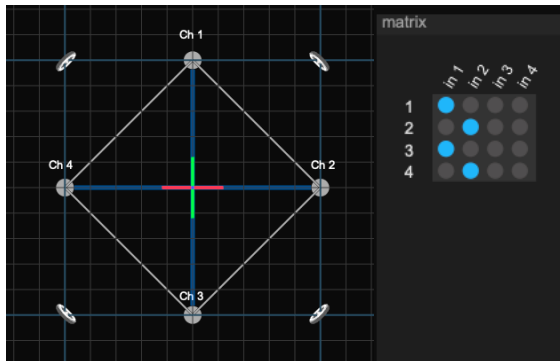


Figure 10 Extrapolation quadripophonique d'une piste à partir d'un signal stéréo

Dans cette logique je groupe mes différents effets afin de limiter le nombre de sources sonores réelles à spatialiser. La figure 11 illustre les matrices d'affectations des différents groupes d'effets. La matrice à droite détermine les affectations directes, la matrice du milieu, les interconnexions internes au patch et la matrice du bas gère les envois vers les groupes auxiliaires. Différents réseaux d'effets peuvent ainsi être associés librement à différents groupes d'effets.

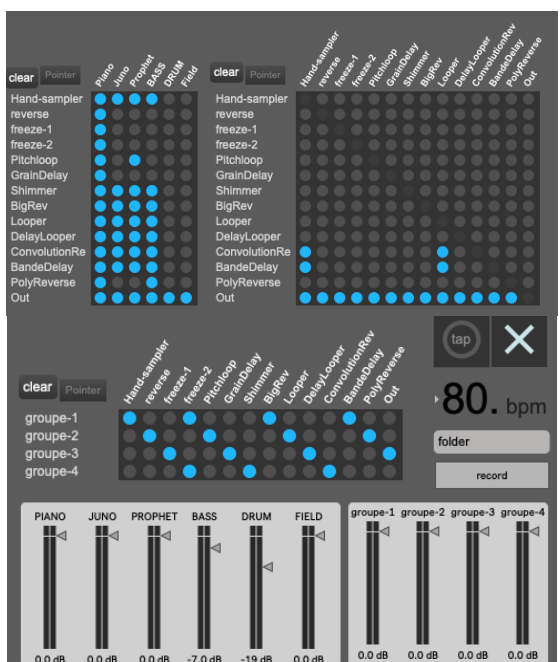


Figure 11 Matrices d'affectations des effets audio du patch Max

Dans une œuvre en direct, les dynamiques spatiales opérées par l'interprète ou le performeur font face à

plusieurs obstacles : complexité technique et temps de préparation entre autres. Il y a une tendance naturelle à user de stéréotypes par facilité : rotation systématique des sources autour du public, déplacement aléatoire extravagant ou répétition des figures spatiales n'en sont que quelques exemples. Ils enrichissent le lexique expressif de la spatialisation sonore mais les gestes d'interprétation liés à la spatialisation peuvent néanmoins être variés sans avoir recours à une virtuosité sophistiquée. La figure 12 présente par exemple des effets activés manuellement de manière momentanée, ils participent à étendre l'instrumentarium avec des typologies sonores dédiées à la spatialisation sonore.

Dès 1971, John Chowning considère comme primordial de chercher un optimum de la composition à adapter aux possibilités du concert [33]. Si bien que pour optimiser ces possibilités, il s'agit de définir un cadre spatial virtuel dans lequel les mouvements pourront être répartis de manière homogène tout en contrastant les points d'écoute.

Pour cette performance mixte, la mise en œuvre du processus de spatialisation le plus simple consiste à animer le *listener*, qui est le centre de référence virtuel du mixage. *Sound Trajectory* permet de contrôler son *listener* avec 6 degrés de liberté. C'est un outil d'expression redoutable capable de créer quelques vertiges acoustiques avec très peu de moyens. Tout le public en profite, car c'est la scène sonore complète qui est alors modulée avec la seule modification de l'emplacement du *listener*. Il faut pour cela que le processeur audio synthétise l'effet doppler avec une réverbération dynamique, modèle de John Chowning qui est resté une référence pour simuler efficacement l'illusion psychoacoustique de la cinétique sonore.

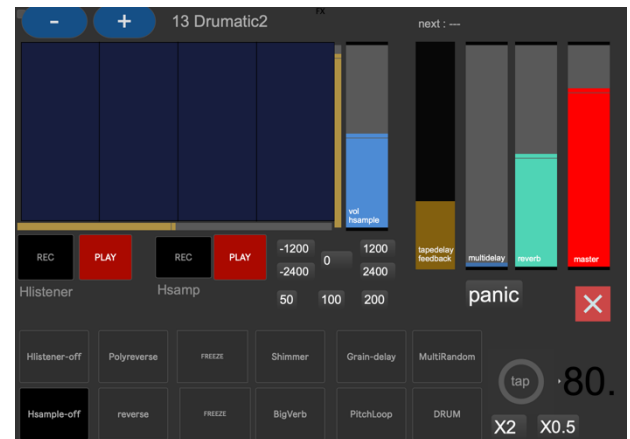


Figure 12 interface mira permettant d'activer à la volée des effets momentanés

6.4. Effet Stéréo

Une majorité de traitements sont bicanaux, le traitement est analogue aux effets spatiaux permis par la stéréo, à la différence que les deux sources seront distribuées librement dans un espace en trois dimensions (XYZ). Ainsi un simple effet délai en ping-pong,

délayant le même signal sur deux canaux avec deux temps de délai différents, créera l'illusion d'un déplacement du son le long d'un vecteur, lui-même déplaçable dans un espace tridimensionnel. Avec d'autres types d'effets, il est envisageable de créer des plans sonores élaborés qui pourront se mouvoir dans l'espace.

6.5. Effets multicanaux.

Les bus multicanaux sont utilisés pour des processus tirant directement parti de configurations immersives octophoniques à minima, parfois divisées en deux quadraphonies, l'une pour le bas et l'autre pour la hauteur. La synthèse granulaire tire particulièrement bien profit de ces configurations sans avoir à envoyer au processeur de nombreuses informations de position, car tout le processing audio est rendu dans le format défini. Des effets polyphoniques en tirent également parti, Polyreverse et Multirandom. Le premier lit à l'envers un buffer en enregistrement circulaire continu, en superposant des extraits de manière aléatoire, mais dans un registre défini (vitesse, longueur et éventuelle synchronisation à un tempo). Chaque voix de polyphonie créée est automatiquement répartie sur les canaux disponibles. Le Multirandom quand à lui, déclenche un panoramique aléatoire à la suite d'une attaque activant l'effet. Ce traitement est particulièrement efficace sur les sons de percussions.

6.6. Groupes auxiliaires

Les groupes auxiliaires peuvent reprendre les signaux de tous les traitements en les répartissant sur quatre groupes distincts. Les groupes sont stéréo et permettent de simplifier le nombre de sources à réellement gérer dynamiquement en live.

6.7. Mixage automatique

Les groupes sont ensuite transmis à *Sound Trajectory* par le biais de connexions inter-applications. Ils seront mixés de manière automatique, c'est-à-dire qu'une méthode semi-supervisée pour chaque groupe est utilisée pour coordonner leurs mouvements dans l'espace. La figure 13 présente le réseau d'interactions possibles au sein du logiciel dont les fonctions sont regroupées sur la page de ST appelée *interactive mapping*. Ces dernières s'articulent en deux classes distinctes et complémentaires, les *actions* et les *events*. Les modules génèrent un signal de contrôle à partir de l'analyse d'une entrée audio, MIDI ou OSC. Les modules *events* proposent une variété de séquences d'automation cyclique ou aléatoire.

La palette d'interaction permet d'agir sur tous les paramètres ayant une incidence sur le placement dans l'espace des sources sonores. Les sorties de contrôle en dehors du logiciel sont envisageables au-delà du domaine sonore à destination d'environnements, MIDI, OSC, DMX et Artnet. Il est ainsi possible d'envoyer des informations de contrôle d'éclairage, de synchro vidéo, et

tous les paramètres courants du logiciel : position des sources, du *listener* ou de groupes de sources, etc...

Le nombre de modules activables est illimité. Plusieurs *events* peuvent être contrôlés par une seule *action*. Imaginons une trajectoire basique sous forme de cercle, un même slider contrôlé depuis une interface peut modifier le diamètre du cercle, sa rotation et son élévation. D'un seul geste, la source se déplacera comme si elle était sur une spirale, mouvement par ailleurs adéquat pour percevoir plus clairement l'élévation du son dans un dispositif de diffusion sonore en 3d.

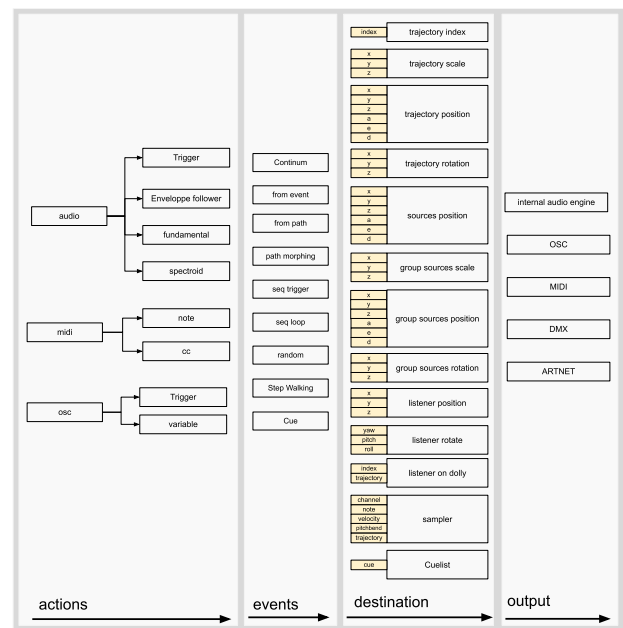


Figure 13 mapping et réseaux d'interactions

De manière plus basique des séquenceurs peuvent faire moduler la position des sources sur les trajectoires à des vitesses variables. De manière plus élaborée, la vitesse de lecture des séquenceurs peut être modulée par une action, elle-même contrôlée par exemple par la brillance d'un signal audio.

Certains modules d'événements aléatoires par « random walk » peuvent créer l'illusion de mouvements browniens plus ou moins naturels.

Tout un échafaudage d'articulations et d'interactions est ainsi possible à la manière de la synthèse modulaire.

7. CONCLUSION

Avec les cas présentés, l'auteur espère avoir mis en lumière quelques caractéristiques propres à l'outil *Sound Trajectory*. Par son aspect semi-modulaire, ce logiciel propose une solution simple d'utilisation, adaptée aux créateurs sonores cherchant à explorer les possibilités d'une écriture interactive de la spatialisation sonore.

Les workflows et modes opératoires de production sont toujours relativement complexes à mettre en œuvre et à maintenir dans le temps. Les perspectives de développement, de documentation et de libération des sources de *Sound Trajectory* visent à assurer une pérennité des concepts et logiques du programme. Le

programme d'origine étant développé en Max, toutes les sources sont disponibles sous forme d'abstractions pouvant être utilisées indépendamment de la structure d'origine du logiciel [34].

Les contraintes inhérentes au développement du standalone pourront être levées avec une utilisation libre au sein de l'environnement Max. Ainsi l'articulation des abstractions et modules du logiciel d'origine permettra la conception de dispositifs originaux et plus facilement adaptables aux besoins spécifiques des projet artistiques.

Un portage sur le modèle des Beap Tools facilitera le prototypage rapide de processus de synthèse et traitements dédiés à la spatialisation sonore.

La réalisation d'abstractions de haut niveau avec une interface ludique profite à une sensibilisation sensorielle à l'écriture et l'interprétation du son dans l'espace.

Ultérieurement, un portage partiel sur Pure Data, permettra une intégration des fonctionnalités au sein de projets intégrant la lib.pd ou utilisant des ordinateurs légers de type Raspberry pour la réalisation de dispositifs singuliers open source.

L'ouverture publique des sources du logiciel *Sound Trajectory* sous forme de package pour max sera annoncée en juin 2023. Les projets discutés dans cet article seront proposés en démonstrations pour accompagner la publication des sources.

8. RÉFÉRENCES

1. Davis, K. « Motive and Spatialization in Thomas Tallis' Spem in Alium », South Central Graduate Music Consortium (SCGMC), University of Virginia, Charlottesville, VA, 2016
2. Barrett, N. « Spatial music composition », in 3d Audio (1st ed.). Routledge., p175 <https://doi.org/10.4324/9780429491214>
3. Quiroz, D. « Gestural control for 3d audio », in 3d Audio (1st ed.). Routledge. p64 <https://doi.org/10.4324/9780429491214>
4. Normandeau, R. « Timbre Spatialization: The medium is the space » Organized Sound, 14(3), 2009, p. 277-285 doi:10.1017/S1355771809990094
5. Pottier, L. « Dynamical spatialization of sound. HoloPhon : a graphic and algorithmic editor for Sigma1 », Proc. of the International Conference on Digital Audio Effects (DAFx), Barcelona, Spain, 1998
6. Normandeau, R. et al. « Spat-GRIS/ServerGRIS, Creative tools for 2D and 3d sound spatialization », Proceedings of the 2018 international computer music conference. Daegu, Korea : The International Computer Music Association, 2018, p. 291-297.
7. Loret, E. « Le Synesthesium », Art Press Novembre 2015
8. Mercier, M. et al. « Synthèse d'interactions multimodales parcimonieuses pour l'écriture de l'oeuvre Iquisme et l'analyse de ses percepts » JIM 2015, Montréal, Canada. hal-03104605
9. Jacquemin, G. et al. « Partitions retroactives avec Iannix » JIM 2014, Bourges, France. hal-03104631
10. Carpentier, T. et al « Twenty Years of Ircam Spat: Looking Back, Looking Forward » in Proc. of the 41st International Computer Music Conference, Denton, TX, USA, Sept. 2015, pp. 270 – 277.
11. Carpentier, T. « Tosca : An OSC Communication Plugin for Object-Oriented Spatialization Authoring », Proc. of the 41st International Computer Music Conference (ICMC), pp. 368 – 371, Denton, TX, USA, Sept.2015.
12. Schacher, J. and Kocher K « Ambisonics Spatialization Tools for Max/MSP » In Proceedings of the International Computer Music Conference. New Orleans, LA, June 2006.
13. Wright, M. et Freed, A. « Open Sound Control: A New Protocol for Communicating with Sound Synthesizers » Proceedings of the International Computer Music Conference 1997, Thessaloniki, Hellas, pp. 101-104.
14. Carpentier, T. « Panoramix: 3d mixing and post-production workstation » in Proc. 42nd International Computer Music Conference (ICMC), Utrecht, Netherlands, Sept 2016
15. Harker, A. and Tremblay, P. « The HISSTools Impulse Response Toolbox: Convolution for the Masses » ICMC 2012
16. Lossius, T. et al. , « Dbap distance-based amplitude panning » in Proc. of the International Computer Music Conference (ICMC), Montreal, 2009.
17. Sèdes, A., Guillot, P., Paris, E., « The HOA library, review and prospects », Proceedings ICMC/SMC, Athènes, Grèce, 2014
18. Kendall, G. S. « The Decorrelation of Audio Signals and Its Impact on Spatial Imagery », Computer MusicJournal, MIT Press, USA, 1995.
19. Vaggione, H. « Décorrélation microtemporelle, morphologies et figurations spatiales », Actes des Journées d'Informatique Musicale, Marseille, France, 2002.
20. Sèdes, A. « Approche musicale de la décorrélation microtemporelle dans la bibliothèque HOA », JIM 2015, Montréal, Canada.

21. Goutmann, P. « Traitement spatial du son par décorrélation des signaux en ambisonie d'ordre élevé » Journées d'Informatique Musicale 2021, AFIM, Jul 2021, Visioconférences, France. hal-03313616
22. Pulkki, V. « Virtual Sound Source Positioning Using Vector Base Amplitude Panning » Journal of the Audio Engineering Society, vol. 45, no. 6, pp. 456 – 466, June 1997.
23. Schnell, S. et al. « MuBu & Friends – Assembling Tools for Content Based Real-Time Interactive Audio Processing in Max/MSP » In International Computer Music Conference (ICMC). Montreal, August 2009.
24. Françoise, J. « Motion-Sound Mapping by Demonstration » Thèse de doctorat, UPM 2015. English. p148 et 204, NNT : tel-01161965
25. Clozier, C. « Gmebaphone, Prédicat » ed Misame, p27, 2022, <https://misame.org/anthologie-gmebimeb/>
26. Clozier, C. « Sur le Gmebaphone/Cybernéphone » ed Misame, <https://misame.org/anthologie-gmeb-imeb/>
27. Clozier, C. « The Gmebaphone Concept and the Cybernéphone Instrument », Computer Music Journal, vol. 25, no 4, 2001, p. 81-90.
28. Barrett, N. and Jensenius A. « The 'Virtualmonium': an instrument for classical sound diffusion over a virtual loudspeaker orchestra » Brisbane, Australia, 1 juin 2016.
29. Webster, C. and al. « Vers un acousmonium en immersion VR en ambisonie 3d et binaural » Journées d'informatique musicale, Oct 2020, Strasbourg, France. hal-02977660
30. Puckette, M. « Combining Event and Signal Processing in the MAX Graphical Programming Environment » Computer Music Journal. (1991a).
31. Barrett, N, « Composition and the 'Live Electroacoustic': sound, time, control and ensemble » <https://www.natashabarrett.org/live.html>
32. Roebel, A. « New approach to transient processing in the phase vocoder » 6th International Conference on Digital Audio Effects (DAFx). London, p. 344-349, 2003
33. Chowning, J. « The Simulation of Sound Moving Sources » Journal of the Acoustical Society of America 1971 : 2-6
34. Mercier, M. « code source de Sound Trajectory » dépôt Github : <https://github.com/maxence-mercier/Sound-Trajectory>

ÉMERGENCES : DE LA POÏÉSIS AU PROTOTYPE

Paul Goutmann

CICM MUSIDANSE EA1572

EUR ArTeC

paul@goutmann.net

RÉSUMÉ

Cet article s'appuie sur l'aspect poétique d'*Émergences*, pièce électroacoustique en ambisonie 3D composée dans le cadre d'une recherche sur les interfaces pour le traitement spatial du son afin de souligner les implications technologiques et musicales de cette recherche-création. L'auteur fait ainsi état d'une pratique de compositeur-chercheur à partir de la troisième itération d'une boucle mêlant développements logiciels et compositions musicales.

1. INTRODUCTION

L'espace est une dimension fondamentale de la pensée musicale contemporaine notamment dans la pratique des musiques électroacoustique, mixte et électronique et ne peut être réduite à la spatialisation de sources ponctuelles [25][1][29][9]. Les modèles de spatialisation du son partagés par les communautés scientifiques et musicales sont dominés par des représentations euclidiennes de l'espace prenant la source ponctuelle comme objet fondamental [7]. Cette approche se retrouve par exemple dans le répertoire avec une pièce comme *Répons* de Pierre Boulez. À l'opposé, le *Prometeo* de Luigi Nono cherche dans la pluralité des voix, la profusion et l'indétermination des positions : le champ diffus [26][11]. Depuis les années quatre-vingt-dix, plusieurs logiciels implémentant les modèles de mise en espace du son rendent accessibles des opérations sur la spatialisation dont la modulation du positionnement des sources et la modélisation acoustique de salle sont les plus communément utilisées [6][2][20][10][18][19]¹. Toutefois, de nombreux travaux ont développé des approches alternatives aux techniques de spatialisation orientées sources ponctuelles [16][17][32][22][4]. Les traitements spatiaux par granulation ou par décorrélation microtemporelle héritant des approches de Iannis Xenakis et d'Horacio Vaggione en sont des exemples. En lien avec de nombreuses années de recherche et de création sur la spatialité du son au CICM, *HOALibrary* développée dans le cadre des projets du LabEx Arts-H2H a contribué au développement de ces approches et à leur prise en main par les musicien.ne.s [24].

Depuis 2020, nous menons des recherches dans le cadre de notre doctorat sur la thématique des interfaces

pour le traitement spatial du son avec une méthode de recherche-création centrée sur une boucle itérative de productions musicales et logicielles. Dans cet article, nous prenons le point de vue du compositeur électroacoustique et faisons état du pan « création » de cette recherche².

En partant des travaux du projet *HOALibrary*, d'abord éprouvés sur le plan sensible, nous avons esquissé des prototypes d'interfaces graphiques constituant notre terrain expérimental de thèse [12]. Mises à l'épreuve par notre pratique de composition, nos interfaces graphiques - jusqu'alors pensées comme objet d'étude central - se sont retrouvées projetées en périphérie de notre recherche. Au cours du processus, nous avons d'abord rencontré des difficultés pour les utiliser musicalement.

Afin de dépasser cette difficulté, nous avons d'une part changé d'échelle et considéré les interactions avec la spatialité du son au sens large, notamment au niveau du code DSP [13] ce qui nous a conduit à élargir notre définition des interfaces. Nous considérons maintenant les interfaces comme l'ensemble des points de contact entre deux systèmes complexes qui régissent (conditionnent, structurent) leurs échanges. Dans notre cas, l'interface entre la spatialité du son et le compositeur est feuilletée, comprenant notamment la partie GUI, mais aussi la partie DSP et la partie *hardware* pour interagir avec le dispositif.

D'autre part, nous avons adapté nos manières de les utiliser pour en découvrir le potentiel. Ce n'est qu'à partir du moment où nous avons cherché à les faire devenir des *instruments* que nous avons pu composer avec. En somme, cette boucle expérimentale nous a conduit à façonner notre *organon* : notre *espace composable*³.

Dans cet article, nous nous appuyons sur l'aspect poétique d'*Émergences*⁴, la troisième itération musicale de notre boucle expérimentale, pour souligner les implications musicales et technologiques de notre recherche-création. Ainsi, nous ouvrons le moteur de notre atelier, « comprendre pour faire et faire pour comprendre »⁵,

2. Nous avons écrit un premier article concernant les prototypes d'interface pour le traitement spatial aux JIM 2021, puis un article sur l'implémentation en Faust de deux traitements spatiaux dans le cadre de la conférence SMC 2022.

3. Le terme *organon* vient du grec ancien, signifiant « instrument de travail », « de musique » et « organe du corps ». L'espace composable est un instrument, c'est l'ensemble des variables sensibles que nos espaces de travail fournissent [27].

4. Version binaural de la pièce : <https://hal.science/hal-04084391>

5. Formule utilisée dans le séminaire « Composition et Re-

1. Ainsi que les GRM Tools, les plug-in IEM et de nombreux autres.

prenons un de nos *prototypes*⁶ afin de montrer ce que cette double posture a déplacé de (et éclairé sur) notre objet d'étude. Dans la première partie, nous présentons le dispositif technique, nous exposons dans un premier temps notre écosystème logiciel, puis les trois éléments principaux du moteur audio utilisé dans (et construit pour) cette pièce. Dans la deuxième partie, nous présentons les idées musicales propres à l'écriture de la spatialité dans cette pièce, nous exposons les trois morphologies spatiales travaillées : texture spatiale, enveloppe d'espace et strate.

2. ASPECT POÏÉTIQUE

Dans cette section, nous présentons les éléments techniques de notre environnement de composition à différentes échelles : le matériau sonore initial (fichiers audio), l'écosystème logiciel, les manières de faire du montage ainsi que les traitements audionumériques utilisés.

2.1. Sources, écosystème logiciel et montage

Notre matériau initial est composé de trois fichiers audio mono :

1. une boucle réalisée à partir d'un son impulsionnel répété ;
2. une séquence de douze secondes composée comportant des échantillons de voix ralentis et des sons de synthèse ;
3. un son de vingt-et-une secondes issu d'un synthétiseur *Pro 2* du constructeur *Dave Smith*.

Ces trois matières sonores sont travaillées dans l'ordre énoncé sans se superposer, à l'exception de la dernière partie de la pièce dans laquelle les matières 1 et 3 se superposent.

Nous utilisons principalement une banque de filtres résonants sur les fichiers mono, du *time stretching*, la ring modulation et la décorrélation dans le domaine ambisonique. Nous décrirons plus largement ces deux traitements dans les sous-parties 2.2 et 2.3

Le montage et le mixage sont faits dans *Reaper* dans le domaine ambisonique⁷. Les parties *DSP* et décodage ambisonique sont faites avec *Max*. Les flux audios transitent entre les logiciels par des bus internes grâce à la fonction *loopback* de l'interface audio réinjectant les sorties vers les entrées de l'interface audio (voir figure 1).

Pour des raisons compositionnelles précisées dans la partie 2.2 et 3.3, nous considérons l'espace global avec

cherche »d' Anne Sèdes en 2020 référant à [21].

6. La notion de prototype fait ici référence à la fixation (composée) d'un travail en cours, une manière de rendre sensible. Pour illustrer cette notion, nous pouvons reprendre les mots d'E.During : « 'Prototype' désigne moins, en effet, une catégorie spéciale d'objets qu'un nouveau régime de production. [...] La question n'est plus alors d'éviter à tout prix de fixer quelque chose [...]. Au contraire, tout le problème est de savoir 'comment s'arrêter'. Comment donner à un projet la consistance et la solidité qui le rendra lisible ? » [8].

7. En l'occurrence dans le domaine des harmoniques circulaires (voir sous-section 2.2).

une superposition de trois plans 2D respectivement d'ambisonie d'ordres 3, 2 et 1. Les fichiers audios sur lesquels nous faisons le montage comportent 15 canaux (7-5-3). Une sommation des différents fichiers est envoyée aux décodeurs ambisoniques pour avoir un *playback*.

Pour le rendu de la pièce, nous décodons chaque plan pour un système de restitution supérieur ou égal à une situation dite régulière, puis encodons à l'ordre 3 en 3D chaque signal associé à un haut-parleur comme une nouvelle source puis stockons les 16 canaux ambisoniques (dans l'ordre suivant l'ACN et avec la normalisation SN3D). Ceci afin de pouvoir aisément décoder la pièce pour des systèmes de diffusion 3D différents de celui sur lequel nous avons composé⁸.

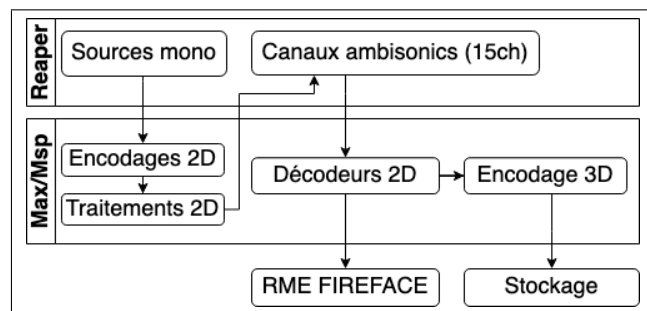


Figure 1. Dispositif logiciel utilisé pour la composition d'*Émergences*

2.2. Layer Based Ambisonics Processes - LBAmP

L'expression *Layer Based Ambisonics Processes* (LBAmP) fait ici référence à une superposition de plans de traitement en ambisonie 2D. Dans cette partie, nous expliquons l'origine de cette approche de l'espace global comme superposition de strates et sa mise en œuvre dans *Émergences*. Ni l'ambisonie, ni les traitements en ambisonie ne sont l'objet principal de cet article, toutefois nous donnons un aperçu très synthétique de ces modèles et invitons les lecteurs curieux à se référer à [14][33].

2.2.1. Format-B et HOA

Le format-B proposé par M.Gerzon et l'ambisonie d'ordre élevé par J.Daniel sont un ensemble de techniques pour l'enregistrement, la synthèse, le traitement et la reproduction de champs sonores. Ces techniques sont basées sur trois étapes : l'encodage, l'optimisation et le décodage. L'encodage consiste en la description de la fonction sphérique (ou circulaire) du champ sonore capté (ou qu'il eût été possible de capté dans le cas d'une synthèse) avec une somme pondérée de fonctions de références appelées harmoniques sphériques (ou circulaires). Le nombre d'harmoniques utilisées pour la décomposition de la fonction associée au champ sonore dépend de l'ordre ambisonique. Plus le nombre d'harmoniques est grand, plus la précision

8. La pièce a été composée au studio d'essai son de la MSH Paris Nord.

de la description sera grande. En 2D, l'encodage utilise $2(N + 1)$ harmoniques et en 3D $(N + 1)^2$ harmoniques, où N est l'ordre. L'étape de décodage consiste à projeter la fonction sphérique (ou circulaire) correspondant au champ sonore sur le dispositif de haut-parleur. Autrement dit, calculer le jeu de gain à appliquer aux haut-parleurs en fonction de la fonction sphérique et du système de diffusion

2.2.2. Traitement en ambisonie

HOALibrary a rendu accessibles les opérations sur les signaux ambisoniques et a introduit des traitements bien connus des compositeurs électroacoustiques dans le modèle ambisonique. Appliquer des traitements en ambisonie dans cet héritage consiste à appliquer les traitements audio numériques directement sur les harmoniques spatiales, c'est-à-dire sur les signaux entre l'étape d'encodage et de décodage. Depuis l'obsolescence de la version *Pd* de la bibliothèque et le portage des traitements en langage Faust, cette approche se retrouve dans *abc.lib* et *hoa.lib* [5][13].

2.2.3. LBAmP

Avant d'avoir implémenté les versions 3D des traitements ambisoniques, nous approchions les traitements 3D comme une superposition de traitements 2D prenant autant de plans que le système de diffusion comporte de couronnes de haut-parleurs⁹.

Dans *Émergences*, nous utilisons cette approche avec la décorrélation et la ring modulation en ambisonie¹⁰. Pour chacun de ces deux traitements, le processus est le même : autant d'encodeurs 2D, de chaînes de traitements puis de décodeurs 2D que de couronnes de haut-parleurs. La figure 2 représente un exemple de chaîne de traitements pour trois couronnes de haut-parleurs superposées. Chaque traitement de la chaîne est piloté par des variables individuelles et par des métaparamètres s'appliquant à tous les traitements avec des coefficients eux-mêmes variables.

Par exemple, dans le cas de la décorrélation telle que nous la proposons dans *hoa.lib*, il y a 4 variables : *fa* le facteur de diffusion, *del* la fenêtre de décorrélation, *tf* la fonction de distribution, *fd* le taux de réinjection. Dans le cas d'une superposition de N strates de décorrélation, nous aurons $fa_1, fa_n, \dots, fa_N, del_1, del_n, \dots, del_N, etc.$

9. Lors d'une résidence de composition au CICM, la compositrice Giulia Francavilla a utilisé *abc.lib* qui n'a pas de décodeur 3D. Pour composer l'espace 3D de sa pièce, la compositrice a mixé différents plans 2D. Lors de cette résidence, nous étions en train de travailler sur l'ajout d'une couche de VBAP au décodeur 2D de *hoa.lib*. C'est cette conjonction qui nous a amené à travailler sur cette approche de la 3D en superposition de plans 2D. Par ailleurs, pour le lecteur souhaitant compiler des décodeurs 3D en langage Faust, voir [15].

10. Dans des versions légèrement modifiées par rapport à celles disponibles en ligne. Les modifications sont principalement pour une meilleure compatibilité avec nos interfaces et pour laisser plus de variables accessibles afin d'expérimenter avant une nouvelle version du traitement.

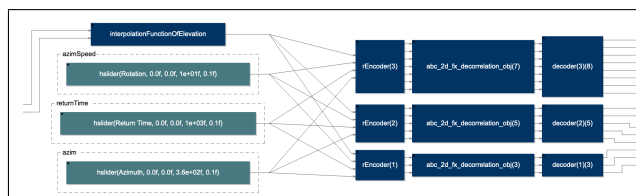


Figure 2. Exemple de chaîne de traitement pour trois couronnes de haut-parleur

Si nous nous concentrons sur les facteurs de décorrélation *fa*, nous pouvons insérer des coefficients et piloter le n^{eme} facteur avec un métafacteur (fa_{meta}) en faisant le produit :

$$fa_n \cdot c_n \cdot fa_{meta} \quad (1)$$

L'idée est de faire des couches de traitement pouvant être manipulées indépendamment. Pour passer d'un plan à l'autre, nous utilisons une portion de sinus pour l'interpolation en entrée. L'envoi du signal de la source vers les trois encodeurs est pondéré en fonction de l'élévation souhaitée (voir figure 3).

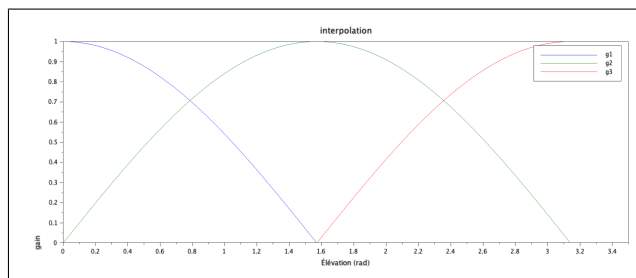


Figure 3. Facteur de gain par encodeur 2D en fonction de l'élévation

Dans un deuxième temps, nous avons testé d'associer des traitements à des plans de haut-parleurs qui ne respecte pas les couronnes. Nous avons considéré des plans ayant une certaine « courbure ». Par exemple, en sélectionnant les haut-parleurs aux numéros pairs pour un premier plan et impairs pour un deuxième plan, une rotation de source sur l'un des deux plans aura un rendu qui ne sera pas une trajectoire linéaire. Cette approche, tout en étant contraignante pour une pratique de la spatialisation, a l'avantage de nous inviter à imaginer des systèmes de figurations spatiales (trajectoires, strates de traitements, etc.) pouvant être intégrés dans un système de composition.

2.3. Décorrélation de décorrélation

La décorrélation micro-temporelle appliquée à la composition réfère à des techniques de décalages micro-temporelles (ou application de filtres all-pass). D'abord présentée par Kendall [16] pour des traitements globaux de type réverbération, largement saisie par Horacio Vaggione [30][31] puis Anne Sèdes [23], elle fut introduite dans le modèle ambisonique dans *HOALibrary* puis finalement dans *abc.lib* et *hoa.lib* en Faust. La décorrélation

en ambisonie n'est pas l'objet de cet article, le lecteur intéressé peut se référer à [13] pour plus de précisions sur les choix d'implémentation.

La décorrélation dans le domaine ambisonique consiste en l'insertion de lignes à retards dans les signaux associés aux harmoniques circulaires ou sphériques tout en pilotant le temps de retard de chaque harmonique par deux paramètres : la fenêtre de décorrélation et le facteur de diffusion. Si le facteur est nul, aucun retard n'est appliqué aux harmoniques. Dans le cas contraire, un retard variable est appliqué à chaque harmonique en fonction de la taille de la fenêtre de décorrélation, du facteur, d'un coefficient lié au numéro de l'harmonique et un coefficient lié au type de distribution parmi vingt-deux types de distributions du retard (linéaire, carré, racine carrée, portion de sinus, etc.).

Afin d'enrichir le résultat perceptif du traitement, nous avons utilisé différents modules de décorrélations en parallèles pour les mêmes signaux ambisoniques d'entrées. Cette approche nous permet d'attribuer à chaque module de décorrélation un paramétrage particulier, et ainsi décorréler les modules de décorrélation. Par exemple, un encodeur ambisonique d'ordre 3 en 2D envoie les 7 signaux associés aux harmoniques circulaires vers n modules de décorrélation ayant tous 7 entrées et 7 sorties. Maintenant, imaginons que ces n modules aient un facteur de diffusion au maximum, une même fenêtre de décorrélation mais des types de distribution de retards différents. Une fois les $7n$ signaux fusionnés et envoyés vers un seul décodeur 2D, la multiplicité des lignes à retards engendre une complexification du résultat largement perceptible.

À la suite de ces essais, nous avons implémenté une approche de métadécorrélation, dans laquelle les modules de décorrélation se décorrélent entre eux. Un étage de décorrélation contrôle le type de distribution des autres modules parmi les 22 types de distributions (4). En utilisant l'objet *preset* et *patternstorage* ainsi qu'un script dans *jsui*, nous pouvons faire des interpolations entre chaque type de distribution pour chacun des 7 modules de décorrélation. Dans la pièce, nous avons choisi de façon arbitraire d'utiliser sept modules de décorrélation pour la couronne au niveau des oreilles, cinq modules pour la couronne centrale et trois modules pour la couronne en haut du dôme.

2.4. Gain le domaine des ondes planes

Le troisième traitement utilisé dans *Émergences* est une approche originale du contrôle de gain dans le domaine des ondes planes. Dans la version *Max* de *HOA-Library*, une série d'objets et de patches permettent de faire des opérations dans le domaine des ondes planes (signaux associés à des haut-parleurs virtuels ou réels) notamment le patch « *hoa.fx.gain* » contrôlable avec l'interface « *hoa.space* ». Ce traitement lorsqu'il permet de contrôler le gain dans la direction de chaque haut-parleur (virtuel).

L'objet *hoa.space* prend la liste des amplitudes des haut-parleurs virtuels en entrée. Pour contrôler cet objet avec un plus haut niveau, nous avons écrit un script (*directivite.js*) permettant de faire des distributions d'ampli-

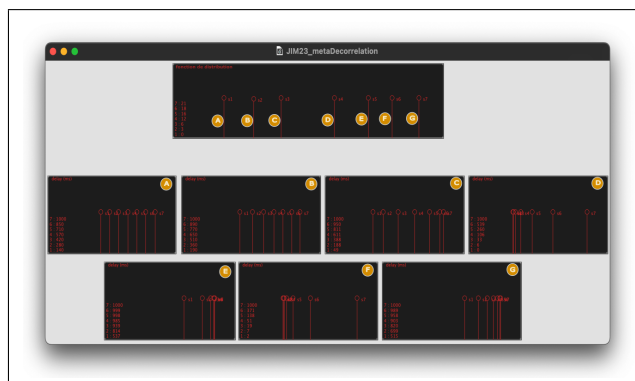


Figure 4. Illustration du contrôle du type de distribution de 7 instances de décorrélation par un étage de décorrélation avec l'interface *ddd.js*

tudes inspirées des diagrammes de directivité des microphones (voir figure 5). Le script prend en entrée un angle et un facteur de directivité. Plus le facteur de directivité est élevé, plus la zone écoutée sera étroite, à l'inverse plus le facteur sera grand, plus la zone écoutée sera large. Cette approche nous permet de contrôler le rendu global d'un plan de traitement avec simplement deux variables.

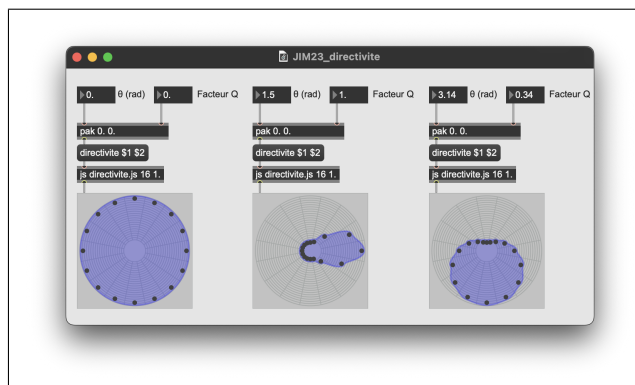


Figure 5. Illustration du contrôle de l'objet *hoa.map* avec le script *directivite.js*

3. MORPHOLOGIES SPATIALES

Dans cette partie, nous exposons les morphologies¹¹ composées dans *Émergences*. Nous nous concentrons ici sur les opérations, manipulations et objectifs musicaux liés à ce que nous nommons *morphologies spatiales*. Dans

11. La notion de morphologie traverse de façon diverse et polysémique la pensée musicale moderne et contemporaine (Schaeffer, Xenakis, Smalley, Bayle, Vaggione) notamment dans les musiques électroacoustiques. Nous employons ici le terme dans le sens de H.Vaggione. Cette notion feuilletée est présente dans la grande majorité de son œuvre théorique et n'est pas l'objet de cet article. Retenons ici que les morphologies sont « les termes qui émergent de faisceaux de relations qui sont à leur tour confrontés à des nouveaux faisceaux de relations déterminant des nouveaux termes et ainsi de suite » [28]. Il s'agit pour nous de pointer vers des entités sonores composables, multi-échelles, émergentes de l'interaction avec d'autres entités.

cette pièce, le paradigme de la source ponctuelle n'est employé que de façon anecdotique. Nous ne considérons pas les objets dans l'espace, mais l'espace lui-même, nous faisons des opérations nativement spatiales [4]. Autrement dit, nous ne faisons pas des opérations sur des morphologies sonores que nous projetons dans l'espace, nous faisons des opérations sur l'espace lui-même : nous travaillons avec et sur des morphologies spatiales qui se déclinent en trois types (textures, enveloppes et strates).

La pièce peut être scindée en trois parties : la première allant jusqu'à 3'50", dans laquelle nous travaillons principalement les textures spatiales et ce que nous appelons les enveloppes spatiales, la deuxième allant jusqu'à 5'10" au cours de laquelle nous travaillons la décorrélation directement sur les signaux envoyés aux haut-parleurs ainsi que du *time stretching*, et la troisième partie dans laquelle nous travaillons la stratification et le traitement par ring modulation.

3.1. Textures spatiales / texturer l'espace

Ce que nous appelons ici « texture spatiale » pointe vers la morphologie spatiale que nous travaillons à partir de la décorrélation de décorrélation en ambisonie (voir 2.3) et la superposition de multiples prises de ces mêmes textures. Nous faisons cette analogie avec la matière, car elles semblent partager certains attributs comme l'agencement d'unités à différentes échelles pouvant être homogène ou hétérogène. Ce que nous cherchons dans le travail de ces textures est de donner à l'auditeur des sensations de variations d'ordre spatial, sans donner la possibilité de pointer aisément vers l'origine de ces variations. Les attributs sonores que nous travaillons sont : (1) la densité, (2) la distribution spectrale, (3) la profondeur, (4) l'agitation.

1. La densité est principalement contrôlée par le nombre de prises que nous superposons, le nombre de modules de décorrélation utilisés en parallèle ainsi que leur taux de décorrélation les uns avec les autres ;
2. La distribution spectrale est liée à la bande de fréquence de chacune des prises que nous superposons et le comportement des filtres en amont et en aval des traitements ;
3. Nous contrôlons la profondeur par la composition de plans sonores - eux-mêmes émergents des assemblages de densité et distribution spectrale - ainsi qu'à la superposition de prises avec des opérations à des échelles temporelles différentes ;
4. L'agitation est construite là aussi de façon empirique comme la quantité (et la qualité) des variations par unité de temps.

Une limitation de cette approche est l'absence de variable sur des zones précises de l'espace. Cette limitation nous a conduit à utiliser le gain dans le domaine des ondes planes (voir 2.4).

3.2. Enveloppe d'espace

La deuxième morphologie spatiale que nous travaillons dans *Émergences* est ce que nous décrivons comme des *enveloppes d'espace* (de 2' à 3'50"). Les opérations permettant de les faire émerger consistent à projeter le champ sonore - dans notre cas des textures spatiales - dans le domaine des ondes planes, et écrire temporellement le comportement des gains dans chaque direction. Pour cela, nous utilisons le processus que nous avons décrit dans la sous-section 2.4. Nous utilisons le terme enveloppe pour plusieurs raisons. D'une part, nous faisons référence au critère utilisé pour l'évaluation de la qualité d'une acoustique de salle : la sensation d'enveloppement (*Listener Envelopment LEV*) [3]. Par là, nous cherchons à pointer vers une sensation d'enveloppement variable par le son dont les deux extrêmes pour nous seraient la source ponctuelle et le champ diffus. Nous faisons aussi référence aux enveloppes d'amplitudes, car nous utilisons une interface graphique de type *piecewise linear function* pour contrôler l'objet *hoa.space* en passant par le script *directivite.js* (voir figure 6).

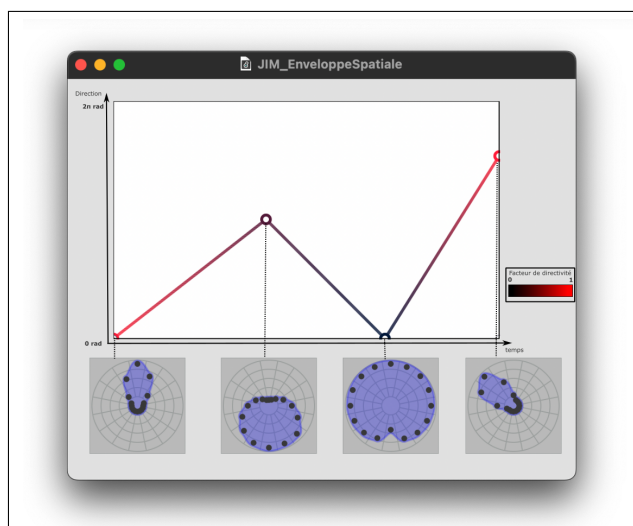


Figure 6. Illustration de l'écriture d'enveloppe spatiale avec *2plf.js*, *directivite.js* et *hoa.map*

Nous cherchons par ailleurs à composer des morphologies spatiales dont le feuilletage dépasse le seuil d'intelligibilité des éléments qui la constituent tout en laissant à l'auditeur la sensation que quelque chose se passe - sans perdre l'attention. Pour cela, nous superposons plusieurs modules d'enveloppement en jouant sur l'écriture de leur comportement temporel. Nous démultiplions les textures, écrivons des enveloppes spatiales pour chacune d'elles, en les mettant en boucle sur des temporalités différentes. Nous cherchons à garder un contrôle sur la globalité, tout en brouillant les pistes de nos actions. En somme, nous cherchons des modalités d'écriture d'un contrepoint spatial.

3.3. Stratification

L'idée de considérer l'hémisphère de haut-parleurs comme un tout ou comme un assemblage de plans (plus ou moins courbes), nous a semblé avoir un potentiel compositionnel à investir. Nous explorons avec cette morphologie les possibilités sensibles par la superposition de plans de traitement aux paramétrages différents. L'idée principale que nous avons explorée est la possibilité de jouer avec les sensations d'élévation de plans entiers. Nous ne cherchons pas à donner la sensation d'une trajectoire ascendante d'une morphologie, mais la transformation de la morphologie spatiale au cours de son élévation. Pour cela, nous avons notamment essayé de paramétrer les trois strates de traitement avec de légères différences puis de faire passer progressivement la boucle de clics d'une strate à l'autre. D'abord de la première strate à la dernière en 1 seconde, puis 10 secondes et enfin 60 secondes. Toutefois, nous n'avons pas réussi à trouver un résultat satisfaisant.

Dans cette pièce, nous avons testé cette idée avec trois traitements :

1. la décorrélation avec différents paramétrages de feedback par strate, ainsi que le nombre de lignes à retard par strate (ce qui a tendance à impacter la sensation de densité) ;
2. la ring modulation avec des fréquences modulantes différentes par strate ;
3. vitesse de rotation différentes par strate.

Ces tentatives ont ouvert un niveau opératoire sur lequel nous commençons à peine à nous aventurer.

4. CONCLUSION

Dans cet article, nous nous appuyons sur l'aspect poétique d'*Émergences* considérée comme un prototype de notre atelier afin de souligner les implications technologiques et musicales de notre recherche-création sur les interfaces. Dans la partie 1, nous avons exposé les difficultés rencontrées lors de la confrontation musicale des esquisses d'interfaces graphiques (*2plf.js*, *dtd.js*). Nous avons expliqué la nécessité que nous avons eu de modifier notre dispositif expérimental et d'adapter nos manières de composer. Nous avons vu par la suite que ce point de tension a mis au centre de nos recherches la notion d'instrument et d'interactions nous amenant à reconsidérer notre définition d'interface.

Dans la partie 2, nous avons présenté les différents éléments constituant la partie technique de notre *espace composable* toujours considéré en mouvement. Nous avons explicité dans la sous-partie 2.1 notre écosystème logiciel et nos manières de faire du montage dans le domaine ambisonique. Puis dans la sous-partie 2.2, après un rappel sur l'ambisonie (2.2.1) et sur les traitements en ambisonie (2.2.2) nous avons présenté l'approche que nous avons appelé *Layer Based Ambisonics Processes*, utilisant chaque couronne de haut-parleur d'un dispositif 3D comme un plan 2D sur lequel nous venons faire des opérations. Par la suite dans les sous-parties 2.3 et 2.4, nous avons explicité

les approches originales de traitement pour atteindre le niveau opératoire souhaité dans cette pièce. D'abord dans l'approche de la décorrélation de décorrélation, utilisant plusieurs modules de décorrélations en parallèle décorrés entre eux (2.3), puis avec une utilisation de la gestion du gain dans le domaine des ondes planes (2.4).

Dans la partie 3, nous avons esquissé un éclairage particulier du concept de morphologie : les morphologies spatiales. Nous avons présenté dans cette section les trois morphologies spatiales principalement composées dans *Émergences* : les textures spatiales, les enveloppes d'espace et la stratification. Elles sont toutes les trois à penser comme des opérations musicales sur l'espace et pas dans l'espace. La première émerge d'un feuilletage de modules de décorrélation et de superposition de prises, la deuxième d'un jeu avec l'écriture temporelle en contrepoint du comportement des textures dans le domaine des ondes planes, et la troisième émerge d'un travail sur la superposition en élévation de plans de traitements.

Toutes ces manières de faire sont les marqueurs d'un trajet articulant recherche, composition et développement. Cette posture du compositeur-chercheur nous permet de construire une approche singulière de la spatialité du son, nous interrogeant sans cesse sur ce qui fait émerger notre instrument. Aujourd'hui, nous souhaiterions formaliser les traitements esquissés dans cette pièce afin de pouvoir les partager. D'autre part, nous souhaiterions continuer l'exploration musicale des strates de traitements, notamment sur les sensations d'élévation des morphologies spatiales. Pour finir, nous souhaitons formaliser et générer de nouvelles morphologies spatiales.

5. REMERCIEMENTS

Ce travail a bénéficié d'une aide de l'EUR ArTeC financée par l'ANR au titre du PIA ANR-17-EURE-0008 et est accueilli à la Maison des Sciences de l'Homme Paris Nord.

6. REFERENCES

- [1] Natasha Barrett. Spatio-musical composition strategies. *Organised Sound*, 7(3) :313–323, 2002.
- [2] Charles Bascou. Adaptive spatialization and scripting capabilities in the spatial trajectory editor holoedi. In *Proceedings of the 7th Sound and Music Computing Conference (SMC)*, pages 404–408, Barcelone, 2010.
- [3] Leo Beranek. *Concert and opera halls : how they sound*. Published for the Acoustical Society of America through the American Institute of Physics, Woodbury, NY, 1996.
- [4] Alain Bonardi. Composer l'espace sonore. In *Revue Francophone d'Informatique Musicale*, volume 7-8 Culture du code, 2020.
- [5] Alain Bonardi. La librairie abclib : un ensemble de codes Faust rassemblant 20 ans de recherche,

- enseignement et création en musique mixte. In *Journées d'Informatique Musicale 2021*, Visioconférences, France, July 2021. AFIM.
- [6] Thibaut Carpentier, Markus Noisternig, and Olivier Warusfel. Twenty Years of Ircam Spat : Looking Back, Looking Forward. In *41st International Computer Music Conference (ICMC)*, pages 270 – 277, Denton, TX, United States, September 2015.
- [7] Julien Colafrancesco. *Spatialisation de sources auditives étendues : applications musicales avec la bibliothèque HOA*. phdthesis, Université Paris 8 Vincennes Saint-Denis, 2015.
- [8] Elie During. Prototypes. *L'étincelle (Ircam)*, 7 :6–12, Juin 2010.
- [9] Yvan Etienne. *De l'espace sonore*. Haute école des arts du Rhin, Les Presses du réel Présence graphique, Strasbourg, France Dijon, France Monts, Frances, 2014.
- [10] Jérémie Garcia, Thibaut Carpentier, and Jean Bresson. Interactive-compositional Authoring of Sound Spatialization. *Journal of New Music Research*, 46(1) :74–86, nov 2016.
- [11] Kevin Gohon. *Critique du discours musical et émergence d'une pensée « mixte » dans les œuvres électroacoustiques de Pierre Boulez et Luigi Nono*. PhD thesis, Université Rennes 2, 2018.
- [12] Paul Goutmann. Traitement spatial du son par décorrélation des signaux en ambisonie d'ordre élevé. In *Journées d'Informatique Musicale 2021*, Visioconférences, France, July 2021. AFIM.
- [13] Paul Goutmann and Alain Bonardi. Approaching Spatial Audio Processing by Means of Decorrelation and Ring Modulation in Ambisonics. In *Sound and Music Computing*. Université Jean Monnet, GRAME, INRIA, 2022.
- [14] Pierre Guillot. *La représentation intermédiaire et abstraite de l'espace comme outil de spatialisation du son*. Theses, Université Paris 8 Vincennes Saint-Denis, December 2017.
- [15] Aaron Heller, Eric Benjamin, Richard Lee, and Pandit Litoral. A toolkit for the design of ambisonic decoders. In *Linux Audio Conferece*, Stanford University, California, 2012. CCRMA.
- [16] Gary S. Kendall. The decorrelation of audio signals and its impact on spatial imagery. *Computer Music Journal*, 19(4) :71, 1995.
- [17] David Kim-Boyle. Spectral and Granular Spatialization with Boids. In *ICMC*, pages 139–142. Michigan Publishing, 2006.
- [18] Pierre Lecomte. Ambitools : Tools for Sound Field Synthesis with Higher Order Ambisonics - v1.0. In *Proceedings of the 1st International Faust Conference (IFC-18)*, Mainz, Germany, July 2018.
- [19] David Ledoux, Robert Normandeau, Olivier Bélanger, and Christophe Lengelé. Vers une approche immersive de la composition musicale par le développement d'outils de spatialisation sonore : ControlGRIS et ServerGRIS. In *JIM 2019 Journées d'Informatique Musicale*, Bayonne, France, May 2019. LaBRI.
- [20] Chikashi Miyama and Götz Dipper. Zirkonium 3.1-a toolkit for spatial composition and performance. In *Proceedings of the International Computer Music Conference*, volume 313, page 312, 2016.
- [21] Jean-Louis Le Moigne. « une fureur sacrée de comprendre pour faire et de faire pour comprendre ». *e-Phaïstos*, IX-1, apr 2021.
- [22] Tapani Pihlajamäki, Olli Santala, and Ville Pulkki. Synthesis of spatially extended virtual source with time-frequency decomposition of mono signals. *J. Audio Eng. Soc.*, 62(7/8) :467–484, aug 2014.
- [23] Anne Sèdes. Approche musicale de la decorrélation microtemporelle dans la bibliothèque HOA. In université de Montréal, editor, *Journées d'Informatique Musicale 2015*, Montréal, Canada, May 2015. Université de Montréal.
- [24] Anne Sèdes, Pierre Guillot, and Eliott Paris. The HOA library, review and prospects. In *International Computer Music Conference | Sound and Music Computing*, Proceedings ICMCISMC, pages 855 – 860, Athènes, Greece, September 2014.
- [25] Marco Stroppa. *Accorder musicalement un espace réel et un espace inventé*. Presses universitaires de France, Paris, 2007.
- [26] Anne Sèdes. L'espace sonore du prometeo de luigi nono. In Giordano Ferrari, editor, *L'opéra éclaté - La dramaturgie musicale entre 1969 et 1984*, pages 223–238. L'Harmattan, 2006.
- [27] Anne Sèdes. La notion d'appareil empruntée à W. Benjamin, appliquée à la création musicale en studio électroacoustique, et à sa diffusion. *Revue Appareil [En ligne]*, 11, 2013.
- [28] Horacio Vaggione. L'approche morphologique. In *Musique électroacoustique : expérience et prospective : actes IV*, pages 140–145, Bourges, 1998.
- [29] Horacio Vaggione. *L'espace composable. Sur quelques catégories opératoires dans la musique électroacoustique*. L'Harmattan, 1998.
- [30] Horacio Vaggione. Composing musical spaces by means of decorrelation of audio signals. In *Proceedings of the DAFx Conference on Digital Audio Effects.*, University of Limerick, December 2001. University of Limerick, Limerick (Irlande).
- [31] Horacio Vaggione. Décorrélation microtemporelle, morphologies et figurations spatiales. In *Journées d'Informatique Musicale*, Marseille, France, May 2002.

- [32] Charles Verron, Mitsuko Aramaki, Richard Kronland-Martinet, and Grégory Pallone. A spatialized additive synthesizer. In *The Inaugural International Conference on Music Communication Science*, page 4, Sidney, Australia, December 2007.
- [33] Franz Zotter and Matthias Frank. *Ambisonics : A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*. Springer, 2019.

APPROCHE DE LA SYNTHÈSE PAR FRONTS D'ONDE (WFS), UTILISANT UNE ANTENNE À FAIBLE DENSITÉ

Raymi Nicolas Morales

Doctorant, Université de Paris 8, MSH Paris Nord
CICM/Laboratoire MUSIDANSE.
Université des Arts, Guayaquil, Équateur
raymi.morales@uartes.edu.ec

Anne Sedès

Université de Paris 8, MSH Paris Nord
CICM/ Laboratoire MUSIDANSE,
Saint-Denis, France
anne.sedes@univ-paris8.fr

RÉSUMÉ

La Synthèse par fronts d'onde permet une distribution de sources sonores virtuelles dans l'espace ayant une interprétation cohérente sur une zone étendue de l'espace de reproduction. Cependant, dans sa conception normale, elle nécessite une grande quantité d'équipements et de puissance de calcul. Cette recherche analyse un montage plus simple utilisant une antenne à faible densité et rendant ainsi le système plus accessible.

1. INTRODUCTION

Actuellement, le système de spatialisation du son le plus répandu et utilisé est le système stéréophonique, qui se base sur une illusion perceptive [1] appelée « source fantôme » [2], en utilisant majoritairement l'indice interaural d'intensité (ILD). Ce système est extrêmement simple et efficace mais il présente un certain nombre de défauts, dont on citera les suivants :

- L'interprétation n'est correcte que dans un espace réduit appelé « sweet spot », ce qui empêche une correcte interprétation pour une partie des auditeurs en condition d'écoute collective ou quand l'auditeur se déplace dans l'espace de reproduction en dehors du sweet spot [3].

- Les sources peuvent uniquement être placées suivant une ligne droite entre les enceintes [4] (on a alors éventuellement recours à l'aide d'une coloration et/ou effet de salle pour donner une sensation de profondeur).

C'est pourquoi il existe toute une série de recherches et développements autour de la diffusion du son pour améliorer la qualité de reproduction spatiale. On peut citer le VBAP/VBIP [5][4], Binaural, Ambisonique/HOA, WFS.

Selon DANIEL [11], « Dans un contexte d'expérience d'écoute très général, l'ITD (Différence interaurale de temps) est un indice plus prégnant que l'ILD ». Rayleigh et Middlebrooks défendent également la « Duplex theory » [12] qui explique que « la localisation du son à basses fréquences repose sur des différences de phase au niveau des deux oreilles, ou ITD, et que la localisation à hautes fréquences repose sur des différences interaurales de niveau sonore (ILD). » [13]. C'est pourquoi nous avons jugé important d'utiliser le principe

de la WFS, qui en reconstruisant les fronts d'onde exploite à la fois l'ILD et l'ITD, et présente des caractéristiques qui favorisent l'écoute collective en créant une localisation cohérente des sources dans une zone étendue de l'espace de reproduction.

Parmi les contraintes de la WFS il y a la complexité de sa mise en place ainsi que la quantité d'équipements nécessaires, ce qui en limite l'accessibilité pour la majorité des créateurs sonores. C'est pourquoi, dans le cadre de cette recherche nous avons réalisé le montage d'une antenne [4] à faible densité d'enceintes, afin d'en analyser le comportement et les limites, puis en déduire les possibilités de création et les enjeux artistiques.

2. WFS

2.1 Principe

La Synthèse de fronts d'onde se base sur le principe de Huygens [4] pour recréer des variations de pression acoustique que générerait une source primaire [6] ou source virtuelle [7] [1], à l'aide d'une ceinture d'enceintes situées tout autour de l'espace de reproduction.

Les principaux avantages de la WFS sont :

- La restitution sonore spatialisée de manière homogène sur une zone étendue de l'espace de reproduction
- Le déplacement de la source virtuelle en profondeur, en l'éloignant au-delà du plan de positionnement de l'antenne d'enceintes ou en la focalisant à l'intérieur de l'espace de reproduction [6].

Chaque source secondaire est alimentée par un signal préalablement traité, modifiant le contenu spectral, l'amplitude et le retard. Les filtres utilisés pour générer les signaux qui alimentent les haut-parleurs sont ceux présentés par [8] et [9] utilisant la synthèse par modèle physique *Model-based Rendering* [8].

Pour réaliser la synthèse de fronts d'onde on utilise une ceinture d'enceintes situées tout autour de l'espace de reproduction, ce qui génère des ondes cylindriques et non pas sphériques. C'est pourquoi on doit réaliser une correction dans la loi d'atténuation et se placer dans le cas de « 2.5D » [3] [10]

2.2 Discrétisation et Aliasing

Selon l'équation de Kirchhoff-Helmholtz [4], pour pouvoir reconstruire correctement les fronts d'onde pour tout le spectre il est nécessaire d'avoir une densité infinie de sources secondaires [10]. Or, dans l'état de l'art actuel, il n'est pas possible d'obtenir une ligne continue de sources secondaires. La reconstruction des fronts d'onde est donc correcte jusqu'à la fréquence d'aliasing donnée par la relation suivante [4] :

$$f_{al} = \frac{c}{2\Delta \sin \alpha_{max}}$$

Dans le cadre de cette recherche on se place dans la situation d'une antenne à faible densité pour en étudier le comportement, en utilisant une antenne avec 8 enceintes théoriquement omnidirectionnelles espacées de 40 cm. Ce qui nous donne une fréquence d'aliasing de 430 Hz pour une source frontale et 860 Hz pour une source à 45°.

Cette approche implique que les fronts d'onde ne sont correctement reconstruits qu'en basses et moyennes-basses fréquences. On s'éloigne donc du concept original de la synthèse de fronts d'onde vu que la plus grande partie du spectre n'est pas correctement synthétisée.

C'est pourquoi au-dessus de la fréquence d'aliasing il se produit des interférences qui génèrent des filtres en peigne qui peuvent être identifiables. Pour résoudre ce problème on peut introduire des décorrélations temporelles [4] pour les fréquences au-dessus de la fréquence d'aliasing et ainsi générer un léger effet de salle

Cependant on peut utiliser les principes de la WFS sur une antenne espacée pour en obtenir des résultats perceptifs intéressants quant à la distribution et localisation spatiale des sources sonores.

3. LOGICIEL DE SYNTHÈSE

Le cahier des charges de cette recherche inclut une grande accessibilité du système. C'est pourquoi nous avons développé un logiciel de traitement multiplateforme qui sera prochainement en libre accès sous forme d'une bibliothèque d'objets qui pourra être intégrée au sein d'une multitude d'environnements de programmation sonore.

Ce traitement des signaux ne nécessite pas une grande puissance de calcul (jusqu'à une certaine limite). La technologie moyenne actuelle offre des capacités suffisantes pour que le traitement puisse être effectué sur n'importe quel ordinateur de gamme moyenne.

L'entrée et sortie des signaux audio s'effectue par des entrées et sorties physiques ou virtuelles, et l'information sur la position spatiale de chaque source sonore, ainsi que le modèle physique de front d'onde à utiliser (source point, onde plane, source focalisée) se transmettent à travers le protocole OSC.

4. ÉVALUATION PERCEPTIVE

Une première maquette a été installée à la Maison des Sciences Humaines à Paris, avec une seule antenne contenant huit enceintes Meyer Sound qui ont été placées à différentes



Figure 1. Maquette WFS installée à la Maison des Sciences Humaines, Laboratoire MUSIDANSE

distances (20, 30 et 40 cm d'espacement entre chacune) pour en analyser les résultats. On a utilisé une série d'échantillons représentatifs comme sources sonores virtuelles comprenant des instruments enregistrés (guitare, piano, percussion, voix, ambiances) et des sons générés par ordinateur.

Au cours des expérimentations nous avons constaté qu'en utilisant le principe de la WFS qui utilise à la fois l'ILD et l'ITD on peut obtenir une plus grande transparence et une meilleure perception de l'espace par rapport aux systèmes qui utilisent uniquement l'ILD.

Les premières observations qui ont été réalisées portent sur le compromis entre la distance séparant les enceintes et la qualité de restitution des sources sonores. On a pu constater qu'une distance de 40 cm peut être acceptable vu qu'elle permet de conserver les propriétés de la WFS pour les basses et moyennes-basses fréquences tout en limitant significativement la quantité d'enceintes nécessaires et rendant le système plus accessible.

Un des problèmes perçus est que si l'on met à contribution l'ensemble des enceintes de l'antenne pour synthétiser une source sonore statique, il est difficile d'identifier précisément la position de la source sonore. Ceci est dû à la faible différence d'intensité entre les sources secondaires qui génèrent des pré-échos, et par effet de précedence confondent l'indice interaural d'intensité qui est l'indice de majeure efficacité pour la localisation spatiale des sources sonores en hautes fréquences.

Quand la source virtuelle se déplace dans l'espace, le sens de déplacement de cette dernière est observable, ce qui démontre clairement l'efficacité de l'ITD, mais ne permet pas d'identifier avec précision la position de la source.

Ce problème a pu être masqué en limitant la quantité de sources réelles mises à contribution pour synthétiser la source sonore. Ceci a permis d'identifier très clairement la position de la source virtuelle vue uniquement en augmentant l'influence de l'ILD. Dans notre configuration de salle, on a pu constater que l'utilisation de l'apport de 4 enceintes pour chaque source permet à la fois d'obtenir une source transparente et de localiser très précisément la source dans l'espace.

La distance entre les enceintes implique qu'il se produise des filtres en peigne au-dessus de la fréquence d'aliasing qui peuvent être masqués pour des sources fixes, mais qui deviennent très évidents pour des sources en mouvement, les

décorrélations temporelles ne suffisant plus. C'est pourquoi en fonction de la vitesse de mouvement il est préférable de limiter la quantité de différences de temps entre les enceintes, même si cela implique que le front d'onde devienne plus plat, car il reste comme objectif d'expérimentation pour la suite de limiter les différences de temps pour les sources virtuelles en mouvement, mais uniquement pour les fréquences au-dessus de la fréquence d'aliasing et ainsi maintenir le front d'onde pour les basses et moyennes-basses fréquences.

On a pu constater également que la localisation des sources virtuelles est plus précise pour les sons percussifs ou pour ceux qui ont une grande dynamique sonore.

4.1 Profondeur et focalisation

Avec une fréquence d'aliasing aussi basse (430 Hz ou 860 Hz) et en se rapprochant du VBAP, il devient difficile de percevoir les sources derrière la ceinture de hauts parleurs, et pour ce qui est de la profondeur, on se rapproche des caractéristiques du système stéréophonique où les sources sont localisées sur l'axe traversant les enceintes.

Pour ce qui est des sources focalisées, l'effet se perd complètement et en échange, on perçoit des artefacts qui ne correspondent pas à l'effet désiré de percevoir une source sonore à l'intérieur de l'espace de reproduction. Il est évident que pour pouvoir reconstruire correctement une source focalisée il est nécessaire d'avoir une plus grande densité d'enceintes.

5. PROJETS EN COURS DE DEVELOPPEMENT

5.1 Salle d'expérimentation sonore

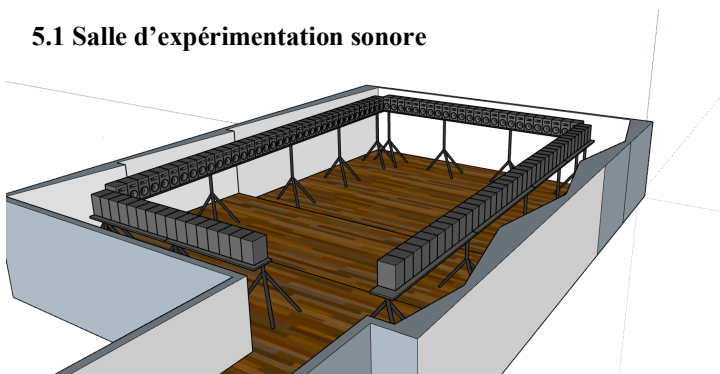


Figure 2. Maquette de l'installation en cours au Centre d'innovation et Production Artistique (Manzana 14), Universidad de las Artes, Guayaquil - Equateur

Pour accueillir ce projet on installera une salle au sein du Centre d'innovation en arts (Manzana14) de l'université de Guayaquil, en Équateur. Elle sera équipée de 40 enceintes JBL 305p mk2 distribuées sur 3 antennes dans une première phase, puis 147 enceintes conformant une ceinture qui englobe tout l'espace de reproduction.

Le système que nous sommes en train de développer va subir une période d'évaluation pour mettre en évidence le rendu, les apports du système à la perception de l'espace et à la position des sources virtuelles et les nouvelles possibilités que cela peut offrir à la création sonore.

Une des premières informations à obtenir concerne les limites de puissance de calcul que peut fournir l'ordinateur qui fait office de DSP, et si nécessaire, envisager la possibilité d'utiliser le GPU pour certains calculs et ainsi fournir de la puissance de calcul supplémentaire au système.

Puis nous réaliserons une série de mesures objectives et analyses subjectives pour déterminer la réponse du système : son homogénéité tout au long du spectre, l'apparition de filtres en peigne dus aux interférences entre les enceintes, et la précision dans l'interprétation des positions spatiales des sources sonores.

Dans une deuxième phase on analysera l'efficacité du rendu pour des sources focalisées (qui se trouvent à l'intérieur de l'espace de reproduction) ainsi que l'impact des réflexions produites par la salle sur la perception de la position des sources sonores virtuelles.

Pour réaliser ces évaluations on mettra la salle et le système à disposition de créateurs pour qu'ils puissent développer plusieurs projets artistiques conçus spécifiquement pour la WFS.

5.2 Synthèse hybride

Utiliser une antenne à faible densité va à l'encontre de ce que requiert la WFS, c'est une condition dégradée de la WFS, cependant, il ne reste pas moins pertinent d'utiliser les principes de la WFS pour avoir une vraie reconstruction du front d'onde dans le bas du spectre et avoir une spatialisation cohérente pour les fréquences aiguës. La solution mise en pratique est la réduction du nombre d'enceintes qui contribuent à la construction de la source virtuelle, ainsi que la réduction de la courbature du front d'onde pour les sources en mouvement.

Pour les fréquences aiguës, qui ne sont pas correctement synthétisées en WFS, il est envisagé de fractionner le spectre en utilisant une autre technologie comme HOA avec compensation de distance ou VBAP, pour éviter des phénomènes de filtre en peigne qui sont très dérangeants pendant le mouvement de sources virtuelles et créer ainsi une synthèse hybride.

5.3 Projets artistiques liés

Il existe plusieurs projets artistiques qui utiliseront le système proposé dans cet article et ainsi aideront au développement et validation du dispositif, parmi lesquels :

- « Les voix de l'enfermement », exposition sonore réalisé par Priscilla Aguirre, qui cherche à sensibiliser le public sur les conditions de vie que subissent les personnes enfermées dans une prison féminine en Équateur. Pour cela on cherche à plonger l'auditeur dans l'univers sonore de la prison afin de représenter la violence et l'oppression qui peuvent être ressenties à travers des sons froids et métalliques qui se superposent aux cris, voix, et autres éléments sonores propres à cet espace si particulier.

- Le compositeur colombien Fredy Vallejos réalisera une pièce acousmatique donnant un rôle protagoniste à l'espace et à la position et déplacements des sources dans l'espace.
- Un ensemble qui mélange instruments musicaux et sons générés par ordinateur, adaptera et interprétera un set de jazz qui permettra d'évaluer le système en conditions de concert.
- Le système fera également l'objet de plusieurs mémoires de fin d'études d'étudiants de l'université des Arts à Guayaquil.

5.3 Méthodes d'évaluation

Dans l'état actuel de développement du projet, on a réalisé une évaluation perceptive subjective. Il est prévu par la suite de procéder à une validation scientifique du dispositif.

Les projets artistiques liés nous permettront d'avoir une quantité suffisante d'utilisateurs du système et d'auditeurs pour obtenir des informations subjectives sur les avantages et défauts du dispositif et nous donneront les indices nécessaires pour l'améliorer.

On recueillera en premier lieu les impressions des créateurs qui auront été invités à pousser le système jusqu'à ses limites. Nous nous focaliserons sur la transparence du rendu fréquentiel des sources sonores virtuelles, la qualité de rendu pour les sources en mouvement et la possibilité d'utiliser les filtres en peigne, conséquence de la discrétisation comme outil créatif pour générer des traitements spatiaux et ainsi développer les ressources créatives propres du système.

On réalisera également une série d'évaluations perceptives avec des auditeurs pour évaluer la perception de l'espace et la précision sur la position des sources sonores dans l'espace.

Dans une deuxième phase on évaluera également avec différents groupes d'auditeurs la perception de la proximité des sources virtuelles qui se trouvent à l'intérieur de l'espace de reproduction.

6. CONCLUSION

La WFS à faible densité d'enceintes est un dispositif qui permet d'obtenir des sources sonores beaucoup plus transparentes et naturelles, ce qui implique une plus grande lisibilité des sources sonores quand il s'agit de scènes sonores complexes [11], un plus grand confort d'écoute et une meilleure immersion sonore par rapport aux systèmes stéréophoniques ou VBAP.

L'avantage avec le système que nous proposons est qu'il peut être obtenu avec une quantité relativement faible d'équipements et de puissance de calcul. C'est pourquoi il s'agit d'un système accessible à la plupart des membres de la communauté artistique, pour pouvoir créer des scènes sonores complexes ayant une grande lisibilité de tout le contenu spectral des sources sonores en évitant le masquage fréquentiel.

Ce système permet également une distribution continue et homogène des sources virtuelles dans l'espace ainsi que des mouvements fluides sans avoir de zones d'ombre dans les trajets des sources.

Les filtres en peigne et les décorrélations temporelles peuvent être utilisés comme traitements spatialisés pour générer des sensations perceptibles par les auditeurs.

Cette recherche a comme but principal de proposer aux créateurs sonores des outils sous plusieurs formes pour faciliter l'exploration de l'espace dans la création sonore tels que:

- Une bibliothèque complète d'objets qui puissent être intégrés dans la plupart des logiciels ou environnements de création sonore pour générer une spatialisation sonore par synthèse de fronts d'onde.
- Un code ouvert en libre accès qui permettra l'appropriation du code par les créateurs et ainsi partager le savoir sous forme de science ouverte.

7. BIBLIOGRAPHIE

1. Caulkins, T. « Caractérisation et contrôle du rayonnement d'un système de Wave Field Synthesis pour la situation de concert », PhD thesis, Université Paris VI – Pierre et Marie Curie, 2007.
2. Frank, M. « Phantom Sources using Multiple Loudspeakers », PhD thesis. University of Music and Performing Arts Graz, 2013.
3. DeBruijn, W. « Applications of Wave Field Synthesis in Videoconferencing », PhD thesis, TU Delft, Delft, Pays-Bas, 2004.
4. Nicol, R. « Restitution sonore spatialisée sur une zone étendue : Application à la téléprésence », Ph.D. thesis, Université du Maine, Le Mans, France, Le Mans, France, 1999.
5. Pulkki, V. « Virtual Sound Source Positioning Using Vector Base Amplitude Panning », J. Audio Eng. Soc, 45, 456 – 466, 1997.
6. Corteel, E. « Caractérisation et Extensions de la Wave Field Synthesis en conditions réelles », Ph.D. thesis, Université de Paris VI, Paris, France, 2004.
7. Merlier, B. « Vocabulaire de l'espace en musiques électroacoustiques. coll. Musique et sciences », éditions Delatour, France, pp.230, 2006.
8. Wierstorff, H. « Perceptual Assessment of Sound Field Synthesis », PhD thesis, TU Berlin, Berlin, Allemagne, 2014.
9. Ahrens, J. « Analytic Methods of Sound Field Synthesis. Springer, Berlin, 2012.
10. Verheijen, E.N.G. « Sound Reproduction by Wave Field Synthesis », PhD thesis, TU Delft, Delft, Pays-Bas, 1997.
11. Daniel, J. « Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia », PhD thesis, Université Paris VI – Pierre et Marie Curie, 2001.
12. Strutt JW. « On our perception of sound direction », Philos Mag 13: 214–232, 1907.
13. Middlebrooks, J. C. « Sound localization », In Handbook of Clinical Neurology Vol. 129 (3rd series) Chapitre 6, 2015.

Actes des Journées d'Informatique Musicale 2023
24-26 mai 2023
Maison des Sciences de l'Homme Paris Nord

**SESSION « FAIRE DE LA MUSIQUE AVEC L'ÉLECTRICITÉ
DU CORPS ET DU CERVEAU »**

L'électromyographie pour la performance et la composition : trois études de cas dans le cadre du projet BBDMI

David Fierro

Alain Bonardi

Guilherme Carvalho

Anne Sèdes

CICM / MUSIDANSE

Université Paris 8

davidfierro

@gmail.com

CICM / MUSIDANSE

Université Paris 8

alain.bonardi

@univ-paris8.fr

RIRRA21

Université Paul Valéry

Montpellier 3

guilherme.carvalho

@univ-montp3.fr

CICM / MUSIDANSE

Université Paris 8

anne.sedes

@univ-paris8.fr

RÉSUMÉ

Dans cet article nous décrivons les méthodologies mises en œuvre et les résultats obtenus lors des études utilisateurs menées dans le cadre du projet BBDMI. Pendant ces études utilisateurs, nous avons testé de nouvelles manières d'interagir avec les traitements sonores temps réel en utilisant comme signaux de contrôle l'information provenant de capteurs EMG placés sur le corps des interprètes. Nous décrivons les protocoles de tests mis en place, les difficultés que nous avons rencontrées lors de nos expérimentations, et les bonnes pratiques que nous en tirons.

Nous détaillons également dans cet article le processus d'adaptation de deux pièces mixtes en remplaçant les signaux de contrôle originaux par des signaux EMG. Nous évoquons les avantages et les défis qui se posent lors de l'écriture instrumentale et l'interprétation de pièces mixtes quand les signaux de contrôle proviennent de l'implication musculaire de l'interprète lors de l'exécution du geste instrumental.

MOTS CLÉS

Etude utilisateur, EMG (électromyogramme), interaction homme-machine, expressivité musicale, prise en main, instrument.

1. INTRODUCTION

Le développement rapide de technologies pour la captation et le traitement de signaux physiologiques est en train de changer la manière dont nous interagissons avec les interfaces numériques de création musicale utilisant notre corps. La possibilité de capter les signaux électriques provenant des muscles ou du cerveau devient chaque jour plus accessible avec des solutions de plus en plus nombreuses et de moins en moins onéreuses.

Ce domaine en pleine expansion est exploré depuis plusieurs années par des chercheurs et artistes, surtout dans le domaine de l'utilisation de l'EMG comme signal de contrôle pour la performance musicale. L'EMG est utilisée dans plusieurs situations musicales allant de la synthèse sonore pilotée par le geste [10], à la musique mixte

associant instrument acoustique, traitement de l'instrument par des processus électroniques et contrôle de ces derniers par l'énergie musculaire. De nombreux travaux s'intéressent à la prise de conscience réflexive de l'EMG par la personne équipée de capteurs [12], sachant que les gestes captés conjuguent souvent plusieurs membres en sollicitant plusieurs muscles.

Le projet BBDMI (Body Brain Digital Music Instruments) financé par l'Agence Nationale de la Recherche associe la Maison des Sciences de l'Homme Paris Nord, le CICM / MUSIDANSE (Université Paris 8), l'Institut du Cerveau et de la Moëlle Epinière, et l'entreprise Soixante Circuits [9]. Il est articulé autour de trois objectifs :

- Faire de la musique à partir des signaux électriques des muscles et du cerveau en s'appuyant sur des technologies EMG (électromyogramme) et EEG (électroencéphalogramme).
- Produire des prototypes instrumentaux de manière ouverte (open software et open hardware) et les documenter dans le cadre des sciences ouvertes (principes FAIR).
- Développer une méthodologie centrée utilisateur incluant aussi bien les musiciens expérimentaux que les praticiens de la pédagogie musicale, les publics empêchés, et ce par les moyens de la pratique et de la médiation artistique.

Pour ce troisième objectif, nous menons un ensemble d'études utilisateurs régulièrement échelonnées au cours du projet (mené sur 42 mois depuis début 2022), comme le montre la figure 1 ci-dessous.

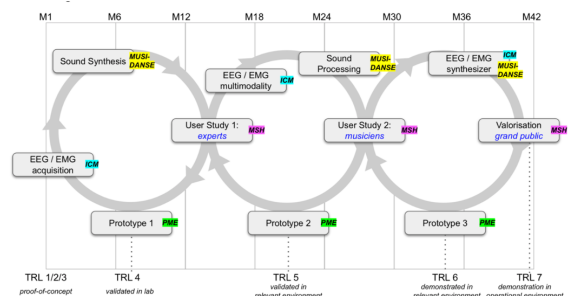


Figure 1. Schéma d'organisation du projet BBDMI en trois cycles de conception de prototypes.

Comme le montre le schéma d'organisation du projet BBDMI, les études utilisateurs vont concerner des intervenants variés, allant des experts du domaine au grand public en passant par des musiciens. Lors de ces études, l'objectif est de prendre en main le prototype d'instrument dans son état de développement au moment de l'atelier mené, ainsi que l'environnement de programmation préprogrammé, pour explorer leur potentiel d'expression musicale. Les résultats des études utilisateurs contribuent à chaque fois à l'inflexion du développement du projet.

Sur un an, nous avons organisé des ateliers d'études utilisateurs avec trois musiciens différents : d'une part deux musiciens professionnels, Amélia Mazarico (guitariste et professeur de guitare, lors de deux séances le 18 octobre 2022 et le 14 février 2023) et Guilherme Carvalho (violoncelliste et enseignant-chercheur, co-auteur de cet article, lors d'une séance le 7 septembre 2022); d'autre part, Laurie Pompidor, étudiante en Licence 3 Musicologie et Pratique de la Musique de l'université Paris 8, également élève en 3^{ème} cycle en guitare au Conservatoire de Châtenay-Malabry (lors de son stage de Licence en juin 2022).

Dans les trois cas, la situation de création musicale relève tout d'abord de la musique mixte : un instrument acoustique joué par un instrumentiste interagit avec des transformations ou synthèses sonores temps réel; mais vient s'ajouter le contrôle de ces processus par des captations EMG grâce à des électrodes placées sur le corps du musicien.

Dans cet article, nous commencerons par décrire les grandes lignes des interactions entre le cycle de prototypage du projet et la prise en main des dispositifs par les musiciens. Puis nous montrerons comment les dispositifs à base de capteurs EMG ont été pris en main par les trois intervenants avec lesquels nous avons mené les études utilisateurs jusqu'à maintenant. Nous analyserons ensuite les relations entre les signaux EMG et le geste musical, puis présenterons plusieurs situations d'appropriation, notamment dans la pièce *Immersion* d'Anne Sèdes pour violoncelle et électronique, ou dans *Fil de Soi 3* d'Alain Bonardi pour guitare et électronique. Nous conclurons cette contribution en évoquant les perspectives de développement du projet BBDMI.

2. PROTOTYPAGE DE L'INSTRUMENT VIA LA PRISE EN MAIN PAR LES MUSICIENS

2.1. Description du dispositif hardware

Le projet BBDMI a pour but le développement de prototypes instrumentaux de manière ouverte (*open software* et *open hardware*) et leur documentation dans le cadre des sciences ouvertes (principes FAIR). Nous avons choisi de commencer

nos travaux par la partie logicielle et musicale, à partir de dispositifs électroniques existants expérimentés dans des situations de jeu instrumental et de création.

Nous développons un système de captation, traitement et sonification des signaux électriques provenant du corps. Pour la captation des signaux EMG nous utilisons la carte EAVI¹ développée par un de nos partenaires. Cette carte nous permet de mesurer les différences de potentiels sur la peau avec des électrodes différentielles. Elle est fixée au bras du musicien grâce à un strap élastique. La carte offre une capacité de 4 canaux EMG avec une résolution de 14 bits et une fréquence d'échantillonnage de 8 KHz. Les signaux EMG sont généralement échantillonnés à des fréquences comprises entre 200 Hz et 1000 Hz. Cette fréquence d'échantillonnage élevée nous a aidé à développer une étape d'extraction de caractéristiques plus efficace [7].

Notre objectif essentiel de recherche est un instrument accessible que l'on doit pouvoir s'approprier facilement, mais possédant une grande versatilité. Le rôle des études utilisateurs est ici très important : il s'agit d'analyser l'ergonomie des interfaces ainsi que l'expérience utilisateur.

2.2. Captation et extraction de caractéristiques

En partant de la carte EAVI munie de ses électrodes pour la captation, nous menons un développement de logiciels open-source, qui concerne pour l'essentiel le traitement des signaux physiologiques et la synthèse sonore, en utilisant le langage FAUST ou l'environnement *Max*, pour de nouveaux développements mais également en intégrant des travaux précédents comme la bibliothèque *abclib* [1].

Nous nous appuyons sur les modules développés dans notre projet de recherche² pour préparer, analyser et finalement utiliser les signaux EMG reçus de la carte EAVI. Les signaux EMG peuvent varier en intensité selon la personne et le muscle capté; pour cette raison il est impératif de calibrer les signaux d'entrées avant qu'ils puissent être utilisés. Une phase de filtrage des bruits est également nécessaire car l'intensité des signaux EMG est comparable à celle des bruits électromagnétiques mesurés sur les électrodes.

Indépendamment des traitements sonores spécifiques à chaque étude utilisateur, nous avons utilisé pour toutes les études des modules de filtrage, de calcul de valeur efficace, de calibration et de suivi d'enveloppe [8].

¹https://www.researchgate.net/publication/333699608_EAVI_EMG_board

²<https://gitlab.huma-num.fr/bbdmi/bbdmi>

2.3. Choix du type et positionnement des électrodes EMG

Les électrodes peuvent être positionnées à plusieurs endroits sur le haut du corps : sur l'avant-bras (face interne et face externe), sur l'épaule, ou à d'autres endroits.

Avoir des électrodes indépendantes (jusqu'à 4 par carte EAVI) nous offre la possibilité de mesurer l'activité sur un muscle spécifique ou sur un ensemble de muscles.

Nous avons choisi des électrodes avec gel qui fonctionnent très bien alors que les électrodes sèches induisent trop de bruit et ne sont pas utilisables. Chaque jeu comporte trois électrodes correspondant à la masse, à la valeur de référence et au signal musculaire.

Les électrodes placés sur l'épaule donnent un signal de la même qualité que celui obtenu sur l'avant-bras (cf. figure 2).



Figure 2. Exemples de positionnement des électrodes lors des trois études menées.

2.4. Geste musical et observation musculaire

Dans le cas du violoncelle et de la guitare, chaque côté du corps est dédié à un aspect du jeu instrumental : du côté droit, l'excitation des cordes; du côté gauche, la sélection des hauteurs des sons produite par le positionnement des doigts sur les cordes. Le choix de côté pour l'avant-bras ou de l'épaule est donc important dans l'idée d'associer ou au contraire dissocier le contrôle de la production instrumentale et la prise de conscience musculaire pouvant conduire à d'autres formes de contrôle.

Lors de chacune des sessions menées, les trois musiciens ont témoigné qu'ils découvraient que tel ou tel muscle participait à tel ou tel geste musical. C'est par exemple le cas des muscles de l'épaule, qui sont rarement associés à la production

instrumentale, et donc rarement mesurés. Cette exploration ouvre de nouvelles façons d'interpréter le geste musical.

Le placement des électrodes doit être pensé en relation avec le suivi du geste musical. La nature du geste implique une relation entre temporalité et engagement musculaire spécifique à chaque mouvement et chaque mode de jeu. Dans ce sens, pour un même geste musical nous pouvons obtenir des signaux différents qui dépendent du placement des électrodes.

L'objectif n'est donc pas seulement d'obtenir un bon signal avec une forte intensité mais de trouver le muscle qui permet le plus haut degré d'expressivité. Le choix de ce muscle changera selon le mode de jeu, le geste désiré et surtout le résultat final pour l'instrumentiste qui considérera ce signal EMG comme signal de contrôle.

3. LA PRISE EN MAIN DE CAPTEURS EMG COMME OUTIL D'EXPRESSION MUSICALE EN SITUATION DE MUSIQUE MIXTE

3.1. Intensité musculaire et maîtrise instrumentale

A travers les trois études menées, la première remarque que nous pouvons faire est que l'intensité musculaire investie dans le geste instrumental dépend du degré de maîtrise acquis. En effet, si l'on compare l'étude menée avec l'élève-guitariste avec celle menée avec la guitariste professionnelle, il apparaît que, sur des gestes identiques, un musicien expérimenté investit moins d'énergie musculaire qu'un jeune musicien en formation. Cette observation rejoint des travaux antérieurs [6] dans le domaine des sciences du mouvement appliquées à l'étude du jeu instrumental : les musiciens experts optimisent leur mouvement de la manière à produire l'effet voulu avec le minimum d'effort et en réduisant les composantes du geste qui sont superflues et inutiles à la réalisation de cet effet.

Le processus de calibration des signaux musculaires proposé dans l'instrument en cours d'élaboration est donc particulièrement important.

A l'inverse, utiliser le signal EMG comme signal de contrôle d'une manière active en situation de musique mixte demande à engager nettement l'effort musculaire, ce qui finalement peut aller contre ce que l'on a longuement appris dans l'enseignement instrumental. Nous avons remarqué que même après une phase de calibration des signaux d'entrée, la dépendance à la création d'une note reste présente lors du jeu instrumental.

3.2. Décorrélacion entre geste instrumental et contrôle EMG

Chaque geste instrumental implique l'utilisation d'un ensemble de muscles précis pendant une durée spécifique. Selon le positionnement des électrodes,

le geste instrumental et le signal capté par les électrodes peuvent être décorrélés, voire complètement indépendants [5]. Nous avons testé trois manières différentes de faire interagir les traitements sonores avec le geste instrumental. Dans chacune de ces configurations, nous obtenons des façons différentes d'interagir avec l'instrument.

Pour la première approche, nous avons placé les électrodes sur le muscle qui était le plus impliqué dans le geste musical. Avec cette configuration, l'instrumentiste ne peut pas éviter la génération du signal et c'est donc de l'écriture instrumentale dans la partition que vient le contrôle du niveau de sortie obtenu par les électrodes à un moment précis.

Une autre façon de lier le geste instrumental et le signal électrique généré est de placer les électrodes sur des muscles que nous pouvons appeler secondaires dans l'exécution du geste musical. Ces muscles peuvent être utilisés au début, à la fin, pendant ou peuvent même être exclus du geste instrumental, avec un peu d'effort de la part de l'interprète.

Dans ces deux premières configurations, le signal EMG généré interagit avec le système de traitement sonore d'une manière passive car les capteurs sont placés sur des muscles nécessaires à la génération du son.

Ainsi, dans l'étude menée avec Laurie Pompidor, nous nous sommes rendu compte de la difficulté à avoir l'épaule détendue pendant le jeu instrumental, et d'une certaine manière la corrélation entre les deux était subie.

Une autre relation possible entre le geste instrumental et le signal électrique généré peut être fondée sur l'indépendance entre les deux. Nous avons placé des électrodes sur des muscles qui n'étaient pas impliqués dans le geste musical, donnant une plus grande indépendance à l'instrumentiste pendant l'exécution d'une pièce. Le fait de pouvoir séparer le geste musical et le signal généré offre plus de possibilités expressives au musicien en même temps qu'il demande une écriture plus détaillée de la part du compositeur pour décrire ce qui est souhaité.

3.3. L'adaptation de la pièce *Immersion* d'Anne Sèdes au contrôle par EMG

La pièce *Immersion* a été composée par Anne Sèdes en 2012 pour violoncelle et électronique en temps réel. Son écriture est liée au développement de la bibliothèque HOA du CICM pour Max et Pure Data, dans le cadre du programme de recherches "La spatialisation du son par les musiciens, pour les musiciens" (Labex Arts-H2H). Dans sa première version, le niveau sonore du son du violoncelle (mesurée par l'objet [*peakamp*~] sur *Max*) pilotait un ensemble de paramètres du moteur sonore.

Des traitements sonores comme la granulation et la décorrélation temporelle [2][11] ont été mis à

l'œuvre dans cette pièce pour créer des textures sonores en transformant le son du violoncelle.

L'utilisation de l'intensité sonore comme signal de contrôle pour les traitements sonores lie d'une manière implicite l'écriture instrumentale au traitement appliqué dans chaque moment précis. Chaque son écrit dans la partition représente une intensité sonore à un moment précis de la pièce. Cette interaction entre l'écriture instrumentale et les traitements sonores n'offre pas une dimension expressive additionnelle à l'interprète car le signal de contrôle dépend directement de ce qui a été écrit dans la partition et de la logique de l'interprétation.

Dans le cadre du projet BDDMI, la compositrice Anne Sèdes a expérimenté une nouvelle forme d'interaction avec les traitements sonores dans une perspective musicale, d'expressivité, de jouabilité et de créativité. Cette nouvelle version remplace l'intensité sonore comme signal de contrôle par le signal EMG provenant des électrodes placés sur le corps de l'interprète. Cette nouvelle configuration change complètement le dynamisme des signaux de contrôle car l'intensité musculaire a un comportement beaucoup plus organique que l'enveloppe du son produite. La figure 3 nous permet d'observer le signal reçu par les électrodes placés sur la partie interne de l'avant-bras droit (palmaris longus) du violoncelliste et l'intensité sonore du son produite pendant qu'il joue des pizzicati successifs.

Pendant nos multiples essais sur des modes de jeu, nous avons remarqué des différences importantes entre l'effort musculaire et l'amplitude du son généré. Pendant que l'interprète joue des notes continues en crescendo, l'intensité sonore augmente rapidement tandis que le signal EMG change légèrement mais avec des modifications liées aussi aux changements d'archet. Dans le cas d'un trémolo serré pianissimo sur des notes graves, l'effort physique impliqué est très important bien qu'un trémolo serré fortissimo sur des notes aiguës génère un signal EMG faible. Nous avons pu remarquer au travers de ces expériences l'importance du placement des électrodes lors du suivi du geste instrumental.

Remplacer l'intensité sonore par l'implication musculaire a ouvert la porte à de nouvelles formes d'expressivité de la part de l'interprète. La globalité du geste instrumental devient donc un paramètre de l'espace composable car il agit en permanence sur le traitement sonore. Cette nouvelle façon d'interagir avec les traitements sonores témoigne d'une nouvelle dimension sensible du jeu instrumental permettant une grande capacité d'expression musicale. Cette capacité a été bien perçue par l'interprète, qui portait une attention plus importante à des aspects de son geste dont il est habituellement peu conscient.



Figure 3. Intensité sonore de l'instrument et signal EMG produit par le muscle, dans le temps.

3.4. De la guitare classique à la guitare sans guitare

Dans notre deuxième étude utilisateur, nous avons travaillé avec la guitariste Amélia Mazarico, guitariste et professeur de guitare qui collabore avec le CICM depuis de nombreuses années.

Nous avons trois objectifs principaux. Le premier était de trouver des placements adéquats des électrodes pour divers modes de jeu, en tenant en compte quelle partie du geste instrumental nous voulions suivre.

Comme deuxième objectif nous voulions expérimenter de nouvelles interactions avec les traitements sonores selon différents modes de jeu utilisés dans la pièce *Fil de Soi 3* composée par Alain Bonardi, en utilisant le signal EMG comme contrôle. Des études précédentes ont démontré les capacités du suivi du geste instrumental par des capteurs EMG comme outil d'expression musicale [3].

Notre troisième objectif était de déterminer les capacités expressives des capteurs EMG comme seul signal de contrôle d'un moteur de synthèse d'une guitare par modèle physique.

3.4.1. Placement des électrodes selon les modes de jeu

Dans cette première partie de notre expérience nous avons essayé de déterminer les muscles impliqués dans différents modes de jeu pour la guitare et de définir dans quelle partie du corps connecter les électrodes EMG pour mesurer l'activité musculaire. Chaque mode de jeu entraîne un ensemble de mouvements synchronisés qui déterminent quels muscles seront utilisés à quel moment.

Nous avons effectué des tests utilisant différents modes de jeu tels que le glissando, le pizz. Bartók, le trémolo, le bottleneck, le tapping, les accords, les arpèges, entre autres. Selon chaque type de mode de jeu nous avons remarqué une utilisation assez distincte des muscles impliqués dans le geste instrumental.

Le choix du positionnement des électrodes a été fait en prenant en compte les muscles impliqués dans l'ensemble des modes de jeu. Nous avons décidé de placer trois électrodes sur le corps de l'interprète, une sur l'épaule droite en essayant de capter le muscle deltoïde, une autre paire d'électrodes sur la partie externe de l'avant-bras

droit près du muscle long extenseur du carpe et une troisième paire d'électrodes sur la partie interne de l'avant-bras gauche sur le muscle palmaris longus.

3.4.2. Signal EMG comme contrôle de traitements sonores

Dans la deuxième partie de notre expérience nous avons rejoué des extraits de la pièce *Fil de Soi 3* en utilisant des traitements sonores contrôlés par les signaux EMG. Avec cette expérience nous voulions évaluer la prise en main du dispositif par l'interprète, son appropriation et la facilité à discerner quel capteur agit sur quel traitement sonore, autant de la part de l'interprète que de la part du public. Nous avons assigné à chaque électrode un traitement différent afin de pouvoir comprendre plus facilement leurs comportements.

Voici les traitements liés à chaque électrode :

- Avant-bras droit : un module d'harmonisation permettant la transposition du son de la guitare.
- Avant-bras gauche : contrôle de la position dans l'espace de la source sonore avec un encodeur ambisonique.
- Épaule droite : facteur de décorrélation dans le module de décorrélation temporelle³ [4].

Pendant notre expérience nous avons obtenu de bons résultats en utilisant les 3 entrées EMG pour contrôler les traitements sonores. En tant que spectateur il est possible de reconnaître avec précision les traitements sonores associés à chaque geste. Il est important de remarquer que l'association du mouvement au traitement sonore n'est pas évidente au premier abord pour l'interprète. Nous avons observé un temps d'adaptation à l'interface qui implique un apprentissage de la part de l'interprète, créant les liens entre geste et transformation sonore.

Une relation intéressante se met en place quand on utilise des signaux EMG pour contrôler un système de spatialisations : un lien intéressant se crée entre le mouvement du corps et le changement de "forme" de l'espace. Cette relation permet à l'interprète de s'approprier plus facilement l'interface et d'en comprendre le fonctionnement.

3.4.3. Synthèse contrôlée par des signaux EMG

Dans cette troisième phase nous avons connecté des signaux EMG à un système de synthèse par modèle physique de la guitare, afin de créer des sons uniques et complexes qui imitent le jeu d'un guitariste. Le système (à base de fonctions en langage FAUST issues de la librairie de synthèse par modèles physiques) est capable de créer des sons qui sont très proches de ceux produits par une

³ Traitement spatial qui retarde les harmoniques sphériques entre l'encodage et le décodage ambisonique, créant un champ diffus.

guitare acoustique, nous permettant de changer la longueur de la corde et la position de pincement.

Grâce à la sensibilité des capteurs EMG il est possible de capter des nuances subtiles dans le mouvement des mains et des doigts du guitariste pour reproduire une grande variété de sons. Cette capacité expressive et sensible de l'interface situe le système comme une extension du corps plus qu'une extension de l'instrument. Avec cette configuration, nous avons pu observer les nuances sensibles générées par le geste chorégraphique de l'instrumentiste dans le moteur sonore. Sans instrument, c'est le corps qui devient instrument et chaque geste influence le son produit.

Selon le retour de l'instrumentiste, jouer sans l'instrument l'invite au mouvement, à marcher, à bouger, à faire de la musique et improviser avec le corps. Cette nouvelle façon d'interagir avec le moteur sonore la fait se sentir plus présente que pendant des concerts de musique mixte sans capteurs EMG.

3.5. Prise en main du système

Lors de nos études utilisateurs nous avons remarqué des similitudes dans la prise en main du système de la part des interprètes. Nous avons remarqué certaines difficultés qui viennent de la nouveauté de ce type de système et du manque d'expérience comme pour tout nouvel instrument.

Une première difficulté que nous avons remarquée a été la prise de conscience par l'interprète des muscles impliqués dans chaque geste instrumental. Sans expérience préalable, il est spécialement difficile de déterminer l'implication musculaire nécessaire à chaque geste. Pour faciliter l'appropriation du système, nous avons déterminé comme une bonne pratique pour nos études utilisateurs de commencer par des *mappings* simples entre le signal EMG et le moteur sonore en utilisant un seul électrode. Une autre méthode que nous avons trouvée assez convaincante dans cette première phase d'appropriation était de montrer visuellement à l'interprète l'effet de l'effort musculaire. En regardant le signal changer en temps réel, l'interprète arrive à comprendre plus facilement la relation entre geste et engagement musculaire.

Après la première phase d'adaptation, la deuxième difficulté que nous avons remarquée est liée à l'interprétation d'une partition en essayant de contrôler en même temps le signal généré par les électrodes. Les interprètes ont constaté une difficulté à suivre la partition tout en étant conscients des signaux EMG générés car cette configuration implique une attention de la part de l'interprète à d'autres dimensions composables et interprétables auxquelles ils n'avaient pas accès auparavant.

Après la prise en main du système en utilisant un seul électrode, nous avons augmenté la quantité de

capteurs EMG que nous avons placés sur le corps des interprètes pour toutes nos études utilisateurs. Chaque nouvelle électrode représente donc une nouvelle entrée du système donnant à l'interprète une autre dimension composable. Cette nouvelle configuration donne aux interprètes plus de capacités expressives mais ajoute un paramètre additionnel à contrôler. Nous avons pu remarquer une difficulté de la part des interprètes à intégrer chaque nouveau capteur. On constate en même temps une amélioration de la courbe d'apprentissage une fois achevée la prise en main d'un seul capteur.

Lors de nos expérimentations nous avons identifié une difficulté de la part des interprètes à maintenir le signal de contrôle à une valeur constante. Dans nos études utilisateurs nous avons lié chaque niveau d'entrée aux paramètres du système. Malgré une possible dextérité acquise par les interprètes, il est assez laborieux de maintenir le signal EMG à une valeur précise. Dans le cadre du projet BBDMI nous développons des modules pour faire face à cette problématique. Cet ensemble de nouveaux modules offre aux interprètes la possibilité de contrôler et maintenir le niveau du signal de sortie à une valeur précise avec plus d'aisance. Ils permettent parallèlement de maintenir le niveau de sortie même avec le muscle complètement relâché.

4. NÉCESSITÉ D'UNE NOTATION POUR LA MUSIQUE EMG

Comme nous l'avons évoqué dans cet article, le positionnement des électrodes joue un rôle très important pendant la captation du geste instrumental et dans les capacités expressives de l'interprète. Dans les cas où les électrodes sont placées sur des muscles qui ne sont pas nécessaires lors du geste instrumental, la question se pose de savoir quel type de geste doit être exécuté à un moment précis de la partition. Cette situation nous amène à la problématique de la notation pour la musique fondée sur des signaux physiologiques.

La partition telle que nous la connaissons aujourd'hui nous permet de décrire les notes à jouer et les modes de jeu à utiliser, par contre nous n'avons pas de méthode établie pour décrire le geste instrumental désiré à un moment précis.

La nécessité d'une notation mais aussi d'un langage commun se fait évidente lorsque que nous voulons décrire dans la partition la manière dont l'interprète devra interagir avec les traitements sonores au travers de son geste. Nous avons constaté une difficulté de communication entre les musiciens et le compositeur/développeur de l'interface car le langage n'est pas partagé. Côté musicien, le langage est basé sur des paramètres musicaux, tandis que du côté du développeur il peut y avoir une préférence à faire référence aux électrodes et aux intensités des signaux mesurés.

Nous envisageons pour nos futurs ateliers deux alternatives possibles. La première étant de décrire sur la partition le niveau d'engagement musculaire voulu en faisant référence aux électrodes par numéro. La partition devra donc spécifier le positionnement des électrodes. Une autre approche est de décrire le geste instrumental désiré en trouvant un langage commun qui puisse d'une manière concise et claire décrire tous les types de gestes envisagés. Cette dernière approche nous offre une manière plus précise de décrire le geste instrumental mais implique une plus grande difficulté lors de la description du geste dans la partition. Une autre possibilité évoquée lors de nos expérimentations, était de penser la partition comme une vidéo et pas comme une image statique.

5. CONCLUSIONS ET PERSPECTIVES

Dans cet article nous avons détaillé les expériences menées lors des études utilisateurs réalisées dans le cadre du projet BBDMI : nous avons expérimenté de nouvelles manières d'interagir avec les traitements sonores utilisant des capteurs EMG placés sur le corps des interprètes. Nous avons mis en évidence des relations assez complexes entre le geste instrumental et les signaux EMG générés, dépendant du positionnement des électrodes sur le corps de l'interprète.

Nous avons constaté que le travail avec les signaux EMG et le geste instrumental appelle une forme d'écriture qui puisse prendre en compte le geste instrumental désiré ainsi que la musique à jouer.

Les résultats obtenus dans nos études utilisateurs sont similaires à ceux des projets de recherche précédents[12]. Les musiciens ont pu expérimenter une nouvelle façon d'interagir avec la musique, apportant ainsi de nouvelles formes d'expression musicale. Ces résultats soulignent l'importance de continuer à explorer les possibilités de capteurs EMG en tant qu'interfaces pour interagir avec la musique d'une manière plus créative et innovante.

6. REMERCIEMENTS

Le projet BBDMI (Body Brain Digital Music Instruments) est financé par l'Agence Nationale de la Recherche sur la convention ANR-21-CE38-0018. Nous remercions Atau Tanaka, Stephen Whitmarsh et Francesco Di Maggio pour les échanges concernant la phase d'études utilisateurs.

7. RÉFÉRENCES

[1] Bonardi, A. La librairie abclib : un ensemble de codes Faust rassemblant 20 ans de recherche, enseignement et création en musique mixte, *Actes des Journées d'Informatique Musicale 2021*, (Visioconférence), France, juillet 2021. AFIM

[2] Duchêne, J. and Goubel, F. 1993. Surface electromyogram during voluntary contraction: processing tools and relation to physiological events. *Critical Reviews in Biomedical Engineering*. 21, 4 (1993), 313–397.

[3] Erdem, C. et al. 2020. Exploring Relationships Between Effort, Motion, and Sound in New Musical Instruments. *Human Technology*. 16, (Nov. 2020), 310–347. DOI:<https://doi.org/10.17011/ht/urn.202011256767>

[4] Goutmann, P., Bonardi, A. Approaching Spatial Audio Processing by Means of Decorrelation and Ring Modulation in Ambisonics. In Romain Michon, Laurent Potier, and Yann Orlarey, editors, *Proceedings of the Sound and Music Computing Conference 2022*, Saint-Étienne, France, June 2022. Université Jean Monnet, Saint-Étienne, GRAME et INRIA.

[5] Mesin, L. et al. 2009. Surface EMG: The issue of electrode location. *Journal of Electromyography and Kinesiology*. 19, 5 (Oct. 2009), 719–726.

[6] Oku, T., Furuya, S., "A Novel Vibrotactile Biofeedback Device for Optimizing Neuromuscular Control in Piano Playing," 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Osaka, Japan, 2019, pp. 1554-1555, doi: 10.1109/VR.2019.8797765.

[7] Phinyomark, A. et al. 2018. Feature Extraction and Selection for Myoelectric Control Based on Wearable EMG Sensors. *Sensors*. 18, 5 (May 2018), 1615. DOI:<https://doi.org/10.3390/s18051615>.

[8] Sèdes, A. Approche musicale de la décorrelation microtemporelle dans la bibliothèque HOA. *Actes des Journées d'Informatique Musicale 2015*, Montréal, Canada, Mai 2015. Université de Montréal.

[9] Tanaka, A. et al. in review. Brain-Body Digital Musical Instrument Work-in-Progress. ISEA2023 (en revue).

[10] Tanaka, A., Intention, Effort, and Restraint: The EMG in Musical Performance. *Leonardo* 2015; 48 (3):298–299. doi:https://doi.org/10.1162/LEON_a_01018

[11] Vaggione, H. Décorrelation microtemporelle, morphologies et figurations spatiales. *Actes des Journées d'Informatique Musicale*, Marseille, France, mai 2002.

[12] Verdugo, F., Ceglia, A., Frisson, C., Burton, A., Begon, M., Gibet, S., & Wanderley, M. M. (2022). Feeling the Effort of Classical Musicians - A Pipeline from Electromyography to Smartphone Vibration for Live Music Performance. *NIME* 2022. <https://doi.org/10.21428/92fb44.3ce22588>

MYCELIUM GARDEN

Diane Schuh
CICM/MUSIDANSE/EDESTA
MSH Paris Nord UAR 3258
schuh.diane@gmail.com

David Fierro
CICM/MUSIDANSE/
MSH Paris Nord UAR 3258
davidfierro@gmail.com

RÉSUMÉ

Ce projet vise à prendre en compte la singularité d'un réseau de mycélium dans un mode d'interaction avec l'humain qui se fonde sur l'écoute et l'attention. Nous construisons un dispositif pour le développement d'interactions sensibles avec le mycélium, organisme facilitant les symbioses, par une interprétation de la spécificité de ses expressions électriques connectées à notre monde musical humain. Pour construire cette interaction, nous utiliserons des algorithmes de Machine Learning pour cartographier le comportement du mycélium, co-composer des sons à partir des variations électriques enregistrées et classées. Dans un deuxième temps et afin de construire une relation interspécifique et sensible avec le mycélium nous construirons un modèle d'IA pour tenter d'entrer en « communication » avec cet être organique non-humain. Ce modèle tentera de conserver la dimension d'altérité dans l'interaction sensorielle. Dans un processus expérimental nous tentons d'entrer en relation avec un organisme non-humain afin d'observer son comportement et ajuster notre comportement en interaction avec lui, interrogeant ainsi les notions d'intelligence et d'adaptation dans un processus de composition à trois agents humain/IA/mycélium. Quelle serait l'intelligence du mycélium comparée à celle du musicien humain et à celle de l'intelligence artificielle : qu'entendons-nous par-là ? Cette installation aborde l'importance de reconnaître l'altérité des êtres non-humains dans le projet écologique de prendre soin du vivant.

1. INTRODUCTION

Ce projet de recherche-crédation est une exploration interspécifique sur les possibilités de communication sensible et de co-composition entre trois agents (mycélium/IA/humain). Ces agents interagissent dans une relation qui ne cherche pas à partager une identité commune mais qui est fondée sur la notion d'ajustement des différences et de reconnaissance de l'altérité. Cette installation traite de l'importance de la reconnaissance de l'altérité des êtres non-humains et non-animaux dans le projet écologique en facilitant les conditions d'émergence d'une dynamique symbiotique [16].

1.1. Écouter la « musique » des cellules vivantes

Depuis 1873, nous savons que les cellules des plantes conduisent des signaux électriques. Burdon-Sanderson a été le premier à identifier le potentiel d'action (PA) dans les végétaux, dont certains mouvements s'activent par stimulation mécano-induite, comme chez les *Dionaea* [7].

Des musiciens et artistes sonores ont utilisé ces propriétés électriques pour construire des dispositifs à partir de systèmes de transduction du signal électrique [2] [4] [8].

1.2. Mycélium et *fungi*, signaux et réseaux symbiotiques

Le mycélium est un organisme souterrain végétatif dont les fructifications sont les champignons. Dans le cas des champignons mycorrhiziens, le réseau peut alors connecter plusieurs systèmes racinaires de végétaux parfois différents [11] [16]. On observe alors une symbiose : une coexistence durable entre deux organismes dont les interactions sont à bénéfices mutuels [16].

Des scientifiques ont mis en évidence des potentiels d'action chez les champignons. En 1995 Olsson et Hansson ont mesuré des signaux électriques dans le mycélium de *Neurospora crassa* [12]. Dehshibi et Adamatzky ont découvert par une analyse multicanale la complexité de systèmes de transmission du signal dans le réseau de mycélium de *Pleurotus ostreatus* [1]. Cette extrême diversité pouvant alors être comparée à la complexité de langages humains, en la surpassant [5].

2. RECONNAITRE L'ALTERITE DU VIVANT : INVENTER DES MODALITES D'INTERACTIONS AVEC DES ETRES ORGANIQUES NON-HUMAINS

2.1. Au-delà de l'anthropomorphisme

À ce jour nous n'avons pas connaissance de l'existence de recherches sur la composition musicale à partir des signaux électriques produits par un réseau de mycélium qui rende compte de la complexité mise en

évidence par Dehshibi et Adamatzky. De plus, ces démarches relèvent souvent de l'anthropomorphisme, voire de l'anthropocentrisme, d'une personnification du vivant non-humain.

Notre projet cherche à développer un dispositif pour des modalités d'interactions sensibles et spécifiques avec le mycélium et ses fructifications, en passant par une interprétation de la singularité de leurs expressions électriques [3] [1].

2.2. Rendre compte des altérités

Notre projet vise à prendre en compte l'altérité, la singularité et la multiplicité d'un réseau de mycélium dans un mode d'interaction avec les humains basé sur l'écoute et l'attention. Il s'agit de rendre audibles les dimensions complexes et invisibles de ces êtres organiques souvent perçus comme inertes et d'inventer un modèle d'interaction qui facilite chez l'humain la sensibilité pour autre que lui. Ce modèle ne cherchera pas à induire une réaction d'empathie anthropomorphique mais tentera de maintenir la dimension d'altérité dans l'interaction non-verbale, créant ainsi un nouvel espace sonore compris comme une hybridation d'individuations symbiotiques, ce qui nous donnera également l'occasion d'examiner et d'approfondir la notion d'individuation symbiotique dans ce travail.

3. METHODOLOGIE

Le projet consiste à construire les conditions d'émergence d'un réseau de mycélium à partir d'une souche récoltée dans le jardin de la Maison des Sciences de l'Homme Paris Nord (MSH PN), à adapter une technologie conçue pour l'homme à un organisme vivant non animal et non végétal, et à trouver une cartographie adéquate de leur PA à l'aide d'un modèle IA, afin de produire une composition musicale qui invite à l'écoute et à l'attention pour le vivant.

3.1. Porter attention au vivant dans un jardin de mycélium

Le cœur du projet est la création d'une installation sonore interactive en co-création avec un réseau de mycélium dont l'émergence, le développement et les conditions de vie ont été organisés par l'humain.

Le développement du jardin de mycélium est planifié sur une période de huit mois. Ces travaux seront réalisés en trois phases. La phase de recherche comprend le recueil du mycélium dans le jardin de la MSH, l'extraction de la souche et le clonage dans des boîtes de Pétri stériles. La deuxième phase comprend la culture en bocaux stériles et la première multiplication du mycélium.



Figure 1. *Agaricus sp.* « Hortus Annae », champignon à partir duquel nous clonons actuellement le mycélium. Photo ©Diane Schuh.

Enfin, la dernière phase comprend la multiplication du mycélium en 3 étapes (de 0,1 m³ à 1m³ selon la taille de l'installation) ainsi que le transport du mycélium du laboratoire à la salle d'exposition, le montage du substrat et de la structure de structure scénographique. Pendant ces 8 mois, nous sommes accompagnés par Spora Studio en tant que consultant. Ils ont en charge le conseil et le suivi technique de la partie organique de l'installation.



Figure 2. Les 16 boîtes de Pétri du clone d'*Agaricus sp.* « Hortus Annae ». Photo ©Stephen Whitmarsh.

Nous sommes actuellement dans la seconde phase. Nous avons identifié et collecté différents champignons présents dans le jardin de la MSH Paris Nord. Le champignon *Agaricus sp.* « Hortus Annae¹ » que nous avons trouvé adéquat (Figure 1) pour notre projet a commencé à être cloné dans des boîtes de Pétri. Dans cette première étape, nous avons eu un succès d'une souche sur six. Nous avons ensuite fait un deuxième transfert qui nous a permis de cloner et développer le mycélium dans seize boîtes de Pétri (Figure 2). Le mycélium croît actuellement à un rythme régulier, validant notre première hypothèse, à savoir la faisabilité

¹ Au moment de l'écriture de cet article nous attendons la fructification du clone afin d'effectuer une analyse ADN pour déterminer l'espèce

d'Agaric que nous avons trouvé au jardin. « Hortus Annae » est le nom que nous lui avons donné.

du clonage d'un champignon extrait du jardin de la MSH Paris Nord.

Nous avons l'intention de passer au début de la phase 3 au moment des JIM 2023 : nous serons en mesure de présenter plusieurs souches de mycélium qui auraient formé des réseaux émergents que nous pourrions monitorer afin de construire le modèle d'intelligence artificielle, le mapping et la composition musicale.

3.2. Enregistrer le signal du mycélium : essais et protocoles

3.2.1. Adapter un dispositif conçu pour l'humain : le choix du matériel

Pour faciliter la création d'un environnement de co-composition entre les trois agents (mycélium/IA/humain) il est indispensable d'établir un protocole de captation et de traitement des signaux qui puisse nous permettre de mesurer les variations électriques intrinsèques du mycélium en évitant des bruits instrumentaux. Les signaux électriques du mycélium peuvent varier en amplitude et en dynamique. Cependant, ces signaux ont une intensité assez faible et une évolution lente dans le temps [17]. Les caractéristiques de ces signaux impliquent un grand défi technique lors de sa captation. Pour répondre à ce défi, nous nous approprions des technologies de captation de signaux EEG et EMG². Ces dispositifs partagent une très bonne relation signal/bruit, une grande capacité d'amplification et la capacité à éviter ou filtrer des bruits instrumentaux [10].

Dans le cadre de l'ANR BDDMI³, des prototypes permettant de mesurer les signaux électriques du muscle et du cerveau (EMG et EEG) ont été développés en 2022. Nous nous appuyons sur ces technologies et l'expertise testée et éprouvée du projet BDDMI pour développer la cartographie sonore de l'activité du mycélium de l'installation. En utilisant la carte EAVI [6] et des électrodes à aiguilles subdermiques, nous avons testé la faisabilité de ces techniques avec le réseau de mycélium [11]. La carte EAVI possède 4 entrées analogiques bipolaires qui sont traitées par un convertisseur analogique-numérique (ADC) avec un bon rapport signal/bruit (S/N) et une forte amplification. La carte transmet le signal d'entrée brut à un taux d'échantillonnage de 8Khz par USB avec un module d'isolation optique entre la carte et l'ordinateur.

Pour notre projet, nous utilisons deux cartes EAVI, ce qui nous permettra de surveiller 8 intersections du réseau de mycélium et donc de rendre compte de la diversité de ses expressions électriques.

3.2.2. Enregistrer les signaux électriques du mycélium

Lorsque nous effectuons des mesures pour détecter les signaux émis par le mycélium, il est important de mener des analyses approfondies pour nous assurer que les signaux détectés proviennent effectivement du mycélium et ne sont pas simplement du bruit ou des signaux électromagnétiques étrangers. Si nous ne faisons pas suffisamment d'analyses pour distinguer les signaux du mycélium des autres sources possibles, nous risquons de détecter des variations qui ne sont pas réellement liées à l'activité du mycélium.



Figure 3. Monitoring du mycélium de la boîte à champignon au moyen de deux types d'électrodes⁴.

Il existe plusieurs méthodes pour s'assurer que les signaux détectés proviennent effectivement du mycélium. Par exemple, nous pouvons utiliser des techniques de filtrage pour éliminer les signaux indésirables, ou encore réaliser des mesures de contrôle pour nous assurer que les variations que nous détectons sont bien spécifiques au mycélium.

Nous avons décidé de deux protocoles de test pour mesurer les signaux électriques du mycélium en essayant d'éliminer au maximum les possibles sources de bruit externes. Dans notre premier protocole nous avons utilisé des électrodes-aiguilles sous-cutanées⁵ et des électrodes en cuivre pour comparer les résultats (Figure 3). Ces mesures ont été prises sur le mycélium de *Pleurotus ostreatus*, notre organisme modèle⁶. Nous avons sélectionné ce type d'électrodes car elles sont stériles ce qui nous permettra de réduire les risques de contamination dans les boîtes de Pétri. De plus, elles sont assez fines et rigides pour pouvoir s'introduire précisément dans le réseau de mycélium.

Pour le deuxième protocole de test nous avons placé le mycélium et un objet de contrôle dans une cage de Faraday pour minimiser les bruits électromagnétiques près des aiguilles. Nous avons utilisé les électrodes-aiguilles sous-cutanées avec protection contre les bruits.

² Electroencéphalogramme/Électromyogramme

³ ANR-21-CE38-0018, <https://bbdmi.nakala.fr/>

⁴ Photo prise en dehors de la cage de Faraday pour les besoins de l'illustration. le 21 décembre 2022.

⁵ « Disposable Subdermal EEG needle electrode » 13 x 0.40mm (27gauge), Stainless Steel, 1.5m (60inch), PVC wire, DIN42802 connector., selon le protocole de Dehshibi et Adamatzky [5].

⁶ En attendant le développement de la souche clonée du jardin, nous avons utilisé une « boîte à champignon » avec du mycélium de *Pleurotus ostreatus* pour tester les protocoles d'enregistrement des PA.

Les aiguilles ont tendance à être plus sensibles aux bruits électromagnétiques car elles n'ont pas de blindage électromagnétique contrairement aux câbles des électrodes.



Figure 4. Monitoring du mycélium cloné depuis le champignon *Agaricus « Hortus Annae »* du jardin de la MSH et monitoring d'un organisme contrôle (éponge). le 16 janvier 2023.

Les trois électrodes ont été connectées sur la surface du mycélium à deux millimètres de profondeur approximativement en essayant de ne pas toucher le substrat car cela pourrait nuire à la qualité du signal et avoir pour conséquence la réduction de l'impédance entre les électrodes.

En parallèle des électrodes connectées au mycélium, nous avons placé une éponge humide utilisée comme objet-contrôle. Les câbles ont été placés l'un à côté de l'autre afin d'avoir le même bruit électromagnétique dans les deux électrodes. Cette configuration nous permet de déterminer *a posteriori* la quantité et le spectre du bruit électromagnétique aux alentours du mycélium. Malgré le fait que l'éponge et le mycélium n'ont pas la même impédance, le signal perçu par les électrodes placés sur l'éponge peut nous donner assez d'informations pour éliminer de possibles perturbations futures en temps réel. Nous comptons implémenter cette étape de filtrage en utilisant des techniques d'apprentissage machine. L'éponge humide a été choisie comme objet de test car elle n'est pas vivante et peut être utilisée comme référence pour les comparaisons avec les signaux électriques du champignon. On pourrait donc déduire que les perturbations présentes dans les deux électrodes font partie du bruit instrumental. La figure 4 montre la boîte de Pétri à côté de l'éponge avec les trois électrodes-aiguilles sous-cutanées.

3.2.3. Les signaux enregistrés

La figure 5a montre le signal enregistré au moyen des électrodes placées sur le mycélium *Pleurotus ostreatus*.

⁷ Le canal 1 est enregistré avec les électrodes en cuivre et le canal 2 avec les électrodes-aiguilles sous-cutanées. Lors de cette expérience, nous avons pu identifier quelques pics dans la forme d'onde. Nous pouvons également les écouter à 0:04 et 0:25 du fichier: https://dianeschuh.fr/01_spikesMushroom48-Brut.wav

Nous pouvons apercevoir des perturbations sporadiques en même temps que la forte présence d'un signal qui, à première vue, ressemble à du bruit électromagnétique. Il est possible que les données collectées présentent un signal complexe qui comprend à la fois des informations utiles provenant du mycélium ainsi que des perturbations ou des interférences causées par le bruit instrumental. Ce bruit instrumental est composé de tout type d'interférence ou de perturbation introduite dans les données par les instruments de mesure utilisés pour les collecter.

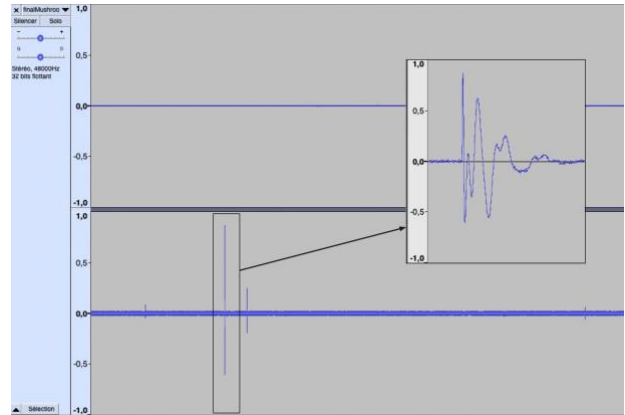


Figure 5a. Visualisation de l'enregistrement du signal de *Pleurotus ostreatus* (boîte à champignon) enregistré le 21 décembre 2022⁷.

La figure 5b montre le signal enregistré sur les électrodes placées sur le mycélium *Agaricus « Hortus Annae »*. Nous remarquons un signal similaire à celui du mycélium *Pleurotus ostreatus* mais avec l'absence des piques de voltage. Nous allons poursuivre nos expériences pour déterminer si ces grandes différences de potentiel proviennent du mycélium. Il est important de noter que la figure 5a montre le signal enregistré lors de la phase de fructification du mycélium *Pleurotus ostreatus*, tandis que la figure 5b montre le signal du mycélium *Agaricus « Hortus Annae »* encore en développement en boîte de petri.

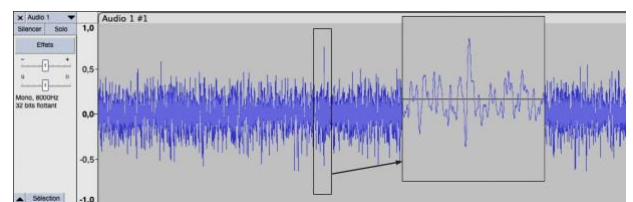


Figure 5b. Visualisation de l'enregistrement du signal d'*Agaricus « Hortus Annae »* (clone du jardin de la MSH), le 16 janvier 2023.

Dû à la nature des possibles signaux électriques générés par le mycélium, il peut être difficile de distinguer les composantes du signal qui proviennent du

Nous restons cependant très prudents quant à l'interprétation que nous pourrions faire de ces pics. Ceux-ci pourraient provenir de mouvements mécaniques du substrat induits par l'humidité, et ne pas être directement liés à l'activité du mycélium.

mycélium de celles qui sont causées par le bruit instrumental, car elles peuvent se chevaucher et avoir des caractéristiques similaires. Cependant, il existe des méthodes d'analyse de données avancées qui peuvent aider à séparer les différentes composantes du signal et à identifier celles qui sont les plus importantes ou les plus représentatives du mycélium. Nous envisageons d'implémenter ces techniques dans la prochaine phase de notre projet.

3.3. Faciliter les communications interspécifiques en utilisant le Machine Learning et l'IA

Dans le cadre de cette recherche expérimentale, nous testerons différents processus de construction d'une communication inter-espèces. Au moment d'écrire cet article, nous entamons une phase d'observation. Pour établir une relation interspécifique avec le mycélium, nous devons observer et enregistrer son activité sur des périodes de plusieurs heures. Pendant cette phase, nous enregistrons et analysons le potentiel d'action du mycélium. Un modèle d'apprentissage automatique basé sur l'apprentissage non supervisé nous aidera à catégoriser et à classer les variations de l'activité électrique du mycélium. Nous serons alors en mesure de déterminer l'échelle de temps et la fréquence des motifs (*patterns*) qui émergent de l'activité électrique du mycélium. Pour cette partie du projet nous nous associons à Hugo Scurto, ingénieur de recherche sur le projet, qui rejoindra l'équipe en avril 2023. Hugo Scurto va développer un modèle d'IA à partir de nos enregistrements. Nous pourrions alors construire un mapping qui permettra au mycélium d'interagir avec les différents moteurs sonores composés. Ces modèles incluent des interpolations linéaires, régressives et d'apprentissage par renforcement [15].

Ce type de machine learning permet aux mappings d'évoluer et de s'adapter en temps réel et en interaction entre les agents [13]. Nous pensons que l'adaptabilité de ce modèle rendra compte de manière plus précise de la spécificité et la diversité de l'activité du réseau de mycélium. En investiguant ces hypothèses, nous tenterons d'examiner la notion de symbiose mais aussi l'ontologie de l'IA.

À la suite de cette phase de recherche, nous serons en mesure d'installer notre dispositif. Les sons co-composés avec l'IA seront modulés par l'activité du réseau de mycélium. La faisabilité d'une interaction en temps réel, compte tenu de la lenteur de l'activité du mycélium, sera une hypothèse à tester lors de cette phase du projet. Les phases de recherche et de co-composition nous permettront de comprendre les différents types de comportements et d'interactions dans une relation mycélium/IA/humain. Cette phase permettra d'expérimenter et d'analyser les différents types de comportement en fonction des types d'organismes impliqués. Peut-on parler ici d'intelligence ou

d'adaptation ? Nous trouverons des stratégies pour rendre compte de la diversité, notamment de la diversité fonctionnelle du mycélium.

3.4. Composer avec le signal du mycélium : synthèse et « archéologie sonore »

Nous nous sommes demandé comment connecter le signal du mycélium à notre monde musical humain ; quelles médiations sensibles et quels choix compositionnels pouvons-nous proposer ?

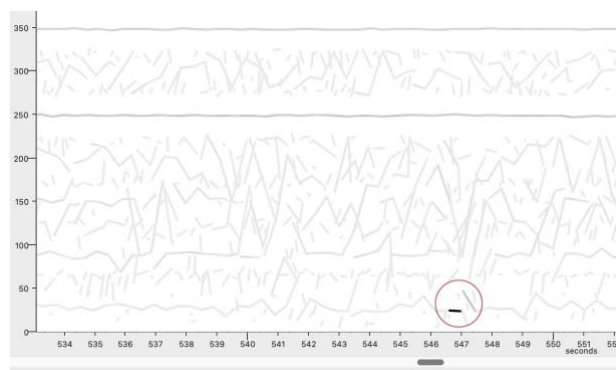


Figure 6. Analyse FFT du signal du mycélium au moyen du logiciel SPEAR, recherche de singularités (entourée en rouge).

Dans cette première phase de recherche-crédation nous avons utilisé l'enregistrement du bruit (le mycélium et son milieu) comme matériau. À partir d'une analyse FFT effectuée dans le logiciel SPEAR nous avons cherché des singularités fréquentielles. Au sein du bruit nous avons pu identifier des intensités rares mais remarquables dans les bandes de fréquences inférieures à 50 Hz (Figure 6).

À partir de cette analyse nous souhaitons développer une poésie sonore inspirée par le mycélium et son milieu : le sous-sol. En utilisant la technique de synthèse par transformation sonographique⁸ (Figure 7), nous avons cherché à mettre en valeur les basses fréquences. Pour cela nous avons travaillé à la composition de contrastes par effacement et réorganisation, en agissant sur la représentation graphique du son. Nous avons testé la composition de pédales faites de basses fréquences mises en contraste avec des textures granulaires et pointillistes dans les hautes fréquences⁹.

Le travail de composition des basses fréquences s'inscrit également dans l'hypothèse de l'établissement d'un « contact » avec le mycélium. La diffusion des bandes sonores composées par la compositrice et modulées par le mycélium se fera sur un système hybride WFS/Ambisonic renforcé par 4 caissons de basses. Nous travaillons ici sur la possibilité d'une réaction de cet organisme aux vibrations des basses fréquences, établissant alors une première boucle de rétroaction par réaction du mycélium à la composition qui sera alors modulée différemment par de nouveaux pics de PA. Cette

⁸ *sonographical transformation* in Roads Curtis [14]

⁹ Deux échantillons sonores composés à partir du bruit du mycélium à écouter: https://dianeschuh.fr/02_48Khz_Mycelium_sonographic.wav et https://dianeschuh.fr/03_48Khz_Mycelium_sonographic.wav

hypothèse sera testée lors de la présentation de l'installation aux JIM23¹⁰.

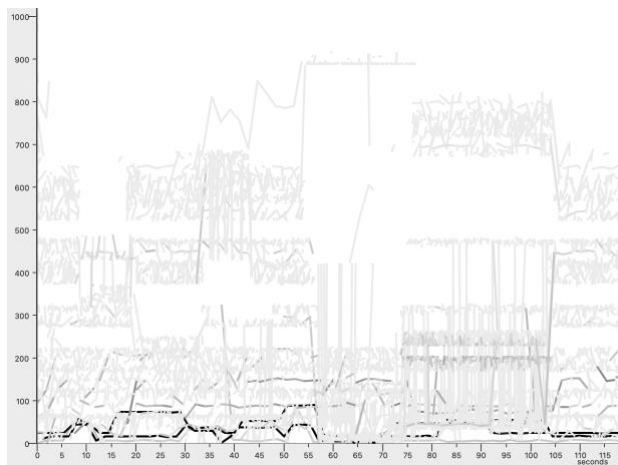


Figure 7. Synthèse par transformation sonographique, par réorganisation des partiels dans le logiciel SPEAR.

Ce travail en cours est une exploration de ce que pourrait être la poésie d'une polyphonie du sous-sol ainsi que le développement d'une composition par ce que nous définissons comme une « archéologie sonore¹¹ ».

Ces échantillons composés serviront de substrat sonore à un patch Max dont les variables seront modulées par l'activité du mycélium¹².

4. PRESENTATION D'UN TRAVAIL EN COURS

Le projet « Mycelium Garden » est soutenu par la MSH Paris Nord et ArTec dans le cadre de son appel à projet « Symbiose ». La MSH Paris Nord accueille ce projet dans ses locaux durant l'année 2023. Pour les JIM 2023, nous proposons de présenter un point d'étape de notre recherche expérimentale. Nous proposons d'exposer une première modalité de présentation de l'installation sonore interactive. Plusieurs boîtes de Pétri avec du mycélium cloné à partir d'*Agaricus « Hortus Annae »* et plusieurs bocaux en verre avec des plantes et leurs réseaux mycorrhiziens transplantés du jardin de la MSH formeront notre jardin de mycélium. Ces organismes seront connectés par des électrodes à deux cartes EAVI, ce qui donnera huit canaux de monitoring des PA. L'activité de ces huit canaux sera interprétée par le modèle d'IA mappé à partir des interactions qu'auront engagé la compositrice et le mycélium lors de la phase de composition de l'architecture. Nous pourrions entendre les premiers résultats de l'installation sonore : la composition mycélium/AI/humain.

¹⁰ Un rendu stéréo d'une première esquisse de 19 pistes composée à partir du substrat de basse fréquences extrait dans SPEAR et transformé dans Max :

https://dianeschuh.fr/MYCELIUM_MULTIPISTE-essai1-render2.wav

¹¹ Nous définissons ici l'archéologie sonore non pas comme une reconstruction de paysages sonores du passé mais comme une manière d'extraire au sein du bruit les singularités de ce que nous imaginons (à ce stade de nos recherches) être l'expression fréquentielle du signal du

Les objets et l'installation développés dans ce projet serviront à repenser la question de la communication inter-espèces, conçue non pas comme un processus basé sur une culture et une intentionnalité communes, mais comme une forme de coproduction de sens par « ajustement de différences » [9]. Ce dispositif invitera également à s'interroger sur les enjeux esthétiques des formes générées automatiquement ou, dans le cas présent, guidées par des processus non-humains, en montrant la fragmentation et la dissolution de l'intentionnalité dans ce type de productions, qui synthétisent et explorent des niches inexplorées de nos espaces culturels. En intégrant le développement IA dans ce contexte, il s'agit aussi de chercher à comparer les intelligences : que voulons-nous dire lorsque nous parlons d'intelligence ? Quelle serait l'intelligence ou la capacité d'adaptation du mycélium par rapport à celle du musicien humain et à celle de l'intelligence artificielle [19] : qu'entendons-nous par-là ?

5. REMERCIEMENTS

Nous remercions les organismes de financement du projet de recherche « Mycelium Garden » : l'EUR ArTec et la MSH Paris Nord, ainsi que l'équipe réunie autour de ce projet : Anne Sèdes, Alain Bonardi, David Fierro et Adrien Zanni du CICM ; Charlotte Janis et Roberto dell'Orco de Spora ; Guillaume Peureux et Julien Schuh du CSLF Nanterre ; Stephen Whitmarsh de l'ICM ; Hugo Scurto d'ArTec.

6. RÉFÉRENCES

1. Adamatzky Andrew, « On spiking behaviour of oyster fungi *Pleurotus djamor* », *Nature, Scientific Reports*, vol. 8, no 1, (2018): 7873.
2. Agapea Saša Spačal, « Infrastructures for Each Other », consulté le 29 novembre 2022, <https://www.agapea.si/en/projects/infrastructures-for-each-other>.
3. Barbara Ricardo, « Electrical currents associated with arbuscular mycorrhizal interactions », *New Phytologist*, vol. 129, n°3, (1995) : 433-438.
4. Darren Moore, Guy Ben-Ary, Andrew Fitch, Nathan Thompson, Douglas Bakkum, Stuart Hodgetts & Amanda Morris « cellF: a neuron-driven music synthesiser for real-time performance », *International Journal of Performance Arts and Digital Media* 12:1, (2016): 31-43.
5. Dehshibi Mohammad Mahdi, Adamatzky Andrew, « Electrical activity of fungi: Spikes detection and complexity analysis », *Biosystems*, vol. 203, (2021): 104373.

mycélium. L'extraction des singularités fréquentielles est le point de départ d'une démarche de réorganisation compositionnelle qui se base sur une poésie du sous-sol.

¹² À ce stade de notre recherche, l'instrument n'a pas encore été composé. Le choix des traitements et les types de variables à moduler sont encore à définir. Ces choix seront précisés lors de la phase de conception de l'IA, en interaction avec le travail de l'ingénieur de recherche.

6. Di Donato Balandino et al., “EAVI EMG board”, (2019).
7. Fromm Jörg, Lautner Silke, « Electrical signals and their physiological significance in plants », *Plant, Cell & Environment*, vol. 30, no 3 (2007): 249-257.
8. Garcia Leslie, « INTERSPECIFICS », consulté le 29 novembre 2022, <https://github.com/interspecifics>.
9. Guillo Dominique, *Les fondements oubliés de la culture: une approche écologique* (Paris, France: Éditions du Seuil, 2019).
10. Metting Van Rijn A.C., Peper A. et Grimbergen C.A., « Amplifiers for bioelectric events: A design with a minimal number of parts », dans *Medical and Biological Engineering and Computing*, vol. 32, n° 3, 1 mai 1994, p. 305-310, [En ligne], <<https://doi.org/10.1007/BF02512527>>, (Consulté le 24 février 2023)
11. Miyauchi Shingo et al., « Large-scale genome sequencing of mycorrhizal fungi provides insights into the early evolution of symbiotic traits », *Nature Communications*, vol. 11, no 1, (2020): 5125.
12. Olsson Stefan, Hansson Bill, « Action Potential-Like Activity Found in Fungal Mycelia Is Sensitive to Stimulation », *Naturwissenschaften*, vol. 82, (1995): 30-31.
13. Radford Alec et al., « Learning Transferable Visual Models From Natural Language Supervision », *arXiv*, (2021), Consulté le 3 décembre 2022, <http://arxiv.org/abs/2103.00020>.
14. Roads Curtis, *Composing electronic music: a new aesthetic*, New York, Etats-Unis d'Amérique, 2015.
15. Scurto Hugo, « Designing With Machine Learning for Interactive Music Dispositifs », (Ph.D. diss. Sorbonne université, 2019).
16. Selosse Marc-André, *La symbiose: structures et fonctions, rôle écologique et évolutif* (Paris, France: Vuibert, 2000).
17. Slayman C, Long W, Gradmann D. 1976. « “Action potentials” in *Neurospora crassa*, a mycelial fungus ». *Biochimica et Biophysica Acta* 426: 732–44.
18. Thompson, Douglas Bakkum, Stuart Hodgetts & Amanda Morris « cellF: a neuron-driven music synthesiser for real-time performance », *International Journal of Performance Arts and Digital Media* 12:1, (2016): 31-43.
19. Turing Alan, « Computing Machinery and Intelligence », dans *Mind*, vol. LIX, n° 236, 1 octobre 1950, p. 433-460, [En ligne], <<https://doi.org/10.1093/mind/LIX.236.433>>, (Consulté le 27 février 2023).

M.I.X.C.S. MUSIC, INTERNET AND COGNITIVE SCIENCES

Pierre-Henri Vulliard

SCRIME / Arts Sciences Information Sarl
pierre-henri.vulliard@u-bordeaux.fr

RÉSUMÉ

Dans cet article, nous présentons une approche intégrée pour la création et la performance musicale, combinant l'électroencéphalographie (EEG), les stations de travail audio numériques (DAW) telles que Ableton Live, Max for Live (M4L) et Serato DJ, et les techniques de classification.

L'objectif principal est d'examiner comment cette approche peut enrichir la création musicale, offrir de nouvelles perspectives pour la recherche sur la cognition musicale et élargir notre compréhension des processus cognitifs impliqués dans la musique. La solution choisie est de créer une boucle de rétroaction adaptative, permettant d'influencer l'état émotionnel et cognitif du musicien pendant la performance, afin d'optimiser les phénomènes psychiques et musicaux.

Nous utilisons un casque EEG Emotiv pour mesurer l'activité cérébrale et des logiciels tels qu'Emotiv BCI et Emotiv Pro pour analyser les signaux EEG et les traduire en échelles émotionnelles et cognitives.

L'API OpenAI est également intégrée pour enrichir le processus créatif en générant des images et des poèmes basés sur l'état émotionnel du musicien.

Les expérimentations préliminaires sont menées avec l'auteur comme utilisateur principal, dans le cadre de performances temps réel en ligne. Cette étude vise à explorer les potentialités de cette approche intégrée pour améliorer la qualité de la composition et de la performance, et ouvrir de nouvelles perspectives dans les domaines artistiques.

1. INTRODUCTION

1.1. Contexte et motivation

La musique est un langage universel qui a le pouvoir d'évoquer des émotions et d'influencer notre humeur. Depuis les premières formes d'expression musicale jusqu'aux développements technologiques modernes, la musique a toujours été un moyen essentiel pour les humains de communiquer et d'exprimer leurs émotions. Cependant, la compréhension des mécanismes sous-jacents qui régissent la relation entre les émotions, les processus cognitifs et la création musicale est encore limitée.

Au cours des dernières décennies, les neurosciences et l'informatique musicale ont fait des progrès

considérables dans l'étude de ces mécanismes, offrant des opportunités uniques pour explorer et exploiter les liens entre les processus cérébraux, les émotions et la création musicale. Les avancées dans les domaines de l'électroencéphalographie (EEG) et des stations de travail audio numériques (DAW) [2] ont permis d'obtenir de nouvelles perspectives sur l'interaction entre le cerveau et la musique, ainsi que sur les effets de la musique sur la situation émotionnelle et mentale des musiciens.

Outre son utilisation dans la recherche sur la cognition musicale, l'EEG a également été utilisé pour créer des compositions et des performances basées sur les réponses cérébrales des musiciens et du public. Par exemple, des compositions basées sur les données EEG ont été créées en utilisant les rythmes et les motifs de l'activité cérébrale pour générer des séquences musicales[3].

Dans ce contexte, cette étude vise à démontrer l'utilisation combinée de l'électroencéphalographie, des stations de travail audio numériques et des méthodes de classification pour optimiser l'état émotionnel du musicien pendant la performance. En intégrant ces technologies et méthodologies, il est possible d'obtenir une compréhension plus approfondie des liens entre les émotions, les processus cognitifs et la musique, et d'explorer de nouvelles formes d'expression artistique basées sur ces liens.

1.2. Objectif de la recherche

L'objectif principal de cette recherche est de créer un système de rétroaction pour optimiser l'état émotionnel du musicien pendant la composition et la performance musicale. Le système combine l'analyse des signaux EEG avec les DAW et les méthodes de classification pour créer une boucle de rétroaction adaptative qui modifie la composition musicale et la performance en fonction des émotions et de l'état cognitif du musicien.

Le système de rétroaction vise aussi à fournir aux musiciens des informations en temps réel sur leur état psychologique, leur permettant d'ajuster leur performance en conséquence. Cette approche intégrée a le potentiel d'améliorer la qualité de la création musicale et de la performance en permettant aux musiciens d'explorer et de maîtriser les interactions entre leur état émotionnel, leur état cognitif et leur expression musicale.

2. MÉTHODOLOGIE

2.1. Acquisition des données EEG

La première étape de cette étude consiste à acquérir les données EEG à l'aide d'un casque EEG Emotiv (Figure 1). Ce dispositif portable non invasif permet de mesurer l'activité électrique du cerveau en temps réel, offrant une méthode fiable pour surveiller les processus cognitifs et émotionnels du musicien pendant la composition et la performance. Les signaux EEG sont enregistrés à partir de plusieurs électrodes placées sur le cuir chevelu, fournissant une représentation détaillée de l'activité cérébrale dans différentes régions du cerveau.



Figure 1. Casques Emotiv

2.2. Analyse des signaux EEG

Une fois les données EEG acquises, elles sont analysées à l'aide du logiciel Emotiv, qui transforme les signaux bruts en paramètres émotionnels et cognitifs, telles que le stress, la relaxation, l'excitation, l'engagement, la concentration et l'intérêt. Ces métriques sont calculées en examinant les modèles d'activité cérébrale et en les comparant aux normes préétablies pour les différentes émotions et états cognitifs. L'analyse des signaux EEG permet ainsi de quantifier et de caractériser les conditions émotionnelles et mentales du musicien tout au long de l'expérience.

2.3. Visualisation des données EEG

Les données EEG analysées sont ensuite visualisées en temps réel sous forme de signaux EEG bruts (Figure 2), ainsi qu'à l'aide d'un modèle de cerveau 3D (Figure 3) qui affiche les fréquences, les couleurs correspondant aux différents types d'ondes (Delta, Theta, Alpha, Beta...). Cette visualisation permet au musicien de suivre l'évolution de ses conditions psychologiques et mentales pendant la performance, et d'identifier les facteurs qui influencent sa créativité et son expression musicale.

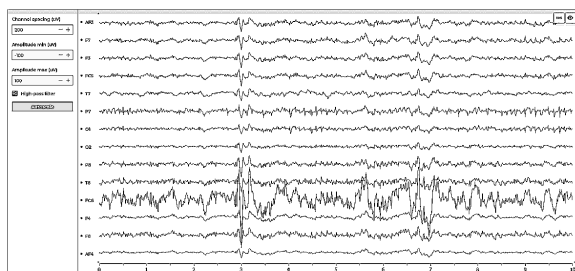


Figure 2. Exemple de fréquences EEG brutes



Figure 3. Fréquences EEG sur un modèle 3D

2.4. Intégration des DAW et mise en place du feedback

Les DAW sont des logiciels qui permettent la composition, l'enregistrement, la production et la diffusion en live. Ils offrent une plate-forme puissante et flexible pour la création musicale, avec des fonctionnalités avancées telles que la manipulation d'échantillons, la synthèse, les effets et le séquençage[4].

L'intégration des DAW et la mise en place du feedback sont essentielles pour créer une boucle de rétroaction adaptative entre les dynamiques émotionnelles et cognitives du musicien et la composition musicale. Dans cette section, nous détaillons les différentes étapes de cette intégration.

2.4.1. Utilisation du protocole OSC

Le protocole Open Sound Control (OSC) est utilisé pour faciliter la communication entre les différents logiciels et composants du système. L'OSC est un protocole de communication basé sur le réseau qui permet l'échange d'informations entre des ordinateurs, des synthétiseurs, des contrôleurs et d'autres dispositifs multimédias. Dans notre approche, l'OSC est utilisé pour transmettre les données EEG et les valeurs émotionnelles et cognitives entre Emotiv BCI et les DAW, tels qu'Ableton Live et M4L.

2.4.2. Intégration de Max for Live (M4L)

Max for Live (M4L) est l'environnement de programmation graphique Max-MSP intégré dans Ableton Live (Figure 4). Il permet aux utilisateurs de créer et d'éditer des instruments, des effets et des dispositifs MIDI personnalisés. Pour cette étude, nous avons développé plusieurs dispositifs Max for Live qui reçoivent les données OSC provenant d'Emotiv BCI et les convertissent en contrôleurs MIDI pour manipuler les paramètres de la composition musicale, ainsi qu'en commandes de contrôle envoyées en UDP pour les autres aspects artistiques et techniques du spectacle.

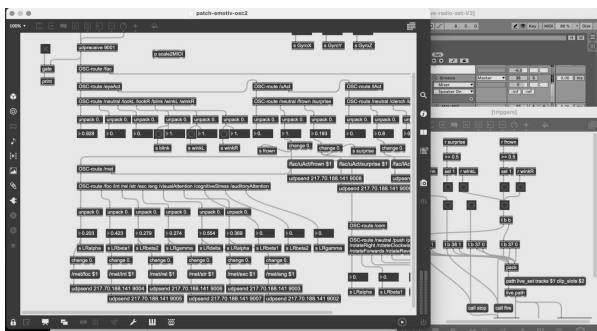


Figure 4. Max-for-Live (M4L)

2.4.3. Mise en place du feedback

Une fois les données OSC reçues dans M4L, elles sont utilisées pour contrôler différents aspects de la composition et de la performance musicale, tels que le déclenchement de sons, la modification des effets audio [4], et la manipulation des paramètres de mixage dans Ableton Live et Serato DJ (Figure 5). En fonction des mesures des processus émotionnels et cognitifs, le système de rétroaction adaptative ajuste la composition pour optimiser le bien-être psychologique et intellectuel du musicien. Cette boucle de rétroaction permet au musicien d'explorer et de maîtriser les interactions entre son état d'esprit et son expression musicale.



Figure 5. Serato DJ

2.4.4. Interactions entre les différentes composantes

L'intégration réussie des DAW, du protocole OSC et de M4L permet de créer une boucle de rétroaction adaptative et interactive. Cette boucle de rétroaction relie les données EEG, les variables émotionnelles et cognitives, les logiciels de traitement et d'analyse des données, les DAW et les dispositifs M4L personnalisés pour créer une expérience musicale dynamique et immersive, basée sur les « états d'âme » du musicien.

Cette intégration devrait permettre aussi d'étudier la cognition musicale d'une manière plus écologique et interactive que les approches traditionnelles basées sur l'écoute passive de stimuli musicaux. Cette approche offre également la possibilité d'examiner les processus cognitifs impliqués dans la création et l'interprétation en temps réel, plutôt que de se concentrer uniquement sur la perception et la mémoire de la musique

2.5. Implémentation du système de classification

Un aspect crucial de cette recherche est l'application d'un système de classification pour améliorer la précision et la pertinence de la rétroaction. Un système de classification est une technique qui emploie des algorithmes d'apprentissage automatique [1] pour identifier des modèles spécifiques dans un ensemble de données et les associer à des catégories prédéfinies. Dans le cadre de cette étude, le système de classification est utilisé pour reconnaître des modèles spécifiques dans les indicateurs émotionnels et cognitifs, et les associer à des paramètres musicaux pertinents [5]. Ce processus permettra de fournir une rétroaction plus précise et personnalisée au musicien, en adaptant la musique en fonction de son état individuel.

2.6. Intégration de l'API OpenAI

Dans cette étude, nous intégrons à titre exploratoire l'API OpenAI pour enrichir le processus créatif et exploiter les capacités de l'intelligence artificielle dans la création musicale. L'API OpenAI est utilisée pour associer des mots-clés aux données émotionnelles et mentales du musicien, et les envoyer via des scripts Python à la plateforme en ligne OpenAI. En fonction des paramètres émotionnels et intellectuels du musicien, l'API OpenAI génère ainsi automatiquement des images et des poèmes qui reflètent les émotions et les pensées du musicien. Ces éléments générés par l'intelligence artificielle peuvent ensuite être intégrés dans la performance, offrant de nouvelles possibilités d'expression artistique et de communication des émotions.

2.7. Expérimentation et évaluation

Afin d'évaluer l'efficacité du système de rétroaction et de déterminer son impact sur la qualité du résultat sonore, plusieurs expériences sont menées avec l'auteur comme utilisateur principal. Bien que les résultats préliminaires ne soient pas encore significatifs, ces expériences permettent de mieux comprendre les défis et les opportunités liés à l'intégration de l'électroencéphalographie, des stations de travail audio numériques et des méthodes de classification dans la création musicale.

Au cours de ces expériences, différents aspects du système sont testés et évalués, notamment la qualité des données EEG, la précision de l'analyse des signaux EEG, la pertinence de la visualisation, l'efficacité de l'intégration des DAW et du feedback, et l'impact de l'API OpenAI sur la création musicale. Ces expériences permettent d'identifier les domaines d'amélioration et de proposer des développements futurs pour optimiser davantage le système de rétroaction et son application dans la composition musicale.

3. RÉSULTATS ET DÉFIS

3.1. Résultats préliminaires

Les expériences préliminaires menées dans le cadre de cette étude n'ont pas produit de résultats significatifs. Toutefois, ces premières tentatives ont permis d'identifier des défis et des domaines d'amélioration pour les expériences futures.

3.2. Défis

L'un des principaux défis identifiés dans cette étude est l'amélioration de la rétroaction grâce à l'implémentation d'un système de classification efficace. De plus, la démonstration actuelle est limitée à un seul utilisateur. Pour évaluer pleinement l'efficacité et l'applicabilité de cette approche, il est nécessaire d'inclure un plus grand nombre de participants et de concevoir des expériences mieux adaptées pour évaluer les impacts de la rétroaction.

4. DISCUSSION

4.1. Implications potentielles

Cette étude présente plusieurs implications potentielles pour les domaines de la musique, de l'informatique musicale et des sciences cognitives. L'intégration de l'EEG, des DAW et des méthodes de classification offre des perspectives prometteuses pour l'amélioration du processus de composition et de jeu instrumental en optimisant l'état émotionnel du musicien. Les implications incluent :

Une meilleure compréhension des liens entre les émotions, les processus cognitifs et la création musicale.

Le développement de nouvelles méthodologies et outils pour la création musicale, basés sur l'interaction entre le cerveau, les émotions et la technologie.

L'exploration de nouvelles formes d'expression artistique et musicale, fondées sur l'adaptation en temps réel à l'état émotionnel et cognitif du musicien.

La possibilité d'utiliser des méthodes similaires pour interagir avec les composantes émotionnelles et mentales des artistes dans d'autres disciplines créatives, telles que la danse, le théâtre ou les arts visuels.

4.2. Développements futurs

Afin de surmonter les défis identifiés et d'améliorer l'efficacité du système de rétroaction, plusieurs développements futurs sont envisagés :

Implémentation d'un système de classification plus sophistiqué : l'utilisation d'algorithmes d'apprentissage automatique avancés et de techniques de traitement du signal pourrait améliorer la précision de la rétroaction et optimiser davantage l'état émotionnel du musicien.

Conduite d'expériences supplémentaires avec un plus grand nombre de participants: cela permettrait d'affiner

le système de rétroaction et de le rendre plus adaptable aux besoins individuels des musiciens

Approfondissement de l'intégration de l'API OpenAI : l'intelligence artificielle pourrait être utilisée pour générer automatiquement des compositions ou des improvisations basées sur les émotions et l'état cognitif du musicien, ou pour analyser et apprendre des modèles musicaux spécifiques associés à différents états émotionnels.

5. CONCLUSION

Cette étude a exploré l'utilisation combinée de l'électroencéphalographie, des stations de travail audio numériques et des méthodes de classification pour créer un système de rétroaction visant à optimiser l'état émotionnel du musicien pendant la composition et la performance musicale. Bien que les résultats préliminaires ne soient pas encore significatifs, cette recherche interdisciplinaire ouvre la voie à de nouvelles explorations et découvertes dans les domaines de la musique, de l'informatique musicale et des sciences cognitives.

Les défis futurs incluent l'amélioration de la rétroaction grâce à l'implémentation d'un système de classification plus sophistiqué, l'inclusion d'un plus grand nombre de participants dans les expériences et l'exploration des possibilités offertes par l'intégration de l'intelligence artificielle dans la création musicale. En surmontant ces défis, cette approche intégrée pourrait offrir de nouvelles perspectives pour la compréhension des liens entre les émotions, les processus cognitifs et la musique, et pour le développement de nouvelles formes d'expression artistique basées sur ces liens.

6. BIBLIOGRAPHIE

1. Briot, Jean-Pierre, Hadjeres, Gaëtan, et Pacht, François. Deep learning techniques for music generation. Springer, 2020.
2. Celerier, Jean-Michaël, Desainte-Catherine, Myriam, et Couturier, Jean-Michel. Rethinking the audio workstation: tree-based sequencing with i-score and the LibAudioStream. In : *Sound and Music Computing Conference*. 2016.
3. Miranda, E. R., & Brouse, A. (2005). Interfacing the brain directly with musical systems: On developing systems for making music with brain signals. *Leonardo*, 38(4), 331-336.
4. Miranda, Eduardo Reck et Castet, Julien (ed.). *Guide to brain-computer music interfacing*. Springer, 2014.
5. Vulliard, P. H., Larralde, J., & Desainte-Catherine, M. (2012). Rétroaction musique-physiologie : une approche selon le paradigme des émotions. In *Journées d'Informatique Musicale*.

IDENTIFICATION PAR EEG ET UTILISATION DANS UN CADRE ARTISTIQUE DES ÉTATS DE CONCENTRATION : VERS UNE SONIFICATION DE L'ATTENTION ?

Armand Bazin
Université Gustave Eiffel
bazin.armand@gmail.com

Ulysse Del Ghingaro
Musidanse/C.et.I (EA1572)
Université Paris 8
ulyssedgo@gmail.com

RÉSUMÉ

Ce projet de recherche fait suite à un premier article publié aux JIM 2020, concluant sur certaines possibilités et limites des technologies d'électro-encéphalogrammes (EEG) dans un cadre de recherche et création. En reprenant une méthodologie proche, ce nouvel article contient une proposition d'un dispositif de sonification des données d'un casque EEG. L'objectif du projet est de développer un environnement sonore interactif en favorisant les rétroactions entre le dispositif de mesure et l'utilisateur, notamment à travers la sonification intuitive de ses états de concentration. Ces états de concentration sont discriminés par apprentissage automatique grâce à la bibliothèque Java *BrainFlow*.

L'article explicite la façon dont les données ont été mesurées et traitées ainsi que les choix de mapping et de contrôle de la création sonore. Le projet étant encore en développement, la partie discussion ouvre sur des pistes d'améliorations concernant la synthèse sonore et les méthodes de classification des états concentrés / relaxés.

1. INTRODUCTION

Ce projet mobilise d'un côté des connaissances d'analyse de données et d'apprentissage automatique et de l'autre une volonté d'intégrer à la composition musicale des réflexions écologiques – en pensant la musique comme le résultat d'interactions composées entre le porteur du casque et son environnement (sonore) à l'image des travaux du compositeur Agostino Di Scipio [1].

Ce qui nous intéresse particulièrement dans l'utilisation du casque EEG est la possibilité de créer une boucle de rétroaction entre l'auditeur, porteur du casque, et le son résultant de la sonification des données du casque. L'idée de départ de ce projet pouvait s'apparenter à une sonification de l'"écoute" de l'utilisateur. Cependant, identifier des réactions d'écoute par EEG s'est révélé trop complexe.

Nous avons donc décidé de nous concentrer sur la mesure de l'état de concentration de l'utilisateur, en cherchant à créer un environnement sonore capable

d'interagir avec sa capacité à se relaxer ou à se concentrer, via un retour sonore intuitif de l'état de concentration détecté par le système. Le dispositif est voulu simple : un environnement sonore, créé à partir d'une synthèse soustractive basée sur le signal reçu du casque, et modifié par le couple relaxation / concentration. Ainsi, un des buts de notre dispositif est de s'approcher d'une sonification de l'attention du porteur du casque.

Le terme de sonification recouvre une grande diversité de démarches pratiquées à la fois par des artistes sonores et des compositeurs. Une de ses définitions est de rendre sensible, grâce au son, certains phénomènes non perceptibles en temps normal [2].

Cette démarche est largement répandue dans les arts sonores écologiques qui cherchent, entre autres, à initier une écoute dite "écologique" chez les auditeurs, c'est-à-dire un mode d'écoute "dans lequel l'attention s'éloigne des choses ou des objets isolés au profit d'une dynamique plus globale centrée sur celle des interactions et des interrelations." [3].

Pour donner deux exemples on peut citer l'artiste Andrea Polli et son projet *Atmospheric weathers works* [4], basé sur de la sonification de données météorologiques ou bien *Le souffle de la Terre* de l'artiste Lorella Abenavoli [5] qui rend audible des données sismiques.

Dans ce projet de recherche, la technique de l'EEG est donc appréhendée comme une outil pour rendre concrètes et sensibles les interactions cognitives entre un auditeur et son environnement sonore selon l'attention qu'il y porte. Cette approche de la composition provient aussi de l'intégration de réflexions tirées de l'écologie de l'attention, élaborée par Yves Citton [6], et de l'écologie de l'esprit, conceptualisée par Gregory Bateson [7]. Ces réflexions ne seront pas développées dans cet article pour nous concentrer exclusivement sur l'aspect technique et sonore du dispositif.

2. MATÉRIEL ET MÉTHODE

2.1. Méthode

Ce travail s'intègre dans une démarche de recherche et création. Étant non spécialistes, notre approche consiste à essayer de s'appropriier artistiquement les possibilités d'un casque électro-encéphalogramme (EEG), tout en essayant de rester au plus proche de la signification des données captées.

Cet article provient de la suite du travail qui avait résulté en une précédente publication [8]. Cette première version avait présenté des tests du casque *Open_BCI* mark IV ainsi qu'un projet de création appelé CAANEVA (création autonome par analyse neuronale d'environnement visuel et auditif).

Le présent article repart des conclusions du précédent pour mettre à jour les avancées dans le domaine des EEG par la firme open source *Open_BCI*¹ et propose un nouveau projet créatif plus en adéquation avec les capacités de ce matériel. Ce nouveau travail se base donc sur les objectifs suivants : simplifier le dispositif aussi bien au niveau matériel – un nombre réduit d'électrode avec une meilleure méthode d'analyse des données – qu'au niveau créatif – une boucle de rétroaction (feedback) sonore simple à appréhender – dans une volonté de permettre plus d'interactivité et de contrôle à l'utilisateur.

Des essais ont été menés avec le casque *OpenBCI* pour identifier la méthode de classification la plus appropriée et les activités correspondantes à l'état "relaxation" et l'état "concentration". Des modifications ont été apportées au code source du logiciel *OpenBCI GUI* afin de récupérer certaines données. Ces données ainsi qu'une Transformée de Fourier Rapide (FFT²) des fréquences d'activation neuronale sont les métriques sur lesquelles la création sonore est basée. Celle-ci est assurée par le logiciel *Max/MSP*.

Nous nous concentrerons ici sur les changements entre la première version et la seconde version du dispositif, le matériel n'étant pas tout à fait le même³.

2.2. Hardware et tests

Au sein de ce projet nous avons utilisé un casque *OpenBCI mark IV* huit canaux avec électrodes sèches de 5 mm. L'objectif étant de mesurer l'état de relaxation ou de concentration de l'utilisateur nous avons placé les électrodes selon la disposition 10-20 en Fp1, Fp2, F3 et F4. "10-20" fait référence aux rapports de distance entre deux électrodes adjacentes qui représente 10% ou 20 % de la longueur et largeur totale du crâne [9].

Les électrodes servant de terre et de référence sont positionnées sur les lobes d'oreille du porteur du casque en A1 et A2 (Figure 1).

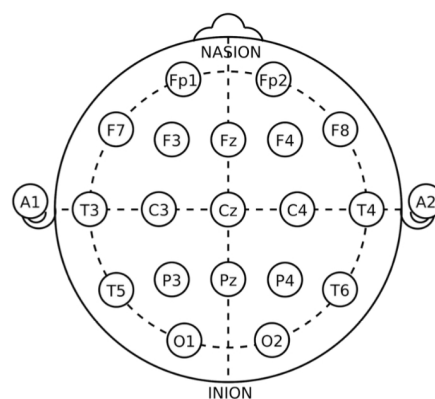


Figure 1. Disposition 10-20.

Les électrodes en F3 et F4 ne sont pas obligatoires pour les mesures mais peuvent servir de données complémentaires. Spécifiquement pour ces électrodes, nous avons rajouté des têtes d'électrodes "dry EEG comb electrodes"⁴ qui permettent d'obtenir un signal convenable malgré la chevelure.

Afin de récupérer les données des électrodes nous avons utilisé une carte d'*OpenBCI* à 4 canaux appelée *Ganglion*. Pour son utilisation en EEG, le pin de chaque canal a été réglé sur *down* comme indiqué dans la documentation *OpenBCI*⁵, passant les électrodes en configuration unipolaire⁶.

La récupération des données par l'ordinateur se fait grâce à une clé USB bluetooth et au logiciel *OpenBCI GUI*.

Les tests du hardware ont été très similaires aux tests menés dans le cadre du premier article, avec l'utilisation du widget de test d'impédance en plus.

Le widget de test d'impédance du ganglion est un outil dont l'utilisation est indispensable avant chaque session. En envoyant un courant dans chaque électrode de façon successive, il permet de calculer leur l'impédance. Cela permet de vérifier si ces dernières sont bien en contact avec la peau et donc si les données qu'elles renvoient ne sont pas des artefacts. Une impédance en dessous de 50 kΩ indique un bon contact de l'électrode. Réussir à réduire l'impédance des électrodes sous ce seuil n'est pas simple du fait de leur très grande sensibilité à tout changement de position.

Pour les électrodes placées en F3 et F4, les *comb electrodes* semblent bien fonctionner car une faible impédance est obtenue rapidement une fois les électrodes en contact avec le cuir chevelu.

Pour les électrodes placées en Fp1 et Fp2, l'opération est plus délicate. Nous avons testé différentes têtes d'électrodes sans changement notable. La position et le

1 <https://openbci.com/>

2 Fast Fourier Transform en anglais

3 Des informations plus approfondies sur les bases de l'EEG sont disponible dans l'article de 2020 [8]

4 <https://shop.openbci.com/products/5-mm-spike-electrode-pack-of-30>

5 <https://docs.openbci.com/AddOns/Headwear/HeadBand/#headband-ganglion-tutorial>

6 https://www.neuro.mcw.edu/meg/index.php/EEG_reference_with_Brainstorm

serrage des électrodes sont à modifier subtilement jusqu'à ce que l'impédance baisse convenablement.

Quant aux références, pour chaque session il nous a fallu attendre environ 5 minutes sans toucher au positionnement des électrodes de référence le temps que leur impédance soit bonne. Ce changement d'impédance peut être dû à la sudation qui améliore la conductivité entre la peau et l'électrode.

Enfin, nous avons pris soin de rassembler les câbles d'électrodes et de bien les fixer au casque pour limiter au maximum leurs mouvements, qui sont source d'artéfacts.

Ces remarques peuvent paraître superflues, mais dans le cadre d'une présentation artistique il est important de noter que ce matériel nécessite un temps de mise en place et de subtils ajustements avant de pouvoir l'utiliser correctement (Figure 2).

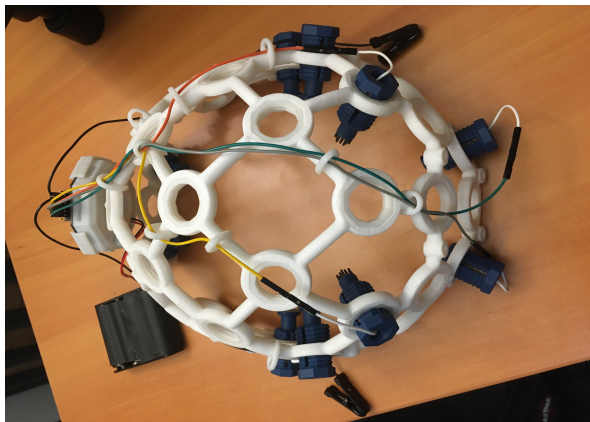


Figure 2. Casque OpenBCI Mark IV.

2.3. OpenBCI GUI et Focus Widget

Pour récupérer les données du ganglion nous avons utilisé le logiciel à interface graphique OpenBCI GUI (version 5.1.0).

OpenBCI GUI est un logiciel libre codé sous Processing (Java). Le fait qu'il soit open source nous a permis d'accéder à son code et de le modifier afin de récupérer une variable initialement non exportée.

Cette variable, appelée *metric value*, correspond au taux de confiance des différents classifieurs⁷ utilisés dans la détermination de l'état de concentration de l'utilisateur lors du port du système, à savoir concentré, neutre ou relaxé.

Ce "taux de confiance" établit la probabilité avec laquelle le classifieur estime l'état mental de l'utilisateur. Elle oscille ainsi entre 0 et 1 pour chacun des deux états, avec une relation inversement proportionnelle telle que :

$$P(\text{Etat concentré}) \approx 1 - P(\text{État Relaxé}) \quad (1)$$

⁷ Les classifieurs sont des algorithmes d'apprentissage automatique, ici en l'occurrence d'apprentissage supervisé ayant pour but de déterminer la catégorie d'appartenance de données. Dans notre cas, les classifieurs sont utilisés pour classer les informations reçues du casque sur une courte période dans une catégorie: concentré ou relaxé.

Une méthode *MetricValueGetter* a été ajoutée au code source pour y accéder.

La *metric value* évolue de façon continue et sans saccades tout en filtrant une partie du bruit et des artefacts⁸. Comparativement aux *times series*⁹, la *metric value* est moins sujette aux petites variations dans le signal, celles-ci n'étant pas forcément interprétées par le modèle comme des données représentatives d'une activité cognitive. Autrement dit, un travail de nettoyage du signal (bien qu'imparfait) est opéré par le modèle. Il se fait lors de l'attribution de la valeur de poids et de biais définissant "l'importance", estimée par le modèle, d'un feature¹⁰ donné dans le résultat de sa classification. Cette propriété de la *metric value* constitue un avantage pour le mapping (Figure 3).

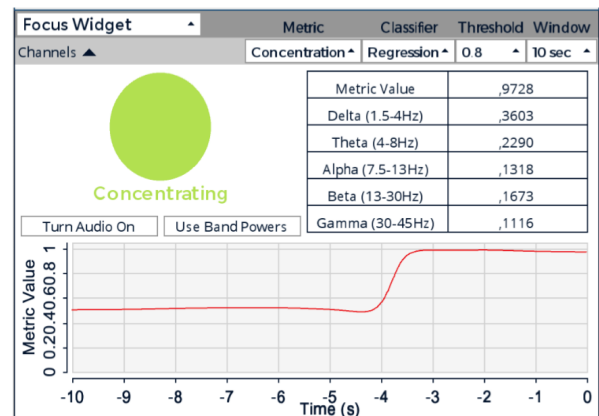


Figure 3. Focus Widget mesurant un état concentré depuis 3 secondes.

2.4. Définitions et utilisation de la bibliothèque BrainFlow

La version 5.1.0 d'OpenBCI fait désormais appel à la bibliothèque Java BrainFlow.¹¹ Cette bibliothèque est spécialisée dans l'analyse et le traitement de signaux biométriques et notamment des signaux EEG. C'est d'après les modèles d'apprentissage automatique de cette bibliothèque qu'est tirée la *metric value* et qu'est, par la même, établit la classification des états relaxé et concentré de l'utilisateur.

Au sein du Focus Widget d'OpenBCI GUI on peut ainsi faire appel à différents classifieurs¹², et accéder à certains paramètres comme la taille de la fenêtre d'analyse ou encore le seuil de confiance

⁸ Les artefacts sont du bruit résultant de phénomènes extérieurs à l'activité cognitive que l'on cherche à mesurer. (ie: clignement des yeux, serrage de mâchoires) ces derniers ont été une problématique non entièrement résolue du précédent projet.

⁹ Il s'agit de la tension électrique mesurée par chaque électrode en fonction du temps.

¹⁰ Données variables utilisées dans l'analyse du modèle.

¹¹ <https://brainflow.org/>

¹² Pour clarifier les interactions entre les différentes parties logiciels : la bibliothèque BrainFlow est utilisée ici par le logiciel OpenBCI GUI, au sein de celui-ci figure une fonction Focus Widget permettant de récupérer les états de concentrations au cours du temps.

nécessaire au classifieur pour considérer un état comme étant relaxé ou concentré.

D'après la documentation d'*OpenBCI*, l'état de relaxation est atteint en fermant les yeux et en faisant "de la méditation". L'état de concentration est atteint les yeux ouverts et en "se concentrant intensément".

Cette description étant un peu floue pour la concentration, nous avons mené nos propres tests pour essayer d'identifier les activités qui correspondent à l'état concentré discriminé par le *Focus Widget*.

Pour tous les tests menés, l'état de relaxation consistait à reproduire les consignes d'*OpenBCI* : fermer les yeux et faire des exercices de respiration.

Pour l'état de concentration, nous avons effectué différents exercices tels que du calcul mental, de la lecture de texte, de la lecture de notes, de la concentration visuelle¹³.

Chaque session de test a consisté en 5 minutes d'état relaxé puis cinq minutes réparties entre les différentes activités de concentration. Ces phases de tests seront discutées dans la conclusion de l'article.

Avant d'aller plus loin, et afin de clarifier nos choix de modèles, nous définirons ici brièvement l'apprentissage automatique.

L'apprentissage automatique, ou *machine learning* en anglais, est une branche de l'intelligence artificielle constituée d'un ensemble de techniques dont le but est d'apprendre à une machine à résoudre des problèmes ou à prendre des décisions en s'appuyant sur des données, sans être explicitement programmée pour cela [10]. En apprentissage automatique on cherche à définir des modèles mathématiques, des algorithmes, pouvant apprendre à partir d'un jeu de données dit d'entraînement, afin de réaliser des prédictions ou des classifications sur de nouvelles données.

Il existe différentes méthodes d'apprentissage pour entraîner un modèle : l'apprentissage supervisé, l'apprentissage non supervisé et l'apprentissage par renforcement [11].

Le type d'apprentissage se choisit en fonction du type de données utilisées et des tâches à effectuer. Dans notre cas, classifier des états cognitif en "concentré" ou "relaxé", l'apprentissage supervisé est la seule méthode appropriée.

Plus spécifiquement, nous avons opté pour un modèle de régression logistique avec comme paramètres une fenêtre d'analyse de 10 secondes, et un seuil de 0,8 pour l'état concentré¹⁴ (0,2 pour l'état relaxé), les valeurs intermédiaires pouvant correspondre à un état "neutre".

Cette méthode statistique d'apprentissage supervisé est utilisée pour prédire si une certaine valeur est vraie ou fausse, en fonction d'autres variables connues, ici les valeurs des différents types d'ondes.

C'est à partir de cette méthode que sont classés les états mentaux en état concentré ou en état relaxé. Dans cette version du projet l'état mental est considéré comme binaire (selon la fonction (1)).

Cette régression utilise une fonction logistique¹⁵ pour modéliser la probabilité d'une réponse binaire – concentré ou relaxé – en fonction des variables explicatives. La régression logistique utilise également une fonction de coût, appelée fonction de perte logistique, pour minimiser l'erreur de prédiction. Cette fonction de coût¹⁶ prend en compte la distance entre les valeurs prédites et les valeurs réelles qui ont été établies lors de la phase labélisation des données par *BrainFlow*.

La régression logistique est donc un algorithme de classification binaire qui estime la probabilité qu'une observation appartienne à une catégorie en utilisant une fonction logistique et une fonction de coût pour ajuster les paramètres de la fonction de prédiction.

En d'autres termes, c'est à partir des données labellisées que l'algorithme va apprendre à différencier un état concentré d'un état relaxé, en apprenant à minimiser son erreur de classification. Une fois l'algorithme entraîné il est déployable sur des données non labellisées.

Entraîner un modèle d'apprentissage automatique signifie trouver le minimum de sa fonction de coût. Cela équivaut à réduire le nombre d'erreurs de classification moyen ce qui revient à améliorer la qualité de notre modèle dans la tâche de distinction des états mentaux concentré/relaxé.

L'entraînement d'un modèle s'établit à partir d'un jeu de données labellisées (qui sont les réponses qu'on cherche à obtenir de notre modèle) et des prédictions de l'algorithme qui après avoir été dans un premier temps complètement aléatoires s'affinent et se précise simplement du fait de la minimisation de la fonction de coût. D'un point de vue technique, cette minimisation de la valeur de fonction de coût (proxy de l'erreur moyenne du modèle) se fait grâce à une descente de gradient [12]. Cette dernière informant le modèle de la participation et de l'importance de chacune des variables dans l'erreur produite par l'algorithme. Afin de bien entraîner un modèle il est important d'avoir un jeu de données d'entraînement représentatif des données réelles et nouvelles auxquelles l'algorithme aura affaire une fois déployé. Si ces données sont trop spécifiques ou trop peu spécifiques et que le modèle est mal paramétré on risque d'avoir un problème de sur-interprétation ou de sous-interprétation amenant à de mauvaises prédictions/classifications de l'algorithme. Dans notre cas, étant données que l'entraînement du modèle s'est fait à partir des données labellisées autres que celles issues des tests de notre utilisateur principal le risque de sur interprétation est significatif. Ré-entraîner un modèle spécifiquement sur notre testeur

13 Exemple de test de concentration visuel réalisé : <https://www.youtube.com/watch?v=EI-b1fC4rVM>

14 Les valeurs de ces deux paramètres furent choisies empiriquement.

15 Egalement appelée fonction sigmoïde.

16 Le résultat de la fonction de coût donne une estimation de l'erreur du modèle.

principal est susceptible d'améliorer la fiabilité de nos classifications, mais également de rendre le modèle plus spécifique et le projet moins diffusable.

Comme précisé précédemment, nous avons opté pour l'algorithme de régression logistique au sein de ce projet. Cet algorithme se base sur une régression non linéaire et cherche à tracer une courbe sigmoïde pour coller au mieux aux données labellisées par *BrainFlow* de façon à maximiser le nombre d'occurrences dans lesquelles:

$$\begin{aligned} f(\text{paramètres des états concentrés}) &= 1 \text{ et} \\ f(\text{paramètres des états relaxés}) &= 0 \end{aligned} \quad (2)$$

Cette étape constitue la phase d'entraînement de notre modèle qui sera ensuite utilisé sur les données mesurées en temps réel afin de les associer à une classe.

Le résultat d'une fonction sigmoïde est toujours compris entre 0 et 1, et correspond à la *metric value*. Un résultat étant égal à 1 signifiant que l'état cognitif mesuré est identifié comme concentré et 0 comme relaxé. Toute valeur intermédiaire représente un certain niveau d'incertitude du classifieur (Figure 4).

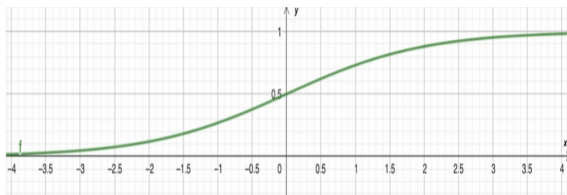


Figure 4. Courbe d'une fonction sigmoïde.

$$f(x) = \frac{1}{1 + e^{-x}} \quad (3)$$

Tous les classifieurs disponibles sur *OpenBCI* sont des algorithmes d'apprentissage supervisé entraînés sur le même jeu de données labellisé, ayant lui-même été constitué par *BrainFlow*. Ce jeu de données n'est accessible que sur le site de *BrainFlow* et n'est pas en accès libre. Il en va de même pour les classifieurs de *BrainFlow* qui, même s'ils sont intégrés à un projet *Open Source*, ne sont pas libres d'accès. Ainsi, les différents classifieurs utilisés dans *OpenBCI* ne sont pas paramétrables, la plupart des paramètres sont inaccessibles même au niveau informationnel¹⁷. D'après la documentation *BrainFlow* il serait cependant possible d'ajouter des modèles ONNX¹⁸ au pipeline.

La phase de tests des différents classifieurs était d'autant plus cruciale dans le projet, du fait de la non

¹⁷ Par exemple, pour l'algorithme K-nearest neighbour (modèle disponible dans *brainflow*) on ne connaît pas la valeur de K et il est impossible de le modifier ce qui pourrait changer drastiquement les performances de ce classifieur.

accessibilité du jeu de données d'entraînement et du paramétrage des modèles.

Les classifications d'états de concentration et de relaxation ont été utilisées comme paramètres de la synthèse et de l'évolution du feedback sonore. L'idée étant de faire en sorte que l'utilisateur ait un retour sonore de son état mental plus intuitif que dans le projet auquel cet article fait suite.

Dans notre projet nous avons défini quatre états: relaxation profonde, relaxation, concentration, et concentration profonde sur lesquels nous reviendrons dans la partie "Aspects artistiques et sonores".

Le logiciel *Max/MSP* a été utilisé afin de réaliser la synthèse sonore, les données du système EEG sont ainsi communiquées à *Max/MSP* en OSC¹⁹ depuis le *OpenBCI GUI*. Pour des raisons d'optimisation, nous avons essayé de répartir le maximum de calculs sur le logiciel *OpenBCI GUI* pour que le patch *Max/MSP* ne se consacre qu'à la synthèse sonore et au mapping.

3. ASPECTS ARTISTIQUES ET SONORES

3.1. Synthèse soustractive et sonification

Pour ce qui est de la création sonore, nous utilisons le logiciel *Max/MSP*, dont la figure 4 présente le mode présentation du patch.

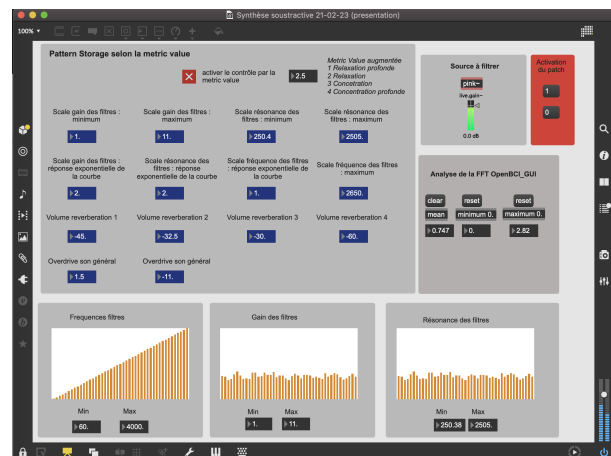


Figure 4. Mode présentation du patch *Max/MSP* utilisé pour la création sonore.

Nous nous basons sur un principe de synthèse soustractive. Cette méthode de synthèse possède l'avantage de produire un son riche plus facilement qu'une synthèse additive; elle procure aussi une approche originale de la sonification, tout en se

¹⁸ *Open Neural Network Exchange*: format d'export de modèle d'apprentissage automatique traduisible depuis la bibliothèque Python spécialisée *Keras*.

¹⁹ Protocole de communication *Open Sound Control*

combinant facilement avec les données d'une FFT. En effet, la FFT nous retournant des fréquences et des amplitudes, nous avons décidé de les associer aux paramètres de filtres en parallèle (objet : *fffb~*) appliqués à un bruit rose. C'est ensuite en modifiant les bornes de cette mise à l'échelle que le résultat sonore sera contrôlé via la *metric value*.

La FFT est calculée par le logiciel *OpenBCI GUI*, qui lui applique un filtre passe bande de 5 à 50 Hz. Elle est ensuite directement communiquée au patch *Max/MSP* avec un taux d'échantillonnage de 200 Hz. Ainsi nous n'utilisons que 45 valeurs de la FFT pour la synthèse sonore. Les fréquences de la FFT allant de 5 à 50 Hz, une partie non négligeable de celles-ci se trouve hors du champ audible (en dessous de 20 Hz) le reste étant extrêmement grave. Nous avons donc décidé de faire une mise à l'échelle linéaire avec 60 Hz comme fréquence minimum, la valeur maximum étant un paramètre variable contrôlé par la *metric value*.

L'amplitude du filtre et sa résonance sont mappées sur l'amplitude de chaque fréquence de la FFT. Ainsi lorsque l'amplitude est basse le son résultant est proche du bruit rose, le filtre laissant passer un grand nombre de fréquences, mais avec un volume faible. Lorsque l'amplitude est grande, le filtre devient plus résonant et son gain augmente, résultant en l'apparition d'une fréquence précise (Figure 5).

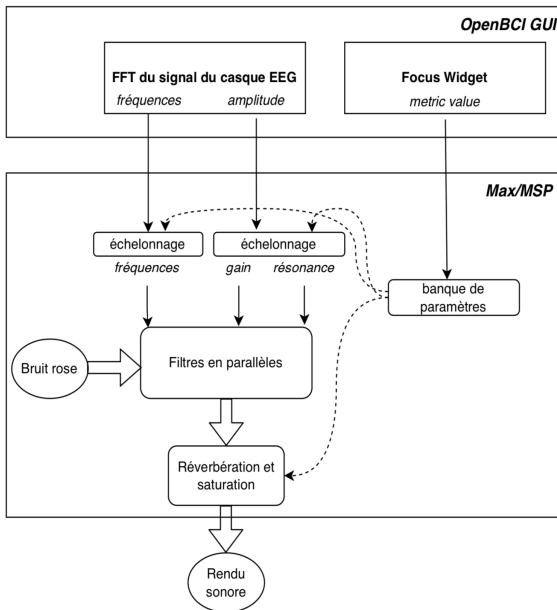


Figure 5. Schéma du fonctionnement du patch *Max/MSP*.

Le résultat d'ensemble s'entend comme un son homogène mais avec un contenu spectral toujours en variation.

Afin d'enrichir la création sonore et de créer des identités sonores distinctes pour chaque état mental, certains traitements sonores ont été ajoutés au résultat de la synthèse soustractive. Il s'agit d'une réverbération et

d'une saturation. La réverbération permet de polir le résultat assez brut de la synthèse sonore et est utilisée de façon intensive dans les états "profonds" pour mieux les caractériser. La saturation (objet : *overdrive~*) n'est présente que dans les états concentrés pour enrichir les sons issus des filtres résonants.

3.2. Contrôle de l'environnement sonore grâce à la *metric value*

Nous avons organisé le contrôle de l'environnement sonore selon quatre états mentaux (relaxation profonde, relaxation, concentration et concentration profonde).

La relaxation et la concentration profonde ne sont pas des états identifiés par le *Focus Widget*, il s'agit d'un ajout de notre part. L'accès à un état dit "profond" se fait en restant dans un même état avec un seuil de 0,2 – *metric value* en dessous de 0,2 pour l'état relaxé et au-dessus de 0,8 pour l'état concentré – pendant vingt secondes. Si l'état de l'auditeur change, le rendu sonore retourne à son précédent état en seulement trois secondes. Ainsi l'auditeur n'a accès aux états "profonds" qu'une fois qu'il maîtrise réellement son attention. L'intégration de ces états s'est faite afin de rendre le dispositif plus interactif mais aussi plus riche musicalement. Ces durées ont été choisies de façon empirique en faisant un compromis pour avoir un rendu musical intéressant sans trop contraindre le porteur du casque, rester vingt secondes d'affilée dans le même état n'étant pas une tâche trop ardue même sans entraînement.

Le contrôle du patch de création sonore par la *metric value* peut être résumé selon le schéma suivant. Les paramètres contrôlant ces états sont sauvegardés dans le patch (objet : *patrstorage*). Pour passer d'un état à un autre, nous utilisons une interpolation linéaire qui suit la *metric value* entre l'état 2 et 3, et une interpolation linéaire en fonction du temps entre les états 1 et 2 et entre les états 3 et 4 (Figure 6).

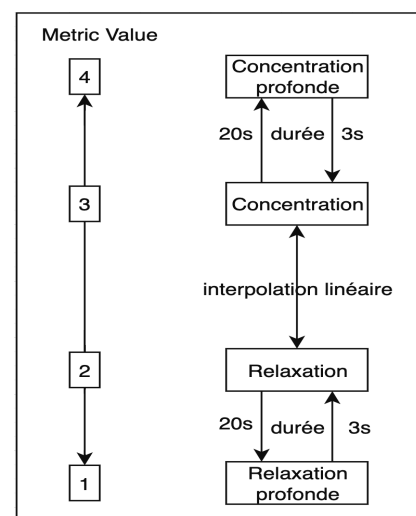


Figure 6. Schéma de la correspondance entre *metric value* et états de concentration à la base du contrôle de l'environnement sonore.

Au niveau sonore, le choix des paramètres ont été pensé pour donner les caractéristiques suivantes à chaque état de concentration. L'état de relaxation profonde correspond à de nombreux sons graves et doux dans une grande réverbération, qui peut faire penser au son de l'océan. L'état de relaxation est proche mais possède une plus grande gamme de fréquences. Quant aux états de concentration, ils sont caractérisés par des fréquences résonantes précises et aiguës, plongées dans une grande réverbération si l'état est "profond" (Tableau 1).

Paramètres	Relaxation profonde	Relaxation	Concentration	Concentration profonde
Gain minimum du filtre	0	1	1	0,01
Gain maximum du filtre	1	2	10	20
Résonance minimum du filtre	0.5	1	1000	3000
Résonance maximum du filtre	1	10	5000	5000
Fréquence maximum du filtre (Hz)	70	300	5000	7000
Overdrive	1	1	2	2,5
Réverbération	longue	courte	courte	longue

Table 1. L'ensemble des valeurs des paramètres qui caractérisent chaque état de l'utilisateur

L'idée englobant ce dispositif est de penser l'interaction, sous la forme d'une boucle de rétroaction, entre l'attention (symbolisée par le couple relaxation / concentration) du porteur du casque et l'environnement sonore, comme une forme musicale émergente. Pour ce faire, l'utilisateur doit pouvoir comprendre le fonctionnement du dispositif, ce qui implique un mapping en modulation temporelle avec lequel il peut se familiariser. Une fois cette étape réussie, il n'est pas exclu de créer une évolution temporelle du mapping – en accentuant les courbes de réponse des paramètres en fonction du temps par exemple – qui enrichira le rendu sonore tout en conservant une boucle de rétroaction claire.

4. DISCUSSION

Le projet est encore en phase de développement à plusieurs niveaux. Des tests portant sur d'autres classificateurs, sur le traitement des données et sur le rendu sonore sont encore en cours au moment de l'écriture de cet article.

Pour ce qui est du traitement des données, une étape de calibrage automatique est en cours d'élaboration. Cette étape semble nécessaire à chaque utilisation du dispositif puisque, selon la position des électrodes et le résultat des tests d'impédance, l'amplitude du signal du casque peut varier légèrement. Elle est aussi nécessaire afin de filtrer certains artefacts qui faussent les valeurs d'amplitude de la FFT, en faisant apparaître des amplitudes bien trop grandes pour la mise à l'échelle. Par exemple, des valeurs d'amplitude de plus de $300 \mu V$ pour des valeurs moyennes de $1 \mu V$.

Ce calibrage consisterait en la création d'une courbe de répartition des valeurs de la FFT sur une certaine durée. D'après les premiers tests effectués cette courbe est proche d'une gaussienne. Cependant, une étude statistique reste nécessaire afin de pouvoir fixer la borne supérieure des valeurs pour la mise à l'échelle. Avec ceci, les amplitudes de la FFT ne pourront plus dépasser cette valeur seuil, ce qui devrait améliorer de façon significative la mise à l'échelle.

Concernant la création de l'environnement sonore, plusieurs pistes d'approfondissements sont possibles. Le son de base utilisé dans la synthèse soustractive est ainsi sujet à évoluer, pour, par exemple, passer d'un bruit rose à un son issu de synthèse granulaire.

De plus, pour renforcer l'identité sonore de l'état de concentration profonde ainsi que sa compréhension par l'auditeur, l'idée d'ajouter un/des opérateur(s) comme un LFO (*Low Frequency Oscillator*) sur le rendu sonore est en voie d'implémentation. Un LFO pourrait par exemple s'activer uniquement dans les dernières secondes avant l'état "profond" afin d'indiquer au porteur du casque de façon claire son changement d'état tout en amenant de la variabilité aux rendus sonores possibles du dispositif.

Dans le même ordre d'idée, le paramétrage de la synthèse soustractive nécessite encore certains tests pour affiner l'esthétique du rendu sonore et son aspect intuitif à retranscrire un état relaxé ou concentré.

Pour ce qui est du contrôle de l'environnement sonore par la *metric value*, une des pistes à explorer est l'intégration, pour certains paramètres, d'interpolations exponentielles ou bien de courbes de réponse dessinées sur mesure afin de sculpter des transitions sonores plus subtiles, chaque paramètre ayant sa propre évolution selon la *metric value*.

Par ailleurs, la recherche d'ERP (*event-related potential*) en lien avec l'environnement sonore reste une partie annexe de ce projet. Les ERP ou potentiels évoqués, sont ainsi des réponses corticales fortes mesurables à la suite d'un stimuli précis. La mise en place de tels stimuli (ici sonore) n'est pour l'instant pas faite dans le projet, mais une recherche en ce sens est en cours.

Enfin, afin d'avoir un meilleur contrôle paramétrique sur les modèles utilisés dans la classification de données, il est envisagé d'importer nos propres

classifieurs au format ONNX à travers *BrainFlow*. En effet, bien que la classification des états concentré/relaxé fonctionne empiriquement, tous les classifieurs disponibles nativement dans *BrainFlow* utilisent des méthodes d'apprentissage supervisé entraînés sur un jeu de données n'étant pas accessible librement. Pour l'heure, il semble impossible de faire un "fine tuning"²⁰ des modèles. Or, même si le système semble marcher en l'état, il existe des variations probables des activités corticales selon les utilisateurs, justifiant une étape de fine tuning. Cette étape pourrait potentiellement amener de sensibles améliorations dans les classifications des modèles. L'utilisation de *BrainFlow* par OpenBCI nous amène à nous poser des questions sur la méthodologie mise en œuvre pour l'entraînement des classifieurs, sur la relative opacité des paramètres de certains modèles ainsi que sur la constitution du jeu d'entraînement et de ses données labellisées.

5. CONCLUSION

Cette seconde version du projet est encourageante vis-à-vis des objectifs fixés. L'environnement sonore réagit de façon convenable aux variations de la *metric value* et celle-ci semble bien corrélée avec l'état mental du porteur du casque selon les tests effectués. De plus, l'environnement sonore créé est suffisamment intuitif pour créer une boucle de rétroaction avec l'utilisateur, transformant le casque en objet plus complexe qu'un simple contrôleur. Bien qu'encore imparfait, le feedback du dispositif est ainsi plus intuitif que dans la première version grâce à la retranscription sonore du niveau de concentration/relaxation. Un autre apport de ce projet comparativement au précédent réside dans la sensible réduction du bruit et des artefacts issues des mesures du casque. Ceci a été rendu possible en utilisant une métrique d'ordre supérieure aux time series, la *metric value* des classifieurs [13].

Finalement, sur le plan théorique, ce projet pose des questions à la fois sur la sonification et l'attention : jusqu'où peut-on associer attention et couple d'état relaxé/concentré ? Quels sont les différents types d'interaction entre le porteur du casque et l'environnement sonore ? Et comment les enrichir ?

Autant de questions auxquelles le projet tentera de répondre lors de son développement.

6. RÉFÉRENCES

[1] Meric, R. et Solomos, M. "Audible Ecosystems, and emergent sound structures in Di Scipio's music. Music philosophy enhances musical analysis", Fourth Conference for Interdisciplinary Musicology (CIM08), Thessaloniki, Greece, 2008.

- [2] Schoon, Andi and Dombois, F. "Sonification in music." in *Proceedings of the 15th International Conference on Auditory Display*, Copenhagen, Denmark, 2009.
- [3] Giomi, A. « Du field recording à l'art de la sonification ». *Filigrane*, Paris, 2022.
- [4] Polli, A. « Atmospherics/weather works: A spatialized meteorological data sonification project ». *Leonardo*, 2005, vol. 38, no 1, p. 31-36, .
- [5] Abenavoli, L. « The Pulse of the Earth and sonification » . *AI & society*, 2012, vol.27, p.277-279.
- [6] Citton, Y. *Pour une écologie de l'attention*, Paris, Seuil, 2014.
- [7] Bateson, G. *Vers une écologie de l'esprit* (Drosso, F., Lot, L., Simion, E., Trad.), Paris, Seuil, 1977.
- [8] Bazin, A., et Del Ghingaro, U. « Projet CAANEVA, Possibilités et limitations de L'usage Des Interfaces Cerveau-Machine Dans un Contexte de Création d'environnement visuel et sonore ». *Acte musical et environnements informatiques : actes des Journées d'Informatique Musicale 2020*, Université de Strasbourg/Faculté des Arts, Strasbourg, 2021
- [9] Jasper, H. "Report of the committee on methods of clinical examination in electroencephalography", *Electroencephalography and Clinical Neurophysiology*, USA, 1958.
- [10] Bishop, C. M. *Pattern recognition and machine learning*. Springer, Berlin, Allemagne, 2006
- [11] Ben-David, S. & Shaleve-Shwartz, S. *Understanding machine learning from theory to algorithms*. Cambridge University Press, Cambridge, Angleterre, 2014.
- [12] Bengio, Y. ; Courville, A. & Goodfellow, I. *Deep Learning*. The MIT Press, Cambridge, USA, 2016
- [13] Nunez, P. L., & Srinivasan, R. *Electric fields of the brain: the neurophysics of EEG*. Oxford University Press, USA, 2006.

²⁰ Ajustement de modèles pré-entraînés à l'aide de données issues de mesures personnelles.

Actes des Journées d'Informatique Musicale 2023
24-26 mai 2023
Maison des Sciences de l'Homme Paris Nord

SESSION « REPRÉSENTATIONS MUSICALES »

ANALYSE ET ESSAI DE RECONSTRUCTION DE STRUCTURES ARBORESCENTES DES GRILLES DE JAZZ

Patrice Thibaud Mathieu Giraud

Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRISTAL
F-59000 Lille, France

RÉSUMÉ

L'harmonie tonale s'organise autour d'enchaînements V/I et d'autres schémas harmoniques. Elle peut se modéliser par des arbres. La Jazz Harmony Treebank (JHT) regroupe ainsi 150 analyses de grilles de jazz sous forme arborescente. Nous analysons ici quelques aspects de la répartition des enchaînements d'accords et schémas suivant les arborescences de la JHT, et nous proposons un algorithme pour reconstruire ces arbres à partir de la grille. Selon les scénarios, cet algorithme regroupe autour de 60% des accords des grilles, avec une précision d'environ 70%, confirmant la prédominance de certains schémas harmoniques dans la structuration de ces grilles.

1. GRILLES

L'harmonie tonale s'organise autour de la gamme et d'enchaînements harmoniques. Elle peut se penser dans une structuration arborescente – que ce soit en considérant les différents niveaux d'analyse schenkérienne [13], ou, plus explicitement, dans les modèles introduits par Lerdahl et Jackendoff dans la *Generative Theory of Tonal Music* [9]. Rohrmeier propose plus précisément une organisation de l'harmonie tonale sous forme d'arbres modélisés par des grammaires, incluant des suites de *prolongations* et de *préparations* [11].

Jazz et grilles. Une représentation des suites d'accords sous forme de *grilles* est couramment utilisée afin de permettre au musicien d'accompagner, d'improviser et d'apprendre de manière plus aisée la suite de progressions harmoniques. En jazz, ces grilles sont souvent complétées de la ligne mélodique pour obtenir une *lead sheet*.

En 1942, George Godwin (directeur d'une station de radio) a imaginé les *Tune-dex*, cartes à destination des professionnels de la radio indiquant, au recto, des informations éditoriales que nous considérons aujourd'hui comme des *méta-données* (compositeur, parolier, droits d'auteur etc.), et, au verso, des *données* musicales succinctes, mélodie et accords [8].

Ces cartes n'étaient pas spécifiquement destinées au jazz, mais des musiciens s'en sont emparées afin de posséder un vaste répertoire pour jouer au débotté n'importe quelle chanson en faisant illusion (*to fake my way through*). Expurgées de leurs données éditoriales, les *lead sheets* ont

été assemblées, indexées (parfois thématiquement), naissance aux premiers *Fake books*, illégaux. L'émergence des Real Books quelques années plus tard et aujourd'hui le partage de grilles par l'intermédiaire de sites spécialisés fournissent de formidables sources à disposition des musiciens.

Jazz et MIR, Corpus. La recherche en informatique musicale (et, plus précisément, en MIR, music information retrieval) peut concerner de nombreux aspects sur le jazz autour des grilles, mais aussi des lead sheets, des solos, d'audio voire de métadonnées. On peut citer les travaux faits autour des transcriptions de solo de la Weimar Jazz Database [10] ou bien l'archive des enregistrements du Montreux Jazz Festival [4].

Les grilles sont un point particulier d'étude. Chemillier a proposé de générer des grilles suivant des grammaires formelles [2]. Shanahan et Broze [14] ont collecté 1200 grilles d'accords depuis le forum *irealb*¹ pour constituer un corpus de 1160 morceaux en format *jazz* (format apparenté au *kern* [7]). Sur 50 ans de jazz (entre 1920 et 1970) que traversent ce corpus, leur étude a notamment montré que les progressions de deux accords les plus couramment observés étaient ii-V, V-I et V/ii-ii [1]. En 2020, Harasim et al. ont analysé 150 de ces grilles (en 3/4 ou 4/4) sous forme d'arbres syntaxiques dans la JHT (Jazz Harmony Treebank) [6].

Chaque arbre est construit à l'aide d'une grammaire qui repère dans la grille les prolongations et les préparations d'accords (partant de la sous-dominante ou de la dominante), dégageant pour chaque morceau des plus hauts niveaux de représentation de l'enchaînement d'accords. On peut ainsi directement relier des accords loin les uns des autres dans la grille. La JHT a été réalisée manuellement, certes en voulant reconstruire prolongations et préparations [12], et également avec une relecture et validation, mais jusqu'à ce jour sans visée algorithmique.

Motivations. Nous étudions ici les arbres de la JHT, que ce soit par une analyse ou une ébauche de reconstruction. Nous avons conscience que différentes versions de grilles coexistent pour une même œuvre et que par ailleurs, le corpus sur lequel s'appuie la JHT repose sur des accords très souvent non enrichis (pas de 9^e, 11^e... ou de précision

1. <http://irealb.com/forums>

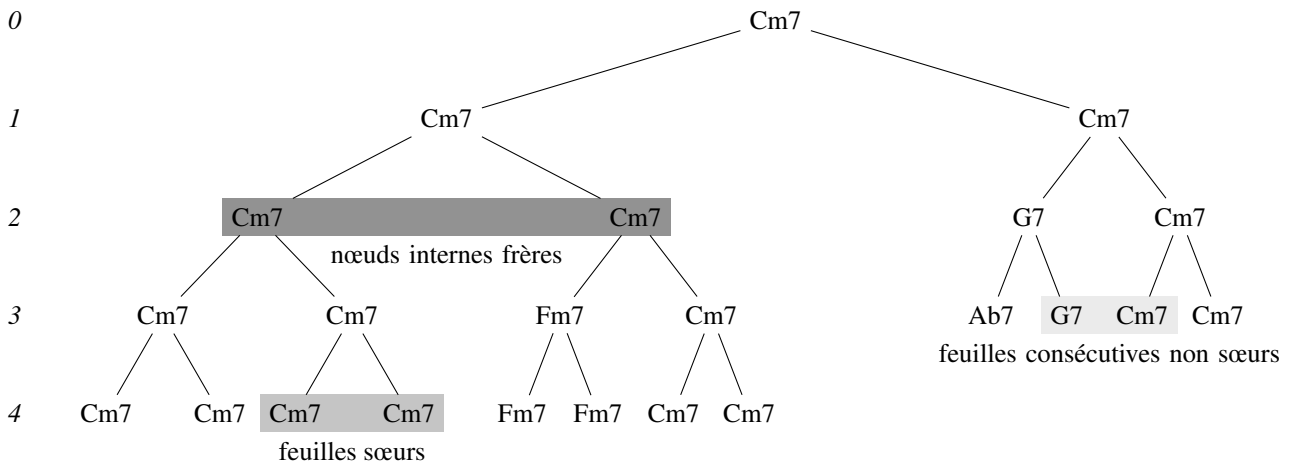


Figure 1 – *Mr. P.C.* (John Coltrane, 1959). Arbre proposé par la JHT (voir <https://github.com/DCMLab/JazzHarmonyTreebank/blob/master/tree-plots.md#mr-pc>), avec la profondeur des nœuds en partant de la racine. La grille ayant servi de référence à la construction de cet arbre provient du corpus *irealPro* [14].

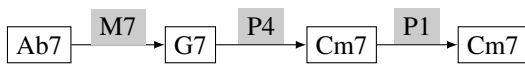


Figure 2 – *Mr. P.C.* (John Coltrane, 1959) : enchaînements d’intervalles entre les fondamentales des quatre derniers accords de la grille.

sur le renversement utilisé par exemple). Ces choix engendrent une perte d’informations sur la richesse harmonique de la musique analysée. L’étude proposée se concentrera donc sur l’analyse des fondamentales des accords, ce qui nous semble un bon point de départ pour comprendre les mouvements harmoniques. Confronter l’analyse humaine proposée par la JHT à une analyse automatique doit contribuer à confirmer ou non des présupposés notamment sur les progressions d’accords et ainsi entamer une réflexion sur la construction de ces arbres.

Contenu. Dans cet article, nous analysons la répartition des enchaînements d’accords en s’appuyant sur les arborescences proposées par la JHT (Section 2), puis nous proposons et évaluons un premier algorithme pour reconstruire ces arbres, en regroupant les accords suivant plusieurs scénarios (Section 3). Analyser et reconstruire ces structures n’est pas une fin en soi, par exemple pour la pratique musicale, mais permet de progresser dans la compréhension des grilles et des modèles d’analyse.

2. ANALYSE DE LA STRUCTURE ARBORESCENTE DES GRILLES

Est-il possible de retrouver des enchaînements d’accords significatifs en analysant les arbres de la JHT ? Est-ce que des règles peuvent être formalisées pour ensuite reconstruire (semi-)automatiquement ces arbres ?

Pour les 150 morceaux de la JHT sont disponibles l’arbre d’analyse, la tonalité d’ensemble, les accords et leur posi-

tion dans la grille. Nous analysons ici les relations entre les nœuds de ces arbres (Figure 1), en se focalisant sur les intervalles entre les fondamentales des accords, selon qu’ils soient portés par des feuilles consécutives, des feuilles sœurs, ou bien selon leur profondeur.

Dans la suite, les intervalles sont représentés à l’intérieur d’une même octave, avec leur nature – principalement juste (P), mineur (m), majeur (M) – et un chiffre compris entre 1 et 7. Par exemple, P4 désigne une quarte juste ascendante (ou son renversement, une quinte juste descendante), comme entre les fondamentales de G7 et Cm7 (Figure 2). P1 désigne un unison (comme entre Cm7 et Cm7, mais aussi entre Cm7 et C6).

Il est enfin à noter que dans la JHT, les répétitions à la fin des grilles ont été supprimées et que, si l’accord de tonique n’était pas présent à la fin, il a été rajouté pour permettre que la racine soit bien le premier degré.

Feuilles de l’arbre consécutives. Pour retrouver l’enchaînement des accords dans la grille, il suffit de parcourir les feuilles de l’arbre reconstruit. Les quartes ascendantes (ou quintes descendantes) P4 représentent 44.0% des intervalles de la JHT entre les feuilles consécutives (Table 1), confirmant les enchaînements sur le cycle de quintes tels que II-V ou V-I et les résultats de Broze et Shanahan sur le corpus élargi [14].

D’autres intervalles se retrouvent. Par exemple, la seconde mineure descendante (ou septième majeure ascendante) M7 s’explique par l’emploi de *substitution tritonique*, en particulier lorsque un V-I est remplacé par un bII-I.

Feuilles sœurs. S’intéresser à la construction de l’arbre, c’est notamment se poser la question des regroupements de feuilles à plus haut niveau. Les enchaînements de P4 et de P1 sont très majoritaires (86.5%) entre feuilles sœurs. On constate aussi que les P1 (principalement des suites du

intervalles	P1	P4	M7	P5	M2	M6	m7	A4	m2	A8	m3	M3	A1	m6	d5	A4
feuilles consécutives	24.0	44.0	5.8	4.2	3.7	4.5	1.8	1.8	1.7	1.7	1.7	1.4	1.2	1.0	0.9	1.8
feuilles sœurs	43.5	43.0	5.6	1.5	1.4	0.7	0.1	0.1	0.9	1.4	0.2	0.3	0.8	0.4	0.2	0.1
nœuds internes frères	33.5	46.5	6.0	1.6	4.7	1.4	0.3	0.3	0.5	1.8	0.9	0.6	0.4	0.8	0.2	0.3
profondeur 1	90.0	4.7	–	–	–	–	–	–	–	–	3.3	–	–	2.0	–	–
profondeur 2	49.6	37.6	1.8	1.1	3.9	1.1	–	–	–	1.4	1.8	0.4	–	1.4	–	–
profondeur 3	43.1	33.4	6.2	2.6	6.8	2.6	0.4	0.6	–	1.0	1.2	0.6	–	1.4	0.2	0.6
profondeur ≥ 4	27.5	51.7	6.7	1.6	4.7	1.3	0.3	0.3	0.6	2.1	0.7	0.7	0.6	0.6	0.3	0.3

Table 1 – Présence dans la JHT des intervalles les plus fréquemment rencontrés en %. Les autres intervalles ont une fréquence inférieure à 0.2%

Algorithme 1 : Reconstruction des progressions harmoniques arborescentes.	
Entrées :	un tableau de feuilles $[f_1 \dots f_n]$, un scénario d'intervalles $[I_1 \dots I_m]$
Sorties :	un tableau d'arbre(s)
Initialiser A à $[f_1 \dots f_n]$	
pour chaque intervalle I du scénario faire	
si $I = P1^4$ alors	
pour toute suite de 4 arbres $[a_i, a_{i+1}, a_{i+2}, a_{i+3}]$	
avec $r = \rho(a_i) = \rho(a_{i+1}) = \rho(a_{i+2}) = \rho(a_{i+3})$, faire	
remplacer dans A ces 4 arbres par un seul arbre $((a_i, a_{i+1}), (a_{i+2}, a_{i+3}))$ de racine r .	
fin	
sinon	
pour toute suite de 2 arbres $[a_i, a_{i+1}]$	
avec $\rho(a_{i+1}) - \rho(a_i) = I$, faire	
remplacer dans A ces 2 arbres par un seul arbre (a_i, a_{i+1}) de racine $\rho(a_{i+1})$.	
fin	
fin	
fin	
Renvoyer A	
Pour un arbre a , $\rho(a)$ donne la tonique de l'accord porté par sa racine.	
Les parcours "Pour toute suite..." se poursuivent de gauche à droite tant qu'il y a des suites d'arbres vérifiant la condition.	

“même accord”) sont généralement privilégiés avant les P4 (Table 1). Il y a même souvent des regroupements de P1 liant quatre accords identiques.

Ceci nous donne un indice fort sur les premières règles à appliquer si on souhaite reconstruire algorithmiquement un tel arbre.

Étude par profondeur. Afin de percevoir la structure a un plus haut niveau, nous avons analysé les intervalles entre tous les nœuds frères, qu'ils soient internes ou non, selon leur profondeur dans l'arbre, en considérant 0 comme la racine (Table 1). Le niveau 1 a une plus forte proportion de P1, ce qui traduit les prolongations fortes des accords correspondants aux tonalités des différents morceaux, y compris le retour à l'accord de I lorsque l'on reboucle sur le premier accord de la grille.

On retrouve des fréquences plus importantes pour les P4 lorsque la profondeur augmente, traduisant l'enchaînement des cycles de quinte couramment rencontrés dans le corpus.

3. VERS UNE RECONSTRUCTION ALGORITHMIQUE DES ARBRES

Prenant en considération les éléments précédents, nous proposons un algorithme de reconstruction d'un arbre, possiblement partiel, étant donné une grille d'accords. Les prochains paragraphes décrivent cet algorithme ainsi que l'évaluation de son efficacité.

Algorithme et scénarios d'intervalles. L'idée de base est de détecter les P4, mais également les P1 ainsi que les autres intervalles. L'algorithme regroupe ainsi au fur et à mesure les accords en testant les intervalles selon un scénario qui est une suite d'intervalles (Algorithme 1). L'algorithme maintient un tableau A de sous-arbres (initialement toutes les feuilles), et, pour un intervalle I du scénario, regroupe de gauche à droite les sous-arbres dont les racines respectent l'intervalle I . En sortie d'algorithme nous obtenons donc un tableau de sous-arbres, éventuellement réduit à un seul arbre lorsque tout a été regroupé.

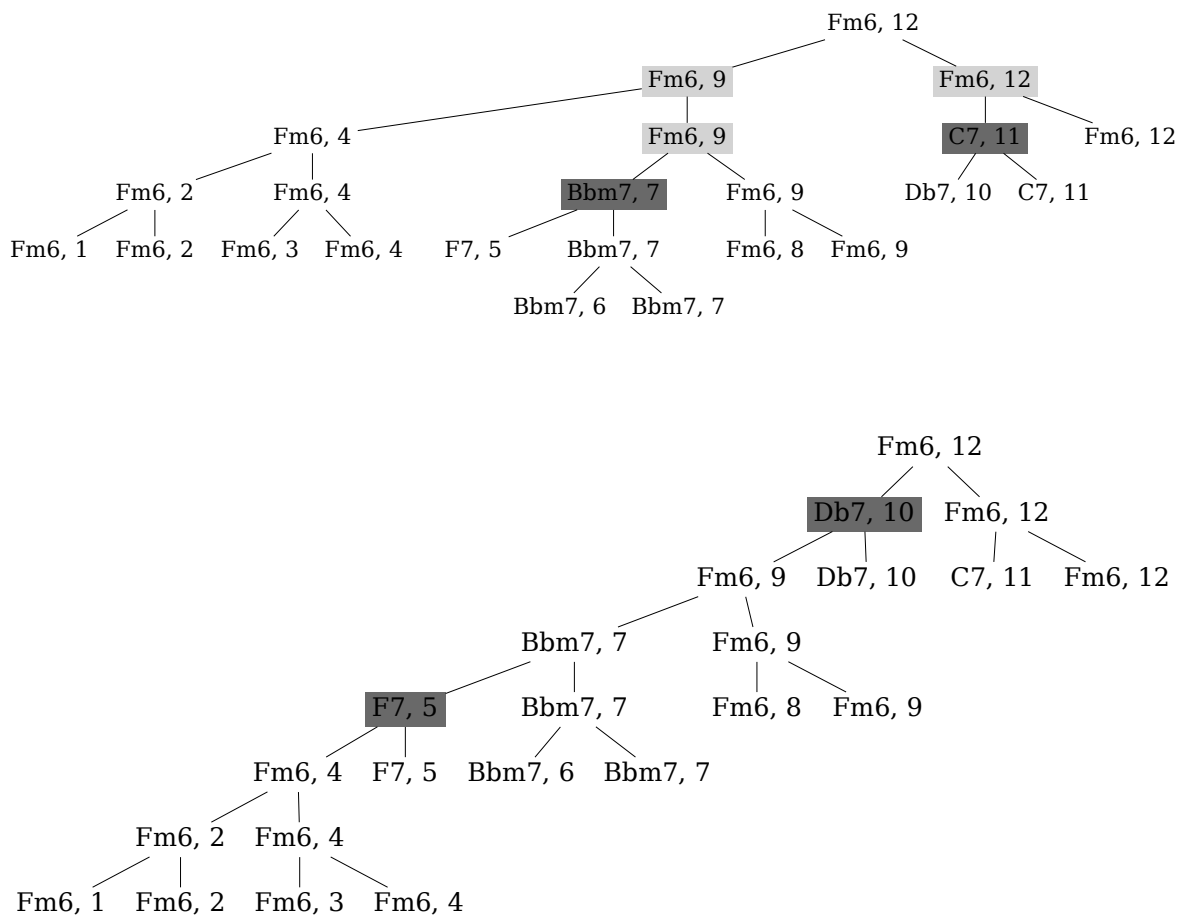


Figure 3 – Analyse arborescente de la grille d’*Interplay* (Bill Evans, 1963). (*haut*). Arbre proposé par la JHT (voir <https://github.com/DCMLab/JazzHarmonyTreebank/blob/master/tree-plots.md#interplay>), avec, pour chaque noeud interne, remontée depuis le fils droit de la dernière position dans lequel l’accord apparaît. (*bas*). Arbre reconstruit par le scénario 4, avec également remontée des positions depuis le fils droit. Deux noeuds (en grisé sombre) sont mal prédits, menant à une précision de 81.8%. D’autres regroupements auraient du être fait à la place de ces noeuds (noeuds grisés sombre dans l’arbre de la JHT), cette différence se propageant sur les noeuds parents (noeuds grisés clairs).

Nous avons retenu quatre scénarios :

- scénario 1 : P1*, P4, P1*
- scénario 2 : P1*, P4, P1*, P4, P1*, [P5, M2, M6], P1*
- scénario 3 : P1*, P4, M7, P1*, P4, [P5, M2, M6], P1*, P4, P1*
- scénario 4 : P1*, P4, M7, P1*, P4, [P5, M2, M6], OTHERS, P1*, P4, P1*

Tous les scénarios font régulièrement intervenir P1*, qui est en fait une suite P1⁴ - P1 : nous cherchons à rassembler d’abord s’ils existent quatre accords identiques deux à deux, puis à regrouper deux par deux ces accords. Ceci traduit la présence non négligeable dans les grilles analysées de P1, mais aussi de trois P1 consécutifs.

Nous avons également regroupé deux ensembles d’intervalles : [P5, M2, M6], de part leur fréquence ainsi que OTHERS, tous les autres intervalles : [A8, m3, m6, M3,

m2, A1, m7, A4, d5, A2, A3, d4, d1]. Les scénarios 1-3 ne prennent pas en considération la totalité des intervalles potentiellement rencontrés, et conduisent à des reconstructions partielles et donc à moins de noeuds que l’arbre de la JHT.

Évaluation. Nous évaluons l’algorithme en comparant les arbres prédits avec ceux de la JHT. Comparer des arbres peut se faire par distance d’édition [5]. Nous proposons ici de repérer les noeuds différents ou plus précisément les noeuds qui n’ont pas été regroupés de la même manière par notre algorithme.

On numérote dans l’ordre de la grille les feuilles puis on fait remonter ces positions à chaque fois du fils droit à son père et ceci jusqu’à la racine pour les 2 arbres (JHT et reconstruit). Un *noeud bien prédit* est un noeud ayant même accord *et* une même position.

Prenons pour exemple *Interplay*, blues mineur de Bill Evans – même si le caractère tonal du blues peut être contesté. Le scénario 4 reconstruit totalement un arbre (Figure 3). La plupart des regroupements sont corrects, mais deux nœuds internes sont mal prédits :

- (*F7,5*) au lieu de (*Bbm7, 7*) : le scénario regroupe par P1* (*Fm6,4*) et (*F7,5*), sans considérer le caractère mineur ou majeur des accords ;
- (*Db7,10*) au lieu de (*C7,11*) : le regroupement M7 aurait du ici être effectué avant le P4 vers (*Fm6, 12*).

Un arbre à n feuilles a $n - 1$ nœuds internes (racine comprise). Ainsi on peut comptabiliser pour les nœuds internes de notre arbre reconstruit (ou de nos sous-arbres), le nombre de nœuds bien prédits m (par rapport à ceux de la JHT) et ceux mal prédits m' . On définit ainsi la *précision* $p = \frac{m}{m+m'}$ et la *sensitivité* $s = \frac{m}{n-1}$ de l'algorithme sur un morceau.

Dans *Interplay*, il y a $n = 12$ feuilles et 11 nœuds internes, tous prédits, avec $m = 9$ nœuds internes bien prédits et $m' = 2$ mal prédits (Figure 3). On a donc ici $p = s = \frac{9}{11} \approx 81.8\%$.

scénario	precision moyenne	sensitivité moyenne
1	0.78	0.56
2	0.72	0.60
3	0.72	0.64
4	0.69	0.68

Table 2 – Sensitivité et de la précision moyenne de la reconstruction sur les 150 grilles de la JHT avec les différents scénarios.

Implémentation et Résultats. L'algorithme 1 a été implémenté en Python, en analysant les fichiers de la JHT (fournis au format *JSON*) et en utilisant *music21* [3] pour exploiter les accords et intervalles. Le code est disponible en licence libre GPLv3+ à l'adresse www.algomus.fr/code.

La Table 2 détaille les précisions et sensibilités moyennes de tous les scénarios. Le scénario 1 est le plus précis (78%)². De plus 28 morceaux (18.6% du total) dépassent les 90% de précision, les nœuds des sous-arbres reconstruits sont très proches de ceux rencontrés dans l'arbre de la JHT.

Les scénarios 2, 3 et 4 prennent en compte plus d'intervalles et prédisent donc plus de nœuds, d'où une sensibilité accrue (68% en moyenne pour le scénario 4), même si c'est parfois au prix d'une baisse de la précision (69%). Par exemple, dans *Interplay*, le nœud (*Db7, 10*) mal prédit

2. Si l'on considérait aussi les feuilles, qui sont bien identifiées entre les arbres reconstruits à ceux de la JHT, la précision serait ici de 89.0% et la sensibilité de 78.1%.

est ainsi la conséquence du M7 dans le scénario 4 et fait baisser la précision.

4. CONCLUSION ET PERSPECTIVES

Les arbres de la JHT révèlent la structuration harmonique des grilles. Les regroupements sont en particulier effectués suivant les P1 et P4, confirmant les prolongations et préparations dans les progressions harmoniques. Rassemblant successivement les accords selon ces intervalles les plus fréquents, l'algorithme proposé, selon les scénarios, arrive à traiter la majorité des accords avec une précision allant de 69% à 78%, et donne ainsi une ébauche de structuration de la grille. Certes, le but n'est pas uniquement de reconstruire automatiquement ces arbres car la vérification de leur pertinence reste à réaliser. Cependant, progresser dans ces algorithmes permet de progresser aussi dans la compréhension du langage musical et de sa modélisation.

Le choix initial de ne s'intéresser qu'aux intervalles entre les fondamentales pouvait paraître surprenant car il mettait de côté la fonction harmonique des accords. Dans plusieurs situations, il est très clair que les regroupements n'auraient pas été les mêmes. Les scénarios proposés ont tenté de tenir compte de cet écueil en intégrant des intervalles caractérisant quelques enrichissements courants en jazz (comme par exemple les substitutions tritoniques). De la même manière, le caractère mineur ou majeur d'un accord n'est pas nécessairement significatif dans l'analyse des enchaînements d'accords.

Il est plus difficile de regrouper les intervalles moins fréquents, en particulier lors de modulations ou de certaines substitutions. Les scénarios pourraient être affinés en ce sens. Plus généralement, il serait possible de générer et d'évaluer des scénarios dans une logique d'apprentissage machine, qui, vu la taille du corpus, serait délicate à effectuer sans sur-apprentissage.

L'évaluation serait aussi à approfondir. D'autres métriques sont possibles, notamment pour préciser la condition selon laquelle un nœud est bien prédit (remontée du fils gauche ou d'un intervalle). Plus généralement, on pourrait estimer une distance d'édition, mais c'est alors un défi de préciser quelles opérations d'édition seraient pertinentes ainsi que leur pondération.

Enfin, l'algorithme actuel ne prend en compte que la succession des accords. Tenir compte de la durée de ces accords, et, plus généralement, de la *forme* du morceau, permettrait probablement d'avoir des regroupements qui seraient structurellement plus pertinents.

Remerciements. Nous remercions Ken Déguernel pour les discussions sur cette étude ainsi que les relecteurs pour leurs remarques pertinentes.

5. REFERENCES

- [1] Yuri Broze and Daniel Shanahan. Diachronic changes in jazz harmony : A cognitive perspective. *Music Perception : An Interdisciplinary Journal*, 31 :32–45, 2013.
- [2] M. Chemillier. Improviser des séquences d’accords de jazz avec des grammaires formelles. In *Journées d’Informatique Musicale (JIM 2001)*, pages 121–126, 2001.
- [3] Michael Scott Cuthbert and Christopher Ariza. music21 : A toolkit for computer-aided musicology and symbolic music data. In *International Society for Music Information Retrieval Conference (ISMIR 2010)*, pages 637–642, 2010.
- [4] Alain Dufaux and Thierry Amsallem. The Montreux Jazz Digital Project : From preserving heritage to a platform for innovation. *Journal of Digital Media Management*, 7(4) :315–329, 2019.
- [5] Serge Dulucq and H el ene Touzet. Analysis of tree edit distance algorithms. In *Combinatorial Pattern Matching (CPM 2003)*, pages 83–95, 2003.
- [6] Daniel Harasim, Christoph Finkensiep, Petter Ericson, Timothy J. O’Donnell, and Martin Rohrmeier. The Jazz Harmony Treebank. In *International Society for Music Information Retrieval Conference (ISMIR 2020)*, pages 207–215. ISMIR, 2020.
- [7] David Huron. Music information processing using the Humdrum toolkit : Concepts, examples, and lessons. *Computer Music Journal*, 26(2) :11–26, 2002.
- [8] 99% invisible. *The Real Book*, episode 438, 2021.
- [9] Fred Lerdahl and Ray Jackendoff. *A Generative Theory of Tonal Music*. MIT Press, 1983.
- [10] Martin Pfeleiderer, Klaus Frieler, Jakob Abe er, Wolf-Georg Zaddach, and Benjamin Burkhart, editors. *Inside the Jazzomat – New Perspectives for Jazz Research*. Schott Campus, 2017.
- [11] Martin Rohrmeier. Towards a generative syntax of tonal harmony. *Journal of Mathematics and Music*, 5(1) :35–53, 2011.
- [12] Martin Rohrmeier. The syntax of jazz harmony : Diatonic tonality, phrase structure, and form. *Music Theory and Analysis (MTA)*, 7 :1–63, 2020.
- [13] Heinrich Schenker. *Der freie Satz*. Universal Edition, 1935.
- [14] Daniel Shanahan and Yuri Broze. iRealPro corpus of jazz standards, 2019.

A CONSTRAINT FORMALIZATION OF FUX'S COUNTERPOINT

Damien Sprockeels, Thibault Wafflard, Peter Van Roy
Université catholique de Louvain
{damien.sprockeels, peter.vanroy}@uclouvain.be
thibault.wafflard@gmail.com

Karim Haddad
IRCAM
karim.haddad@ircam.fr

RÉSUMÉ

La composition musicale assistée par ordinateur est un domaine en plein essor depuis plusieurs décennies maintenant, particulièrement dans le domaine de l'intelligence artificielle appliquée à la composition musicale. Cet article vise à contribuer à ce domaine en offrant un outil mathématique et informatique formalisant le contrepoint à deux voix selon Johann Joseph Fux à l'aide de la programmation par contraintes. Cet outil s'inscrit dans un projet plus ambitieux qui a pour but de créer un outil complet basé sur la programmation par contraintes afin d'assister les compositeurs dans leur processus de création, en ne nécessitant aucunes notions en informatique.

1. INTRODUCTION

The last decades have seen a great increase in the use of computers for musical composition. Applications are endless : from software aiming to facilitate the composer's work by providing an intuitive interface to write scores or combine musical samples into a piece (LogicPro [16], Sibelius [25], OpenMusic [20]), to using machine learning to generate music fitting a given description or a certain style (MusicLM [1]). Our work uses an approach based on constraint programming. We observe that there is a close correspondence between the theory of musical harmony and the abilities of constraint programming. In this way, we aim to build a tool that allows the composer to express musical ideas which are then implemented in terms of constraints. The tool uses a constraint solver to generate musical solutions that satisfy the constraints. This allows the composers to think exclusively about musical ideas as the tool is in charge of expressing these ideas as a score.

Constraint programming Constraint programming is a field in computer science that allows for the specification of a set of rules, called constraints, to be applied to a set of variables modelling a problem. A solver then searches for solutions for the problem that respect the constraints specified. Ever since the invention of constraint programming, it has been an appealing idea to encode musical concepts as constraints. Many concepts can be expressed in this way, as previous work has shown ([28], [29], [5]). Constraint programming has the ability, given an appropriate interface, to allow the composer to specify their musical ideas at a high level of abstraction and it allows an



Figure 1. Example counterpoint in the style of Fux [14].

iterative process where the composer can refine his or her musical ideas depending on the musical solution given by the constraint solver.

Music composition using constraints Constraint programming has often been applied to composition [23, 19, 17, 21], but this has not yet led to a practical tool usable by composers. There are two main reasons for this : first, past research was mainly focused on investigating how to express musical rules as constraints and second, the software tools that were developed required skills in programming and mathematics. A third problem that we have recognized in our own work is that in order to be practically useful to a composer, the tool must incorporate knowledge of the desired musical style. If it does not have such knowledge, the composer must first create it from scratch which makes it much harder to use. That is the main drawback of our own previous work ([2], [6], [3]). Indeed, the prototypes we developed are limited to a small set of constraints and do not support a coherent style. Despite this, they give useful proofs of concept and they can give interesting solutions to musical problems. The present work goes beyond these tools and is a first step towards a tool that supports a coherent musical style.

Choice of the Fux theory Musical styles are complex and few have been analyzed to a degree that would allow to easily formalize them as constraints. As our starting point, we have therefore chosen a relatively simple style that has been defined thoroughly, namely musical counterpoint. We have chosen the theory of counterpoint as presented by Johann Joseph Fux in his classic work *Gradus Ad Parnassum*, which was first published in 1725 [13]. Figure 1 gives an example of fifth species counterpoint for two voices in the Fux theory. We have chosen the Fux theory for three main reasons. First, it is presented in a clear and progressive way that steps from simple to more complex forms of counterpoint. Second, it is defined

with a strict and comprehensive set of rules that make it straightforward to formalize in a mathematical language. Third, it is considered as a classic work and many of its rules are still present in various musical styles that are popular today. We have chosen the Fux counterpoint theory over other counterpoint theories, such as Jeppesen [15], Dubois [26], and Kœchlin [4], because of its simplicity, clarity, and completeness. Mikael Laurson gives a formalization of counterpoint based on the Bach Chorals in the context of the PWConstraints library of PatchWork [19].

Previous attempts at modeling Fux counterpoint Several earlier attempts at modeling Fux counterpoint have been made, some even using constraint programming. The first attempt we could find was Schottstaedt [24], who used a series of conditional statements to ensure a generated counterpoint would satisfy Fux’s rules. His work provides a preference system for each rule as well as a relatively good efficiency (however, the fifth species could in some cases be computationally heavy). The main drawbacks were that the search for a solution was not customizable and that his work required computer science knowledge to be useable by a composer. Ovans and Davidson [22] provide an interactive interface where the composer can select a note and observe dynamic changes in the available possibilities for the remaining notes such that the result satisfies Fux’s rules for the first species of two voice counterpoint. It does not however allow to generate counterpoint for other species. Cope [7] used a learning approach to generating first species counterpoint. He does not implement every rule for this species that Fux gives, but rather uses an evaluation based on six criteria representing the main rules as well as the examples from Fux’s treatise which are used to train the model. Torsten Anders [28], in his Ph.D. work on Strasheela presents a constraint formalization of the first species of the Fux theory for two voices. We go beyond this to formalize all five species of Fux two-voice counterpoint. More recently, Herremans and al. [8] developed a tool allowing to generate fifth species counterpoint using a variable neighbourhood search algorithm to try and maximize an objective function. Our work goes a step further by analysing different translations of Fux’s treatise and by allowing complete freedom for the composer when it comes to how strictly they want to follow the rules.

Ongoing project This paper is one step of a long-term collaborative project on computer-aided composition using constraint programming between UCLouvain and IRCAM. We briefly summarize our past work. In 2020, Baptiste Lapière [2] implemented Rhythm-Box, a proof of concept tool to explore constraints for rhythm. Rhythm-Box was implemented in the OpenMusic visual programming platform for music composition [20] using the Gecode constraint solver [10]. This work was accompanied by the development of GiL [11], an interface between Gecode and OpenMusic. In a next step, Damien Sprockeels developed Melodizer [6], a proof of concept tool applying

constraint programming towards the modeling of simple musical rules for pitch. This work was presented at the IRCAM Forum in March 2022. Finally, we extended Melodizer to combine rhythm and pitch and to support polyphony, giving Melodizer 2.0 [3]. Melodizer 2.0 has an improved software architecture that supports a hierarchical approach to create music by combining constraint blocks. Future work aims to develop practical tools that target different musical styles, as explained in Section 5.

Contributions This paper makes three contributions :

- A mathematical formalization of the Fux theory of counterpoint for two voices. For brevity, the present paper presents the full formalization of the third species and explains how the other species are formalized. The full formalization is expected to be available by June 2023 in the master’s thesis of Thibault Wafflard [27].
- An implementation of this formalization in OpenMusic using the Gecode constraint solver. This implementation is available on github as explained in Section 4.1.
- An evaluation of this implementation. We compare solutions found by the constraint solver with the counterpoints suggested by Fux in his treatise. We also show how it is possible to generate new counterpoints using the formalization.

Structure of the paper The paper is organized as follows :

- Section 2 summarizes the Fux theory of two-voice counterpoint.
- Section 3 gives the formalization of the third species of the Fux theory for two-voice counterpoint.
- Section 4 explains the implementation and evaluates it with respect to the Fux treatise.
- Section 5 explains how we intend to continue this work towards a full-fledged tool for music composition.
- Section 6 recapitulates the work and gives some conclusions.

2. FUX THEORY OF COUNTERPOINT

The Gradus Ad Parnassum by Johann Joseph Fux is considered as a classic reference for counterpoint [13]. Famous composers such as Haydn, Beethoven, and Mozart are all said to have studied it. The treatise is organised in a didactical and progressive manner. It is presented as a conversation between a master, Aloys, and his student, Joseph, where Joseph is discovering and applying the rules of counterpoint aided by Aloys. The treatise starts with two-voice counterpoint, which is presented stepwise in five progressive forms, called “species”. The elaboration of a counterpoint starts with a theme given by the composer, called the cantus firmus. The rules of counterpoint are then used to generate additional voices, called counterpoints. In the first species, both the cantus firmus and

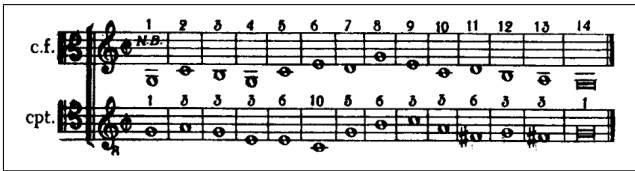


Figure 2. Example counterpoint for the first species [14].

counterpoint consist of measures of whole note against whole note. After a didactic conversation between master and student, the treatise introduces the second species, where the counterpoint consists of two half-notes per measure against a whole note cantus firmus. This leads to the third species of counterpoint, where we have four quarter notes against a whole note, the fourth species which is essentially the first species with syncopation, and finally into the fifth species which is called “florid counterpoint” and is the result of the four other species combined. The treatise then adds a third voice and repeats the same process, and does it again for four voices. We now briefly summarize the rules for each species of two-voice counterpoint. Section 3 then gives the formalization of the third species.

Terminology To define his rules, Fux occasionally uses terms that are no longer used today. He uses the terms “thesis” and “arsis”, the former referring to the downbeat, and the latter referring to the upbeat. In $\frac{4}{4}$ signature, they correspond to the first and third beat respectively. He also uses the term “diminution” to refer to a melodic interval of a third that is filled by a diatonic note. When using the term “perfect consonance”, he refers to unisons, fifths and octaves. When he uses the term “imperfect consonance”, he refers to thirds and sixths.

First species For the first species, the rules are as follows. All intervals between the counterpoint and the cantus firmus must be consonances, and imperfect consonances should be preferred; The first and last harmonic intervals of the counterpoint must be a perfect consonance; The key tone is tuned according to the first note of the cantus firmus; A unison is only allowed on the first and last notes; In the next to last measure, a major sixth or a major third must exist depending if the cantus firmus is in the lower or upper part; The melodic intervals can not be greater than a minor sixth; There can not be direct motion to reach a perfect consonance; There cannot be a *battuta* octave, i.e., an octave reached by a lower voice going up and an upper voice going down more than a third skip; Contrary motion is always preferred to oblique or parallel motions. Figure 2 shows an example of two voice first species counterpoint from Fux’s treatise.

Second species The rules from the first species regarding harmony are applied on the notes on thesis, while the rules regarding melody are applied to all notes. Rules regarding motion are either modified or more complex. Additionally, the following rules are added : Notes on arsis



Figure 3. Example of counterpoint (second species) [14].



Figure 4. Example counterpoint for the third species [14].

cannot be dissonant unless there is a diminution; In addition to the rules for the next to last note from the first species, the thesis note of the next to last measure should be an interval of a perfect fifth, and if that is not possible, an interval of a sixth must be used; If the two voices are getting so close that there is no contrary motion possible without crossing each other, then the melodic interval of the counterpoint can be an octave leap; Two consecutive notes cannot be the same; If the melodic interval of the counterpoint between the thesis and the arsis is larger than a third, then the motion is perceived on the basis of the arsis note. Figure 3 shows an example of two voice second species counterpoint.

Third species For the third species, the rules are as follows. If five notes follow each other by joint degrees in the same direction, then the third note must be consonant; If the third note of a measure is dissonant then it must be a diminution and the second and the fourth notes must be consonant; It is best to avoid the second and third harmonies of a measure to be consonant with a one degree melodic interval between them; In addition to the rule of the next to last note of the first species, if the cantus firmus is in the upper part, then the first note of the next to last measure should be a minor third above the cantus firmus; Each note and its two beats further peer are preferred to be different; The motion is perceived on the basis of the fourth note. Figure 4 shows an example of two voice third species counterpoint.

Fourth species For the fourth species, dissonance can appear in the thesis but the arsis must be consonant. Indeed this species inherits most of the rules from the first species but shifted by half a measure, corresponding to the



Figure 5. Example counterpoint (fourth species) [14].



Figure 6. Example counterpoint for the fifth species [14].

Interval	Unison/Octave	Second		Third		Fourth
Type	Perfect	Minor	Major	Minor	Major	Perfect
Value	0	1	2	3	4	5
Interval	Tritone	Fifth	Sixth		Seventh	
Type	$\sharp 4 / b 5$	Perfect	Minor	Major	Minor	Major
Value	6	7	8	9	10	11

Table 1. MIDI values of the intervals over an octave range.

delay resulting from syncopation. The rules for this species are as follows. Arsis harmonies must be consonant; Dissonant harmonies must be followed by the next lower consonant harmony; If the cantus firmus is in the lower part then no second harmony can be preceded by a unison/octave harmony; If the cantus firmus is in the upper part, then no harmonic seventh interval can occur; For the rule of species 1 about the next to last measure to be satisfied, if the cantus firmus is in the lower part, then the harmonic interval of the thesis note must be a seventh, and if the cantus firmus is in the upper part, then the harmonic interval of the thesis note must be a second; Thesis half notes should be the same as the preceding arsis half note and be linked to it; Each arsis note and its two measures further peer are preferred to be different. Figure 5 shows an example of two voice fourth species counterpoint.

Fifth species The fifth species is the result of the combination of the previous four species and more particularly species 3 and 4. It can be seen as small pieces of these species combined into a single counterpoint. It is then no longer a question of new rules but rather of the previous rules so that they can all merge and apply to the notes depending on the species from which they come. Figure 6 shows an example of two voice fifth species counterpoint.

3. FORMALIZATION OF THE THIRD SPECIES FOR TWO-VOICE COUNTERPOINT

Presenting the formalization for all species would be too long for this paper. We choose to detail the formalization of the third species because it is more interesting than the others and in addition it reuses rules from the first species. We exclude the fifth species since it is essentially just the union of all the rules of the other species. The complete formalization will be available in June 2023 as a master's thesis [27]. To make it easier to represent intervals mathematically, we will be using MIDI notation. Table 1 presents the main intervals together with their MIDI value.

It is important to realize that Fux's treatise is a natural-language text that does not always present rules in a way that is easy to transform into mathematical expressions.

Indeed, some rules are considered as obvious by Fux and are not mentioned explicitly but all the examples present in his treatise seem to follow them. Some rules deal with composer preferences, which requires a cost function in the formalization. Assigning values to those costs can alter the accuracy of the representation of Fux's ideas, especially because preferences are subjective, which is why in the implementation discussed in section 4 the costs are adjustable by the user and not considered to be fixed. Finally, some rules differ depending on the edition and translation of the treatise. For these rules, we cross-referenced three different versions (French [12], English [14], and Latin [13]) and used a critical analysis of Fux's examples to determine the most appropriate formulation. For example, rule 9 in section 3.3 varies depending on the translation.

In order to express the musical rules mathematically as clearly as possible, we first define some common constants and variables.

3.1. Constants

m is the number of notes in the cantus firmus. Since the cantus firmus consists exclusively of whole notes, it is also the number of measures of the cantus firmus and by extension of the counterpoint.

$$m \in \mathbb{N} \quad (1)$$

n is the number of notes in the counterpoint.

$$n \in \mathbb{N} \\ n = 4 * (m - 1) + 1 \quad (2)$$

Since the counterpoint in the last measure is a whole note.

Cons_p is the set of perfect consonances in the form of MIDI values representing the interval between two notes.

$$Cons_p = \{0, 7\} \quad (3)$$

Cons_{imp} is the set of imperfect consonances, also in the form of MIDI values representing the interval between two notes.

$$Cons_{imp} = \{3, 4, 8, 9\} \quad (4)$$

Cons is the set of all consonances.

$$Cons = Cons_p \cup Cons_{imp} \quad (5)$$

Cf is an array of size *m* representing the cantus firmus.

$$Cf = (a_1, a_2, \dots, a_m) \quad (6)$$

For all notes *a_i* we have *a_i* ∈ {0, 1, ..., 127}. *a_i* corresponds to the MIDI value of note *i* of the cantus firmus.

j is the index of a measure in the cantus firmus.

$$j \in [1, m] \quad (7)$$

I is an index excluding the last measure of the cantus firmus.

$$l \in [1, m - 1] \quad (8)$$

3.2. Variables

Cp is an array of size n representing the notes of the counterpoint.

$$Cp = (b_1, b_2, \dots, b_n) \quad (9)$$

For all notes b_i we have $b_i \in \{0, 1, \dots, 127\}$. b_i corresponds to the MIDI value of note i of the counterpoint. It is possible to access the values of the counterpoint on the basis of the index of the measure. For example, for measure j , the notes of the counterpoint are

$$\{Cp[4 * (j - 1) + 1], Cp[4 * (j - 1) + 2], \\ Cp[4 * (j - 1) + 3], Cp[4 * (j - 1) + 4]\} \quad (10)$$

H is a $m * 4$ array where m is the number of measures and the number of notes in the cantus firmus. It represents the harmonic interval between the counterpoint and the cantus firmus in absolute value.

$$\forall i \in \{1, 2, 3, 4\} \quad H[j, i] = |Cp[4 * (j - 1) + i] - Cf[j]| \quad (11)$$

M is an array of size $n-1$ representing the melodic interval between notes of the counterpoint.

$$\forall i \in [1, n - 1] \quad M[i] = |Cp[i + 1] - Cp[i]| \quad (12)$$

P is an array of size $m-1$ representing the relative motion of the counterpoint and the cantus firmus between two consecutive measures.

$$P = (c_1, c_2, \dots, c_{m-1}) \quad (13)$$

For all elements c_i , we have $c_i \in \{0, 1, 2\}$ where a value of 0 represents contrary motion, a value of 1 represents oblique motion and a value of 2 represents parallel motion.

3.3. Rules

It is important to mention that some rules are considered as obvious by Fux and are not mentioned explicitly in his treatise. Nevertheless, they are important and will be marked with a star * in the rest of this section. The following rules are valid for all species :

- The number of measures of the counterpoint is the same as for the cantus firmus*.
- The counterpoint must have the same time signature and tempo as the cantus firmus. The time signature must be $\frac{4}{4}$ *
- The counterpoint must be in the same key and mode as the cantus firmus*. Fux determines the key with the first note of the cantus firmus, and the mode is such that there are no accidentals. However, depending on the key, some notes can be borrowed from the major or minor mode of that key for the counterpoint.

— Chromatic melodies are forbidden* :

$$\forall k \in [1, n - 2] \\ \neg(M[k] = 1 \wedge M[k + 1] = 1 \wedge Cp[i] \neq Cp[i + 2]) \quad (14)$$

Rules inherited from the first species We now give the rules specific to the third species. The first eight rules are inherited from the first species and are only applied to the first note of the measures, except for the melodic rules (rule 6 and preference 2) that are applied to all notes.

Rule 1 All harmonic intervals must be consonances.

$$H[j, 1] \in Cons \quad (15)$$

Rule 2 The first and last harmonic intervals of the counterpoint must be a perfect consonance.

$$H[1, 1] \in Cons_p \wedge H[m, 1] \in Cons_p \quad (16)$$

Rule 3 The key tone is tuned according to the first note of the cantus firmus such that the first and last notes in the bass must be the tonic.

$$Cp[1] < Cf[1] \implies H[1, 1] \bmod 12 = 0 \\ Cp[m] < Cf[m] \implies H[m, 1] \bmod 12 = 0 \quad (17)$$

Rule 4 The counterpoint and the cantus firmus cannot play the same note at the same time except in the first and last measure.

$$\forall i \in [2, m - 1], \forall k \in [1, 4] \quad H[i, k] \neq 0 \quad (18)$$

Rule 5 If the cantus firmus is in the lower part, the harmonic interval of the penultimate note must be a major sixth; if the cantus firmus is in the upper part, the harmonic interval of the penultimate note must be a minor third.

$$Cp[4 * (m - 1)] > Cf[m - 1] \implies H[m - 1, 4] = 9 \\ Cp[4 * (m - 1)] < Cf[m - 1] \implies H[m - 1, 4] = 3 \quad (19)$$

Rule 6 Melodic intervals cannot exceed a minor sixth interval unless they are an octave.

$$\forall i \in [1, n - 1] \quad M[i] \notin \{9, 10, 11\} \quad (20)$$

Rule 7 Perfect consonances cannot be reached by direct motion.

$$H[l + 1, 1] \in Cons_p \implies P[l] \neq 2 \quad (21)$$

Rule 8 In the start of any measure, an octave cannot be reached by the lower voice going up and the upper voice going down more than a third skip*.

$$\begin{aligned}
UpperFirst &:= \max(Cp[4 * (l - 1) + 4], Cf[l]) \\
LowerFirst &:= \min(Cp[4 * (l - 1) + 4], Cf[l]) \\
UpperSecond &:= \max(Cp[4 * l + 1], Cf[l + 1]) \\
LowerSecond &:= \min(Cp[4 * l + 1], Cf[l + 1]) \quad (22) \\
&UpperSecond - UpperFirst \geq 4 \\
&\wedge LowerFirst - LowerSecond \geq 4 \\
&\implies H[l + 1, 1] \neq 12
\end{aligned}$$

This equation seems quite complicated but is actually straightforward. The first four lines of the equation aim to identify the upper and lower notes of the voices for the end of the previous measure and the beginning of the current measure. The remaining two lines say that if, between the fourth beat of measure l and the first beat of measure $l+1$, the upper voice goes up by a major third or more and the lower voice goes down a major third or more, the harmonic interval between the cantus firmus and the counterpoint on the first beat of measure $l+1$ can not be an octave.

Rules specific to the third species The following rules are specific to the third species. Some rules for this species are subject to interpretation. For example, some rules are stated differently depending on the language the treatise is translated in. When there is room for interpretation, our decision will be clearly explained.

Rule 9 If five notes follow each other by joint degrees in the same direction, then the harmonic interval of the third note must be consonant.

$$\begin{aligned}
&\forall k \in [1, n - 4] \\
beat &:= ((k + 2) + 3) \bmod 4 + 1 \\
measure &:= ((k + 2) - 1) \div 4 + 1 \\
isJoint(Cp[k], Cp[k + 1], Cp[k + 2], \\
&Cp[k + 3], Cp[k + 4]) \quad (23) \\
&\implies H[beat, measure] \in Cons
\end{aligned}$$

Beat and measure are just mathematical expressions to get the beat and measure corresponding to the note $k+2$ of the counterpoint. The function *isJoint* has been introduced to make the equation easier to read. This function returns true if the difference between adjacent arguments is smaller than or equal to 2 and if they are all in increasing or decreasing order. This rule differed among the three translations [14, 13, 12], so we used common sense and logical deduction to interpret it as best we could.

Rule 10 If the third note of a measure is dissonant then it must be a diminution and the second and the fourth notes must be consonant.

$$\begin{aligned}
&H[j, 3] \notin Cons \implies \\
&H[j, 2] \in Cons \wedge H[j, 4] \in Cons \quad (24)
\end{aligned}$$

Interval	Unison/Octave	Second		Third		Fourth
Type	Perfect	Minor	Major	Minor	Major	Perfect
Value	1	0	0	1	1	2
Interval	Triton	Fifth	Sixth		Seventh	
Type	♯4 / ♭5	Perfect	Minor	Major	Minor	Major
Value	forbidden	2	2	forbidden	forbidden	forbidden

Table 2. Cost values for melodic intervals

$$\begin{aligned}
&H[j, 3] \notin Cons \implies \\
&|Cp[4 * (j - 1) + 4] - Cp[4 * (j - 1) + 3]| \leq 2 \quad (25)
\end{aligned}$$

$$\begin{aligned}
&H[j, 3] \notin Cons \implies \\
&|Cp[4 * (j - 1) + 3] - Cp[4 * (j - 1) + 2]| \leq 2 \quad (26)
\end{aligned}$$

$$\begin{aligned}
&H[j, 3] \notin Cons \implies \\
&Cp[4 * (j - 1) + 4] > Cp[4 * (j - 1) + 3] \\
&> Cp[4 * (j - 1) + 2] \quad (27) \\
&\vee Cp[4 * (j - 1) + 4] < Cp[4 * (j - 1) + 3] \\
&< Cp[4 * (j - 1) + 2]
\end{aligned}$$

The first equation ensures that the first and third note are consonant, the second and third equations ensure that the intervals are seconds, and the fourth equation ensures that the intervals are in the same direction. The second and third equation together make sure that it is a diminution.

Rule 11 In the penultimate measure, if the cantus firmus is in the upper part, then the harmonic interval of the first note should be a minor third.

$$Cp[4 * (m - 2) + 1] > Cf[m - 1] \implies H[m - 1, 1] = 3 \quad (28)$$

3.4. Preference modelled with costs

In addition to the rules presented previously, Fux gives a series of preferences both explicitly and implicitly. We model preferences mathematically using a cost function. This section gives the mathematical expressions for the preferences and then gives a value for the cost function to evaluate each preference. The solver then minimizes the sum of the costs. Since Fux does not give values for the preferences and does not provide a hierarchy between them, the values presented here are an attempt at modelling Fux's preferences as closely as possible. However, Fux often decides to break the stated preferences to create more original counterpoints.

Preference 1 Imperfect consonances are preferred to perfect consonances for the first note of each measure.

$$\begin{aligned}
cost(H[j, 1] \in Cons_{imp}) &= 0 \\
cost(H[j, 1] \in Cons_p) &= 1 \quad (29)
\end{aligned}$$

Preference 2 Tritone intervals are forbidden.

$$\forall i \in [1, n - 1] \quad cost(M[i]) = 6 = \infty \quad (30)$$

Preference 3 Contrary motions are preferred to other types of motion.

$$\begin{aligned} \text{cost}(P[j] = 0) &= 0 \\ \text{cost}(P[j] = 1) &= 1 \\ \text{cost}(P[j] = 2) &= 2 \end{aligned} \quad (31)$$

As a reminder, $P[j] = 0$ means the motion of the voices getting to measure j is a contrary motion, $P[j] = 1$ means it is oblique motion and $P[j] = 2$ means it is parallel motion. It is important to note that Fux does not explicitly make the distinction between oblique and parallel motion, in fact this omission is one of the main criticisms of his work. We decided to make such a distinction because it is generally observed in other counterpoint treatises, and because these costs are adjustable so the distinction can be ignored if desired simply by giving oblique and parallel motion the same cost.

Preference 4 The second and third harmonies of a measure are preferred to not be consonant with a one degree melodic interval between them*.

$$\begin{aligned} \text{cost}(H[j, 2] \in \text{Cons} \wedge H[j, 3] \in \text{Cons} \\ \wedge |M[4 * (j - 1) + 2]| < 2) &= 2 * m \end{aligned} \quad (32)$$

Again, the cost is high to represent a “last resort”.

Preference 5 Each note and its two beats further peer are preferred to be different*.

$$\begin{aligned} \forall j \in [1, n - 2] \\ \text{cost}(Cp[j] = Cp[j + 2]) &= 1 \end{aligned} \quad (33)$$

Preference 6 Melodic intervals are preferred to be small. This preference is not clearly explained, so the way we interpreted it was to give a lower cost to smaller intervals (i.e. seconds and thirds) compared to the other intervals. Table 2 gives the preference we assigned for each interval based on the example that Fux provides.

4. IMPLEMENTATION AND EVALUATION

4.1. Implementation

The mathematical formalization detailed in section 3 as well as the other species for two voice counterpoint were implemented using the Gecode constraint solver [10] so that we could compare the results of our mathematical model with the examples given by Fux in his treatise. This implementation was done inside of OpenMusic [20] in the form of a library. The final version of the software will support a simple user interface to avoid the need to write Common Lisp code. This library can be found here : github.com/sprockeelsd/Melodizer. This code is available under the free license GPL-3.0. At the moment, counterpoint can only be generated based on $\frac{4}{4}$ time signature cantus firmi.

To use our implementation, one has to download Gecode[10], GiL [11] and OpenMusic [20]. All details regarding installation are on the github page. To use the mathematical functions representing Fux’s counterpoint, one has to create a Voice object in OpenMusic to represent the cantus firmus. Then, a fux-cp function must be called with this cantus firmus as a first attribute and the number representing the species as a second optional attribute. The default value will generate a counterpoint of the first species. Evaluating the fux-cp function will only create the problem, so in order to search for solution one has to use the search-next-fux-cp function after locking the evaluation (press b) of the fux-cp function.

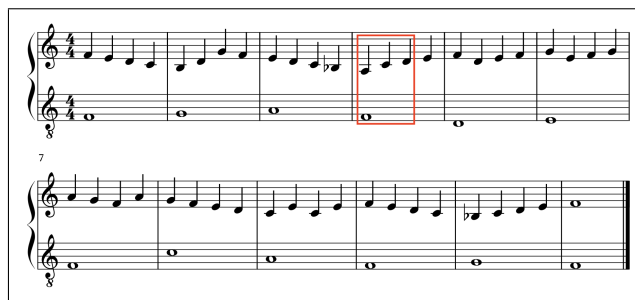


Figure 7. Cantus firmus and corresponding counterpoint as presented by Fux.

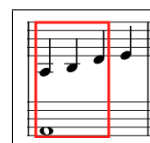


Figure 8. Measure 4 of the counterpoint from figure 7 respecting preference 4.

4.2. Comparison with Fux

In this section we compare the results produced by our tool to the counterpoints that Fux himself presents in his treatise for the same cantus firmi. We only present the counterpoints for one cantus firmus in this paper since they all lead to the same conclusions. The cantus firmus we chose is in the “mode of Fa” and is twelve measures long. Figure 7 shows this cantus firmus together with the counterpoint that Fux gives in his treatise. In order to make sure our solver indeed considers this counterpoint as a solution, a constraint saying that the solution must be equal to Fux’s counterpoint was added to the set of constraints detailed in section 3. For all cantus firmi we tested, the solver could indeed find Fux’s solution.

However, the solution given by Fux is usually far from the first solution suggested by our solver. Indeed, for the cantus firmus in figure 7, the solver finds that this counterpoint has a cost of 56 using the values described in section 3. Figure 9 presents the first solution found by the

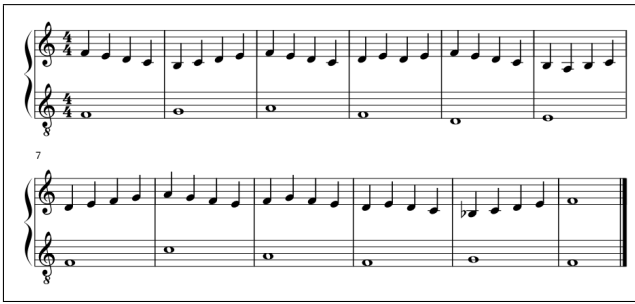


Figure 9. Cantus firmus from Fux’s treatise and corresponding counterpoint generated by our solver.

solver ¹. It is the one with the lowest cost (though other solutions might have an equivalent cost) and has a cost of 24. This can be explained by the fact that Fux establishes preferences but does not seem to follow them closely. Preference 4 is a good example. In measure 4, the harmonic interval goes from a third to a fifth to a sixth (highlighted in red in figure 7), when according to his preference the skip of a third should have been between the second and third note of the measure, hence the harmonic interval should have moved from a third to a fourth and then to a sixth. Figure 8 shows this measure when the intervals respect preference 4. In his treatise, Fux does not provide a hierarchy between the different preferences that he mentions, and neither does he provide cost values for them as his treatise does not aim to produce a mathematical formalization. The costs we decided to assign to different rules were therefore done in a best effort approach, resulting in a high cost for this specific scenario.

For the solution generated by our solver, the main factor increasing the cost is due to preference 5 being “broken” multiple times. As we do not provide a hierarchy between the different costs, since Fux himself doesn’t provide one, we do not have control on what rules are prioritized over others. A solution to this problem we implemented is that all costs are adjustable by the user. This is in our opinion the best approach, as Fux himself changes his preferences for each cantus firmus.

A final observation is that the last measure of the counterpoint the solver generates is often very similar to the one from Fux. That is because the penultimate measure has strict rules, meaning that fewer possibilities remain.

4.3. Examples of new compositions using the default costs inferred from Fux’s examples

Let us now use our solver to generate counterpoints to a brand new cantus firmus that we create ourselves using the costs inferred from the examples. In this section, we present the first counterpoint generated by the solver, hence the one with the smallest cost. Figure 10 starts with a cantus firmus in the mode of G to generate

1. This example can be listened to here : <https://www.youtube.com/watch?v=nDt060Os3VI>



Figure 10. Counterpoint generated based on a cantus firmus in the mode of G (not from Fux’s treatise).



Figure 11. Counterpoint generated based on a cantus firmus in the mode of C (not from Fux’s treatise).

a counterpoint ². This counterpoint has a cost of 20 using the values presented in section 3. We can see that a lower cost tends to limit the originality in the counterpoint, since it tries to respect the preference rules as closely as possible. Indeed, all melodic intervals in figure 10 are seconds resulting in an exclusively stepwise melodic progression, which technically respects Fux’s preferences but one could argue that the melody lacks character. Similarly, the notes tend to stay in the typical mode of G and no accidentals are used except for the penultimate measure where the seventh of the scale of G major is present due to the constraints on that specific measure.

Figure 11 shows another counterpoint generated based on a cantus firmus we wrote ourselves ³. The cost for this one is even lower, and is equal to 14. We can see very clearly how melodies respecting the preferences as much as possible tend to have similar traits and can be seen as unoriginal. Two solutions are offered to counter this problem : the first is that costs are adjustable by the user, meaning that one can influence how important each customizable aspect of the melody are. The second, which is still a work in progress, is to use a branch-and-bound solver to allow for adaptation for the search of solutions. A branch-and-bound solver allows to add an additional rule every time a solution is found. Thus, if the user is not satisfied by the solution the solver offers, in the future they will be able to select a degree of difference that the next solution will have with respect to the previous solution.

2. This example can be listened to here : <https://www.youtube.com/watch?v=zB69kJr6oA0>

3. This example can be listened to here : <https://www.youtube.com/watch?v=Jjya5IpCnTw>

Parameter	step-cost	third-cost	fifth-cost	min-skip-percentage
Value	0	0	1	0.3

Table 3. Modified parameters used to generate the counterpoint in figure 12.

Parameter	step-cost	third-cost	fourth-cost	fifth-cost	min-skip-percentage
Value	0	0	0	1	0.2

Table 4. Modified parameters used to generate the counterpoint in figure 13.

4.4. Experimenting with costs to generate more interesting compositions

In order for our solver to generate more interesting melodies, we implemented a couple of optional rules as well as some computational optimizations. First, we added a constraint to force a percentage of intervals to be skips, to encourage the solver to find counterpoints that step away from the original preferences. This already made the solutions suggested by the solver more interesting to us because it was no longer just conjoint intervals. We also added the rule that each disjoint interval must be followed by a conjoint interval in the opposite direction. This rule comes from Gallon and Bitsch [18] and not from Fux's treatise, but we thought it would make sense to add a rule for skips since we were allowing for more of them. This percentage can be adjusted by the user. Finally, we allowed the user to select a cost for each interval to guide the solver towards what they want. Adding those rules, especially the first one, increased the time for the solver to find a solution since it was going against its heuristic. The solver aims to minimize the costs as close as possible to Fux's preference, and using more skips increases the cost.

To remedy this problem, we decided to compute the minimal cost that the forced skips were inducing. This allows the solver to avoid searching for solutions with a cost that is lower than this minimal cost. As a result, the solver can find solutions relatively quickly for a reasonable number of skips (around 25% of skips). Figure 12 gives the solution for the cantus firmus taken from Fux that figure 9 shows ⁴. Table 3 gives the modified costs and parameters we used to produce this counterpoint. It is clearly different than the counterpoint generated without playing with the costs, using a lot more skips than the previous solution. Thirds are more present than other disjoint intervals because of the costs we assigned to each of them. Though it can be said that it does not follow Fux's preferences as closely, it is in our opinion more interesting and it still follows Fux's rules. Figure 13 shows another example of a counterpoint in the mode of A with modified costs ⁵. Table 4 shows the costs used for this example. The presence of fourth intervals gives the melody yet another flavour.

4. This example can be listened to here : <https://www.youtube.com/watch?v=SQkb1Mr4yNA>

5. This example can be listened to here : <https://www.youtube.com/watch?v=96NunBGW38I>

Figure 12. Counterpoint generated by the solver using the costs discussed in section 4.4 based on the cantus firmus in figures 7 and 9.

Figure 13. Counterpoint generated by the solver using the costs displayed in table 4.

5. FUTURE WORK

The ultimate goal of our project is to develop a practical tool for music composition based on constraint programming. We consider that a practical tool needs to satisfy three properties : (1) it allows a composer to express musical ideas without expertise in mathematics or programming, (2) it embodies specific knowledge about a desired musical style, and (3) it is computationally efficient for the desired compositions. The formalization and solver presented in this paper are a step towards the efficient embodiment of musical knowledge for one style. Our next steps will extend the formalization to other musical styles and will add an interface that requires only musical knowledge on the part of the composer. One concrete step that we are currently working on is to formalize the rules applicable to popular rock music in the period from the 1960s to the 1990s, based partly on the analysis done in the Ph.D. thesis of Drew Nobile [9].

6. CONCLUSION

This work presents a mathematical formalization of the Fux theory of two-voice counterpoint and an implementation of this theory as a constraint satisfaction problem. The implementation is done in the OpenMusic platform by means of an interface with the Gecode constraint solver. We have chosen the Fux theory because of its simplicity and relative importance. This work also provides a critical analysis of this implementation, by comparing its results to the counterpoints presented by Fux himself

in his treatise, and by looking at counterpoints generated from brand new cantus firmi. This work is part of a long-term project in computer-aided music composition with constraint programming. The ultimate goal is to make a practical tool that can be used by composers targeting one or more important musical styles.

7. REFERENCES

- [1] Andrea Agostinelli, Timo Denk, Zalán Borsos, Jesse Engel, Mauro Verzetti, Antoine Caillon, Qingqing Huang, Aren Jansen, Adam Roberts, Marco Tagliasacchi, Matt Sharifi, Neil Zeghidour, and Christian Frank. “MusicLM : Generating Music From Text”, *arXiv*, 2023.
- [2] Baptiste Lapière. “Computer-Aided Musical Composition”. Master’s thesis. UCLouvain, June 2020.
- [3] Chardon Clément, Diels Amaury, and Gobbi Federico. “Melodizer 2.0 : A Constraint Programming Tool For Computer-aided Musical Composition”. Master’s thesis. UCLouvain, June 2022.
- [4] Charles Kœchlin. “Précis des règles du contrepoint”. Heugel & Cie. Paris, 1926.
- [5] Charlotte Truchet. “Contraintes, Recherche Locale et Composition Assistée par Ordinateur”. PhD thesis. Université Pierre et Marie Curie, 2004.
- [6] Damien Sprockeels. “Melodizer : A Constraint Programming Tool For Computer-aided Musical Composition”. Master’s thesis. UCLouvain, Jan. 2022.
- [7] David Cope. “A Musical Learning Algorithm”. *Computer Music Journal* 28(3), 2004, pages 12-27.
- [8] Dorien Herremans and Kenneth Sørensen. “Composing fifth species counterpoint music with a variable neighbourhood search algorithm”. *Expert Systems with Applications* 40(16), 2013.
- [9] Drew Nobile. “A Structural Approach to the Analysis of Rock Music”. PhD thesis. Graduate Faculty in Music, City University of New York, 2014.
- [10] Gecode : Generic Constraint Development Environment. <https://www.gecode.org>. Visité le 19/02/2023.
- [11] GiL = Gecode interface Lisp. <https://github.com/sprockeelsd/GiL>
- [12] Johann Joseph Fux. “Gradus ad Parnassum”. Johann Joseph Fux. French. Ed. by Gabriel Foucou. Trans. Latin by Simone Chevalier, 2019.
- [13] Johann Joseph Fux. “Gradus ad Parnassum”. Latin. Ed. by Johann Peter van Ghelen. 1966 New York Broude Bros reprint. Vienna, 1725.
- [14] Johann Joseph Fux. “The study of counterpoint : from Johann Joseph Fux’s Gradus ad Parnassum”. English Trans. by Alfred Mann. Norton & Company. New York - London, 1965.
- [15] Knud Jeppesen. “Counterpoint : The polyphonic style of the sixteenth century”. Dover Publications, Inc. New York, 1992.
- [16] LogicPro - Apple, <https://www.apple.com/fr/logic-pro/>. Visité le 19/02/2023.
- [17] François Pachet and Pierre Roy. “Musical Harmonization with Constraints : A Survey”. *Constraints Journal*, Kluwer Publisher, 6(1) :7-19, 2001.
- [18] Noël Gallon and Marcel Bitsch. “Traité de contrepoint”. Durand & Cie, 1964.
- [19] Mikael Laurson. “PatchWork PWConstraints”. IR-CAM Documentation, Oct. 1996.
- [20] OpenMusic - IRCAM, <http://repmus.ircam.fr/openmusic/home>. Visité le 19/02/2023.
- [21] Örjan Sandred. “Constraint-Solving Systems in Music Creation”, Chapter 12, *Handbook of Artificial Intelligence for Music*, 2021, pages 327-344.
- [22] Russell Ovans and Rod Davison. “An Interactive Constraint-Based Expert Assistant for Music Composition”, Ninth Canadian Conference on Artificial Intelligence, 1992.
- [23] Camilo Rueda. Magnus Lindberg, Mikael Laurson, Georges Bloch, and Gérard Assayag. “Integrating Constraint Programming in Visual Musical Composition Languages”. *ECAI 98 Workshop on Constraints for Artistic Applications*, Brighton, 1998.
- [24] Bill Schottstaedt. “Automatic species counterpoint”. Report No. STAN-M-19, Center for Computer Research in Music and Acoustics, May 1984.
- [25] Sibelius - Avid, <https://www.avid.com/fr/sibelius>. Visité le 19/02/2023.
- [26] Théodore Dubois. “Traité de contrepoint et de fugue”. Heugel & Cie. Paris, 1901.
- [27] Thibault Wafflard. “Formalizing Fux’s Theory of Musical Counterpoint Using Constraint Programming”. Master’s thesis. UCLouvain, June 2023 (to appear).
- [28] Torsten Anders. “Composing Music by Composing Rules : Design and Usage of a Generic Music Constraint System”. PhD thesis. School of Music & Sonic Arts, Queen’s University Belfast, 2007.
- [29] Torsten Anders and Eduardo Miranda. “Constraint Programming Systems for Modeling Music Theories and Composition”. *ACM Computing Surveys* 43 (Oct. 2011), 30 :1-30 :38.

UNE APPROCHE ÉLECTROACOUSTIQUE DES PAVAGES RYTHMIQUES À PARTIR DES CANONS D'ENTRÉES RÉGULIÈRES

Matías Rosales

Doctorant à l'Université de
Strasbourg et à la HEAR
Mcfernandezr13@gmail.com

RÉSUMÉ

L'article présente une nouvelle solution algorithmique pour générer des pavages rythmiques à partir d'un processus canonique. Il s'agit d'une technique de construction de canons rythmiques mosaïques avec des entrées régulières, que j'ai déjà utilisée dans mes dernières compositions instrumentales. L'article suggère également une nouvelle utilisation de cette technique appliquée à la musique électroacoustique, inspirée des concepts développés par Horacio Vaggione tels que le micro-temps et la décorrélation spatiale et qui pourra être utilisée dans des futurs projets compositionnels autour des musiques électroniques et mixtes. Bien que les pavages rythmiques par des canons d'entrées régulières soient des modèles nés de la musique instrumentale, ils peuvent en effet également être utilisés pour composer de la musique électroacoustique et cela grâce à l'influence des conceptualisations élaborées par Horacio Vaggione.

1. INTRODUCTION

L'utilisation de modèles mathématiques en composition musicale permet d'aborder le processus créatif en amont de sa représentation graphique ou sonore. Ainsi, des concepts formels peuvent être établis à l'avance, ce qui permet au compositeur d'avoir une structure solide pour guider la création musicale tout en ouvrant de nouvelles perspectives théoriques.

Dans mon cas, les modèles m'aident à créer des textures polyphoniques, avec leurs propres contraintes, directions évolutives, et leurs dégradations. Dans cet état de l'art, des concepts éloignés de la réalité musicale peuvent être appliqués de différentes manières selon la réalité matérielle. Par exemple, on peut les appliquer à un orchestre symphonique, ou à un ensemble de chambre, à un instrument soliste, ou bien à un dispositif électronique. Dans chaque cas le compositeur

décidera comment mettre en œuvre les résultats dans un contexte musical, avec différents degrés de flexibilité et d'échelles temporelles. À partir d'un même modèle mathématique, nous pouvons générer des résultats musicaux extrêmement divers, en gardant toujours des caractéristiques communes basées sur leurs fortes relations logiques.

L'article se compose de trois parties : la première explique le pavage rythmique à partir des canons ayant des entrées régulières, à travers l'implémentation personnelle que j'ai développée avec Moreno Andreatta pour générer des canons mosaïques en OpenMusic ; la deuxième partie discute le modèle de pavage rythmique à partir des canons d'entrées régulières dans une perspective compositionnelle ; la troisième et dernière partie propose une application des canons mosaïques ayant des entrées régulières dans le domaine de la musique électroacoustique et en relation étroite avec des concepts développés par Horacio Vaggione.

2. CANONS MOSAÏQUES

L'idée d'utiliser des canons rythmiques basés sur des entrées régulières des voix remonte à Oliver Messiaen qui avait essayé de créer une mosaïque rythmique à partir de motifs rythmiques obtenus par concaténation de rythmes non-rétrogradables. Avec la répétition de toutes les voix en canon, l'axe du temps est rempli progressivement à partir d'une unité temporelle minimale décidée par le compositeur avec la règle que deux unités temporelles n'existent jamais au même instant, générant ainsi une polyphonie continue où la perception d'un contrepoint fragmenté devient une mélodie unique. Comme l'affirme Messiaen dans un passage repris dans l'ouvrage *Traité de Rythme, de Couleur et d'Ornithologie* : « Il résulte de tout cela que les différentes sonorités se mélangent ou s'opposent de manières très diverses, jamais au même moment ni au même endroit comme cela se passe dans un vrai carillon. C'est du désordre organisé » [1].

Bien qu'Olivier Messiaen n'ait pas été en mesure de formaliser et résoudre ce problème à l'aide d'outils mathématiques et informatiques, c'est la première fois dans l'histoire de la musique de tradition écrite que nous trouvons l'idée de générer des pavages de l'axe du temps à partir de répétitions d'un même rythme décalé dans le temps selon le principe du canon. Cette idée ouvre la porte à un nouvel usage des canons dans la création musicale qu'on appellera par la suite « canons rythmiques mosaïques » [2]. Historiquement, les canons avaient été construits à partir des relations harmoniques où le rythme était en arrière-plan. Cette nouvelle proposition ne concerne que le rythme et laisse un espace de liberté au compositeur qui peut ainsi adjoindre aux entrées des hauteurs définies, des bruits, des timbres, etc.

Prenant inspiration de la démarche de Messiaen, plusieurs types de canons rythmiques mosaïques ont été proposés, dont les canons de Vuza qui représentent l'une des constructions les plus utilisées en musique contemporaine [3]. Ces canons ont été utilisés par des compositeurs tels Fabien Lévy, Georges Bloch, Mauro Lanza, Daniele Ghisi.

Dans sa recherche autour des modes ou des gammes musicales à transpositions limitées, Olivier Messiaen a conçu une série de techniques combinatoires qu'il a en quelque sorte tenté de transposer au domaine rythmique.

Bien qu'Oliver Messiaen n'ait pas réussi à développer cette technique de façon précise, d'une certaine manière, il y a une homologie entre l'intuition de Messiaen dans le domaine des pavages rythmiques et la relation entre la synthèse granulaire et le *concret PH* de Xenakis [4]. Grâce à son approche mathématique novatrice et à l'utilisation de techniques du studio de la musique concrète, Xenakis a anticipé l'émergence d'un nouveau type de synthèse sonore ainsi qu'une logique morphologique du son. De la même manière, Messiaen a ouvert la voie à une nouvelle morphologie des textures temporelles tout en introduisant une nouvelle conception du contrepoint.

3. IMPLEMENTATION DANS OPENMUSIC.

L'algorithme à été développé à partir des objets natifs d'OpenMusic [5] et par des objets codés par l'équipe Représentations musicales [6] qu'on trouve dans la bibliothèque MathTools d'OpenMusic. Mon objectif était de parvenir à modéliser en OpenMusic la classe des canons rythmiques mosaïques à la base de l'intuition d'Olivier Messiaen. Le modèle de ce contrepoint rythmique me permettra de l'appliquer à des cas simples, comme à des canons longs avec un grand nombre de voix ; et par ailleurs de proposer un nouveau type de canon qui n'avait pas fait l'objet d'une étude computationnelle systématique en OpenMusic.

La conceptualisation de ce modèle a été faite à partir d'une petite mosaïque rythmique calculée à la main, ce qui a permis de construire l'algorithme pour pouvoir générer ensuite des relations plus difficiles à obtenir manuellement.

Pour générer une mosaïque ayant des entrées régulières, nous devons fournir le nombre des pulsations individuelles que chaque canon aura, ainsi que le nombre de voix. La mise en œuvre doit principalement tenir compte du nombre de pulsations par voix et du nombre de voix.

Le nombre des voix va générer le motif rythmique de la manière suivante : le nombre de pulsations sera converti en une liste entre zéro et la valeur indiquée moins un ; cette liste sera permutée grâce à l'objet `permut-random`, qui nous donnera toutes les possibilités de motifs rythmiques que nous pouvons utiliser pour faire une mosaïque rythmique par canon d'entrées régulières quel que soit le nombre des voix, à partir de deux.

À l'aide d'une abstraction itérative, la liste permutée sera traitée de manière à redistribuer ses éléments dans un cercle en fonction d'un module. Ce module correspondra au nombre d'éléments contenus dans la liste.

Prenons l'exemple d'une liste permutée (2 1 0 4 3) générée à partir d'une liste de cinq éléments débutant à zéro. Pour redistribuer cette liste dans l'espace circulaire, nous devons la considérer comme une séquence ascendante de points. Ainsi, cette liste sera réorganisée en (2 6 10 14 18) en fonction du module correspondant au nombre d'éléments de la liste initiale, dans ce cas module cinq. Ceci permettra de réorganiser la liste de manière circulaire selon un certain module.

L'algorithme crée des pavages rythmiques avec des entrées régulières, même lorsque les figures rythmiques sont irrégulières, comme dans le cas d'un pentagone. Ainsi, l'algorithme transforme des entrées rythmiques irrégulières en une texture rythmique régulière en subdivisant les temps en sous-divisions égales. Cela permet de créer des textures rythmiques complexes tout en maintenant une régularité dans les subdivisions temporelles.

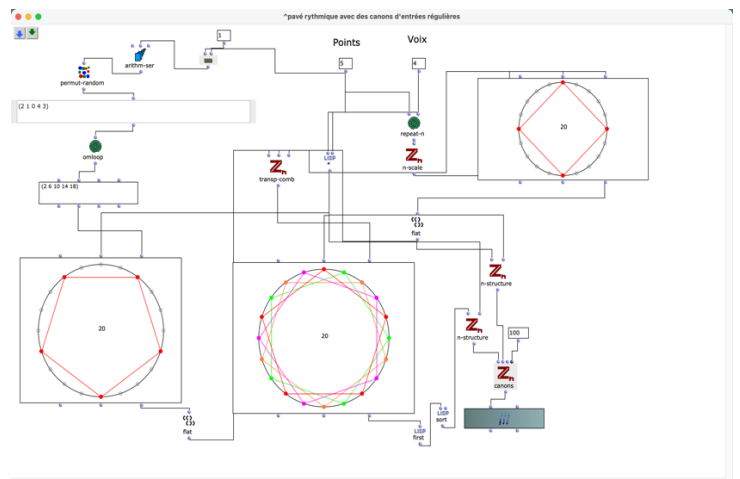


Figure 1. Algorithme pour représenter les canons mosaïques avec des entrées régulières.

Par la suite l'objet `n-structures` construira la structure intervallique d'un sous-ensemble donné du groupe cyclique à partir de la liste et de la multiplication entre la liste et le numéro de voix. Cette multiplication nous donnera la durée totale de notre pavage à partir de la première entrée de la première pulsation de la première voix, et ce jusqu'à la fin de la dernière pulsation du canon.

Finalement l'objet `canons` traduira les informations données à partir de notre motif numérique, en durées de temps. Le nombre de voix donnera les informations nécessaires pour indiquer l'entrée temporelle de chaque voix. L'objet `canons` a également un autre facteur externe à l'ensemble du processus décrit, qui sera très utile pour notre recherche vers une musique micro-temporelle.

L'entrée optionnelle de l'objet `canons` permet de choisir l'unité minimale de notre mosaïque rythmique, permettant de travailler des infimes morceaux de mosaïque à la taille de la durée souhaitée où la milliseconde sera l'unité minimale.

Prenons maintenant un autre exemple de construction d'un canon rythmique mosaïque, cette fois à trois voix, chaque voix ayant quatre pulsations. L'algorithme va construire un canon mosaïque à condition que toutes les voix rentrent de façon régulière sans jamais se chevaucher. Le cœur du programme a une composante aléatoire qui, à chaque évaluation, nous permettra de visualiser une nouvelle solution possible pour construire des canons avec quatre pulsations qui représenteront les possibilités limitées par les valeurs d'entrée. Si nous avons quatre pulsations et trois voix, on aura une mosaïque composée de douze unités et quand la troisième voix du contrepoint entre, ou même parfois avant, on arrive à remplir tous les espaces rythmiques comme Messiaen le souhaitait.

Les objets de la bibliothèque `MathTools` permettent également de visualiser géométriquement le motif rythmique à l'aide d'une représentation circulaire ainsi que la mosaïque complète et l'entrée de chaque voix (figure 1, de gauche à droite).

Dans la figure 2 la mosaïque a été répétée 3 fois, pour mieux visualiser le pavage rythmique, tandis que les hauteurs ne sont pas importantes dans cet exemple. La ligne pointillée sur la grille temporelle indique à partir de quel instant la mosaïque est réalisée.



Figure 2. Canon rythmique mosaïque à partir de trois voix qui rentrent de façon régulière.

La figure 3 montre le résultat du même algorithme mais appliqué à des mosaïques beaucoup plus complexes et longues. Dans ce cas on obtient un canon rythmique mosaïque à 12 voix, chaque voix ayant 28 points d'attaque. Un traitement micro-temporel nous permet d'utiliser des petites pièces de mosaïque, qui à cause de leur courte durée, auront une quantité très dense de pulsations par secondes qui génèreront un continuum dans l'espace temporel, une sorte de micro mosaïque, à l'échelle de cinq-millisecondes.

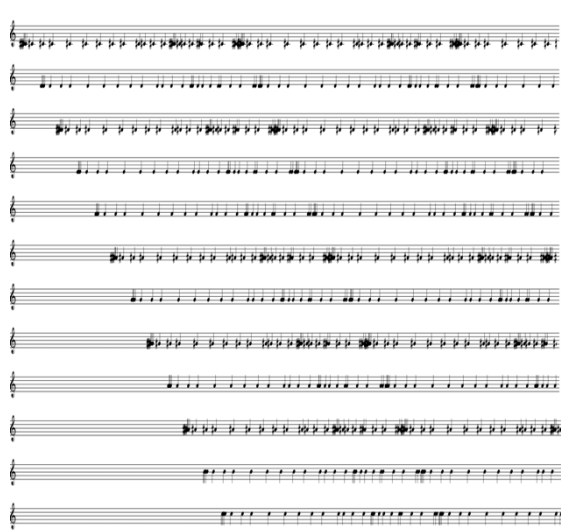


Figure 3. Canon rythmique mosaïque à partir de trois voix qui rentrent de façon régulière.

4. CANONS AVEC DES ENTREES REGULIERES ET COMPOSITION MUSICALE.

Le modèle mathématique est un outil pour aider à la création, sur lequel les décisions arbitraires font en effet partie du processus de création. Avant que le résultat numérique ne soit traduit en représentation musicale, on doit pouvoir se permettre de sculpter la matière, de l'infime détail au changement complet de la structure d'origine. La relation avec laquelle Horacio Vaggione aborde l'usage de l'informatique est pertinente en ce sens. « Vaggione pose en fait la question de la formalisation : l'interaction entre écriture directe et calcul algorithmique est synonyme d'interaction entre formalisation et artisanat. Il n'y a donc pas opposition (ou dialectique) entre deux concepts antagonistes, mais complémentarité, qui se fructifie par leur interaction. » [7].

La créativité joue un rôle au moment de l'élaboration de l'algorithme, elle aussi nous permet d'élaborer d'autres algorithmes de manière novatrice afin de concevoir des contradictions et des ruptures

avec le modèle original. Par exemple notre mosaïque en tant que structure complexe stable et rigide pourrait être dégradée à cause du temps qui passe, car d'autres facteurs et paramètres provoquent sa propre destruction.

La figure 4 est le résultat d'un algorithme construit pour détériorer le modèle original, basé sur un simple modèle d'étirement audio. Imaginons de faire un enregistrement audio de cette mosaïque et de changer progressivement la vitesse de reproduction du fichier avec un étirement dynamique de la bande magnétique. Cette opération modifiera le rythme et les hauteurs progressivement ; l'harmonie initiale est basée sur les moyennes partielles d'un spectre inharmonique, générée aussi avec OpenMusic. En conséquence de l'étirement dynamique de notre bande magnétique virtuelle, l'harmonie-timbre sera progressivement modifiée en ce qui concerne les hauteurs.

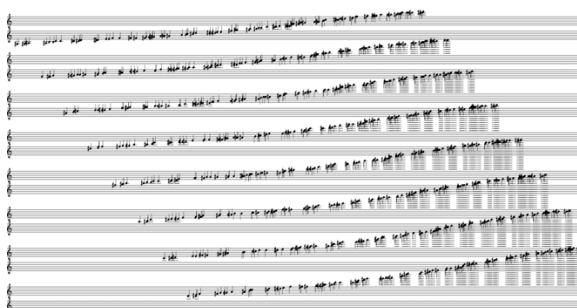


Figure 4. Distorsion du pavage.

Puisque le traitement rythmique dans ma musique a toujours été basé sur l'instabilité rythmique, liée au swing, aux contre-temps et à l'illusion d'une quasi-régularité, j'ai dû réfléchir pour incorporer un élément rythmiquement isométrique à mes propres compositions. Ainsi, le contraste engendré par les mosaïques rythmiques avec les textures de contretemps me permet de créer des situations contrastées de manière efficace, allant d'un moment rythmiquement souple à une mosaïque solide : un contrepoint sans temps forts et où la répétition de subdivisions régulières est évitée, en opposition à une solide entrée d'un canon à trois voix qui sera orchestré petit à petit par l'ensemble entier. Plus tard, les voix se rejoindront pour conclure la pièce avec une homorythmie puissante, représentée par un pavage constitué d'une seule voix [8]. Ou encore, une autre solution possible pour générer des tensions est de superposer un pavage rythmique à un tissu rythmique et harmonique différent. [9].

5. VERS UNE PROPOSITION ELECTRONIQUE DE PAVAGES RYTHMIQUES PAR DES CANONS D'ENTREES REGULIERES.

Bien qu'il existe des compositeurs qui ont utilisé des mosaïques rythmiques pour composer de la musique électronique, le modèle de référence a été

principalement celui basé sur les canons de Vuza. Contrairement aux canons décrits dans cet article, ceux-ci génèrent des mosaïques à partir des entrées irrégulières. Plusieurs utilisations des canons de Vuza sont décrites, par exemple, par Fabien Levy [10]. Par ailleurs, nous ne trouvons pas de canons d'entrées régulières interprétés d'une façon électronique dans le répertoire musical.

Sans avoir l'intention de chercher une interprétation dogmatique de ses écrits, j'ai essayé d'utiliser des concepts développés par Horacio Vaggione pour créer un système numérique entre différents logiciels me permettant de générer de la musique électronique dans le cadre de mes nouvelles pièces de musique mixte.

Une première approche multi-temporelle et micro-temporelle aux pavages rythmiques permettra d'obtenir des mosaïques avec différents niveaux de temporalités ou de profondeurs [11]. On pourrait s'imaginer de complexes superpositions de mosaïques aux temporalités différentes ; nous pouvons superposer de grandes mosaïques avec des unités minimales très longues et des petites mosaïques faites par des unités grandes et toujours sans chevauchements. Les résultats auront une dimensionnalité plus complexe que la proposition initiale du modèle, qui se déroule dans une seule dimension temporelle.

Cela serait semblable aux pavages hyperboliques d'Escher [12], où la continuité de la mosaïque est construite par des morceaux de tailles variables.



Figure 5. Anges et démons de M.C. Escher

Les canons mosaïques ont été conçus à partir d'une problématique de la musique instrumentale et n'ont été théorisés d'un point de vue mathématique que successivement. Pour cette raison, la plupart des exemples historiques font usage d'une temporalité qui reste liée à la musique instrumentale. Avec les musiciens, il est possible de créer une variété infinie de mosaïques, mais avec la limitation de la taille de notre pièce de mosaïque lié à la temporalité humaine.

Un exemple d'approximation plus conforme à la pensée de Vaggione serait d'explorer un univers de mosaïques composé de pièces ayant des temporalités microscopiques impossibles à exécuter par des interprètes. Nous pourrions ainsi créer un canon polyphonique où chaque microparticule, placée à une distance infime de la suivante, engendrerait une continuité dans l'espace temporel et produirait un pavage exhaustif de la totalité de la temporalité, à l'image d'une mosaïque construite à l'échelle d'un pixel d'image de haute résolution. Dans le procédé musical qu'on vient de décrire nous ne différencions plus l'unité minimale, celle-ci étant plutôt perçue comme une forme continue.

J'ai construit deux patches pour synthétiser-écouter-crée ces textures impossibles à réaliser avec des instruments musicaux. Pour me rapprocher des petites unités, j'ai utilisé deux objets classiques de Max-MSP [13] : `Click~` et `Groove~`.

D'une part, l'objet `Click~` génère des impulsions électriques à partir d'un déclenchement qui représente déjà une unité de temps minimale ; d'autre part, l'objet `Groove~` est un objet de lecture d'échantillons en boucle qui utilise les informations audios stockées dans un autre objet appelé `Buffer~`. `Groove~` sera paramétré pour reproduire les durées minimales possibles.

Grâce à l'environnement `bach` [14] de Max/MSP, je peux exporter les différentes représentations musicales générées avec `OpenMusic` à l'aide du modèle des pavages par des canons avec des entrées régulières vers Max/MSP. La représentation musicale est visible depuis une interface conviviale, très rapide, et facile à manipuler ; le `bach.roll` enverra en temps réel les informations nécessaires pour synthétiser avec les objets exposés nos pavages rythmiques.

5.1 Patch avec `Click~`

Le patch fonctionnera comme un synthétiseur polyphonique pour reproduire les `bach.roll`. Chaque entrée de note déclenchera une impulsion électrique avec une riche variété de propriétés, qui seront contrôlées par des mouvements stochastiques et aléatoires en temps réel générés à l'aide de l'objet brownien codé par Karlheinz Essl [15]. Cela permettra une évolution continue des paramètres de la génération sonore, créant ainsi une variété sonore et une texture dynamique dans la reproduction des `bach.roll`.

En partant des paramètres initiaux de l'objet `Click~`, tels que la table d'ondes contrôlée par une fonction aléatoire activée à chaque détection d'une nouvelle voix dans notre objet polyphonique, chaque voix aurait un timbre unique dès son apparition, permettant ainsi de les différencier les unes des autres dès leur genèse.

En outre, l'harmonie que chaque modèle propose sera interprétée grâce à un filtre résonateur qui soustraira les fréquences désirées de l'impulsion électrique-sonore, comme dans une synthèse soustractive, également avec la possibilité d'avoir différentes tailles de filtres résonnants qui seront aussi contrôlés par un mouvement aléatoire. Cela donnera une riche variété de possibilités entre le seuil du bruit et de la hauteur clairement perceptible.

Le volume sonore de chaque particule sera contrôlé par son même déclenchement qui actionnera un mouvement brownien redimensionnant le signal. Ce système affecte directement chaque voix de manière indépendante.

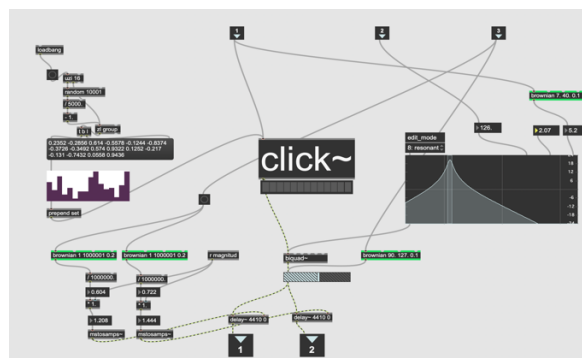


Figure 6. Une voix du synthétiseur basée sur des impulsions électriques. Chaque voix sera multipliée dans un système polyphonique pour pouvoir synthétiser l'information donnée par `OpenMusic`.

5.2 Patch avec `Groove~`

En utilisant aussi les informations directes du `bach.roll`, la fréquence de chaque hauteur affectera la vitesse de lecture de chaque fichier audio, dans une relation proportionnelle où l'octave, supérieure représente une vitesse de reproduction double. La vitesse augmentera proportionnellement pour chaque hauteur écrite dans notre partition virtuelle, toujours en relation directe avec la fondamentale de notre harmonie conçue avec `OpenMusic`. En même temps, un nouveau générateur de mouvements aléatoires changera la direction de la lecture : normale ou vers l'arrière. Le volume de chaque particule sera contrôlé de la même façon que dans l'exemple avec `Click~`.

Avec une commande externe on peut transposer en même temps toutes les fréquences liées à une fréquence fondamentale choisie aléatoirement.

Ce patch a moins de variables pour modifier le timbre, puisque chaque fichier est un timbre complexe déjà capté par un enregistreur portable. J'ai enregistré en stéréo différents sons complexes et continus qui ont attiré mon attention, comme le bruit généré par un réfrigérateur ou le bruit du système électrique dans certaines parties de la maison, ou encore différents sons liés à l'écoulement d'eau dans une maison. L'algorithme associe aléatoirement à

chaque voix un fichier audio précis dans toute la collection des sons collectés. Puis, j'ai converti tous les signaux stéréo en mono pour après leur donner une nouvelle qualité spatiale.

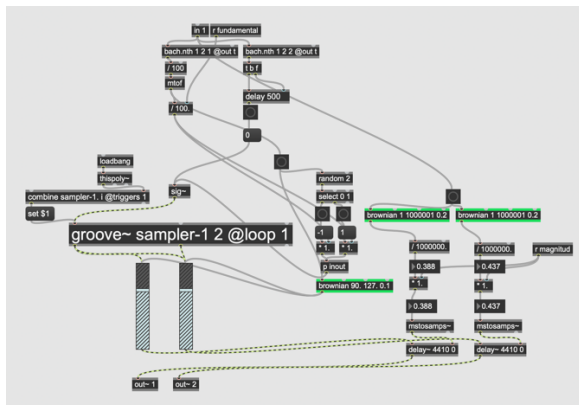


Figure 7. Une voix du synthétiseur basée sur l'objet Groove~.

6. DECORRELATION MICRO-TEMPORELLE DES PAVAGES RYTHMIQUES AVEC DES CANONS D'ENTREES REGULIERES

La décorrélation micro-temporelle est une technique permettant de créer une sensation d'espace et de localisation sonore en multipliant un son avec de légères variations de phase. Cette technique nous permettra d'interagir avec la perception humaine, étant donné que le seuil de perception de la micro-relation en tant qu'espace sonore plutôt que séquences des sons estimés entre 5 μ sec et 1,5 millisecondes [16]. Cette méthode permet donc de créer un espace sonore plus complexe et réaliste, en utilisant des variations microscopiques de phase pour donner l'impression d'une localisation spatiale plus précise.

En tant que compositeur, Horacio Vaggione accorde une importance particulière à l'utilisation musicale de la décorrélation micro-temporelle. Il explore l'idée que des déphasages plus importants entre chaque haut-parleur (soit dans une décorrélation acoustique ou numérique) peuvent générer des sensations auditives différentes et plus riches du point de vue perceptif.

Dans notre cas je vais doubler chaque source sonore pour pouvoir écouter la décorrélation micro-temporelle en stéréo. Chaque canal aura des retard micro-temporels générés par l'objet `delay~` et contrôlés par deux mouvements browniens indépendants. Le temps de retard dans l'objet `delay~` peut être spécifié en échantillons et déterminé par la fréquence d'échantillonnage de notre réglage en Max-MSP. Pour pouvoir travailler avec des micros-déphasages en relation aux unités temporelles de fraction de seconde. Pour cela j'ai utilisé l'objet `mstosamps~` afin de convertir un nombre en millisecondes et contrôler ainsi un certain nombre d'échantillons en relation à la fréquence

d'échantillonnage actuelle de notre réglage audio en Max-Msp.

Comme dans les cas précédents, chaque note de notre interface `bach.roll` déclenchera une action dans le flux audio qui suivra. Ensuite, nous utiliserons un mouvement stochastique pour moduler le taux de micro-décalage, ce qui génèrera une décorrélation progressive et permettra de ressentir le mouvement continu de chaque particule sonore de manière indépendante.

Dans une texture musicale complexe telle que celle que nous avons créée ci-dessus, constituée d'un tissu de mosaïques rythmiques, la décorrélation micro-temporelle est une technique efficace. Étant donné que nous travaillons avec de nombreuses sources sonores qui ne sonnent pas simultanément, grâce aux contrastes internes des mosaïques rythmiques et aux entrées régulières, la décorrélation permet une meilleure distinction entre les différentes particules sonores et améliore la clarté de la texture globale.

L'utilisation de la technique de décorrélation micro-temporelle permet une meilleure distinction entre les différentes particules sonores. Cette technique est particulièrement intéressante dans les situations où plusieurs voix sont activées avec des valeurs de décorrélation différentes. En l'utilisant dans notre texture musicale complexe constituée d'un tissu de mosaïques rythmiques, nous sommes en mesure d'améliorer la clarté de la texture globale et de ressentir le mouvement continu de chaque particule sonore de manière indépendante. Comme Horacio Vaggione l'affirme clairement : « [...] les techniques de décorrélation micro-temporelle sont intéressantes pour la composition de musique électroacoustique quand elles sont utilisées dans une situation où il y a beaucoup de sources parallèles, où plusieurs canaux sont activés simultanément avec des valeurs de décorrélation différentes, engendrant un « tissu » morphologique complexe. » [16].

Enfin, la relation spatiale entre chaque haut-parleur ou casque de notre stéréophonie, grâce à la décorrélation micro-temporelle et au volume dynamique de chaque voix, génère un espace virtuel très dynamique, dans notre micro-mosaïque polyphonique

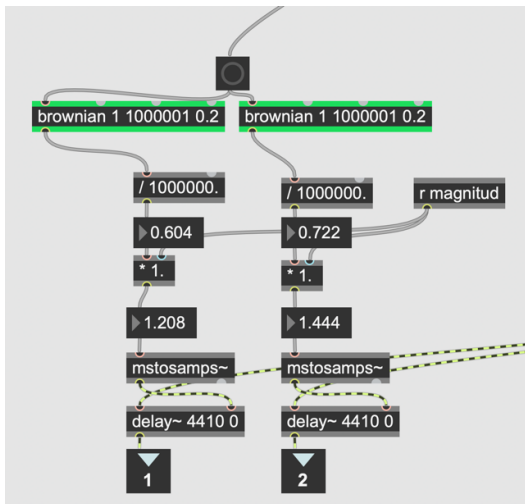


Figure 8. Décorrélation micro-temporelle stochastique

7. CONCLUSION

Un algorithme construit sur la base d'un modèle mathématique, en l'occurrence les mosaïques rythmiques par canons d'entrées régulières, peut générer des données structurales que nous pouvons appliquer à différents niveaux temporels. Un modèle pensé et créé à partir de la musique instrumentale peut aussi nous offrir un tissu morphologique électroacoustique complexe, en dépassant le seuil autour de la note, du macro-temps au micro-temps [17]. Cette expérience montre comment la modification de l'échelle temporelle d'un modèle musical peut mener à des résultats inattendus et intéressants. Les modèles musicaux peuvent être conçus pour résoudre des problèmes spécifiques à une échelle temporelle donnée, mais en les modifiant, on peut parfois découvrir de nouvelles façons de les utiliser et de nouvelles dimensions à explorer. Cela souligne l'importance de la flexibilité et de l'expérimentation dans la création musicale et dans l'élaboration des algorithmes, ainsi que la nécessité de rester ouvert aux résultats imprévus et aux nouvelles idées qui peuvent surgir de ces expériences.

L'approche algorithmique-compositionnelle de Horacio Vaggione m'a permis de repenser l'espace sonore en tant que partie intégrante de la morphologie sonore elle-même, plutôt que comme un traitement externe, comme on le pense actuellement, la plupart du temps. Cette réorientation a ouvert de nouvelles voies de recherche pour moi, en explorant la spatialisation, la diffusion de l'électronique et l'émission sonore par les instruments acoustiques et modulaires, tout en cherchant à créer des expériences auditives uniques et personnelles.

La figure 9 représente le processus final que j'effectue et dans lequel la première mosaïque sera construite à partir d'un canon à 16 voix, chaque voix ayant 60 attaques, avec une subdivision de l'unité

minimale de 5 millisecondes. Les fréquences données seront construites à partir d'un spectre harmonique légèrement dilaté, ayant comme fondamentale la fréquence de 65,40639 Hz. Cette mosaïque sera synthétisée par le patch basé sur Click~. La mosaïque a été répétée 15 fois

Parallèlement, une autre mosaïque a été construite avec un canon à 16 voix, chaque voix ayant 56 attaques, avec une subdivision de l'unité minimale de 7 millisecondes. Les fréquences données seront construites à partir d'un spectre harmonique légèrement compressé, ayant la fréquence de 92,498604 Hz comme fondamentale. Cette mosaïque sera synthétisée par le patch basé sur Groove~. La mosaïque a été répétée 9 fois.

Les nombres 5 et 7 ont été choisis en tant que nombres premiers pour représenter les tailles des pièces de mosaïque dans le but de minimiser les moments où deux sons sont présents simultanément. Ainsi, en répétant la mosaïque construite à partir de pièces de 5 millisecondes 7 fois, il y aura une synchronisation des motifs rythmiques. De même, la voix construite à partir de pièces de 7 millisecondes sera répétée 5 fois jusqu'à la synchronisation, qui correspond au plus petit multiple commun de 5 et 7.

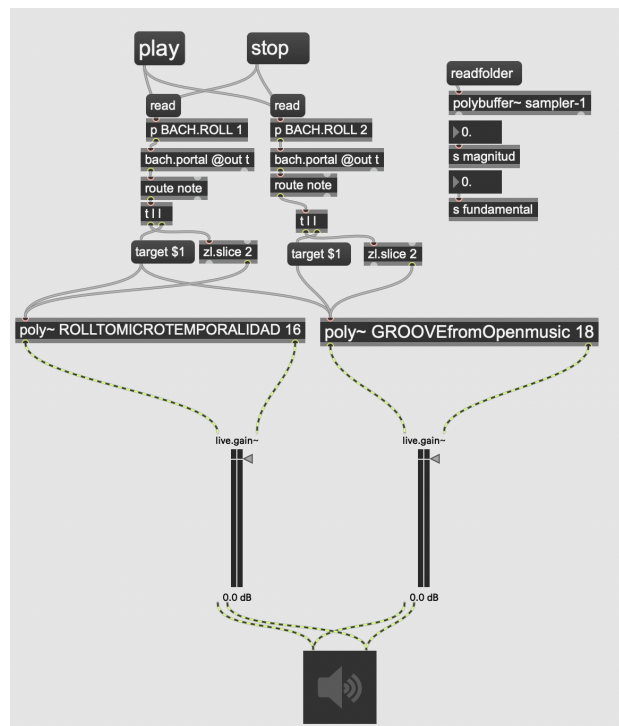


Figure 9. Réalisation électroacoustique finale de deux pavages rythmiques micro-temporels à partir des canons d'entrées régulières.

Dans cet exemple, je n'ai pas utilisé de dilatation ou compression temporelle sur les pavages rythmiques issus de canons, afin de pouvoir percevoir un résultat plus clair avec l'algorithme.

Le matériau sonore de cette mise en œuvre électronique des mosaïques rythmiques par des canons

d'entrées régulières représente le résultat souhaité et sera un outil précieux pour la création de mes prochaines œuvres de musique mixte. Il me permettra également de superposer des canons avec des micro-temporalités différentes, dans lesquelles aucun événement ne sonnera simultanément avec un autre. Cette technique offre une riche variété de possibilités pour l'exploration des textures sonores et des structures rythmiques.

L'utilisation de petites pièces de mosaïque crée une texture sonore continue, car la mosaïque a été conçue pour occuper l'espace temporel de la manière la plus fine possible. Les sources sonores initiales sont toutes des bruits provenant d'impulsions électriques ou d'enregistrements sonores, donnant ainsi une qualité bruitiste à la texture. Grâce à leurs divisions temporelles distinctes, chaque pavage rythmique peut être aisément identifié et différencié de l'autre.

Bien que la texture soit bruitée et chaotique, nous pouvons apprécier les mouvements spatiaux et fréquentiels répétitifs. La force des répétitions des voix du canon permet d'instaurer un certain ordre dans cette texture chaotique, même si les variables spatiales et dynamiques sont constamment en mouvement.

8. LIENS

Voici les liens permettant d'écouter un exemple sonore et avoir accès aux patches OpenMusic et Max/MSP.

Exemple sonore :

https://soundcloud.com/matias-fernandez-rosales/deux-pavages-rythmiques-microtemporelles-a-partir-des-canons-dentrees-regulieres/s-QPorzvAVPlj?si=760ff6ed2a7d491f88228d578d177a6a&utm_source=clipboard&utm_medium=text&utm_campaign=social_sharing

Patch OpenMusic :

<https://drive.google.com/file/d/1hSEdZfQ4VgUICsxhWba1jj3wiOlbcNiD/view?usp=sharing>

Patch Max/MSP :

<https://drive.google.com/file/d/1KwY8emDQAoxq981BthOb8-7xu1KBOLwU/view?usp=sharing>

9. RÉFÉRENCES

- Messiaen, O. *Traité de Rythme, de Couleur et d'Ornithologie*, Paris, 1995. Édition Musicales, p. 46.
- Andreatta, M. « De la conjecture de Minkowski aux canons rythmiques mosaïques », *L'Ouvert*, mars, p. 51-61, 2007.
- Andreatta, M., Agon C. et Vuza D. « Analyse et implémentation de certaines techniques compositionnelles chez Anatol Vieru », *Actes des Journées d'Informatique Musicale*, Marseille, 2002.
- Xenakis, I, *Concret PH (ouvre musicale pour bande seule)* Salabert, 1958.
- OpenMusic <http://repmus.ircam.fr/openmusic/home>
- L'équipe Représentations musicales <https://www.ircam.fr/recherche/equipes-recherche/repmus>.
<https://openmusicproject.github.io/openmusic/download.html>.
- Solomos, M. Horacio Vaggione. 2021. fihal-03153991f. p. 5.
- Rosales, M. *Seeds* (ouvre musicale pour ensemble instrumental) Halle, 2022.
- Rosales, M. *Le Sable du Sahara* (ouvre musicale pour orchestre à vents) Strasbourg, 2021.
- Lévy, F. "Three Uses of Vuza Canons", *Perspectives of New Music*, Volume 49, N° 2 (Summer 2011), p.4.
- Solomos, M. *Espaces Composables, essais sur la musique et la pensée musicale d'Horacio Vaggione*. L'Harmattan, Paris, 2007.
- Escher, M. C. (1960). Angels and Devils - Circle Limit IV [Œuvre d'art visuel].
- Max/MSP <https://cycling74.com>
- Bach <https://www.bachproject.net>
- Brownian www.essl.at
- Vaggione, « H. *Décorrélation microtemporelle, morphologies et figurations spatiales.* » *Actes des Journées d'Informatique Musicale*, Marseille, 2002, p. 9.
- Solomos, M. *Espaces Composables, essais sur la musique et la pensée musicale d'Horacio Vaggione*. L'Harmattan, Paris, 2007.

LIVESCALER : CONTRÔLER EN LIVE L'HARMONIE D'UN MORCEAU DE MUSIQUE ÉLECTRONIQUE

Alice Rixte

Université de Bordeaux
alice.rixte@u-bordeaux.fr

RÉSUMÉ

Dans le contexte de l'Electronic Dance Music (EDM), de nombreux artistes utilisent des techniques de *DJing* pour performer leurs propres productions sur scène, se privant ainsi de l'accès à la structure interne de leurs morceaux, et en particulier de l'équivalent de leur partition : les fichiers MIDI joués par des instruments virtuels. De plus, si l'artiste remixe ou interprète sa propre production, le nombre de pistes pouvant être contrôlées simultanément est limité sans un outil adapté.

Cet article présente le logiciel LiveScaler, qui permet de contrôler en live l'harmonie et la hauteur des notes de tous les instruments virtuels d'un morceau de musique électronique. Un ensemble restreint de transformations de l'espace des hauteurs de notes, les *transformations affines*, est introduit. Ces transformations sont appliquées à tous les flux MIDI d'un morceau composé préalablement. Une implémentation utilisant Max MSP en combinaison avec Ableton Live (Max for Live) est proposée. Une attention particulière est portée aux questions de contrôle, de *mapping* et de mise en pratique dans le cadre de l'EDM.

1. INTRODUCTION

Aux origines de la *club culture*, le rôle des DJs était surtout de jouer des enregistrements musicaux pour un public, en général dansant. Avec l'évolution rapide de la technologie et en particulier l'omniprésence de l'audio-numérique à partir des années 2000, de nombreux DJs sont devenus à la fois compositeurs, producteurs et performeurs de leurs propres morceaux. C'est notamment le cas des artistes évoluant dans le milieu de l'Electronic Dance Music (EDM) ¹.

Ces artistes composent le plus souvent leur musique à l'aide d'une station audionumérique (Digital Audio Workstation ou DAW) qui leur permet de combiner sampleurs, synthétiseurs et enregistrements pour créer un morceau de musique complet. Recréer en live un tel morceau, composé souvent de plusieurs dizaines de pistes, reste difficile et conditionné par les contrôles proposés par les DAW. Pour leur performance live, la plupart des artistes vont ainsi choisir entre interpréter certaines pistes spécifiques à l'aide de synthétiseurs ou sampleurs, remixer en

1. *Electronic Dance Music* (EDM) est un terme parapluie regroupant de nombreux genres de musique électronique tels que la house, la techno, la trance, la drum n bass, le dubstep, etc.

live un morceau préparé au préalable ou utiliser des techniques de *DJing* leur permettant de mixer entre eux des rendus audios en leur appliquant divers effets [14],[22]. Dans ces contextes, modifier le rythme ou l'harmonie d'un morceau joué en live est difficilement réalisable dans ces contextes. En effet, cela requerrait un accès à la structure interne du morceau, ce que le *DJing* ne permet pas, ou bien de contrôler toutes les pistes simultanément, ce qui est précisément ce que nous proposons de faire ici.

Cet article présente LiveScaler, qui permet de modifier en live la structure harmonique d'un morceau de musique électronique ². En partant d'un morceau composé au préalable, LiveScaler permet d'appliquer des transformations MIDI à l'ensemble des instruments virtuels. LiveScaler conserve l'ensemble des caractéristiques du morceau initial mais agit sur la hauteur des notes, permettant de changer en live l'harmonie du morceau ou de générer de nouvelles mélodies. En particulier, l'accent est mis sur la possibilité d'utiliser LiveScaler dans un DAW, ici Ableton Live ³, tout en minimisant les contraintes et les connaissances techniques nécessaires pour son utilisation. Dans Ableton Live, LiveScaler peut s'utiliser aussi bien dans le mode *session* que dans le mode *arrangement*, et s'ajoute à toutes les pistes MIDI que l'on souhaite être impactées par les transformations.

Cet article est organisé de la manière suivante : nous commencerons par définir les transformations affines, un ensemble restreint de transformations de l'espace des hauteurs de notes. Ensuite, nous présenterons LiveScaler, qui implémente l'application de ces transformations en live à un nombre arbitraire d'instruments virtuels. Enfin, nous décrirons la manière dont nous avons utilisé LiveScaler dans le cadre d'une performance live d'EDM. Nous terminerons par une comparaison de LiveScaler avec les travaux et outils existants.

2. TRANSFORMATIONS DE L'ESPACE DES HAUTEURS

Dans cette section nous présentons les transformations de l'espace des hauteurs que nous utilisons dans LiveScaler. L'idée principale est d'associer à chaque note une nou-

2. Une vidéo de démonstration est disponible à l'adresse suivante : youtu.be/Cn0HBgWS5Pw

3. Une implémentation de LiveScaler avec Max For Live est disponible en libre accès sur Github : github.com/autonym8/LiveScaler

velle note, qui sera jouée en lieu et place de la note initiale par l'ensemble des instruments du morceau de musique joué. LiveScaler permet d'appliquer deux types de transformations. D'une part les *transformations périodiques sur un intervalle*, que nous présenterons après avoir défini l'espace des hauteurs de note sur lequel agissent nos transformations. D'autre part les *transformations affines*, dont nous décrirons les propriétés musicales. Enfin, nous terminerons en restreignant l'écart de hauteur entre la note initial et la note transformée afin de résoudre de potentiels problèmes de tessiture.

2.1. Espace des hauteurs

Dans cet article, nous nous plaçons dans l'espace des hauteurs linéaire. Celui-ci est obtenu à partir de la demi-droite réelle positive des fréquences, en fixant une origine arbitraire de fréquence α (par exemple 440 Hz). Ici, nous appellerons cette fréquence l'*ancree*. À partir d'une fréquence $f > 0$, on détermine une hauteur p par

$$p = \log_2(f/\alpha)$$

On obtient alors la droite réelle : $p \in \mathbb{R}$. En discrétisant cette droite, on peut obtenir une note n à partir de la hauteur p par

$$n = \lfloor \beta p \rfloor$$

avec $\beta \in \mathbb{N}^*$ le nombre de divisions de l'octave. On obtient ainsi le tempérament à division multiple β -TET⁴. En particulier, lorsque $\beta = 12$ et $\alpha = 8.1758$ Hz⁵, on retrouve le codage des hauteurs MIDI.

Ainsi, pour définir un tempérament multiple, nous avons besoins de deux paramètres : l'*ancree* α et le nombre de divisions de l'octave β , que nous appelons *base*. On définit alors

$$\begin{aligned} T\langle \alpha, \beta \rangle : \mathbb{R}_+^* &\rightarrow \mathbb{Z} \\ : f &\mapsto \lfloor \beta \log_2(f/\alpha) \rfloor \end{aligned}$$

Dans cet article, nous nous intéresserons à l'*espace des hauteurs linéaires* discrétisé $\mathbb{Z} = T\langle \alpha, \beta \rangle(\mathbb{R}_+^*)$. Par abus de langage, on appellera ici la fonction $T\langle \alpha, \beta \rangle$ le *tempérament égal* d'*ancree* α et de base β . En pratique, α correspondra toujours à une note MIDI, et on se permettra d'écrire $\alpha = C_5$ pour indiquer que la fréquence correspondant à la note C_5 est associée à 0 dans \mathbb{Z} .

Nous définissons alors une *transformation de gamme* comme une fonction qui à toute hauteur de note associe une nouvelle hauteur de note a priori quelconque, autrement dit une fonction de \mathbb{Z} dans \mathbb{Z} . Cette transformation est *a priori* indépendante de la manière dont \mathbb{Z} est relié à l'espace des fréquences.

2.2. Transformations périodiques sur un intervalle

Nous définissons ici un ensemble de transformations des hauteurs facilement implémentable : les transforma-

4. TET signifie Tone Equal Temperament en anglais

5. 8.1758 à 0 est la fréquence (ici arrondie) correspondant à la note C_{-1} et la note MIDI 0.

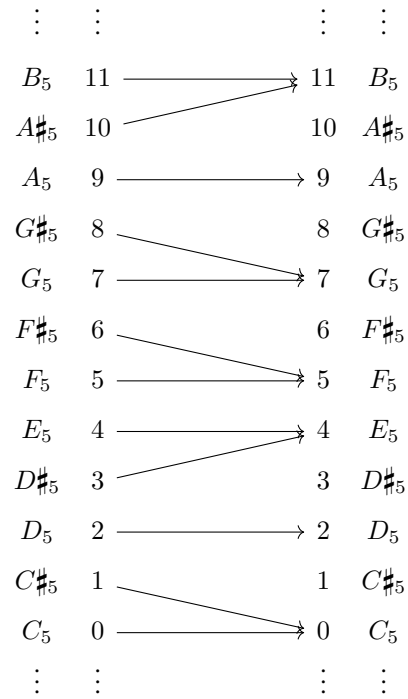


Figure 1: La quantisation vers la gamme majeure est périodique sur l'octave : le même motif est répété sur chaque octave.

tions périodiques sur un intervalle positif $n \in \mathbb{N}$. Ces transformations s'obtiennent en donnant l'image de toutes les notes se situant dans l'intervalle i , typiquement une octave ($i = 12$), puis en répétant ce motif verticalement en additionnant (ou soustrayant) l'intervalle considéré. Plus précisément, pour une transformation $X : \mathbb{Z} \rightarrow \mathbb{Z}$, X est périodique sur l'intervalle i lorsque la fonction $Y : n \mapsto X(n) - n$ est périodique de période i . Autrement dit,

$$Y(iq + n) = Y(n)$$

avec $0 \leq n < i$ et tout $q \in \mathbb{Z}$.

On peut par exemple définir une transformation qui quantiser les 12 demi-tons chromatique vers une gamme de notre choix, par exemple la gamme majeure dans la Figure 1.

2.3. Transformations affines

Nous présentons ici les *transformations affines*, c'est-à-dire les fonctions de la forme $A\langle \mu, \tau \rangle : n \mapsto \mu n + \tau$ avec μ le *mode* de la transformation affine et τ sa *transposition*. Nous allons classifier ces transformations en fonction de μ , qui détermine, en fonction du mode départ, le mode dans lequel on se trouvera après avoir appliqué la transformation.

Les transformations affines ont la propriété importante de préserver les classes de hauteur⁶ (au sens de Forte

6. En effet, pour toute base $\beta \in \mathbb{N}^*$, $\forall n_1, n_2 \in \mathbb{Z}, n_1 \equiv n_2 \pmod{\beta} \implies \mu n_1 + \tau \equiv \mu n_2 + \tau \pmod{\beta}$. Avec $\beta = 12$, on obtient le résultat pour les classes de hauteurs dodécaphoniques.

[15]) c'est-à-dire que si deux notes sont identiques à l'octave près, alors elles le seront toujours une fois la transformation affine appliquée.

La suite de cette section étudie plusieurs exemples afin de donner au lectorat un aperçu de leur expressivité. Dans l'ensemble de ces exemples, on se placera dans le tempérament égal $T\langle C_5, 12 \rangle$: la note C_5 correspondra ainsi à l'entier 0, $C\sharp_5$ à 1, C_6 à 12, B_4 à -1 , C_4 à -12 , etc. Nous nous concentrerons ici sur les cas $\mu = 1$, $\mu = -1$ et $\mu = 2$.

2.3.1. Transpositions

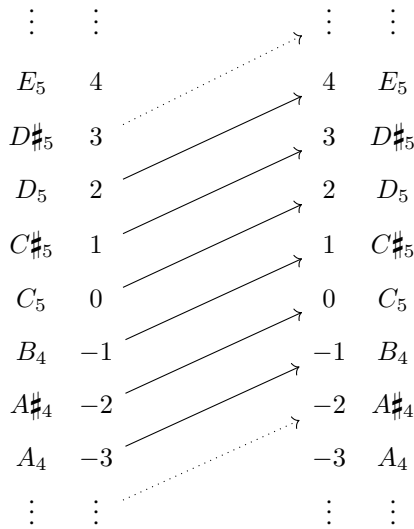


Figure 2: La transformation $A\langle 1, 2 \rangle : n \mapsto n + 2$ correspond à la transposition d'un ton vers l'aigu

Lorsque $\mu = 1$, les transformations affines $A\langle 1, \tau \rangle : n \mapsto n + \tau$ permettent de représenter toutes les transpositions possibles (voir Figure 2).

2.3.2. Inversions

Lorsque $\mu = -1$, les transformations $A\langle -1, \tau \rangle : n \mapsto -n + \tau$ permettent de passer d'une gamme majeure à un mode de Mi et réciproquement (voir Figure 4). La Figure 3 illustre la manière dont la triade C_5, E_5, G_5 est transformée par $A\langle -1, 4 \rangle$ en la triade E_5, C_5, A_4 . De même, l'image de la triade A_4, C_5, E_5 est G_5, E_5, C_5 . Comme les transformations affines préservent les classes de hauteur, on peut affirmer plus généralement que $A\langle -1, 4 \rangle$ transforme l'accord C en Am et Am en C.

Remarquons de plus qu'en changeant l'ancre, en choisissant $\alpha = G_4$ par exemple, l'image de l'accord G par $A\langle -1, 4 \rangle$ sera Em. La Table 1 explicite les images des accords de la gamme de Do majeur par $A\langle -1, 4 \rangle$ lorsque $\alpha = C_n$ et des accords de la gamme de Sol majeur lorsque l'on change l'ancre pour $\alpha = G_n$, avec n un entier quelconque.

Ainsi, en se basant sur l'image de l'accord majeur dont la tonique correspond à l'ancre, nous pouvons associer les

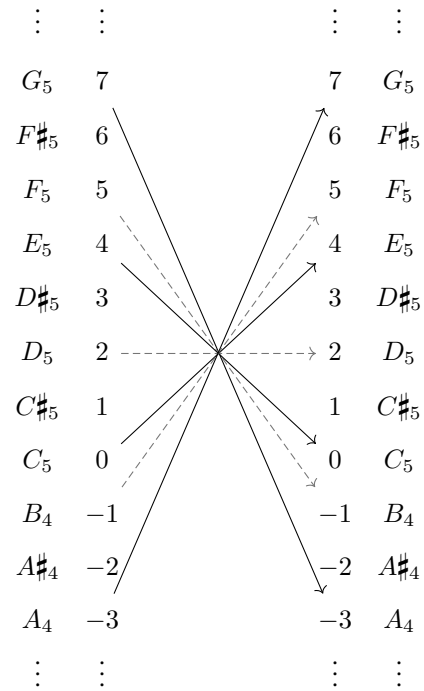


Figure 3: La transformation $A\langle -1, 4 \rangle : n \mapsto -n + 4$ avec $\alpha = C_5$ et $\beta = 12$

Pour des raisons de lisibilité seules les flèches de la gamme de Do majeur ont été tracées. On notera le passage de l'accord C à Am et de Am à C.

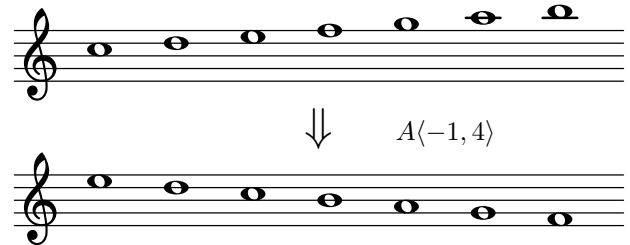


Figure 4: L'image de la gamme de Do majeur par $A\langle -1, 4 \rangle$ est un mode de Mi

C	\mapsto	Am	G	\mapsto	Em
Dm	\mapsto	G	Am	\mapsto	D
Em	\mapsto	F	Bm	\mapsto	C
F	\mapsto	Em	C	\mapsto	Bm
G	\mapsto	Dm	D	\mapsto	Am
Am	\mapsto	C	Em	\mapsto	G
B $^\circ$	\mapsto	B $^\circ$	F $^\circ$	\mapsto	F $^\circ$

Table 1: Images des accords de la gamme de Do majeur par $A\langle -1, 4 \rangle$ pour $\alpha = C_n$ (à gauche) et de la gamme de Sol majeur pour $\alpha = G_n$ (à droite), pour $n \in \mathbb{Z}$

transformations affines pour laquelle $\mu = 1$ ou $\mu = -1$ au degré de l'image de cet accord dans la gamme. Par exemple, $A\langle -1, 4 \rangle$ peut être associée au degré vi car elle associe le sixième degré mineur à l'accord majeur dont la tonique est l'ancre (Am pour C, Em pour G, ...). On obtient ainsi une notation plus intuitive que la notation mathématique, dont la table 2 donne un aperçu.

Degré	Transformation affine
I	$A\langle 1, 0 \rangle$
ii	$A\langle -1, -3 \rangle$
iii	$A\langle -1, -1 \rangle$
IV	$A\langle 1, 5 \rangle$
V	$A\langle 1, 7 \rangle$
vi	$A\langle -1, 4 \rangle$
vii	$A\langle -1, 6 \rangle$

Table 2: Correspondances entre triades d'une gamme majeure et transformations de gamme

Remarquons que, de manière générale, les inversions affines se comportent moins bien sur les modes non naturels. Par exemple, $A\langle -1, 4 \rangle(G\sharp) = G\sharp$ donc l'image de la gamme de La mineur harmonique par $A\langle -1, 4 \rangle$ contient les notes C, D, E, F, G, $G\sharp$, B, ce qui ne correspond pas à un mode standard de la musique tonale. C'est une des faiblesses des transformations affines.

2.3.3. Transformation vers un mode à transposition limitée

Lorsque $\mu = 2$ ou $\mu = -2$, les transformations affines envoient n'importe quelle gamme vers une gamme apparentée à une gamme par tons (voir Figure 5). Les transformations affines peuvent donc permettre de sortir du cadre de la musique tonale.

Contrairement aux inversions et aux transpositions, cette transformation n'est pas bijective : l'image de $A\langle 2, 0 \rangle$ contient exactement 6 classes de hauteurs qui correspondent aux 6 notes d'une des deux gammes par tons. Il est intéressant de noter que l'image d'une gamme majeure ou mineure naturelle par $A\langle 2, \tau \rangle$ contient les 6 notes de la gamme par tons (voir Table 3). Ce n'est pas le cas pour la gamme mineure harmonique dont l'image par $A\langle 2, \tau \rangle$ ne contient que 5 classes de hauteur.

C	↦	C	C	↦	C
D	↦	E	D	↦	E
E	↦	$G\sharp$	$E\flat$	↦	$F\sharp$
F	↦	$A\sharp$	F	↦	$A\sharp$
G	↦	D	G	↦	D
A	↦	$F\sharp$	$A\flat$	↦	E
B	↦	$A\sharp$	$B\flat$	↦	$G\sharp$

Table 3: Image des gammes de Do majeur (à gauche) et Do mineur naturel (à droite) par $A\langle 2, 0 \rangle$

De manière plus générale, lorsque le coefficient modal μ n'est pas premier avec β , on obtient un mode à transpo-

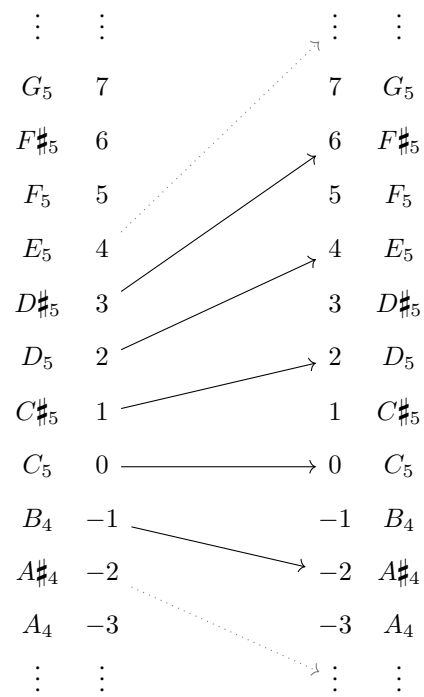


Figure 5: La transformation $A\langle 2, 0 \rangle : n \mapsto 2n$ permet d'obtenir des gammes apparentées à une gamme par tons.

sition limitée dont les notes sont séparées par un intervalle de μ demi-tons. Ainsi, pour $\mu = 4$, on obtient un mode composé de 3 classes d'hauteurs, séparées par des tierces majeures (4 demi-tons).

La Table 4 résume les différents types de transformations qu'offrent les transformations affines, ainsi que le nombre de classes de hauteur dans l'image de ces transformations⁷. Les transformations affines bijectives - leur image contient 12 classes de hauteurs - correspondent aux automorphismes $F\langle u, j \rangle$ du groupe T/I décrits par [19].

μ	Type de transformation	Classes de hauteur
-1	Inversions majeur/mineur	12
0	Constante	1
1	Transpositions	12
-2,2	Gamme par tons	6
-3,3	Tierces mineures	4
-4,4	Tierces majeures	3
-5,5	$F\langle 7, \tau \rangle, F\langle 5, \tau \rangle$	12
-6,6	Tritons	2
12	Octaves	1

Table 4: Classification des transformations affines en fonction de leur coefficient modal μ

⁷. Le nombre de classes de hauteur dans l'image de $A\langle \mu, \tau \rangle$ est égal à $\frac{\beta}{\mu \wedge \beta}$ où \wedge dénote le pgcd de deux entiers.

2.4. Restriction de l'écart de hauteur

Jusqu'ici, nous avons présenté les transformations affines en nous concentrant sur leur action sur les classes de hauteur. Dans la pratique, si nous appliquons directement $A\langle -1, 4 \rangle$ à la note A_4 qui correspond au La 440Hz et à la note MIDI 69, on obtient $A\langle -1, 4 \rangle(69) = -65 = G_{-6}$, qui est bien trop grave pour être audible.

Afin de ne pas trop s'éloigner de la tessiture de l'instrument, ou même du spectre auditif, nous allons restreindre l'écart entre la note initiale et son image. Soit $X : \mathbb{Z} \rightarrow \mathbb{Z}$ une transformation de gamme quelconque. Définissons à partir de X une nouvelle transformation $X\langle \delta^-, \delta^+ \rangle$ de sorte que

$$-\delta^- \leq X\langle \delta^-, \delta^+ \rangle(n) \leq \delta^+$$

où δ^+ (resp. δ^-) est l'intervalle montant (resp. descendant) maximum entre la note initiale et son image.

Soit $\delta = \delta^+ - \delta^-$. Dans la plupart des cas on souhaite que $\delta = \beta = 12$, mais il peut être intéressant de choisir par exemple $\delta = 2\beta$, pour préserver le caractère montant ou descendant d'une ligne mélodique.

Soit $r = |X(n) - n| \bmod \beta$ le reste de la division euclidienne de $|X(n) - n|$ par β . On pose alors

$$X\langle \delta^+, \delta^- \rangle : n \mapsto \begin{cases} n + r & \text{si } r \leq \delta^+ \\ n + r - \delta & \text{sinon} \end{cases}$$

Nous pouvons maintenant appliquer cette restriction de l'intervalle de hauteur à nos fonctions affines. On obtient alors un sous-ensemble de transformations de gamme de la forme $A\langle \mu, \tau, \delta^-, \delta^+ \rangle$. Ce sont exactement ces transformations, combinées avec les tempéraments $X\langle \alpha, 12 \rangle$, qui sont implémentées dans LiveScaler.

3. IMPLÉMENTATION DES TRANSFORMATIONS AFFINES : LIVESCALER

Nous allons à présent nous intéresser à l'implémentation des transformations affines que nous venons de présenter afin de pouvoir les appliquer en live à l'ensemble des flux MIDI qui composent le morceau joué.

3.1. Architecture de LiveScaler

LiveScaler fonctionne à la manière d'un orchestre dont le DJ serait le chef. Chaque piste MIDI contenant un instrument virtuel (synthétiseur, sampleur, etc.) est un instrumentiste de l'orchestre. On souhaite que sur un geste du DJ, chaque instrument virtuel interprète différemment sa partition, c'est à dire le flux MIDI qu'il reçoit. Dans le cadre de LiveScaler, le DJ envoie les paramètres d'une transformation affine à tous les instruments simultanément et ceux-ci doivent appliquer cette transformation dès qu'ils la reçoivent.

L'implémentation de LiveScaler est donc séparée en deux outils interdépendants :

1. une interface (appelée *Conductor*, en référence à l'analogie avec l'orchestre) qui récupère les entrées

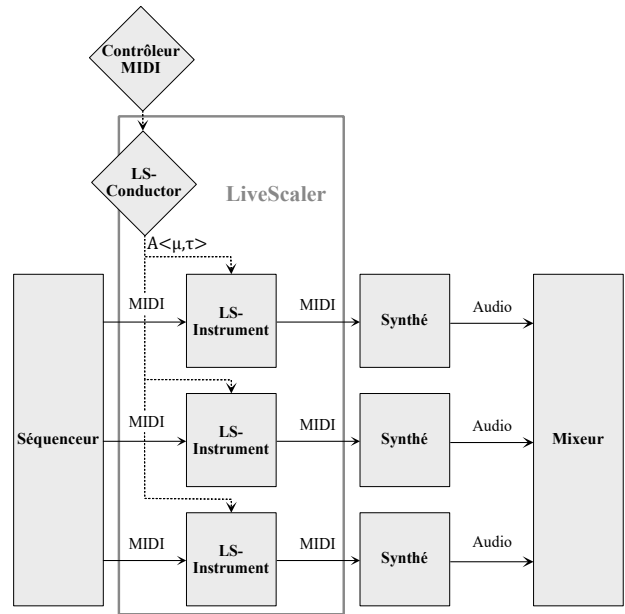


Figure 6: Architecture de Live Scaler

de l'utilisateur (ici le DJ) et les convertit en paramètres d'une transformation affine puis envoie ces paramètres à toutes les instances de *Instrument*.

2. un plug-in MIDI appelé *Instrument* qui transforme le flux MIDI entrant en appliquant à toutes les notes la transformation affine dont les paramètres ont été reçus de *Conductor*.

La Figure 6 illustre l'architecture globale de LiveScaler.

On distingue également les paramètres *locaux*, qui sont propres à chaque instrument, et les paramètres *globaux*, qui sont reçus du chef d'orchestre et donc communs à tous les instruments. Dans le cadre des transformations affines, les paramètres μ , τ , α et β sont globaux, ils correspondent dans une certaine mesure à l'harmonie actuelle du morceau. Quant à δ^- et δ^+ , ils sont locaux et peuvent être adaptés à la tessiture de l'instrument.

3.2. Quand appliquer les transformations ?

Lorsque *Instrument* reçoit la commande d'appliquer une nouvelle transformation de gamme, celle-ci est sensée prendre effet immédiatement et être appliquée à toutes les notes reçues jusqu'au prochain changement de gamme. Lorsque l'instrument n'est pas en train de jouer, cela ne pose aucune difficulté : il appliquera la nouvelle transformation au prochain message MIDI qu'il recevra. Il se peut en revanche que l'instrument soit déjà en train de jouer une note. LiveScaler propose 4 manières de réagir dans une telle situation :

1. *Stop* : toutes les notes en train d'être jouées sont instantanément arrêtées en envoyant un message Note-off pour chaque note en cours. L'instrument reprendra son jeu, en appliquant la nouvelle transformation, au prochain message MIDI qu'il recevra. Cette

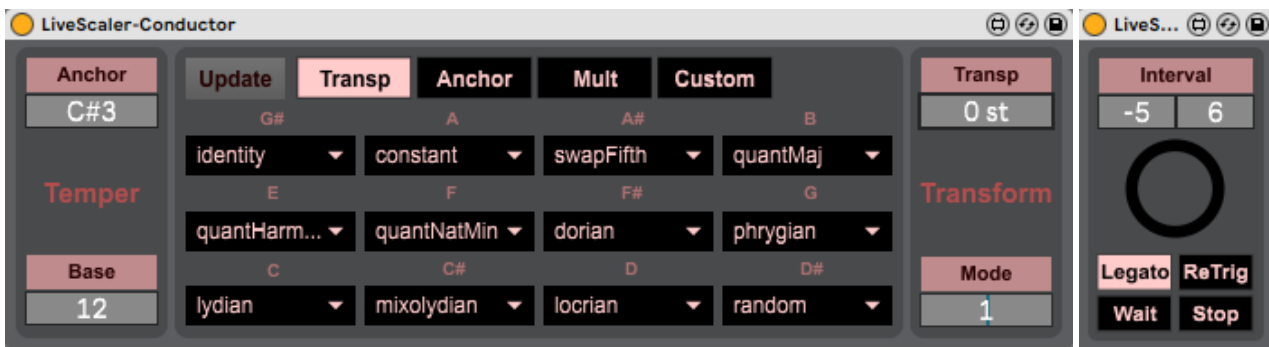


Figure 7: Interface graphique de LiveScaler (*Conductor* à gauche et *Instrument* à droite)

option est particulièrement adaptée aux instruments dont la durée des notes est courte.

2. *Legato* : chaque note en cours est stoppée et instantanément remplacée par son image par la nouvelle transformation. Si l'instrument virtuel est paramétré sur *Legato*, alors les changements de gamme déclencheront des legatos.
3. *ReTrigger* : agit sur le même principe que *Legato* à la différence qu'un court délai est introduit entre la fin de la note en cours et la note transformée, forçant une nouvelle attaque, même si l'instrument virtuel est en mode legato.
4. *Wait* : les notes en cours continuent d'être jouées telles quelles. Si elles ne se sont pas arrêtées avant, elles seront stoppées lorsque la prochaine note sera jouée, à partir de laquelle la nouvelle transformation prendra effet.

Le choix entre ces quatre modes se fait de manière locale, deux instances de *Instrument* pourront donc réagir différemment.

3.3. Implémentation avec Max for Live

L'objectif principal de l'implémentation proposée était de pouvoir expérimenter le plus rapidement possible sur les transformations en tant que musicienne. Max for Live est une intégration du langage de programmation graphique Max MSP à Ableton Live. On peut aisément communiquer avec les différentes instances du logiciel, ce qui permet dans LiveScaler à *Conductor* de contrôler les *Instruments* avec une faible latence⁸. La station audio numérique Ableton Live étant particulièrement populaire pour composer et produire de la musique dans le milieu de l'EDM, c'est donc naturellement que nous avons choisi Max for Live pour une première implémentation.

La Figure 7 montre l'interface graphique de LiveScaler. L'implémentation actuelle permet d'appliquer les transformations affines et les transformations périodiques sur un intervalle. Pour ces dernières, il suffit de renseigner manuellement l'image de chaque note de l'intervalle considéré dans un fichier pour y avoir ensuite accès via LiveScaler.

8. En moyenne, LiveScaler introduit une latence inférieure à 1 ms.

4. ESCAPE : UNE PERFORMANCE AVEC LIVESCALER

Afin de tester la pertinence des transformations proposées, j'ai utilisé LiveScaler pour créer *Escape*, une performance live de trance psychédélique. J'ai notamment pu proposer cette performance en publique pour la soutenance de mon mémoire de master en Septembre 2019. Dans cette section, j'expliquerai la manière dont j'ai procédé pour cette mise en pratique et je donnerai mes impressions subjectives en tant que musicienne et utilisatrice de LiveScaler.

4.1. Contrôle live dans Escape



Figure 8: Contrôleurs MIDI utilisés pour la performance live : à gauche le Push 2 par Ableton (contrôle de la structure du morceau) et à droite ATOM par Presonus (contrôle de l'harmonie du morceau).

Pour *Escape*, j'utilise deux contrôleurs MIDI distincts (voir Figure 8) :

- le Push 2 par Ableton, qui est conçu spécialement pour contrôler Ableton Live. Il me permet de recréer en live la structure d'*Escape* en déclenchant ses différentes parties.
- l'ATOM de Presonus, composé d'une grille de 4 × 4 touches qui me permettent de déclencher les transformations de LiveScaler.

Mod	II	vii	3
++	V	iii	2
--	IV	ii	M↔m
Hist	I	vi	4

Figure 9: Mapping de LiveScaler sur un contrôleur MIDI à 4×4 touches

La figure 9 illustre la manière (le plus souvent appelée *mapping*) dont les touches du contrôleur sont associées aux transformations de LiveScaler.

Voici le détail de l'action des différentes touches :

- I, vi, IV, II, V, iii, II, vii : les deux colonnes centrales déclenchent instantanément les transformations décrites précédemment. Elles sont organisées par relatives mineures/majeures.
- Hist : LiveScaler garde en mémoire un court historique des transformations précédemment appliquées. Appuyer sur Hist permet de déclencher une des transformations de cet historique. Des combinaisons de la touche Hist et des touches Hist, M↔m, 2, 3, et 4 permettent de naviguer dans cet historique⁹.
- ++, -- : applique $\tau = \tau + 1$ (resp. $\tau = \tau - 1$). En pratique, combiner la touche ++ (resp. --) avec une des transformations des colonnes centrales, on transpose cette transformation d'un demi-ton vers le haut (resp. vers le bas).
- M↔m : applique $\tau = \tau + 7\mu$ et $\mu = -\mu$. En pratique, combiner M↔m avec une des transformations centrales permet de passer d'une transposition à une inversion et réciproquement : combiner M↔m avec I (resp. ii, iii, IV, V, vi, vii) donnera la transformation i (resp. II, III, iv, v, VI, VII) et réciproquement.
- 2, 3, 4 : applique $\mu = 2\mu$ (resp. $\mu = 3\mu$, $\mu = 4\mu$). On obtient ainsi les modes à transposition limitée décrits précédemment.
- Mod : en combinant Mod avec une des transformations, on indique à LiveScaler qu'on souhaite modu-

9. Pour plus de détail, se référer au manuel de LiveScaler

ler l'harmonie de notre morceau vers cette nouvelle gamme.

Ainsi, on pourrait imaginer harmoniser en live une instrumentation jouant de manière répétée sur l'accord C. Par exemple, si on veut reproduire la suite d'accord de la chanson *Summer Nights* de la comédie musicale *Grease*¹⁰ en commençant par indiquer à LiveScaler qu'on est dans une tonalité de Ré majeur (Mod + II). Puis, une fois l'instrumentation lancée, on appuiera successivement tous les deux temps sur I- IV- V- IV.

Puis, arrivés au moment tant attendu de la modulation d'un demi-ton vers le haut, on indique à LiveScaler

I - IV - V - (++ + M↔m + vi) - (Mod + ++ + I)

pour repartir joyeusement sur I- IV- V- IV, mais cette fois dans une tonalité de Mi bémol majeur. On obtiendrait ainsi la progression harmonique suivante :

... - D - G - A - G - D - G - A - B♭ - E♭ - A♭ - B♭ - A♭ - ...

4.2. Processus de composition

J'ai composé *Escape* dans le but de le jouer avec LiveScaler. C'est un morceau de trance psychédélique (psytrance)¹¹ composé :

- d'une mélodie minimaliste durant 2 mesures jouée par un synthétiseur dont le son se rapproche d'un métallophone éthéré
- d'une *rolling bass* classique devenue une des signatures de la *psytrance*, et de plusieurs autres basses sur une unique note pédale produisant une ligne de basse riche dans sa texture et son timbre
- d'une rythmique séquencée à l'avance indépendante de LiveScaler (pistes audio)
- de quelques effets sonores (*risers*, *downshifters* ...) eux aussi typiques de la *psytrance*, que je déclenche ponctuellement à l'aide du Push.

L'objectif était d'une part de partir d'un morceau simple, basé sur un unique accord (ici Am) et d'enrichir son harmonie et sa ligne mélodique avec LiveScaler; d'autre part de proposer un morceau typique d'un genre de musique électronique populaire très codifié (ici la *psytrance*). Le choix du genre n'est pas anodin : je souhaitais utiliser un outil de contrôle live expérimental pour un morceau de musique qui, lui, n'a rien d'expérimental. Pour moi, il s'agit plus avec LiveScaler d'explorer de nouvelles modalités live que de nouveaux horizons musicaux.

Pour autant, LiveScaler peut tout à fait intervenir dans le processus de composition. On peut par exemple l'utiliser pour trouver des variations sur une mélodie ou une arpegge en expérimentant avec les différents changements de gamme, puis consolider le MIDI une fois qu'on a trouvé

10. Si, comme pour moi, cette chanson a tendance à rester dans votre tête, je suis (presque) désolée

11. La trance psychédélique est souvent appelée *psytrance*, le lectorat curieux pourra écouter par exemple l'album *The Gathering* (1999) du duo israélien *Infected Mushroom*.

une idée satisfaisante. On obtient alors un processus créatif incrémental et prône à la sérendipité partant d'une mélodie ou d'une progression d'accords simple qu'on enrichit ensuite avec LiveScaler.

5. TRAVAUX CONNEXES

Les transformations affines sont directement inspirées de la théorie transformationnelle (pour une introduction généraliste du point de vue mathématique, voir [5], et du point de vue musicologique, lire [6]). La théorie transformationnelle prend ses racines dans la *Set-Theory*, qui se concentre sur la notion de classe de hauteur, c'est à dire un ensemble de notes identiques à l'octave près [15]. Elle propose une approche plus algébrique que la *Set-Theory*, en se concentrant, entre autre sur la notion de transformation entre ensembles de classes de hauteurs [18].

Les transformations affines sont particulièrement proche des automorphismes du groupe T/I proposés par David Lewin [19]. L'unique différence avec ceux-ci est que les transformations affines ne sont pas nécessairement bijectives et autorisent donc un coefficient modal qui ne soit pas nécessairement premier avec 12. Les transformations affines sont donc une vision plus appliquée (l'aspect algébrique est passé sous silence) des automorphismes proposés par Lewin et Klumpenhouwer, tout en proposant quelques transformations supplémentaires sortant du système tonal.

Plusieurs outils s'appuient plus ou moins explicitement sur les représentations de la théorie transformationnelle, en particulier OpenMusic, qui propose une aide à la composition directement inspirée de celle-ci [4], [3]. Depuis une dizaine d'années, OpenMusic essaie de concilier l'approche hors du temps (approche guidée par les demandes) de l'aide à la composition avec l'approche temps-réel propre à la performance [10], [11]. Plus récemment, Bach propose lui aussi cette approche hybride mais en partant de Max MSP, un langage de programmation fondamentalement temps-réel et guidé par les données [2].

Les outils évoqués ci-dessus offrent une grande flexibilité pour appliquer des transformations potentiellement bien plus sophistiquées que les transformations affines, et sont *a priori* tous capables de le faire en live. Pour autant, cela demanderait un grand travail préparatoire de programmation et de *mapping* avant d'arriver à un résultat fluide. C'est exactement ce travail qui est fait par LiveScaler, mais sur un nombre restreint de transformations, ici jugées pertinentes. LiveScaler sacrifie donc la flexibilité dans le choix des transformations au profit d'une utilisation immédiate, et sans connaissances de programmation requises, pour faire de la musique live.

Une approche intermédiaire offrant plus de flexibilité, mais moins d'immédiateté est celle du *live coding*. Pendant une performance de *live coding*, le musicien utilise un langage de programmation dédié pour coder en live un morceau de musique [8]. En particulier le langage Tidal [21] permet de manipuler et transformer des motifs en

live. Il reprend notamment les transformations de Laurie Spiegel [25] et plus particulièrement les transpositions et inversions (qui correspondent aux coefficients modaux 1 et -1 dans le paradigme des transformations affines proposé ici).

Si une des revendications initialement associées à la pratique du *live coding* était de s'affranchir des contraintes et rigidités des stations audionumériques telle qu'Ableton Live [12], l'ajout de langages de scripting ainsi que la possibilité de contrôler les stations numériques¹² avec des langages de programmation graphique haut niveau¹³ semble avoir développé leur usage dans les pratiques de musique algorithmique live [13]. Ces pratiques, dans lesquelles s'inscrit le présent article, permettent de combiner d'une part la flexibilité et la liberté qu'offrent un langage de programmation et d'autre part l'accès aux outils de production commerciaux utilisés par l'industrie de la musique. C'est particulièrement important dans le cadre de l'EDM, qui utilise intensivement ces outils de production sophistiqués [16].

L'idée d'utiliser un morceau composé au préalable comme matière première à laquelle on applique des transformations est particulièrement développée, et théorisée par Louis Bigo et Darrell Conklin [7]. En analysant au préalable l'harmonie d'un morceau, ils proposent ainsi de la transformer ensuite afin d'obtenir une variation du morceau. LiveScaler se démarque de cette approche par deux principaux aspects : LiveScaler ne nécessite aucune analyse préalable et se concentre sur le live. Une fois de plus la contrepartie est une moins grande flexibilité dans le choix des transformations.

Le logiciel EmoteControl [23] est sans doute celui qui se rapproche le plus de LiveScaler. Il permet un macro-contrôle en live de paramètres tels que le tempo, l'articulation, la hauteur de note, etc. En particulier, il propose une inversion du mode (qui correspond au coefficient modal -1). LiveScaler offre une bien plus grande variété de transformations de gammes et de flexibilité sur la performance live. Une collaboration serait ici particulièrement intéressante : le contrôle de l'articulation ou du timbre seraient particulièrement pertinents à ajouter à LiveScaler.

6. CONCLUSION

Nous avons présenté LiveScaler, qui propose de nouvelles modalités pour la musique électronique live, ainsi qu'un jeu de transformations MIDI : les transformations affines. LiveScaler permet d'appliquer ces transformations affines en live. En particulier, cet outil peut être utilisé dans le contexte de l'EDM, proposant ainsi une nou-

12. On peut par exemple piloter Ableton Live et FL Studio avec Python, Logic et Bitwig avec JavaScript.

13. On trouve dans les stations audionumériques commerciales de plus en plus de langages de "*patching*" permettant de contrôler le logiciel ou de créer des plugins audio de manière modulaire : voir par exemple Max for Live pour Ableton Live, FL FlowStone pour FL Studio, The Grid pour Bitwig.

velle alternative ou un complément pour la performance live en musique électronique.

Plusieurs améliorations techniques pourraient être proposées pour améliorer LiveScaler. En particulier, il serait pertinent de le rendre compatible avec n'importe quel DAW, et pas seulement Ableton Live. Une solution sera de développer un plugin VST ou LV2 pour LiveScaler. De plus, le protocole MIDI étant contraignant, rendre LiveScaler compatible avec un protocole plus flexible tel que OSC [26], MPE (Midi Polyphonic Expression), Midi 2.0 ou encore MP [17], qui serait particulièrement adapté à cette application.

Bien que LiveScaler propose déjà la possibilité d'ajouter manuellement des transformations de gamme périodiques sur un intervalle, ce mécanisme est laborieux et mérite d'être amélioré. Utiliser des transformations sur un espace diatonique serait également intéressant et soulèverait la difficulté de connaître la tonalité dans laquelle on se trouve. Enfin, nous aimerions pouvoir agir sur d'autres paramètres que la hauteur des notes, en particulier le rythme, le timbre, ou même contrôler simultanément des transformations musicales et vidéos. C'est sur ces axes que seront concentrées nos recherches futures, tout en restant sur le même paradigme de performance live que celui proposé ici.

6.1. Remerciements

Je tenais à remercier ici David Janin et Martin Laliberté, mes encadrants de thèse, pour leur relecture et leurs nombreux conseils. Merci également à Chloé Lavrat, pour avoir pris le temps de me lire et pour nos nombreuses discussions sur le sujet, toujours très inspirantes.

7. REFERENCES

- [1] Agostini, A., Ghisi, D., Giavitto, J. "Programmer en Max avec bell", *Journées D'informatique Musicale (JIM)*, 2020.
- [2] Agostini, A., Ghisi, D., Giavitto, J. "Programming in style with bach", *Perception, Representations, Image, Sound, Music*, pp. 257-278, 2021.
- [3] Andreatta, M., Agon, C. "Formalisation algébrique des structures musicales à l'aide de la Set-Theory : aspects théoriques et analytiques", *Journées D'Informatique Musicale*, 2003.
- [4] Andreatta, M., Agon, C. "Implementing algebraic methods in OpenMusic", *ICMC*, 2003.
- [5] Andreatta, M. "Calcul algébrique et calcul catégoriel en musique : aspects théoriques et informatiques", *Le Calcul De La Musique*, pp. 429-477, 2008.
- [6] Andreatta, M. "Une introduction musicologique à la recherche «mathémusicale» : aspects théoriques et enjeux épistémologiques", *Circuit*, **24**, 51-66, 2014.
- [7] Bigo, L., Conklin, D. "A viewpoint approach to symbolic music transformation", *Music, Mind, And Embodiment : 11th International Symposium, CMMR 2015, Revised Selected Papers 11*, pp. 213-227, Plymouth, UK, 2016.
- [8] Blackwell, A., Cocker, E., Cox, G., McLean, A., Magnusson, T. "Live coding : a user's manual", MIT Press, 2022.
- [9] Bresson, J., Agon, C., Assayag, G. "OpenMusic : visual programming environment for music composition, analysis and research", *Proceedings Of The 19th ACM International Conference On Multimedia*. pp. 743-746, 2011.
- [10] Bresson, J., Giavitto, J. "A reactive extension of the openmusic visual programming language", *Journal Of Visual Languages & Computing*, **25**, pp. 363-375, 2014.
- [11] Bresson, J., Bouche, D., Carpentier, T., Schwarz, D., Garcia, J. "Next-generation Computer-aided Composition Environment : A new implementation of OpenMusic", *International Computer Music Conference*, 2017.
- [12] Collins, N., McLean, A., Rohrhuber, J., Ward, A. "Live coding in laptop performance", *Organised Sound*, **8**, pp. 321-330, 2003.
- [13] Collins, N., McLean, A. "Algorave : Live performance of algorithmic electronic dance music", *Proceedings Of The International Conference On New Interfaces For Musical Expression*, pp. 355-358, 2014.
- [14] Ferreira, P. "When sound meets movement : Performance in electronic dance music", *Leonardo Music Journal*, **18**, pp. 17-20, 2008.
- [15] Forte, A. "The structure of atonal music", Yale University Press, 1973.

- [16] Fraser, A. "The spaces, politics, and cultural economies of electronic dance music", *Geography Compass*, **6**, pp. 500-511, 2012.
- [17] Goudard, V., Genevois, H. "Mapping modulaire de processus polyphoniques", *Journées D'Informatique Musicale*, 2017.
- [18] Lewin, D. "Generalized musical intervals and transformations", New Heaven : Yale University Press, 1987.
- [19] Lewin, D. "Klumpenhouwer networks and some isographies that involve them", *Music Theory Spectrum*, **12**, pp. 83-120, 1990.
- [20] Livingstone, S., Muhlberger, R., Brown, A., Thompson, W. "Changing Musical Emotion : A Computational Rule System for Modifying Score and Performance", *Computer Music Journal*, **34**, pp. 41-64, 2010.
- [21] McLean, A., Wiggins, G. "Tidal-pattern language for the live coding of music", *Proceedings Of The 7th Sound And Music Computing Conference*. pp. 331-334, 2010.
- [22] Magana, J. "Performance in EDM-A Study and Analysis of DJing and Live Performance Artists", 2018.
- [23] Micallef Grimaud, A., Eerola, T. "EmoteControl : an interactive system for real-time control of emotional expression in music", *Personal And Ubiquitous Computing*, **25** pp. 677-689, 2021.
- [24] Morris, R. "Composition with pitch-classes", *Yale University*, 1987.
- [25] Spiegel, L. "Manipulations of musical patterns", *Proceedings Of The Symposium On Small Computers And The Arts*, pp. 19-22, 1981.
- [26] Wright, M. "Open Sound Control : an enabling technology for musical networking", *Organised Sound*, **10**, pp. 193-200, 2005.

VERS UNE ARCHITECTURE LOGICIELLE POUR L'ENSEIGNEMENT HYBRIDE DE L'HARMONIE

Anaïs BINET

Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400 Talence, France
anaïs.binet@labri.fr

Myriam DESAINTE-CATHERINE

myriam.desainte-catherine@labri.fr

RÉSUMÉ

L'harmonie est une discipline qui fait partie intégrante du cursus de musicologie, mais aussi du cursus en conservatoire. Le projet présenté dans cet article a pour objectif la conception d'un outil pédagogique pour l'enseignement hybride de cette discipline. Nous recherchons des solutions techniques à des problématiques exprimées à la fois par les enseignants et les élèves.

1. INTRODUCTION

L'harmonie est une combinaison de plusieurs éléments qui « vont bien ensemble ». Dans le domaine de la musique et plus particulièrement dans la théorie musicale occidentale, l'harmonie est l'art d'étudier la construction des accords. On parle alors de l'étude de la dimension verticale de la musique.

L'outil informatique est de plus en plus présent dans les salles de cours et possède certains atouts pédagogiques. Nous chercherons, dans cet article, à proposer des idées d'architecture pour l'enseignement de l'écriture musicale dans les universités et les conservatoires. Nous n'aborderons dans cet article ni les questions sociopolitiques relatives à notre projet ni la façon dont il sera évalué.

Par sa nature interdisciplinaire, ce projet s'appuiera sur des travaux issus de domaines divers tels que la musicologie et l'informatique.

1.1. Motivation

Les problématiques rencontrées par les enseignants auxquelles nous nous intéressons sont les suivantes :

- La difficulté de corriger individuellement le travail de tous les élèves
- Savoir si les notions sont comprises par les élèves
- Aborder l'entièreté du programme dans un temps donné

En effet, l'enseignant est confronté à des classes parfois nombreuses. Dans de telles conditions, il ne peut réaliser que très peu de corrections individualisées qui lui sont coûteuses en temps de cours. Cela ne lui permet donc pas toujours de cerner les notions incomprises par les élèves car un échantillon trop restreint n'est pas représentatif des acquis de l'ensemble de la classe.

Les problématiques rencontrées par les élèves auxquelles nous nous intéressons sont les suivantes :

- Évaluer sa propre progression
- Cerner ses points faibles et ses points forts
- Savoir si la solution trouvée pourrait être juste même si elle n'apparaît pas dans le corrigé réalisé par l'enseignant
- Avoir besoin d'aide sur les exercices en dehors des cours
- Avoir besoin de s'entraîner davantage en autonomie
- Se détacher des règles lors de projet créatif

En effet, la question des corrections individualisées impacte aussi beaucoup les élèves. Elles sont un bon indicateur de progressions et mettent en perspective leurs lacunes ainsi que leurs forces. De plus, en musique comme dans beaucoup d'autres matières, il n'existe pas une solution unique à un problème donné. Or, dans l'enseignement traditionnel, si l'enseignant n'a pas corrigé individuellement le travail de chaque élève, celui-ci n'aura que la correction unique de l'enseignant pour procéder à son autoévaluation. Cette méthode lui permet seulement d'analyser les différences entre son travail et l'unique correction à sa disposition.

Une autre problématique rencontrée par les élèves est de devoir attendre le cours suivant pour avoir la correction quand ils rencontrent des difficultés à faire l'exercice. Dans ce cas, l'élève mémorise et comprend beaucoup plus difficilement les notions. Parfois, il suffirait d'une seule intervention de l'enseignant pour aider l'élève. Cependant, l'enseignant ne peut fournir un accompagnement individualisé à chaque élève entre les cours.

Enfin, les corrections uniques de groupe des exercices d'harmonie ne permettent pas aux élèves

d'explorer toutes les possibilités et d'avoir suffisamment d'exemples. Cela peut créer des difficultés, lors d'exercices de créativité, à se détacher des règles strictes enseignées.

1.2. État de l'art

Plusieurs outils pour l'enseignement de la musique existent déjà. Certains s'adressent aux enseignants et élèves en écoles/conservatoire de musique pour l'enseignement de la théorie musicale et même de la pratique instrumentale. D'autres sont davantage à destination de personnes souhaitant se professionnaliser dans la musique et des professionnels de la musique.

Il existe beaucoup d'outils pour l'enseignement et l'apprentissage du solfège, parfois appelé la formation musicale (FM). Un des piliers de la formation musicale est le travail de l'oreille en passant de la reconnaissance de hauteur et de rythme à l'analyse musicale. EarMaster [17] est par exemple un outil pour apprendre à reconnaître les hauteurs de note et les rythmes afin d'être capable de relever des thèmes, des œuvres. Cet outil ne s'arrête pas là et propose aussi des exercices de lecture, de rythme et chantés. Il permet donc de s'entraîner en toute autonomie aux fondamentaux de la formation musicale. Il existe une version du logiciel pour les écoles de musique et les conservatoires dotée d'une interface enseignant. L'enseignant pourra alors créer ses propres exercices. On peut citer également Deepsolfège [9], Solfeggio [12] et Troubadour [14] comme logiciel pour l'apprentissage de la théorie musicale.

D'autres outils ont aussi été conçus pour l'apprentissage d'un instrument. Les élèves entre les cours ont parfois besoin de conseils et doivent malheureusement attendre le cours suivant pour que l'enseignant débloque la situation. Afin de répondre à cette problématique dans l'apprentissage du violon, Digital Violin Tutor (DVT) [20] a été conçu. Il repose sur la visualisation 3D du geste à l'aide d'un avatar qui reproduit les mouvements enregistrés en vidéo, d'un tracking des erreurs affichées sur un piano roll et une animation 2D représentant les positions des doigts sur la touche du violon. On peut citer également d'autres outils pour l'apprentissage du piano et de la guitare avec Yousician [8] ou encore de la flûte avec le robot anthropomorphe WF-4 [16].

Un certain nombre de ces outils peuvent être utilisés en toute autonomie. Cela peut engendrer plusieurs difficultés dans le processus d'apprentissage. L'élève doit être organisé et pratiquer régulièrement sans sollicitation extérieure. Cependant, seul un certain pourcentage des élèves y parviendra. De plus, les élèves peuvent ressentir de la lassitude face à des difficultés auxquelles l'outil ne saurait répondre. L'intervention d'un enseignant est souvent essentielle dans le processus d'apprentissage. C'est un guide sur le chemin de l'acquisition du savoir et

des compétences. Un outil seul ne peut pas se substituer à un enseignement traditionnel, mais un outil peut aider les enseignants. On parle donc d'enseignement hybride.

2. SCENARIO D'UTILISATION

Nous allons dans cette partie présenter un scénario possible d'utilisation du système dans une classe de licence à l'université. On peut observer le schéma fonctionnel d'utilisation du système sur la figure 3.

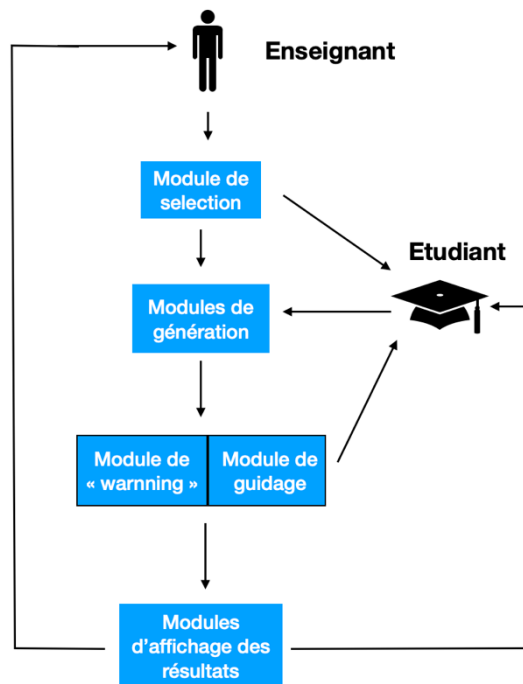


Figure 1. Schéma fonctionnel d'utilisation

Dans un premier temps, l'enseignant remplit tous les paramètres dans l'interface qui lui est dédiée. Dans notre exemple l'enseignant choisit ces paramètres :

- Compositeur : J. S. BACH
- Notion à inclure : l'accord de quintes et ses deux renversements
- Notion à exclure : septième de dominante, septième d'espèce, neuvième et leurs renversements respectifs avec et sans fondamentale
- Guidance : pas à pas

Puis le logiciel lui propose des exercices répondant à ses critères issus du catalogue fourni dans le système. L'enseignant peut alors choisir celui qu'il souhaite. Ensuite il vérifie que tous les élèves sont connectés sur l'interface élèves et prêts à faire l'exercice.

Les élèves peuvent donc commencer l'exercice. Le système leur propose une liste avec des solutions possibles, dans lequel sont présentes des réponses fausses, pour le premier accord. S'ils choisissent une réponse

incorrecte, un message s'affiche leur expliquant pourquoi la réponse est incorrecte, c'est-à-dire quelles sont les règles qui sont violées. Les élèves ayant donné une réponse fautive peuvent alors faire une autre proposition. Chaque fois que l'élève se trompe, le système incrémente une liste dans laquelle chaque élément correspond à une notion. Cela permet ensuite la génération du diagramme qui sera envoyé à l'enseignant pour son suivi pédagogique une fois que tous les élèves auront terminé l'exercice. Au bout d'un certain nombre d'essais infructueux, le système complète avec une réponse juste pour permettre à l'élève de continuer.

Dans le cas où l'enseignant ne choisirait pas la guidance pas à pas et que le système détecterait un trop grand nombre d'erreurs avant même de corriger l'intégralité du devoir de l'élève, sera affiché un message invitant l'élève à demander de l'aide via le module de guidage intégré. Que la guidance pas à pas ou une autre soit choisie par l'enseignant, le système enregistre le temps passé par chaque élève sur le devoir et son avancée, ce qui pourra permettre à l'enseignant, en analysant les données, de détecter le manque de travail, mais surtout la démotivation éventuelle d'un ou plusieurs élèves.

Le module de correction admet donc différents réglages qui permettent à l'enseignant de s'adapter au niveau des élèves dans la limite de ses choix pédagogiques. Ainsi, réglé au maximum, il apporte des indications à l'élève sur son travail même s'il est entièrement juste.

L'enseignant peut procéder à une correction globale de l'exercice au tableau en insistant sur les notions révélées par le système comme étant les moins comprises. Il a le choix entre : élaborer seul la correction, en faire générer une par le système qu'il explique ou bien encore la construire avec la participation de la classe. Le système lui laisse sa liberté pédagogique dans son entièreté.

3. L'HARMONIE, UNE DISCIPLINE MUSICALE

Nous allons, avant d'aborder la question de l'architecture, définir la discipline de l'harmonie et cerner son intégration dans le cursus des élèves afin de décrire les paramètres que nous prenons en considération dans notre projet.

3.1. Définition de l'harmonie musicale

L'harmonie est une discipline qui fait partie intégrante de la formation de toute personne se destinant à un métier touchant à la musique. Étudier l'harmonie permet de s'approprier les langages et styles de la musique tonale occidentale, ce qui la rend incontournable pour tous les interprètes, compositeurs, professeurs de musique ou d'éducation musicale, musicologues, etc.

Dans le cadre de cette discipline sont enseignées les nombreuses règles d'écriture de la musique. Pour ce faire, les enseignants peuvent notamment s'appuyer sur des traités d'harmonie. C'est celui écrit par RAMEAU (Jean-Philippe) [15] qui est encore très largement utilisé.

C'est une discipline à la fois théorique et pratique. Il y a deux principaux formats d'exercices : dans l'un de ces formats, l'enseignant donne une ligne de basses chiffrées à l'élève qui doit alors en faire une réalisation généralement pour chœur à 4 voix. Dans ce format sont travaillées les notions d'écriture mélodique et la réalisation à 4 voix avec les difficultés et règles que cela implique. Dans l'autre format, l'enseignant donne à l'élève essentiellement la mélodie. Sont alors travaillées les notions d'enchaînement harmonique, de formes et de réalisation pour chœur ou pour piano.

Un certain nombre d'ouvrages regroupant des exercices par compétences existent et sont souvent utilisés par les enseignants. Par exemple, à l'université de Bordeaux Montaigne, M. PELAPRAT (Emmanuel) utilise les ouvrages de RAYNAUD (Jean-Claude). On y retrouve à la fois des exercices progressifs, mais aussi des textes pour l'étude des styles de certains compositeurs comme MOZART ou BACH.

Le cours d'harmonie est souvent mis en lien avec celui d'harmonie au clavier, parfois appelé : cours d'accompagnement piano. En effet, cela permet aux élèves d'expérimenter les notions théoriques du cours d'harmonie, mais aussi de les aider à les comprendre par leurs manipulations.

3.2. Son intégration dans le cursus musical

L'harmonie est un prolongement du solfège, autrement appelé formation musicale. Elle est à étudier conjointement à l'analyse.

On observe une différence significative d'interprétation par les élèves quand ils ont suivi des cours d'harmonie. Elles influent directement la pratique de l'instrument. Les élèves comprenant mieux le langage musical et sa construction jouent de manière différente. Leurs interprétations sont plus détaillées, plus précises, plus proches de ce qui est écrit sur la partition, beaucoup moins mécaniques et donc plus musicales.

Parce qu'il est difficile d'enseigner ce que l'on ne comprend pas pleinement, l'harmonie est un des piliers de la formation des enseignants en musique. Elle leur apporte des compétences nécessaires pour la construction de leurs cours. Cela va de l'harmonisation d'une chanson, pour une classe de collège, jusqu'à l'enseignement de la discipline elle-même à des élèves.

Pour devenir compositeur, il faut préalablement étudier le travail des compositeurs du passé. Cela passe certes par les matières d'histoire de la musique et l'analyse, mais aussi par l'harmonie. Effectivement, l'aspect de discipline pratique permet d'approfondir les notions, mais aussi de se les approprier afin de pouvoir les restituer par la suite.

Au-delà de la restitution, cela pourra permettre de s'en servir pour chercher et innover, et ainsi être créatif par rapport aux autres compositeurs. C'est un phénomène qui a pu être observé tout au long de l'histoire de la musique. Par exemple, les compositeurs de la période classique ont étudié les compositeurs de l'époque baroque, et se sont approprié leurs manières de composer pour ensuite les faire évoluer afin de créer ce que les musicologues appellent la musique classique. Et cet exemple vaut pour tous les styles au cours de l'histoire, comme le romantisme, etc.

Pour résumer, l'harmonie n'est pas une fin en soi, mais un outil, une base solide pour tous les métiers de la musique.

4. UNE CLASSE HYBRIDE D'HARMONIE

L'objectif de ce projet est de proposer un outil pédagogique pour l'harmonie qui s'insère dans une démarche de construction de classe hybride et d'enseignement hybride de l'harmonie.

Un enseignement hybride est un mode de formation qui mêle à la fois des cours en présentiel et une part d'e-learning. Une classe hybride est un environnement où l'enseignant va pouvoir alterner entre enseignement classique et enseignement à l'aide d'un ou plusieurs outils informatiques (tablette, ordinateur...).

Dans ce projet, l'e-learning permettra de répondre à des besoins réels chez les élèves. En effet, ils ont parfois besoin entre les cours d'indications, de pistes, afin d'effectuer les devoirs donnés à faire à la maison. Alors que dans un enseignement classique, l'élève doit attendre que l'enseignant corrige au tableau pour pouvoir recopier la correction, parfois sans vraiment la comprendre. Un des objectifs de notre outil sera de leur donner un coup de pouce lorsqu'ils expriment le besoin afin de débloquer un problème et ainsi leur permettre de faire l'intégralité de l'exercice.

Dans le cadre de l'e-learning sera mis également à disposition des élèves un ensemble d'exercices sélectionnés par l'enseignant dans une base de données afin de leur permettre de s'entraîner en toute autonomie. Chaque élève pourra donc apprendre à son rythme à la fois de manière individuelle et collective. C'est ce que l'on nomme la pédagogie différenciée.

Dans l'enseignement classique de l'harmonie où l'enseignant ne corrige qu'un petit échantillon d'exercices réalisés par les élèves, il n'a pas suffisamment de données pour savoir si les notions sont comprises quand les élèves n'expriment pas leurs incompréhensions, leurs difficultés en classe. Pour apporter une réponse à ce problème, l'outil effectuera une correction individualisée de tous les exercices, puis, dans un second temps, elles pourront être transcrites en statistiques et envoyées sur l'interface enseignant sous la forme de diagrammes. Ainsi, d'un seul

coup d'œil, l'enseignant saura quelles notions il doit réexpliquer en classe.

Cet outil devrait permettre à l'enseignant de gagner du temps sur le cours tout en permettant non pas à un petit groupe seulement, mais à tous les élèves, de bénéficier d'une correction individualisée. Cela permettra notamment une correction de groupe, en classe, plus active. En effet, les élèves auront sous les yeux une correction individualisée leur permettant de partager à l'oral à l'ensemble de la classe la solution qu'ils ont trouvée et ainsi alimenter la correction faite au tableau. Cela favorisera la participation à l'oral et aidera à instaurer une équité et un climat de confiance entre tous les élèves. L'enseignant ne fera donc pas une correction en fonction seulement des solutions qu'il a lui-même envisagées, mais pourra les combiner avec celles des élèves.

Grâce à cet outil, les élèves pourront tous bénéficier d'une correction individualisée dans laquelle leurs erreurs seront expliquées. Cela leur permettra de constater leur progression. De plus, l'outil leur permettra d'observer davantage de solutions, d'exemples en fonction de leur propre et unique façon d'aborder l'harmonie. Ce qui participera également à alimenter leurs créativité propres.

L'outil pourra, en fonction des paramétrages réalisés par l'enseignant, corriger pas à pas le devoir de l'élève, permettant à ce dernier d'éviter des erreurs qui pourraient se répercuter sur la suite de l'exercice le rendant parfois totalement faux. L'enseignant aura donc le choix entre : permettre aux élèves d'être guidés pas à pas, à la demande ou pas du tout en fonction de ses choix pédagogiques.

L'outil devrait permettre de favoriser la créativité des élèves en présentant un nombre plus grand d'exemples inspirés des styles des compositeurs étudiés. Des solutions jugées fausses par les traités d'harmonie, mais usitées par les compositeurs pourront alors être abordées. Nous espérons ainsi inverser la tendance et permettre que la discipline d'alimente la créativité des élèves.

Pour mener à bien ce projet, il nous faudra chercher quelles solutions techniques sont les plus adaptées.

5. L'INTELLIGENCE ARTIFICIELLE POUR LA GENERATION MUSICALE

La première musique générée par ordinateur est apparue en 1957. Cette œuvre est une mélodie de 17 secondes intitulée The Silver Scale générée par un logiciel de synthèse sonore nommé « Music I ». Mais c'est plus tard la même année que la première partition, La Suite Illiaque, est composée par un ordinateur [1]. C'est un des premiers exemples de composition algorithmique, utilisant des modèles stochastiques (chaînes de Markov) pour la génération d'œuvres musicales.

Puis, au début des années 1960, XENAKIS (Iannis) continue sur la même voie et expérimente des modèles de composition stochastique [19], dans sa composition intitulée *Atrées* en 1962. L'idée consistait à utiliser des calculs rapides pour calculer diverses possibilités à partir d'un ensemble de probabilités conçues par le compositeur afin de générer des échantillons de pièces musicales à sélectionner.

À la fin des années 1980, COPE (David) développe le système nommé « Experiments in Musical Intelligence » (EMI) qui a la capacité d'apprendre à partir d'un corpus de partitions d'un compositeur un grand nombre de règles afin d'en reproduire le style [6]. La même année, Kemal Ebcioglu développe dans le cadre de sa thèse un système expert pour harmoniser des chorales dans le style de J. S. Bach intitulé « CHORAL ». Son système se base sur trois principes : l'encodage d'un très grand nombre de connaissances du style musical de J. S. Bach, des contraintes pour générer des solutions, mais également corriger les générations inacceptables avec un retour en arrière, et pour finir, l'utilisation d'heuristiques afin de hiérarchiser les choix de l'algorithme pour que des solutions soient préférées à d'autres, car plus acceptables.

Plus récemment, d'autres programmes comme Omax [2] pour l'improvisation musicale s'appuient eux sur un système d'oracle de facteur. Puis émerge la notion d'apprentissage avec le machine learning et le deep learning qui permettent encore à l'heure actuelle d'aller toujours plus loin dans le domaine et de créer des programmes toujours plus performants comme les transformers. Deep-Bach [10] par HADJERES (Gaetan) et le Doodle [11] de Google sont l'exemple de l'application de méthodes d'apprentissage à la génération de musique.

6. ASSOCIER DES IA POUR NOURRIR LA CREATIVITE

Pour répondre au problème soulevé précédemment, nous allons devoir trouver des solutions techniques, notamment pour la correction des exercices d'harmonie, la génération de propositions d'harmonisations ou la complétion de solutions des élèves, l'explication et l'évaluation de réponses à un exercice d'harmonie. En effet, le logiciel proposera les fonctionnalités suivantes :

- Permettre à l'enseignant de faire des choix pédagogiques de manière interactive pour adapter le niveau de l'aide du logiciel.
- Apporter de l'aide aux élèves en proposant des générations ou des complétions d'harmonisations d'un chant donné ou de réalisations d'une basse chiffrée.
- Corriger en décrivant et expliquant musicalement le résultat généré ou les exercices des élèves.

- Proposer plusieurs modes de guidage pour les élèves dans la réalisation de l'exercice, selon leur niveau.
- Proposer des visualisations représentatives de la progression des élèves au moyen d'un classement des erreurs dans les notions correspondantes.
- Utiliser plusieurs corpus pour faire découvrir la manière d'écrire des compositeurs du passé et ainsi alimenter la créativité des élèves.

Il apparaît donc que le système devra être doté d'un certain nombre de modules, notamment un module de génération destiné à compléter les exercices à la place des élèves pour leur faire des propositions, un module de correction destiné à exhiber les règles utilisées dans une solution, ou à expliquer pourquoi une solution est fautive.

6.1. Construction du processus de génération

Pour répondre à ces besoins et au regard des exemples de l'application de l'intelligence artificielle pour la génération de musique nous avons envisagé plusieurs choix techniques.

Pour permettre au professeur de sélectionner des compositeurs selon son choix pédagogique, nous envisageons de construire une base de données qui regroupera un ensemble de partitions musicales classées en corpus étiquetées en fonction de leur compositeur, chaque compositeur ne faisant pas partie du même courant stylistique.

Nous souhaitons privilégier le format Music XML qui conserve la nature des intervalles à la place du format MIDI pour construire notre base de données. En effet, si l'on considère par exemple l'intervalle entre les notes « DO » et « MI# » on a une tierce augmentée. Mais, si l'on considère l'intervalle entre les notes « DO » et « FA » on a une quarte juste. Or, les valeurs MIDI de « MI# » et « FA » sont identiques.

Notre système doit être capable de reproduire la manière d'écrire des compositeurs à partir d'un corpus. Aussi nous envisageons d'utiliser une technique d'apprentissage profond. Cette technique a déjà fait ses preuves pour sa capacité à compléter des séquences déjà existantes à l'aide de l'inpainting. Cela permettra de répondre au besoin de guidage des élèves et à la correction. Cependant, le système doit s'inclure dans une démarche pédagogique et donc être capable d'expliquer la génération obtenue. Or, la problématique de l'explication pour les réseaux de neurones est un sujet qui est encore à l'état de recherche. Nous envisageons donc pour cette tâche d'expérimenter un couplage entre un réseau de neurones et un système à base de règles. Le système à base de règles devra être en mesure d'analyser les productions du générateur pour en extraire les règles utilisées dans les

enchaînements ou bien pour en détecter des erreurs. Ce module pourra aussi être utilisé pour corriger les productions de l'élève. La **Figure 2** schématise l'enchaînement des modules. Entre le générateur et le correcteur se situe un module d'analyse qui est destiné à étiqueter harmoniquement la séquence.

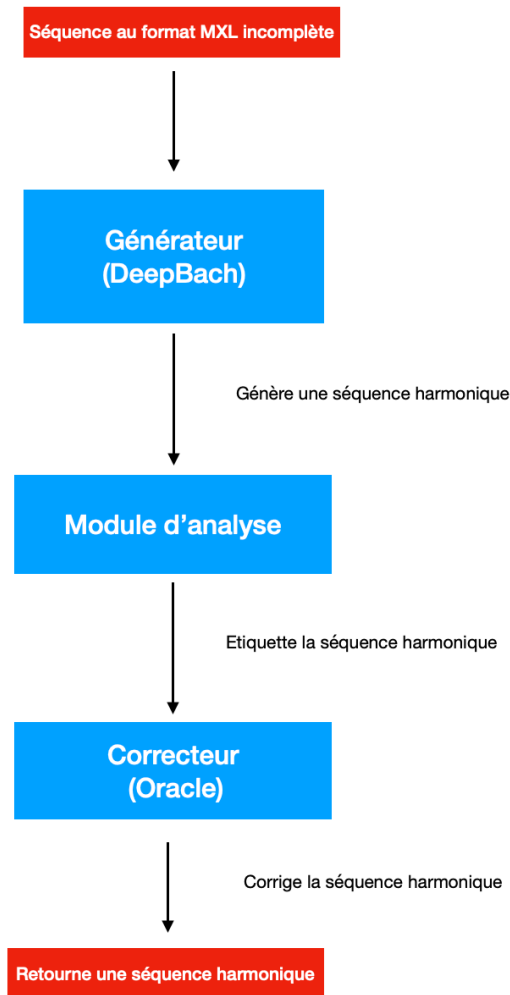


Figure 2. Processus de génération

Ce module nous servira également pour analyser le travail des élèves pour que notre système soit interactif.

Nous avons fait le choix d'utiliser comme base pour le générateur le réseau de neurones DeepBach créé par Gaetan Hadjeres [10] qui peut prendre en entrée des fichiers XML. Ce qui nous a amené à choisir ce réseau en particulier est le fait qu'il est doté de la capacité de faire de l'inpainting par opposition à d'autres réseaux de neurones que nous avons étudiés. Le programme ayant été développé dans le cadre d'une thèse au sein du Laboratoire LIP6 dans le Département ASIM de l'Université Pierre et Marie Curie, sous licence MIT, nous pourrions l'utiliser, le modifier et le réutiliser afin de garantir que notre projet soit open source. De plus, il

existe un plugin pour l'intégrer au logiciel open source et gratuit d'édition de partition, MuseScore.

Par opposition, le correcteur prendra la forme d'un oracle de facteurs. Nous le construirons à l'aide du module Python conçu par WANG (Cheng-i) [18].

Tous les modules constitutifs de notre architecture pourront être utilisés de manière indépendante pour l'apprentissage (exemple : le module d'analyse), mais aussi pour de nouveaux projets informatiques car ils seront tous open source.

6.1.1. Module d'analyse harmonique

Il y a différentes techniques de notation de l'harmonie en musique. En effet, il y a la notation baroque qui s'appuie sur un chiffrage des accords en chiffres arabes associés à des chiffres romains pour les degrés, mais aussi la notation anglo-saxonne que l'on utilise souvent en jazz.

Nous avons conçu un module d'analyse qui permet de réaliser une analyse harmonique de fichiers de partitions XML. Ce module est en cours de finition. Il est représenté par le schéma de la **Figure 3**. Il est disponible sur GitHub [4]. Ce module sera utilisé pour construire les corpus analysés qui seront en entrée du système.

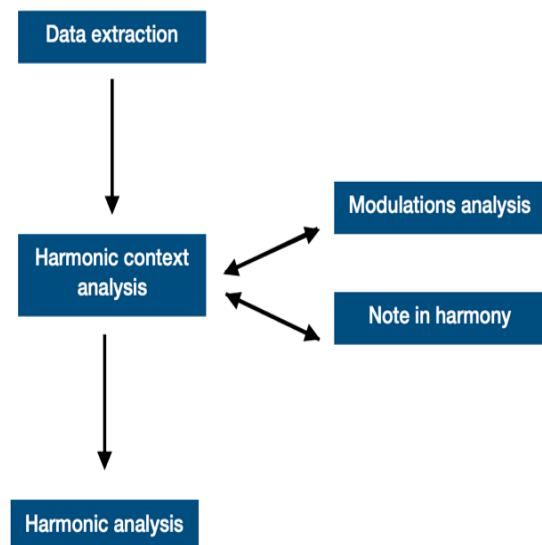


Figure 3. Module d'analyse de partition MXL

Dans un premier temps, nous extrayons les données musicales d'un fichier au format Music XML, c'est-à-dire les hauteurs et les rythmes des différentes parties (voix). Puis, nous normalisons les données afin d'obtenir une liste des accords avec un pas de temps d'une double croche (un quart de temps).

Pour analyser les accords, il est nécessaire de connaître le contexte dans lequel ils sont joués. Il arrive régulièrement que les notes jouées simultanément ne fassent pas partie de la même harmonie. On dit qu'il y a

des notes étrangères. Il faut alors les identifier. Il faut aussi identifier les notes qui font partie de l'harmonie, mais qui sont jouées avant ou après. On les appelle les notes réelles.

Après avoir analysé les notes étrangères et les avoir remplacées dans le tableau de chaque accord par les notes réelles, le système va réduire le jeu de données aux éléments essentiels. Nous n'avons plus besoin d'observer les accords avec un pas de temps d'une double croche. On va par exemple garder un accord sur quatre si la fréquence harmonique est à la noire ou encore un sur deux si la fréquence est à la croche, etc.

Afin de pouvoir donner le chiffrage baroque d'un accord, il nous faut aussi connaître la tonalité au moment où il est joué. En effet, le degré d'un accord dépend de la tonalité et pourra être un premier degré dans une tonalité donnée et un cinquième degré dans une autre tonalité. Or, en harmonie l'enchaînement des degrés est important et est le sujet de nombreuses règles. Le système crée alors un tableau avec la tonalité dans laquelle est joué chaque accord.

Le contexte harmonique étant connu, il est alors possible de commencer à analyser les accords. Le module commence par construire un tableau avec le chiffrage anglo-saxon de tous les accords. Puis, il fait de même avec le chiffrage baroque, mais dans deux tableaux : un pour le chiffrage et un pour les degrés.

Ce module retourne pour finir toutes ces listes : une pour les modulations, une pour les degrés, une pour le chiffrage baroque et une dernière avec le chiffrage anglo-saxon. C'est cette dernière que nous utiliserons comme entrées de l'oracle.

6.1.2. *Module de correction*

Dans le cadre d'un mémoire de Master [3], nous avons réalisé une méthode d'extraction de règles à partir d'un corpus constitué des sonates de MOZART. Nous avons tout d'abord dû analyser l'ensemble du corpus afin de l'annoter avec le chiffrage anglo-saxon et les degrés. À partir de ces partitions annotées, nous avons reporté dans un tableau les séquences possibles de deux accords consécutifs. C'est-à-dire que nous avons noté l'accord et la note de la mélodie au temps t , mais aussi l'accord qui le précédait, et ceci pour l'ensemble du corpus. Cependant, nous n'avons pas reporté l'intégralité des enchaînements. Nous avons fait le choix d'exclure tous les enchaînements qui n'étaient pas dans une tonalité voisine à celle du morceau afin de garder les enchaînements les plus courants. Ce travail a débouché sur la conception d'une application mobile Learning Harmoniser [13]. Celle-ci permet d'harmoniser note à note une mélodie dans le style de MOZART.

Dans le cadre de notre projet, nous avons fait le choix de réemployer cette méthode, mais en l'automatisant à l'aide de l'outil informatique. Dans le cadre d'un

prototype, nous allons utiliser un corpus constitué des chorales de Jean-Sébastien Bach. Nous utiliserons le module d'analyse précédemment cité afin d'annoter les partitions et ne plus le faire manuellement. Cela va représenter un gain de temps au vu de la taille du corpus qui est constitué de plus 300 pièces. Aussi, cela facilitera, dans l'avenir, le processus, ce qui permettra d'ajouter plus facilement d'autres corpus d'autres compositeurs au système. Pour automatiser le processus de recherche des séquences harmoniques dans une liste d'accords anglo-saxonne, nous allons construire un oracle à l'aide du module Python conçu par WANG (Cheng-i) [18].

Nous avons fait l'expérience de nous limiter à des séquences de deux accords consécutifs dans l'application mobile développée dans le cadre du mémoire de master précédemment cité. Cependant, certaines règles de l'écriture musicale nécessitent de prendre en compte plusieurs accords précédents. Nous prévoyons donc de distinguer les enchaînements harmoniques entre séquences courtes (deux accords) et séquences longues (dont un exemple est l'anatole qui est constituée de quatre accords).

6.1.3. *Module de génération*

Le réseau de neurones que nous allons utiliser pour notre projet est DeepBach [10].

C'est un modèle destiné à modéliser des chorales dans le style de Bach. Il contraste avec de nombreuses autres approches de composition automatique de musique qui tendent à composer séquentiellement. On peut également interagir avec le programme. En effet, l'utilisateur peut contraindre la génération en imposant des contraintes telles que des notes, des rythmes ou des cadences. Un plug-in de « DeepBach » dans l'éditeur MuseScore est disponible et l'interaction avec DeepBach est ainsi plus facile et accessible.

Le système DeepBach utilise une méthode MCMC (plus précisément une procédure d'échantillonnage de type Gibbs). Les méthodes de Monte-Carlo par chaînes de Markov, ou méthodes MCMC, utilisent des systèmes de simulation complexes, et génèrent des distributions de probabilité.

Contrairement aux modèles basés sur les RNN, DeepBach n'échantillonne pas de gauche à droite, ce qui permet d'appliquer des contraintes. Dans notre projet les contraintes sont soit le sujet de l'exercice soit sa réalisation incomplète par un élève de ce même exercice.

6.2. *Interactivité du système*

Pour répondre aux besoins que nous avons précédemment cités, le système doit pouvoir être interactif.

Afin que l'enseignant puisse cadrer l'utilisation de l'outil et l'utiliser, nous avons fait le choix de concevoir un module par le biais duquel il pourra exprimer ses choix pédagogiques. Nous l'avons appelé « module de sélection ». Il permettra d'agir sur un certain nombre de paramètres, entre autres le compositeur, les notions qui devront être présentes dans l'exercice et celles qui ne le devront pas. Il permettra également de décider si le système guidera pas à pas les élèves, ou à leur demande, ou pas du tout. L'enseignant pourra également agir sur le taux de propositions dérogeant aux règles des traités, mais usitées par le compositeur choisi, que proposera le module de génération, et qu'acceptera le module de correction. Alors, une liste d'exercices répondant à ces paramètres, sélectionnés dans le catalogue que nous aurons conçu, sera proposée. L'enseignant pourra alors choisir celui qu'il souhaite faire faire aux élèves.

Une fois le travail réalisé par les élèves, l'enseignant aura accès par le biais de l'interface à un diagramme représentant les notions comprises et celles qui devront être retravaillées en cours. Il pourra également regarder le temps passé par les élèves à travailler entre les cours.

L'interactivité est aussi importante pour les élèves. Effectivement, nous avons prévu de concevoir un module de guidage et un modèle de correction qui leur permettra d'interagir avec le système. Le module de guidage, en fonction des paramètres choisis par l'enseignant, pourra proposer des solutions à l'élève, à chaque étape de l'exercice ou bien quand il en fera la demande. Dans tous les cas de figure, l'outil ne fournira pas la réponse juste à l'élève, mais il pourra par exemple lui proposer une liste de solutions. On pourrait envisager que dans cette liste de propositions se trouvent des réponses fausses comme dans un QCM. Nous nous inspirerons notamment pour concevoir cette fonctionnalité des logiciels de correction orthographique tels qu'Antidote qui, à l'inverse des correcteurs orthographiques des logiciels d'éditions de texte, ne fournissent pas la réponse, mais invitent la personne à se questionner pour trouver la réponse seule et ainsi en tirer un apprentissage.

Le module de correction quant à lui permettra de bénéficier d'une correction individualisée sans intervention d'un enseignant. Cela permettra à l'élève, par exemple, de s'entraîner en dehors des cours sur d'autres exercices que ceux donnés par l'enseignant, en toute autonomie.

7. CONCLUSION

Ce projet n'en est qu'à son début. Nous avons pour le moment développé le premier module, celui d'analyse, qui est déjà disponible sur GitHub [4]. Cependant, nous sommes encore dans la phase de validation de ce module

à l'aide des analyses des chorals de Bach par CZARNECKI (Christopher) [7].

La prochaine étape sera de construire l'oracle qui nous permettra d'extraire les règles d'enchaînements des corpus et ainsi pouvoir expérimenter le processus de génération avec DeepBach. Cela participera à la construction du premier prototype sur les chorals de BACH. Pour l'élaboration du catalogue d'exercices nous utiliserons les premiers volumes d'exercices écrits par CHALLAN (Henri) [5].

Nous espérons que notre projet permettra à la fois de faire progresser la pédagogie de l'harmonie mais également d'initier les élèves plus tôt dans leur cursus, idéalement dès le 2^e cycle, c'est-à-dire à partir de la 4^e année de musique environ (en fonction des établissements).

Afin de valider et/ou d'améliorer le système, nous prévoyons de faire tester notre architecture par divers publics : des étudiants de l'université de Bordeaux Montaigne, des élèves en classe d'harmonie du conservatoire de Bordeaux et des élèves de 2^e cycle de formation musicale du conservatoire de Mérignac.

8. REMERCIEMENTS

Nous tenons à remercier Messieurs François PACHET, Gérard ASSAYAG, Jean-Pierre BRIOT et Pierrick LEGRAND, pour leurs précieux conseils et leur aide dans la définition du projet.

Nous souhaitons également remercier Monsieur Pascal PISTONE, Professeur à l'Université de Bordeaux Montaigne en musicologie, pour son co-encadrement du projet.

Et pour finir, nous remercions l'équipe du Studio de Création et de Recherche en Informatique et Musiques Expérimentales (SCRIME) pour le soutien apporté dans ce projet.

9. REFERENCES

1. Di Nunzio A. (Réalisateur) (s. d.-a) Lejaren Hiller - Illiac Suite for String Quartet [1/4]. Consulté 23 juin 2021 : <https://www.youtube.com/watch?v=n0njBFLQSk8>
2. Assayag G., Bloch G., Chemillier M, OMax-Ofon, 2006.
3. Binet A., L'Intelligence artificielle pour la composition musicale [Université Bordeaux Montaigne]. 2022 <https://doi.org/10.13140/RG.2.2.35830.06722>
4. Binet A., analysis harmonies, 2023. https://github.com/doremijazz/analysis_harmonies
5. Challan H. 380 BASSES ET CHANTS DONNES, vol 1A. Consulté le 19 octobre 2022.

6. Cope D., The Algorithmic Composer A-R Editions, Inc, 2000.
7. Czarnecki C. J.S. Bach 413 Chorales: Analyzed. SeeZar Publications, 2013.
8. Eli A. C., « Yousician » - Music Education Platform Information and Communication Technology in Musical Field, 2017.
9. Ferreira D., Haworth B. DeepSolfège: Recognizing Solfège Hand Signs Using Convolutional Neural Networks In G. Bebis, V. Athitsos, T. Yan, M. Lau, F. Li, C. Shi, X. Yuan, C. Mousas, & G. Bruder (Éds.), Advances in Visual Computing Springer International Publishing, 2021.
10. Hadjeres G., Pachet F., Nielsen F. DeepBach: a Steerable Model for Bach Chorales Generation, 2017.
11. Huang C. A., Hawthorne C., Roberts A., Dinculescu M., Wexler J., Hong L., Howcroft J. The Bach Doodle: Approachable music composition with machine learning at scale, 2019.
12. Kiraly Z. Solfeggio 1: A Vertical Ear Training Instruction Assisted by the Computer International Journal of Music Education, 2003.
13. Learning harmonizer – Applications sur Google Play (s. d.-k). Consulté 24 février 2023, à l'adresse <https://play.google.com/store/apps/details?id=com.a.learningharmonizer&hl=fr>
14. Pesek M., Vucko Ž., Šavli P., Kavcic A., Marolt M. Troubadour: A Gamified e-Learning Platform for Ear Training IEEE Access, 2020.
15. Rameau J. P. Traité de l'harmonie réduite à ses principes naturels Slatkine reprints, 1986.
16. Solis, J., Bergamasco, M., Chida, K., Isoda, S., & Takanishi, A. (2004m) The anthropomorphic flutist robot WF-4 teaching flute playing to beginner students, 2004.
17. Thebaut C. Questioning Conservatories Cultural Legitimacy: An Investigation of Dutch and French Networks of Music Education.
18. Wang C. vmo - Python Variable Markov Oracle Library, 2022. <https://github.com/wangsix/vmo>
19. Xenakis I. Formalized Music. Thought and Mathematics in Composition. (Pendragon Press), 1963.
20. Yin, J., Wang, Y., & Hsu, D. Digital violin tutor: An integrated system for beginning violin learners, 2005.

ÉTUDE COMPARATIVE DES ARTICULATIONS DU JEU AU PIANO, AU METAPIANO ET AU MIDIFILEPERFORMER AU MOYEN DES S-LANGAGES

Myriam Desainte-Catherine

Univ. Bordeaux, CNRS, Bordeaux INP,
LaBRI, UMR 5800, F-33400 Talence, France
myriam@labri.fr

Jean Haury, Bernard P.Serpette

SCRIME, Inria, Bordeaux
jeanhaury@gmail.com
Bernard.Serpette@inria.fr

Sylviane R. Schwer

USPN, CNRS, LIPN UMR 7030
F-93430 Villetaneuse, France
schwer@lipn.fr

RÉSUMÉ

Dans cet article, nous modélisons de façon algébrique le fonctionnement du Metapiano et du MidiFilePerformer pour comparer leurs propriétés respectives en terme d'expressivité et de difficulté en faisant référence à celle du piano. Pour ce faire, nous utilisons une même partition pour donner concrètement tous les détails des articulations musicales permises par ces deux systèmes et nous quantifions un certain nombre de propriétés. Cette formalisation permet de montrer notamment que la principale différence entre les deux modèles réside dans le traitement des suspensions qui nécessite un peu d'informations spatiales pour le MidiFilePerformer.

1. INTRODUCTION

Les systèmes musicaux interactifs (SMI) Metapiano [8] et MidiFilePerformer [4] combinent le « hors-temps » d'une « partition interactive », (une séquence numérique de notes de musique), avec le « temps réel » des mouvements d'un musicien sur les touches d'un clavier. Le Metapiano n'utilise que quelques touches d'un clavier numérique quelconque. Il peut être joué par une main seulement, quelques doigts, voire avec un seul. Il intègre en sa mémoire les notes des partitions qu'il exécute sous les commandes expressives de l'interprète sur ses touches. Seules les hauteurs des notes sont codées et mémorisées en une structure ordonnée des relations mélodiques, harmoniques et contrapuntiques entre les notes. L'interprète n'ayant plus à sélectionner les hauteurs sur le clavier, met en sons la structure de hauteurs en jouant sur le nombre restreint de touches. Connaissant « d'oreille » la pièce à jouer, elle ou il en exécute rythmes, tempo, articulations, accents dynamiques et agogiques pour l'interpréter. Le MidiFilePerformer est l'application des principes du Metapiano aux fichiers MIDI, les innombrables partitions numériques téléchargeables en ligne. Ces midifiles sont séquencés pour être joués automatiquement en associant les timings des rythmes et des durées aux informations de hauteur des notes. Le MidiFilePerformer ne prend en compte que la séquence ordonnée « hors-temps » des hauteurs de notes du midifile faisant abstraction de leurs données temporelles. C'est au « performer » de gérer en

« temps réel » sur les touches d'un clavier numérique les rythmes, tempo, accents dynamiques et agogiques pour son interprétation. Les SMI présentés fonctionnent sur des transcriptions numériques d'une partition imprimée de référence. Le Metapiano utilise la structure atemporelle de hauteurs qui se suivent ou se superposent, le MidiFilePerformer agit sur la séquence de hauteurs sans rythmes ni durées du midifile. Dans les deux transcriptions on considère chaque note comme un intervalle temporel borné par un début et une fin ou suspension. La succession des notes met en relation la fin d'une note avec le début de la suivante. Selon le placement de la suspension, l'enchaînement des notes peut être fait selon trois modalités : la fin de note précède le début de la suivante, la fin est synchrone avec le début, la fin succède au début. Dans le premier cas la note s'arrête juste avant la suivante, introduisant un silence entre les deux notes. Dans le deuxième cas il y a contiguïté des notes, l'une se terminant au moment où l'autre commence. Dans le troisième cas la note s'arrête tout de suite après le début de la suivante, les notes se superposent un instant. Ainsi les enchaînements peuvent se faire détachés ou liés selon le mode de jeu de l'interprète sur les touches. Les compositrices et compositeurs classiques ont noté systématiquement en marge des notes de leurs partitions des signes d'articulation pour lier deux notes ou au contraire les détacher l'une de l'autre. Vers 1800, on peut lire dans une méthode de piano anonyme : « les notes liées sont réunies par un signe nommé coulé. . . Lorsqu'il réunit deux notes, la seconde note doit être entendue avant de quitter la première et perdre la moitié de sa valeur en silence. » Il y a donc un court chevauchement des deux notes et un silence d'articulation est introduit avant la prochaine note à venir. On verra dans la suite comment les signes d'articulation conditionnent l'espace de jeu entre suspensions et émissions de notes. Précisons que cet article se limite à l'étude des articulations et ne traite pas des infimes désynchronisations permises par le jeu pianistique car les SMI que nous étudions ici conditionnent la synchronisation exacte.

Dans cet article nous commençons par présenter le modèle algébrique qui permet de formaliser les SMI pour mieux les décrire et les comparer. Nous considérons ensuite les exécutions de ces SMI qui constituent une interprétation d'une partition. Nous convenons dans cet ar-

ticle qu'il s'agit du sous-ensemble des exécutions qui conservent l'ordre des débuts des notes. Cette étude se focalise donc sur la gestion des suspensions des notes et des articulations qu'elle permet à l'interprète de contrôler en temps réel. Pour comparer les SMI, nous introduisons des propriétés d'interactivité, d'expressivité et de difficulté que nous quantifions. Celles-ci sont définies à partir de dénombrements des exécutions des SMI et des commandes de leurs dispositifs en prenant le piano pour référence.

2. MODÈLE ALGÈBRE

Dans cette section, nous introduisons les définitions et les notations qui sont utilisées pour la formalisation.

2.1. Définition d'un système musical interactif

Un SMI (système musical interactif) est un quadruplet $\langle is, \mathcal{I}(s), \mathcal{D}, \mathbf{m}(s) \rangle$ où :

- is est une partition interactive, c'est-à-dire un couple $\langle s, i \rangle$, où s est un squelette de partition et i en est la partie interactive ;
- $\mathcal{I}(s)$ est une fonction associant à la partition s un ensemble d'exécutions qui seront dites *correctes* ;
- \mathcal{D} est le langage du dispositif d'entrée ;
- $\mathbf{m}(s)$ est un mapping associant une exécution de s à un mot de \mathcal{D} .

Tous ces éléments sont définis dans les sous-sections qui suivent.

2.2. Squelette d'une partition

Les événements musicaux provenant d'une partition, tels que nous les étudions, n'ont pas de temps stricts. La seule chose que l'on définit est une relation d'ordre : un événement est avant un autre. Nous nous focalisons sur des événements correspondant au début de quelque chose ou à une fin. Chaque début n'a qu'une et une seule fin qui lui correspond. Inversement, chaque fin n'a qu'un seul début qui lui correspond. Cette unicité peut être garantie par construction en associant les débuts et les fins dans une même structure. On peut remarquer que, dans une partition manuscrite ou imprimée, cette association est déjà présente, car le symbole d'une note définit à la fois son début (position temporelle dans la partition) et sa fin (la durée est explicite, noire, croche...).

Soient deux domaines, D celui des débuts et S celui des suspensions. Nous utiliserons le terme de *suspensions* plutôt que *fins de notes* car celles-ci seront contraintes par un ordre partiel qui conduira à plusieurs positions temporelles possibles. Cette étude porte justement sur les placements des suspensions lors de l'exécution d'une partition. À l'inverse, les débuts de notes seront toujours ordonnés entre eux totalement, exception faite des synchronisations qui sont impossibles à réaliser manuellement.

Le domaine qui nous intéresse est défini comme le produit cartésien des débuts et des suspensions : $D \times S$.

Pour tout domaine T , on définit l'ensemble des s-mots sur T , $SW(T)$, comme étant l'ensemble des listes d'ensembles non-vides d'éléments de T ($SW(T) \subset List(Set(T))$). Si les ensembles peuvent être vides on notera SW^* .

Un squelette de partition p est un s-mot sur $D \times S$: $p \in SW(D \times S)$.



Figure 1. Les cinq premières mesures du Scherzo de la Sonate Op. 26 de Beethoven.

Les S-langages [6] permettent de noter les synchronisations au moyen de s-lettres. Une s-lettre est un ensemble non vide de lettres noté entre parenthèses. Un s-mot est une suite de s-lettres. Un s-langage est un ensemble de s-mots. Le squelette de la partition de la figure 1 est composé de 10 s-lettres synchronisant les débuts des événements verticaux, qu'il s'agisse de notes ou d'accords, et de 22 lettres correspondant aux 22 notes apparaissant au total, toutes occurrences comprises. Les intervalles temporels sont notés au moyen des noms des notes associées. Par exemple $\langle A4, \overline{A4} \rangle$ pour la première note. Voici donc ce squelette dans lequel tous les débuts des notes sont ordonnés selon la partition :

$$sb = (\langle A4, \overline{A4} \rangle) (\langle B4, \overline{B4} \rangle) (\langle F4, \overline{F4} \rangle) (\langle A4, \overline{A4} \rangle) (\langle C5, \overline{C5} \rangle) (\langle D5, \overline{D5} \rangle) (\langle E4, \overline{E4} \rangle) (\langle G4, \overline{G4} \rangle) (\langle E5, \overline{E5} \rangle) (\langle G3, \overline{G3} \rangle) (\langle B4, \overline{B4} \rangle) (\langle A3, \overline{A3} \rangle) (\langle C5, \overline{C5} \rangle) (\langle F3, \overline{F3} \rangle) (\langle A4, \overline{A4} \rangle) (\langle F5, \overline{F5} \rangle) (\langle B3, \overline{B3} \rangle) (\langle A4, \overline{A4} \rangle) (\langle D5, \overline{D5} \rangle) (\langle E3, \overline{E3} \rangle) (\langle G4, \overline{G4} \rangle) (\langle E5, \overline{E5} \rangle)$$

Pour discriminer les occurrences d'une même note dans un s-mot, on indiquera le rang chronologique de l'occurrence chaque fois que ce sera nécessaire. Ainsi, la suite des $A4$ sera notée de la façon suivante : $[A4, 1][\overline{A4}, 1][A4, 2][\overline{A4}, 2][A4, 3][\overline{A4}, 3][A4, 4][\overline{A4}, 4]$ ou bien s'il n'y a pas d'ambiguïté, $A4\overline{A4}A4\overline{A4}A4\overline{A4}A4\overline{A4}$.

2.3. Exécution d'une partition

Pour représenter l'exécution d'une partition, nous considérons uniquement l'ordre des événements, et non les durées. Ainsi, une exécution est représentée par un s-mot sur les lettres représentant les événements de début et de suspension des notes. Pour exprimer ces s-mots, nous introduisons deux opérations, le développement qui permet de séparer les événements de début et de suspension d'une même note, et l'étirement qui permet de déplacer les suspensions vers la droite selon l'ordre temporel indiqué sur la partition.

2.3.1. Développement d'un squelette

Le développement dvp d'un squelette est une fonction $(SW(D \times S) \rightarrow SW^*(D + S))$ définie par itération sur

une fonction $\text{sq} (Set(D \times S) \rightarrow Set(D + S))$. Un développement consiste à déstructurer les éléments du produit cartésien $D \times S$ et à introduire, une fois sur deux, une s-lettre vide (d'où le type SW^*) qui ultérieurement, par étirement, recevra éventuellement des suspensions. Pour bien identifier cette s-lettre vide, nous la noterons $\bar{()}$.

$$\begin{aligned} \text{dvp}(\bar{()}) &= () \\ \text{dvp}(s1 \cdot \bar{sw}) &= \text{sq}(s1) \cdot \bar{()}\text{dvp}(sw) \\ \text{sq}(\emptyset) &= \emptyset \\ \text{sq}(\langle d, s \rangle \cdot sw) &= d \cdot s \cdot \text{sq}(sw) \end{aligned}$$

Les situations temporelles des notes du squelette de la partition sont représentées sur la figure 2. Sur le diagramme de gauche, les notes sont indiquées par un point rassemblant leur début et leur fin. Ces points sont appelés des *granules temporelles* et seront utilisés lors de la présentation du fonctionnement du Metapiano 3.1.3. Nous nous intéressons particulièrement à l'ordre des débuts car il sera conservé dans toutes les exécutions que nous considérons comme étant des interprétations de la partition. Sur la figure, les granules sont positionnées selon leur début. Les lignes verticales réunissent les événements synchronisés mais les durées ne sont pas représentées. En effet, les colonnes ne sont pas de durées égales.

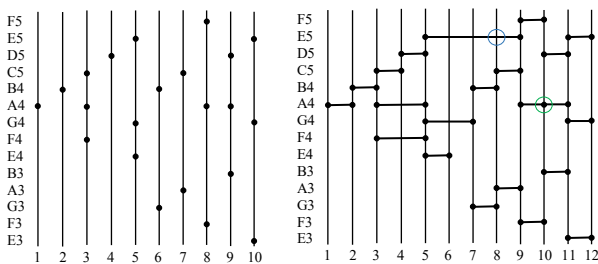


Figure 2. À gauche, les situations temporelles relatives des notes de la partition vues comme granules temporelles et à droite le développement avec continuité et contiguïté.

Propriétés du développement : Le développement respecte les quatre propriétés suivantes : (1) le contenu (2) l'ordonnement des débuts (3) le fait que les suspensions sont toujours après leurs débuts respectifs (4) une alternance de s-lettres contenant au moins un début et de s-lettres n'en contenant pas.

Le développement du squelette de notre partition est donc le suivant :

$$\text{dvp}(sb) = (A4\bar{A4})\bar{()}(B4\bar{B4})\bar{()}(F4\bar{F4}A4\bar{A4}C5\bar{C5})\bar{()}(D5\bar{D5})\bar{()}(E4\bar{E4}G4\bar{G4}E5\bar{E5})\bar{()}(G3\bar{G3}B4\bar{B4})\bar{()}(A3\bar{A3}C5\bar{C5})\bar{()}(F3\bar{F3}A4\bar{A4}F5\bar{F5})\bar{()}(B3\bar{B3}A4\bar{A4}D5\bar{D5})\bar{()}(E3\bar{E3}G4\bar{G4}E5\bar{E5})\bar{()}$$

2.3.2. Représentation de l'ordre induit par le rythme

Il y a essentiellement deux façons de considérer le temps : soit comme une relation entre deux instances de l'univers de discours, qui les situent l'un par rapport à l'autre, soit comme propriété d'une instance, qui indique sa durée. Dans le premier cas, nous parlerons du temps événementiel, dans le second cas, du temps durée, que

cette durée soit objective ou subjective, absolue ou relative. Nous nous intéressons dans ce travail au temps événementiel, qui correspond dans le domaine musical aux articulations entre les émissions des notes, qui sont nos objets de discours.

Aristote (Physique, V, 3) a fait une étude des différents types de relations temporelles existant entre des événements duratifs considérés comme objets primitifs et non comme des intervalles de la droite réelle constituée de points sans dimension. Il a pu ainsi distinguer deux relations qui n'existent pas dans l'ensemble des nombres réels mais indispensable dans le domaine musical : *contiguïté* versus *continuité*, suivant que deux événements en contact – l'un se terminant au moment où l'autre commence – partagent une borne commune tout en conservant leur individualité ou bien se fondent en un nouvel événement qui est leur somme.

Les troisième et quatrième occurrences de A4 dans la partition sont contiguës. Elles ont une borne commune mise en évidence par un cercle sur le schéma de droite de la figure 2. Cela signifie qu'il n'est pas possible d'insérer entre elles un événement de même type, aussi petit soit-il.

La liaison des deux premières occurrences de E5 dans la partition les rend continues, elles sont représentées par un seul segment sur le schéma de droite de la figure 2 où un cercle entoure l'absence du point.

2.3.3. Étirement du développement d'un squelette

Pour transformer notre squelette de façon à représenter l'ordre induit par une exécution, nous introduisons une opération d'étirement des suspensions vers la droite.

Un étirement est une réécriture (\rightarrow_e) dans $SW^*(D + S)$ qui déplace une suspension vers la droite :

$$s \in S, (\dots (s \cdot w) w' \dots) \rightarrow_e (\dots w (s \cdot w') \dots)$$

L'étirement préserve les quatre propriétés du développement. La relation d'étirement est confluente et fortement normalisante.

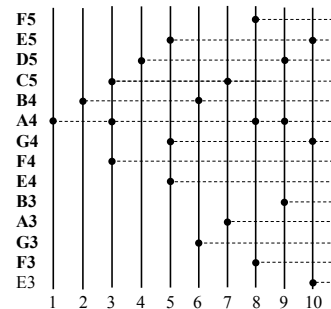


Figure 3. Situations temporelles relatives des notes du squelette de la partition avec étirements.

Pour coder l'ordre induit par l'exécution de la partition, il s'agit d'étirer le développement du squelette pour amener les suspensions à leur place dans l'ordre chronologique des événements. Voici l'exemple correspondant à la

figure 3 : $(A4)(\overline{B4})(\overline{A4})(F4A4C5)(\overline{D5})(\overline{E4G4E5})(\overline{B4})(G3B4)(\overline{C5})(A3C5)(\overline{A4})(F3A4F5)(\overline{A4D5})(B3A4D5)(\overline{E5G4})(E3G4E5)(\overline{E3G4E5F4E4B4G3B4A3C5F3F5B3})$

Considérons un autre exemple que nous noterons *sbr* et qui est obtenu par fermeture transitive de l'étirement appliqué à *dvP*(*sb*) et que nous allons utiliser par la suite :

$sbr = (A4)(\overline{A4})(B4)(\overline{B4})(F4A4C5)(\overline{C5})(D5)(\overline{D5F4A4})(E4G4E5)(\overline{E4})(\overline{G4G3B4})(\overline{G3B4})(A3C5)(\overline{A3C5E5})(F3A4F5)(\overline{F3A4F5})(B3A4D5)(\overline{B3A4D5})(E3G4E5)(\overline{E3G4E5})$

Les situations temporelles relatives des notes conformément à ce mot sont représentées sur le schéma de droite de la figure 2.

La composition d'un développement et de la fermeture transitive des étirements définit une fonction des squelettes $SW(D \times S)$ dans un ensemble de $SW^*(D + S)$. A partir d'un squelette *sb* comportant *n* s-lettres s_i , l'ensemble des étirements possibles du développement a un cardinal de $\prod_{i=0}^{i=n-1} (2(n-i) - 1) * ||s_i||$

où $||s||$ représente le cardinal de l'ensemble *s*.

Appliqué à notre partition, l'ensemble des étirements a un cardinal de $19 \times 1 \times 18 \times 1 \times 17 \times 3 \dots \times 3 \times 3 \times 3 \times 1 \simeq 3,25 \times 10^{14}$.

Notons que l'étirement a conduit à inclure la lettre $\overline{G4}$ dans la s-lettre (*G3B4*) pour assurer l'alternance des s-lettres contenant des débuts et celles n'en contenant pas et pour conserver le nombre de lettres. On verra plus loin que le relâchement de la commande permettra de suspendre la note *E4* pour interpréter le silence qui la suit, alors que la suspension de la note *G4* sera automatique conduisant à un *esatto legato* avec les notes suivantes.

2.3.4. Espace d'exécution

Une exécution d'une partition est dite *correcte* si elle respecte les consignes indiquées sur la partition. L'ensemble des exécutions correctes est un ensemble de s-mots qui est donné avec la partition et noté $\mathcal{I}(s)$. On parlera aussi d'*espace d'exécution*. La figure 4 représente la partition comportant les consignes d'articulation pour l'exécution des cinq premières mesures du Scherzo de Beethoven.



Figure 4. Les cinq premières mesures du Scherzo de Beethoven restreintes aux consignes d'exécution du legato et staccato.

Les consignes pour l'exécution concernent la rythmique, les articulations et la dynamique. Concernant les consignes rythmiques, les durées seront laissées à l'appréciation de l'interprète et notre codage ne prend en compte

que l'ordre temporel des suspensions induit par les durées. Quant aux articulations, elle seront codées par des relations temporelles d'intervalle.

2.3.5. Représentation de l'ordre temporel induit par les consignes d'articulations

Musicalement la contiguïté de deux notes différentes conduit à deux sortes de consignes d'articulations pour l'exécution de la partition : *legato* ou *staccato*. Le *legato* cherche à lier le plus possible les notes qui se suivent, le *staccato* au contraire les détache plus ou moins les unes des autres. Le *legato* consiste donc à lier les sons successifs de sorte qu'il n'y ait pas de silence entre eux. Le *staccato*, au contraire, est obtenu en introduisant des pauses minimales entre les notes. Les deux techniques permettent d'articuler la phrase musicale. Dans le jeu *legato*, les instruments à cordes frottées, les vents et les voix réalisent une continuité entre les notes successives dans un « *esatto legato* ». Les instruments polyphoniques à touches quant à eux, pouvant superposer les sons, jouent *legato* en réalisant un chevauchement très court des sons au passage d'une note à la suivante. Dans le jeu *staccato*, pour tous les instruments, les notes jouées sont détachées les unes des autres en perdant une partie de leur valeur pour un silence équivalent.

Ces considérations sur les articulations se rapportent aux travaux de [9, 10]. Elles se fondent sur les positions relatives entre intervalles, développées dans le cadre des algèbres relationnelles initiées au début du XX^e siècle par [12], développées par [11] et popularisées dans le domaine du raisonnement temporel en intelligence artificielle par [1] dont nous reprenons la terminologie : *before*, *meet*, *overlap*, *start*, *during*, *end*, *equal*. Selon cette terminologie, deux notes liées au clavier par un *legato* sont en relation *overlap*, deux notes liées par un *esatto legato* sont en relation *meet* et deux notes articulées par un *staccato* sont en relation *before*. Les notes dont les débuts sont parfaitement synchronisés sont en relation *start*, les notes dont les suspensions sont synchronisées en relation *end*. Enfin, il faut noter que les synchronisations exactes qui sont codées dans le s-mot *sbr* obtenu par développement et étirements du squelette pour représenter le rythme de la partition dans le paragraphe 2.3.3 ne sont pas des consignes pouvant être respectées exactement par les interprètes. En pratique, les doigts du pianiste exécutent bien souvent d'infimes désynchronisations privilégiant les notes graves sur les aiguës par exemple. Nous considérerons dans la suite, que les débuts des notes qui sont synchronisés admettent ce seul ordre.

Sur la figure 4, on ne considère que les signes d'articulation situés au-dessus des notes, liaisons ou points. Les liaisons ou coulés indiquent que les notes sont jouées sans séparation *legato* ; les points, au contraire, indiquent que les notes sont jouées détachées les unes des autres *staccato*, car séparées par des silences. Pour expliquer le codage, procédons mesure par mesure.

Mes. 1 $(A4)(B4)(\overline{A4})(\overline{B4})$, les deux notes A4 et B4 sont liées par un coulé. Pour lier deux notes successives au cla-

vier, il faut relâcher la première touche après avoir enfoncé la deuxième. Donc, la s-lettre $(\overline{A4})$ doit apparaître après la s-lettre $(B4)$. L'exécution correcte du coulé s'écrit : $(A4)(B4)(\overline{A4})(\overline{B4})$. Lors du passage de A4 à B4, il y a chevauchement des sons créant le *legato*, mais également une dissonance entre les sons que l'interprète évalue et minimise par un *overlap* court.

Mes. 2 $(F4A4C5)(D5)(\overline{C5})(\overline{D5})$, l'accord $(F4A4C5)$ comporte trois touches à jouer le plus synchroniquement possible. Puis D5 est jouée seule $(F4A4C5)(D5)$. C5 et D5 sont liées par un coulé et sont exécutées comme précédemment en *overlap* : $(F4A4C5)(D5)(\overline{C5})(\overline{D5})$. F4 et A4 durent la mesure entière et sont liées aux notes du début de la 3ème mesure. A4 n'a pas de signe de coulé comme F4, mais de façon implicite on lui en attribue un.

Mes. 3 $(E4G4E5)(\overline{F4A4})(\overline{E4})(G3B4)(\overline{G4})(\overline{G3})(\overline{B4})$, l'accord $(E4G4E5)$ comporte trois nouvelles touches à jouer le plus synchroniquement, et est suivi de la suspension en *overlap* $(F4A4)$, des deux notes tenues de la 2^e mesure. La touche E4 est ensuite relâchée pour faire entendre le silence du 2^e temps $(\overline{E4})$. Au 3^e temps, le pianiste joue le plus synchroniquement deux touches, $(G3B4)$, il relâche rapidement $(\overline{G4})$ la touche précédemment jouée en début de mesure pour exécuter, en *overlap*, le *legato* entre les deux notes G4 et B4, puis il relâche $(\overline{G3})$ pour exécuter le *staccato* de cette note, et enfin relâche $(\overline{B4})$ le plus tard possible avant la fin de la mesure pour ménager un bref silence d'articulation avant le début de la 4^e mesure. Pendant ce 3ème temps de la 3^e mesure, 3 relâchements de touches ordonnés et non synchrones se succèdent : $(\overline{G4})(\overline{G3})(\overline{B4})$. Le pianiste a toute liberté de s'écarter des consignes d'exécution, il peut étirer les relâchements différemment et les exécuter dans un autre ordre.

Mes. 4 $(A3C5)(\overline{A3})(F3A4F5)(\overline{C5E5})(\overline{F3})(\overline{A4F5})(B3A4D5)(\overline{B3A4D5})$, deux types d'articulations sont confrontées : points de *staccato* à la main gauche, coulés de *legato* à la main droite. Les touches $(A3C5)$ sont jouées (la note E5 n'est pas rejouée, elle est maintenue depuis le début de la 3^e mesure par la liaison de prolongation qui met les deux E5 en continuité). La touche A3 est relâchée seule $(\overline{A3})$ pour exécuter le point de *staccato*. Puis, trois nouvelles touches sont jouées $(B3A4F5)$, avec un point de *staccato* pour la main gauche et des fins de coulés *legato* pour la main droite. Comme constaté précédemment dans la 3^e mesure, 3 moments de relâchements de touches ordonnés et non synchrones se succèdent : $(\overline{C5E5})(\overline{F3})(\overline{A4F5})$, le premier pour clore les *overlaps* de *legato*, le deuxième pour exécuter la note *staccato*, le dernier pour ménager un prompt silence d'articulation avant l'accord qui va venir. Ici les trois touches attaquées simultanément $(B3A4D5)$ et relâchées *staccato* également simultanément $(\overline{B3A4D5})$ sont en relation harmonique et en relation d'intervalles *equal*.

Mes. 5 $(E3G4E5)(\overline{E3G4E5})$, cet accord est exécuté de la même façon. Le pianiste a cependant toute liberté de s'écarter des consignes d'exécution, il peut étirer les relâchements différemment et les exécuter dans un autre ordre.

Réunissons les s-mots successifs établis pour chaque mesure. Le langage décrivant l'espace d'exécution de la partition, c'est-à-dire l'ensemble des exécutions correctes selon les consignes écrites et les mouvements des touches sous les doigts du pianiste est défini de la façon suivante :

$$\mathcal{I}(\text{sb}) = (A4)(B4)(\overline{A4})(\overline{B4})(F4A4C5)(D5)(\overline{C5})(\overline{D5})(E4G4E5)(\overline{F4A4})(\overline{E4})(G3B4)(\overline{G4})(\overline{G3})(\overline{B4})(A3C5)(\overline{A3})(F3A4F5)(\overline{C5E5})(\overline{F3})(\overline{A4F5})(B3A4D5)(\overline{B3A4D5})(E3G4E5)(\overline{E3G4E5})$$

Ce s-mot représente l'exécution idéale respectant toutes les consignes d'articulation indiquées sur la partition. Si nous avons considéré que les synchronisations pouvaient donner lieu à toutes les permutations possibles, nous aurions obtenu un ensemble dont le cardinal peut être calculé par la formule suivante :

$$1!1!1!1!3!1!1!1!1!3!2!1!1!2!1!1!1!2!1!1!1!3!2!1!2!3!3!3! = 8\,957\,952$$

Mais ce nombre n'a aucun sens en pratique car il ne représente pas vraiment le nombre d'exécutions qu'un interprète peut exécuter sur un piano. Aussi, dans la suite, nous considérerons que l'ensemble $\mathcal{I}(\text{sb})$ est réduit au seul s-mot donné ci-dessus, et que ce s-mot lors de l'exécution au piano donne lieu à un seul mot, induit par un seul ordre temporel pour les synchronisations comme nous l'avons mentionné dans le paragraphe 2.3.5. Nous admettrons donc que $|\mathcal{I}(\text{sb})| = 1$.

2.3.6. Dénombrement des exécutions possibles

Nous nous situons dans le cas de l'interprétation et non de l'improvisation, ainsi nous ne considérons que des exécutions de la partition qui ne modifient pas l'ordre de la séquence des débuts des notes. Les autres exécutions ne sont pas considérées comme des exécutions de cette partition. Nous nous intéressons donc à l'ensemble de toutes les permutations des débuts et des suspensions, qui respectent l'ordre des débuts. Nous proposons deux façons de définir cet ensemble.

Une première proposition consiste à considérer la fermeture transitive du développement du squelette de la partition pour l'opération d'étirement. Les situations temporelles relatives des notes sont représentées sur la figure 2. Cela revient à considérer qu'il est impossible d'insérer des suspensions au milieu d'événements qui sont synchronisés. Ce dénombrement a été réalisé dans le paragraphe 2.3.3 :

$$|\mathcal{A}^*(\text{sb})| = 19 \times 1 \times 18 \times 1 \times 17 \times 3 \dots \times 3 \times 3 \times 3 \times 1 \simeq 3,25 \times 10^{14}$$

Une deuxième proposition consiste à considérer qu'au piano, il est possible d'insérer des suspensions à l'intérieur des synchronisations et que donc les s-lettres contenant des débuts peuvent elles aussi être cassées en plusieurs s-lettres. Ces objets s'apparentent à des involutions sans points fixes qui peuvent être dénombrées par la formule suivante où n est la demi-longueur du mot : $(2n-1)(2n-3)\dots \times 5 \times 3 \times 1$. Comme la partition comporte 22 notes, nous avons 44 événements (22 débuts et 22 suspensions) :

$$|\mathcal{A}^*(\text{sb})| = 43 \times 41 \times 39 \times \dots \times 5 \times 3 \times 1 \simeq 5,6 \times 10^{26}$$

Il conviendrait en toute rigueur d'enlever les s -mots correspondant aux suspensions d'une occurrence de note qui passent au delà des occurrences suivantes. Car le pianiste doit relâcher une touche pour pouvoir l'enfoncer à nouveau. Le nombre de possibilités est alors compris entre $1,3 \times 10^{18}$ et $7,3 \times 10^{22}$. Mais nous nous dispenserons de faire ce calcul, car le résultat serait du même ordre de grandeur.

2.4. Partition interactive

Soit s une partition, sq son squelette et soit sra un s -mot obtenu par développement et étirements successifs à partir de sq pour correspondre à l'ordre induit par le rythme de la partition s et au contrôle des articulations souhaité. Une partition interactive est un coloriage de sra c'est-à-dire un couple :

$$is = \langle sra, i \rangle$$

où i est un s -mot inclus dans sra ¹. Le s -mot i est dit *s-mot d'interaction* et les lettres de i sont appelées des *points d'interaction*. Ces lettres seront associées à un dispositif de commande pour activer des événements de la partition.

On définit le degré d'interactivité d'une partition interactive is par la quantité suivante :

$$di(is) = \frac{|i|}{|s|}$$

c'est-à-dire le nombre de points d'interaction divisé par le nombre total de lettres de s . Ce nombre correspond à la quantité d'actions que l'interprète doit effectuer pour jouer la partition. Ainsi un degré d'interactivité nul indique qu'il n'y a pas de points d'interaction, le système joue tout seul, alors qu'un degré d'interactivité de 1 indique que tous les événements de la partition, qu'ils soient des débuts ou des fins, sont interactifs et peuvent être activés par l'interprète.

2.5. Langage du dispositif d'entrée

On considère dans cet article des dispositifs tels que des boutons ou des touches de clavier qui vont permettre d'activer les points d'interaction. Deux actions sont possibles sur ces dispositifs, l'enfoncement et le relâchement. Ainsi pour décrire le langage d'une touche, on associe une lettre x pour l'enfoncement et \bar{x} pour le relâchement.

2.6. Mapping

Soit $is = \langle s, i \rangle$ une partition interactive obtenue à partir d'un squelette sq et soit \mathcal{D} le langage du dispositif d'entrée, le mapping précise la correspondance entre un mot du dispositif d'entrée et une exécution de la partition au moyen de la partition interactive. On introduit les définitions suivantes.

¹. i est inclus dans sra s'il est obtenu en supprimant des lettres de sra .

2.6.1. Commande

Une commande est un mot coloré du langage du dispositif d'entrée \mathcal{D} . En effet, la correspondance entre le mot du dispositif d'entrée et le s -mot d'interaction nécessite parfois de considérer que certaines lettres du mot du dispositif sont inactives.

2.6.2. Mot de commande

Soit $\langle w, wi \rangle$ une commande, le mot wi est appelé mot de commande.

2.6.3. Mapping de commande

Un mapping de commande est une fonction totale associant une exécution de $\mathcal{I}(s)$ à une commande : $\mathbf{m}(sq) : \mathcal{C}(s) \rightarrow \mathcal{I}(s)$, où $\mathcal{C}(s)$ est inclus dans l'ensemble des commandes et représente le domaine de $\mathbf{m}(sq)$. On notera $\mathcal{I}(\mathbf{m}(sq))$ le codomaine de $\mathbf{m}(sq)$, c'est-à-dire l'ensemble des exécutions que les commandes permettent d'exécuter.

Rappelons que les mots w et wi sont des séquences de lettres correspondant à des enfoncements ou des relâchements de touches. Chaque lettre du mot de commande wi est associée à un point d'interaction de i pour activer les événements associés dans la partition.

Remarque. Pour un squelette sq , la conception de la fonction $\mathbf{m}(sq)$ nécessite de faire plusieurs choix. On remarque que pour des raisons d'ergonomie musicale, il est préférable d'associer des enfoncements de touches à des débuts de notes plutôt qu'à des fins de notes. De plus, dans l'association des lettres de wi avec celles de i , le respect de l'ordre chronologique est préféré pour permettre le jeu en temps réel.

3. ESPACES D'EXÉCUTION DU METAPIANO ET DU MIDIFILEPERFORMER

Pour présenter les deux espaces, nous allons utiliser le même squelette sb (donné au paragraphe 2.2) de la partition qui est représentée sur la figure 4. Mais il faut préciser qu'en pratique, le MidifilePerformer prend en entrée un fichier MIDI pour l'analyser automatiquement. Le résultat dépend donc de la façon dont le fichier MIDI code la partition. Le plus souvent dans les fichiers MIDI les notes sont synchronisées au moyen d'un séquenceur ce qui correspond au codage de notre étude. Nous supposons donc que le MidifilePerformer a en entrée un fichier MIDI qui le conduit au squelette sb .

3.1. Partition interactive

Nous allons considérer la même partition interactive pour les deux systèmes pour mieux les comparer. Précisons toutefois que la partition interactive du Metapiano est obtenue par analyse musicale, avec une séparation en voix, et des choix manuels pour pouvoir réaliser les articulations prescrites. Alors que celle du MidifilePerformer est

obtenue par un algorithme [4] appliqué à un fichier MIDI. On supposera ici à nouveau que le fichier MIDI conduit au même résultat que l'analyse manuelle du Metapiano et donc à la même partition interactive qui est décrite ci-dessous.

3.1.1. Coloriage en gras des lettres de débuts de notes

Remarquons que les débuts de notes qui sont dans une même s-lettre de sb sont synchronisés. Ainsi il suffit de colorer un seul début dans chaque s-lettre pour pouvoir activer l'ensemble des débuts qui y apparaissent. Considérons le coloriage suivant des lettres de début dans le squelette sb :

$$((\mathbf{A4}, \overline{A4})(\mathbf{B4}, \overline{B4})(F4, \overline{F4})(A4, \overline{A4})(\mathbf{C5}, \overline{C5})(\mathbf{D5}, \overline{D5}))$$

$$((\mathbf{E4}, \overline{E4})(G4, \overline{G4})(E5, \overline{E5}))((G3, \overline{G3})(\mathbf{B4}, \overline{B4})(A3, \overline{A3})(\mathbf{C5}, \overline{C5}))$$

$$((F3, \overline{F3})(\mathbf{A4}, \overline{A4})(F5, \overline{F5}))((B3, \overline{B3})(A4, \overline{A4})(\mathbf{D5}, \overline{D5}))((E3, \overline{E3})$$

$$(G4, \overline{G4})(\mathbf{E5}, \overline{E5})).$$

Concernant les suspensions, il conviendra de colorer une suspension par s-lettre pour la même raison, mais pour les déterminer, il faut se référer au rythme et aux articulations de la partition. Cette opération sera présentée en dernier directement sur le mot sra.

3.1.2. Codage des articulations liées

Pour construire le s-mot sra qui va nous servir de référence pour définir la partition interactive, il faut se référer au s-mot sbr (voir paragraphe 2.3.1) qui donne déjà les synchronisations des suspensions selon le rythme de la partition. Considérons maintenant les articulations qui sont indiquées sur la partition. Celles-ci doivent pouvoir être exécutées par l'interprète. Ainsi, en raison de la liaison *legato* avec $E4$, le $\overline{F4}$ de la 2^e mesure ne doit pas être synchronisé avec la s-lettre $(\overline{D5A4})$. Il est alors déplacé dans la s-lettre suivante ($E4G4E5$). Ainsi, l'ordre entre les suspensions en question est ici modifié pour pouvoir articuler la liaison.

De plus, dans la 4^e mesure nous avons deux liaisons *legato* simultanées, l'une de continuité entre le $E5$ commençant à la 3^e mesure et le $F5$ de la 4^e mesure, et l'autre entre $C5$ et $A4$ tous les deux dans la 4^e mesure. Ces deux liaisons ne sont pas exécutées de la même façon, la première étant sur un temps beaucoup plus long que l'autre. Seule la liaison entre $C5$ et $A4$ sera exécutée par l'interprète et l'autre sera fixée. En conséquence, la lettre $\overline{E5}$ est déplacée dans la s-lettre contenant $F5$. En appliquant les opérations d'étirement adéquates, on obtient le mot suivant :

$$(A4)(\overline{A4})(B4)(\overline{B4})(F4A4C5)(\overline{C5})(D5)(\overline{D5A4})(\overline{F4E4G4E5})(\overline{E4})$$

$$(\overline{G4G3B4})(\overline{G3B4})(A3C5)(\overline{A3C5})(\overline{E5F3A4F5})(\overline{F3A4F5})(B3A4$$

$$D5)(\overline{B3A4D5})(E3G4E5)(\overline{E3G4E5}).$$

3.1.3. Codage des notes piquées

Enfin, le s-mot d'interaction doit permettre aussi d'exécuter correctement les notes piquées, qui n'ont pas été prises en compte jusqu'ici. Notamment il s'agit des notes piquées qui sont synchronisées avec des notes non piquées, telles que le $G3$ de la 3^e mesure, le $A3$ et le $F3$

de la 4^e mesure. Les suspensions de ces notes ne doivent pas être synchronisées avec les noires qui sont jouées en même temps mais doivent se produire avant. Comme il est impossible de conserver l'alternance des s-lettres, celles contenant des débuts et celles n'en contenant pas, en ajoutant pour ces trois notes une suspension seule dans une s-lettre, on utilise ici la notion de *granule temporelle* introduite au paragraphe 2.3.1 pour les représenter. La granule est ponctuelle et contient en elle-même l'information de suspension de la note dans le futur à une distance temporelle déterminée à l'avance. Les granules sont notées en ajoutant le symbole $^\circ$ à la fin de la lettre. Il convient alors de retirer les lettres correspondant aux suspensions des notes représentées par des granules. On obtient alors :

$$\mathbf{sra} = (A4)(\overline{A4})(B4)(\overline{B4})(F4A4C5)(\overline{C5})(D5)(\overline{D5A4})(\overline{F4E4G4E5})$$

$$E5)(\overline{E4})(\overline{G4G3^\circ B4})(\overline{B4})(A3^\circ C5)(\overline{C5})(\overline{E5F3^\circ A4F5})(\overline{A4F5})$$

$$(B3A4D5)(\overline{B3A4D5})(E3G4E5)(\overline{E3G4E5}).$$

3.1.4. Codage de la partition interactive

Il ne reste plus qu'à colorier le s-mot sra décrit ci-dessus pour obtenir la partition interactive. Il convient de se conformer au coloriage des lettres de début proposé au paragraphe 3.1.1 et de colorier les lettres de suspension à raison d'une par s-lettre ne contenant pas déjà une lettre de début colorée. Pour mieux s'y retrouver, on choisit à chaque fois de colorier une lettre de suspension correspondant à une lettre de début colorée. Ainsi, on obtient la partition interactive suivante :

$$is = (\mathbf{A4})(\overline{A4})(\mathbf{B4})(\overline{B4})(F4A4C5)(\overline{C5})(\mathbf{D5})(\overline{D5A4})(\overline{F4E4G4E5})$$

$$(\overline{E4})(\overline{G4G3^\circ B4})(\overline{B4})(A3^\circ C5)(\overline{C5})(\overline{E5F3^\circ A4F5})(\overline{A4F5})(B3A4$$

$$D5)(\overline{B3A4D5})(E3G4E5)(\overline{E3G4E5}).$$

3.2. Dispositif d'entrée

Pour pouvoir comparer le Metapiano et le Midifile-Performer, nous allons leur associer le même dispositif d'entrée. Il s'agit d'un clavier numérique quelconque pour jouer avec les cinq doigts d'une main.

3.2.1. Cas du Metapiano

Nous considérerons dans cet article que les touches du Metapiano sont génériques. En effet, parmi les informations provenant du dispositif d'entrée, seules les informations de vélocité d'enfoncement de touche et de vélocité de relâchement de touche sont prises en compte, la vélocité donnant la dynamique aux notes qui sont déclenchées ou relâchées. Or nous ne considérons pas ces informations dans cet article. Ainsi, le choix de la touche enfoncée ou relâchée ne change pas le résultat sonore, toutes les touches ont le même effet. Notons \mathcal{D}_{M_p} le langage du dispositif, il s'agit de l'ensemble des mots de Dyck de hauteur maximale 5. Pour rappel, un mot de Dyck w a les propriétés suivantes :

- Tous les préfixes de w ont un nombre de lettres \bar{x} inférieur ou égal au nombre de lettres x .
- Le mot w a autant de lettres x que de lettres \bar{x} .

3.2.2. Cas du MidifilePerformer

Les touches du clavier du MidifilePerformer ne sont pas génériques car il faut mémoriser la correspondance entre les touches enfoncées et les notes de la partition de façon à stopper la bonne note. Les lettres associées aux touches enfoncées sont alors numérotées. Le langage d'une touche est défini par l'expression rationnelle suivante :

$$T_x^1 = (x\bar{x})^*$$

Dans le cas de 5 doigts, le langage du dispositif est un shuffle des mots de T_x^1 :

$$\mathcal{D}_{M_f} = \sqcup_{i=1}^5 T_{x_i}^1$$

3.3. Mapping de commandes

Le mapping définit l'effet sur la partition des actions sur le clavier en associant un mot de commande au s-mot d'interaction de la partition interactive.

Considérons une partition du s-mot is qui est la partition interactive considérée au paragraphe 3.1.4 selon que le point d'interaction est un début ou bien une suspension. Notons a_i les s-lettres dont le point d'interaction est un début et \bar{a}_i celles dont le point d'interaction est une suspension :

$a_1 = ([\mathbf{A4}, 1])$	$\bar{a}_1 = ([\bar{\mathbf{A4}}, 1])$
$a_2 = ([\mathbf{B4}, 1])$	$\bar{a}_2 = ([\bar{\mathbf{B4}}, 1])$
$a_3 = ([F4, 1][A4, 2][\mathbf{C5}, 1])$	$\bar{a}_3 = ([\bar{\mathbf{C5}}, 1])$
$a_4 = ([\mathbf{D5}, 1])$	$\bar{a}_4 = ([\bar{\mathbf{D5}}, 1][\bar{A4}, 2])$
$a_5 = ([F4, 1][\mathbf{E4}, 1][G4, 1][E5, 1])$	$\bar{a}_5 = ([\bar{\mathbf{E4}}, 1])$
$a_6 = ([G4, 1][G3^\circ, 1][\mathbf{B4}, 2])$	$\bar{a}_6 = ([\bar{\mathbf{B4}}, 1])$
$a_7 = ([A3^\circ, 1][\mathbf{C5}, 2])$	$\bar{a}_7 = ([\bar{\mathbf{C5}}, 1])$
$a_8 = ([E5, 1][F3^\circ, 1][\mathbf{A4}, 3][F5, 1])$	$\bar{a}_8 = ([\bar{\mathbf{A4}}, 3][\bar{F5}, 1])$
$a_9 = ([B3, 1][A4, 4][\mathbf{D5}, 2])$	$\bar{a}_9 = ([\bar{\mathbf{B3}}, 1][\bar{A4}, 4][\bar{\mathbf{D5}}, 2])$
$a_{10} = ([E3, 1][G4, 2][\mathbf{E5}, 2])$	$\bar{a}_{10} = ([\bar{\mathbf{E3}}, 1][\bar{G4}, 4][\bar{\mathbf{E5}}, 2])$

Le principe est de conserver lors de l'exécution l'ordre des s-lettres correspondant aux a_i entre elles ainsi que l'ordre induit par le fait que les débuts de notes sont activés avant leur suspension. Cette dernière propriété est impliquée par les touches du dispositif utilisé qui nécessitent d'être enfoncées avant de pouvoir être relâchées.

Certains mots produits par l'interprète sur le dispositif d'entrée peuvent mener à une exécution erronée selon le synthétiseur utilisé. En effet, un mot commençant par : $xx\bar{x}\bar{x}\dots$ va activer les s-lettres suivantes : $a_1 a_2 a_3 \bar{a}_1 \dots$. Les lettres a_1 et a_3 comportant chacune un début de la même note A_4 , le mot va conduire à réactiver cette note alors qu'elle est déjà en train de jouer, ce qui peut conduire à une erreur au moment du relâchement. Ainsi, pour éviter d'avoir à prendre en compte ce problème, nous nous plaçons dans le cas où les occurrences de notes sont différenciées.

3.3.1. Cas du Metapiano

Le principe du mapping du Metapiano est de fonctionner avec des touches génériques et de donner à l'interprète la possibilité d'exécuter les articulations. Le mapping du

Metapiano préserve l'ordre des s-lettres \bar{a}_i lors de l'exécution. Le mapping du Metapiano est effectué manuellement.

Examinons maintenant les mots de commandes. Ceux-ci sont des mots de Dyck de hauteur 5 maximum de longueur 20. Voici un exemple de mot de commande :

$$w = xx\bar{x}\bar{x}xx\bar{x}\bar{x}\bar{x}\bar{x}\bar{x}\bar{x}\bar{x}\bar{x}\bar{x}\bar{x}\bar{x}\bar{x}\bar{x}\bar{x}\bar{x}\bar{x}\bar{x}\bar{x}\bar{x}$$

Voici le déroulé complet du jeu de la partition avec le mot w , en notant en gras les points d'interaction :

$$\begin{aligned} x &\rightarrow \mathbf{A4}, x \rightarrow \mathbf{B4}, \bar{x} \rightarrow \bar{\mathbf{A4}}, \bar{x} \rightarrow \bar{\mathbf{B4}}, x \rightarrow \mathbf{C5}F4A4, x \rightarrow \mathbf{D5}, \bar{x} \rightarrow \bar{\mathbf{C5}}, \bar{x} \rightarrow \bar{\mathbf{D5}}A4, x \rightarrow \bar{F4}E5\mathbf{E4}G4, \bar{x} \rightarrow \bar{\mathbf{E4}}, x \rightarrow \mathbf{B4}G3^\circ\bar{G4}, \bar{x} \rightarrow \bar{\mathbf{B4}}, x \rightarrow \mathbf{C5}A3^\circ, \bar{x} \rightarrow \bar{\mathbf{C5}}, x \rightarrow F5F3^\circ\mathbf{A4}\bar{E5}, \bar{x} \rightarrow \bar{\mathbf{A4}}F5, x \rightarrow \mathbf{D5}B3A4, \bar{x} \rightarrow \bar{B3}A4\bar{\mathbf{D5}}, x \rightarrow \mathbf{E5}E3G4, \bar{x} \rightarrow \bar{E3}G4\bar{\mathbf{E5}} \end{aligned}$$

Soit $m_{mp}(sb)$ la fonction de mapping, le s-mot activé est le suivant :

$$m_{mp}(sb)(w) = a_1 a_2 \bar{a}_1 a_2 a_3 a_4 \bar{a}_3 \bar{a}_4 a_5 \bar{a}_5 a_6 \bar{a}_6 a_7 \bar{a}_7 a_8 \bar{a}_8 a_9 \bar{a}_9 a_{10} \bar{a}_{10}$$

Certaines articulations peuvent donc être décidées lors de la performance, au risque de ne pas respecter les consignes d'exécution. Notamment, les deux premières notes peuvent être jouées *legato* comme dans le mot w mais elles peuvent aussi être jouées *staccato* avec un mot commençant plutôt comme ceci : $x\bar{x}\bar{x}\bar{x}$.

3.3.2. Cas du MidifilePerformer

Le principe du MidifilePerformer est de conserver l'association entre la touche enfoncée et la s-lettre a_i jouée pour que l'interprète puisse contrôler l'ordre des s-lettres de suspension en choisissant la touche à relâcher. Ainsi, l'ordre des s-lettres \bar{a}_i n'est pas conservé en général, cela dépend de l'exécution de l'interprète.

Les mots de commande sont des shuffles de mots de Dyck avec un alphabet de lettres indicées. Voici un exemple de mot de commande qui va donner le même résultat que dans le cas du Metapiano :

$$w' = x_1 x_2 \bar{x}_1 \bar{x}_2 x_1 x_2 \bar{x}_1 \bar{x}_2 x_1 \bar{x}_1 x_1 \bar{x}_1 x_1 \bar{x}_1 x_1 \bar{x}_1 x_1 \bar{x}_1$$

En effet, dans ce mot, les lettres barrées apparaissent chronologiquement conformément aux lettres non barrées de même indice. Soit $m_{mf}(sb)$ la fonction de mapping du MidifilePerformer, le s-mot activé par ce mot de commande est donc le même que pour le Metapiano : $m_{mf}(sb)(w') = m_{mp}(sb)(w)$. En revanche, le mot de commande suivant donnera une exécution différente :

$$w'' = x_1 x_2 \bar{x}_2 \bar{x}_1 x_1 x_2 \bar{x}_1 \bar{x}_2 x_1 \bar{x}_1 x_1 \bar{x}_1 x_1 \bar{x}_1 x_1 \bar{x}_1 x_1 \bar{x}_1$$

En effet, au début de ce mot, on a inversé l'ordre entre \bar{x}_1 et \bar{x}_2 . La touche x_2 est donc relâchée avant la touche x_1 . La lettre \bar{x}_2 étant associée à la s-lettre \bar{a}_2 , il en résultera que la note B_4 sera stoppée avant la note A_4 . Le s-mot activé par ce mot de commande est le suivant : $m_{mf}(sb)(w'') = a_1 a_2 \bar{a}_2 \bar{a}_1 a_3 a_4 \bar{a}_3 \bar{a}_4 a_5 \bar{a}_5 a_6 \bar{a}_6 a_7 \bar{a}_7 a_8 \bar{a}_8 a_9 \bar{a}_9 a_{10} \bar{a}_{10}$.

Dans ce dernier exemple, les deux premières notes de la partition sont en relation *during*. Il est à noter que cette relation est impossible à réaliser avec le Metapiano car elle implique la permutation de l'ordre des suspensions de ces s-lettres, ici \bar{a}_1 avec \bar{a}_2 .

3.4. Espaces d'exécution

En fonction des mappings que nous avons définis, il est maintenant possible d'exprimer les espaces d'exécutions des deux systèmes.

3.4.1. Cas du Metapiano

Soit $\mathcal{I}_{M_p}(\text{sb})$ l'espace d'exécution du Metapiano. Pour énumérer toutes les exécutions possibles du Metapiano, nous admettons que la fonction $\mathbf{m}_{\text{mp}}(\text{sb})$ est bijective. En effet, une exécution est un mot écrit avec les lettres a_i et \bar{a}_i pour i variant de 1 à 10. Si l'on oublie les indices de ces lettres, les images par $\mathbf{m}_{\text{mp}}(\text{sb})$ correspondent exactement aux mots de Dycks que sont les commandes (en substituant les x par des a). De plus, la numérotation des lettres a_i et \bar{a}_i est elle-même bijective car elle consiste à numéroter de façon univoque les lettres de gauche à droite.

L'ensemble des mots du dispositif d'entrée est l'ensemble des mots de Dyck de longueur 20. Le cardinal de cet ensemble est donné par le nombre de Catalan : $\mathcal{C}_{10} = 16\,796$. Cependant, nous nous sommes limités aux 5 doigts d'une main, donc l'ensemble qui nous intéresse est l'ensemble des mots de Dyck de hauteur limitée à 5. Ce nombre peut être calculé au moyen de la suite récurrente suivante [2] : $T(n+1, k) = \sum_{i=0}^n T(i, k)T(n-i, k-1)$ où n est la demi-taille des mots et k est la limitation en hauteur et $T(0, k) = 1$ si $k \geq 0$ et 0 sinon. Ainsi le cardinal de cet ensemble est réduit à :

$$|\mathcal{I}_{M_p}(\text{sb})| = 14\,041$$

3.4.2. Cas du MidifilePerformer

Soit $\mathcal{I}_{M_f}(\text{sb})$ l'espace d'exécution du MidifilePerformer. De même que pour le dénombrement des exécutions du Metapiano, nous considérons l'ensemble des mots de commande en admettant que la fonction $\mathbf{m}_{\text{mf}}(\text{sb})$ est bijective pour des raisons similaires aux arguments donnés pour le Metapiano. La différence réside dans le fait que nous avons des mots de Dyck indexés du fait que les touches enfoncées sont différenciées. Ces objets ont été dénombrés au paragraphe 2.3.6. Cependant, compte tenu du fait que l'on considère une main à 5 doigts, il faut limiter la hauteur de ces mots. Si on note D_i la série des chemins de Dyck étiquetés, mais partant de la hauteur i et finissant en i aussi (sans jamais plonger sous ce niveau), en regardant le premier retour au niveau i , on a : $D_i = 1 + tD_{i+1}(i+1)D_i$ ou encore $D_i = \frac{1}{1 - t(i+1)D_{i+1}}$. Comme la hauteur à un instant donné correspond au nombre de touches enfoncées, on écrit ces équations pour $i = 0, \dots, 4$, puis on complète avec $D_5 = 1$. La série recherchée est $D_0 = \frac{1-14t+33t^2}{1-15t+45t^2-15t^3}$ qu'il convient de développer en une série - qui est rationnelle - pour obtenir le coefficient du terme t^{10} : $D_0 = 1 + t + 3t^2 + 15t^3 + 105t^4 + 945t^5 + 9\,675t^6 + 104\,175t^7 + 1\,141\,425t^8 + 12\,578\,625t^9 + 138\,877\,875t^{10}$. On obtient donc :

$$|\mathcal{I}_{M_f}(\text{sb})| = 138\,877\,875$$

4. COMPARAISON ET DISCUSSION

Pour comparer les deux systèmes, on introduit quelques propriétés. Au degré d'interactivité introduit par le paragraphe 2.4, on ajoute deux autres propriétés pour exprimer l'expressivité et la difficulté d'exécution. Ensuite on dénombre les ensembles de mots pour calculer les quantités et obtenir le tableau comparatif de la figure 5.

4.1. Définition des SMI

On notera le Metapiano et le MidifilePerformer comme suit :

$$\text{mfP} = \langle \text{sbi}, \mathcal{I}(\text{sb}), \mathcal{D}_{M_f}, \mathbf{m}_{\text{mf}}(\text{sb}) \rangle$$

Pour les deux systèmes, les partitions interactives sont les mêmes, ainsi que l'espace d'exécution de la partition. Les dispositifs d'entrée sont les mêmes au niveau matériel, mais les langages considérés diffèrent puisque le Metapiano a des touches génériques alors que le MidifilePerformer a des touches différenciées. Les fonctions de mapping sont donc différentes ainsi que leurs espaces d'exécution.

Introduisons maintenant le piano sous forme de SMI pour mieux le comparer avec les deux autres SMI.

$$\text{mp} = \langle \text{sbi}_p, \mathcal{I}(\text{sb}), \mathcal{D}_{M_p}, \mathbf{m}_{\text{mp}}(\text{sb}) \rangle$$

Où :

- $\text{sbi}_p = \langle \text{sb}, \text{sra} \rangle$: la partition est entièrement interactive car il s'agit pour l'interprète d'activer tous les événements de début et tous les événements de fin de la partition, c'est-à-dire 44.
- L'ensemble des exécutions correctes de la partition est le même que pour les deux autres SMI.
- $\mathcal{D}_{M_p} = \bigsqcup_{i=1}^{88} T_{x_i}^1$: le clavier comporte 88 touches numérotées. Nous considérerons que pour le piano, le nombre de doigts n'est pas limité, afin de prendre pour référence une situation idéale d'utilisation de l'instrument.
- La fonction de mapping \mathbf{m}_{mp} ne dépend pas de la partition mais uniquement des touches enfoncées ou relâchées, auxquelles elle associe la note correspondante.

4.2. Propriétés d'un SMI

Soit \mathbf{s} une partition, soit $\text{ms} = \langle \text{is}, \mathcal{I}(\mathbf{s}), \mathcal{D}, \mathbf{m} \rangle$ un SMI, soit n la taille des mots de commande de $\mathcal{C}(m)$ et soit \mathcal{D}^n l'ensemble des mots de \mathcal{D} de taille n : $\mathcal{D}^n = \{w \in \mathcal{D}, |w| = n\}$.

4.2.1. Degré de difficulté

On définit le *degré de difficulté* d'un SMI par le rapport entre le nombre de commandes incorrectes que l'interprète peut jouer sur le dispositif d'entrée, c'est-à-dire le nombre de commandes qui n'ont pas d'image par la fonction de mapping, et le nombre de commandes total qui peuvent être jouées sur le dispositif.

$$\text{dd}(\text{ms}) = \frac{|\mathcal{D}^n| - |\mathcal{C}(m)|}{|\mathcal{D}^n|}$$

L'ensemble des commandes et l'ensemble des exécutions sont égaux dans le cas du Metapiano et du MidifilePerformer. On parvient à des degrés de difficulté nuls selon notre définition. Les cardinaux de ces ensembles ont été calculés au paragraphe 3.4.1 pour le Metapiano et au paragraphe 3.4.2 pour le MidifilePerformer.

$$\bullet \text{dd}(\text{mtp}) = \frac{|\mathcal{D}_{Mp}^{20}| - |\mathcal{C}(m_{mp})|}{|\mathcal{D}^{20}|} = \frac{14\,041 - 14\,041}{14\,041} = 0$$

$$\bullet \text{dd}(\text{mfp}) = \frac{|\mathcal{D}_{Mf}^{20}| - |\mathcal{C}(m_{mf})|}{|\mathcal{D}^{20}|} = \frac{138\,877\,875 - 138\,877\,875}{138\,877\,875} = 0$$

On constate que pour les deux SMI, toutes les commandes exécutées par l'interprète sont justes.

Nous nous sommes placés dans le domaine des fonctions de mapping. En effet, les mots joués par un interprète sont tous corrects. Cependant, les commandes pourraient avoir une taille ne correspondant pas à la taille de la partition, en étant ou trop courts ou trop longs. Dans le cas d'un mot trop court, la partition ne sera pas jouée jusqu'au bout, les SMI attendant que la commande soit complétée. Dans le cas d'un mot trop long, les SMI arriveront à la fin de la partition alors que les touches ne sont pas toutes relâchées. Dans ce cas, les deux SMI réaliseront les suspensions manquantes et les débuts que l'interprète aurait activés en trop seront ignorés.

Pour calculer le nombre de commandes que l'interprète peut jouer au piano, il faut dénombrer toutes les combinaisons de 22 notes parmi les 88 touches du piano, avec répétition et les multiplier par tous les étirements possibles. Pour calculer le degré de difficulté, il convient ensuite de soustraire le nombre d'exécutions de la partition et de diviser ce nombre par le nombre total de commandes². Le résultat de ce calcul est proche de 1.

4.2.2. Degré d'expressivité

On définit le *degré d'expressivité* d'un SMI par le rapport entre le nombre d'exécutions possibles pour le SMI et le nombre d'exécutions de la partition au piano.

$$\text{de}(\text{ms}) = \frac{|\mathcal{I}(m)|}{|\mathcal{I}(s)|}$$

Rappelons que les cardinaux de ces ensembles ont été calculés au paragraphe 3.4.1 pour le Metapiano, au paragraphe 3.4.2 pour le MidifilePerformer et que le cardinal de $\mathcal{A} * (sb)$ que nous utilisons est le premier du paragraphe 2.3.6.

$$\bullet \text{de}(\text{mtp}) = \frac{|\mathcal{I}_{Mp}(sb)|}{|\mathcal{A}*(sb)|} = \frac{14041}{3,25 \times 10^{14}} = 4 \times 10^{-10}$$

$$\bullet \text{de}(\text{mfp}) = \frac{|\mathcal{I}_{Mf}(sb)|}{|\mathcal{A}*(sb)|} = \frac{138\,877\,875}{3,25 \times 10^{14}} = 4 \times 10^{-6}$$

$$\bullet \text{de}(\text{mp}) = \frac{|\mathcal{A}*(sb)|}{|\mathcal{A}*(sb)|} = 1$$

4.2.3. Degré d'interactivité

Les degrés d'interactivité des deux SMI sont les mêmes. C'est-à-dire que l'interprète aura le même nombre d'actions à réaliser pour exécuter la partition.

2. $\frac{\binom{n+k-1}{k!} \times 3,25 \times 10^{14} - 3,25 \times 10^{14}}{\binom{n+k-1}{k!(n-1)!} \times 3,25 \times 10^{14}} \simeq 1$, avec $n = 88$ et $k = 22$

Propriétés	Metapiano	MidifilePerformer	Piano
Interactivité	0.45	0.45	1
Expressivité	4.10^{-10}	4.10^{-6}	1
Difficulté	0	0	1

Figure 5. Tableau comparatif des propriétés des SMI par rapport au piano.

La partition *sb* ayant 22 notes au total, elle comporte 44 lettres correspondant aux débuts et aux suspensions de ces notes. Rappelons que la partition interactive *is* comporte 20 points d'interaction. On obtient la valeur suivante pour le degré d'interactivité : $\text{di}(\text{is}) = \frac{|i|}{|sb|} = \frac{20}{44} = 0.45$. Concernant le piano, son degré d'interactivité est de 1 car tous les événements doivent être activés par l'interprète.

4.3. Discussion

L'évaluation des propriétés confirme que le MidifilePerformer est plus expressif que le Metapiano. En effet, le MidifilePerformer permet plus d'articulations grâce au choix des suspensions à réaliser qui est laissé à l'interprète. En effet, les deux SMI permettent de réaliser les relations temporelles *before*, *meet*, *finish* et *overlap*, alors que le MidifilePerformer permet de surcroît de réaliser des *during*, ce qui conduit à un nombre remarquable de possibilités supplémentaires par rapport au Metapiano.

Notons de plus que $\mathcal{I}(sb)$ est inclus dans les espaces d'exécution des deux SMI. Ce qui signifie que l'interprète peut réaliser l'interprétation idéale consignée dans la partition. Mais il peut aussi s'en éloigner, et c'est ce que mesure le degré d'expressivité.

Les deux systèmes sont de difficulté nulle, au sens où toute commande donne lieu à une exécution de la partition considérée comme une interprétation, plus ou moins fidèle aux consignes, de la partition. En revanche, le piano a une difficulté importante du fait qu'il est possible de ne respecter ni l'ordre des notes, ni les notes elle-mêmes. Ce qui donne lieu à de nombreuses exécutions incorrectes. Au contraire du piano, nos deux SMI sont donc accessibles à tous les interprètes.

Remarquons que certains mots pourraient ne pas fonctionner correctement avec certains synthétiseurs. Par exemple, dans le mot *aaā*, le fait que le début de la deuxième occurrence de la note *a* arrive avant que la première soit relâchée pourrait créer un dysfonctionnement si ce cas n'est pas traité correctement par le synthétiseur.

Les calculs proposés pour évaluer les propriétés donnent des résultats cohérents avec ce qui est attendu sur ces exemples. Cependant leur quantification demande à être affinée pour rester dans une fourchette plus facilement interprétable. Le degré d'interactivité semble facile à appréhender. Mais le fait que le Metapiano soit 10000 fois moins expressif que le MidifilePerformer l'est beaucoup moins. Ce nombre correspond aux possibilités supplémentaires données par le MidifilePerformer de choisir la note qui va être stoppée.

Rappelons en effet que toutes ces mesures s'appliquent

à des SMI pour une certaine partition et que les valeurs de ces propriétés peuvent varier selon la partition considérée. Serait-il envisageable de séparer la difficulté d'un SMI en deux valeurs indépendantes, l'une liée plutôt à la partition et l'autre au mapping ? Cela permettrait de s'abstraire de telles dépendances pour avoir des mesures génériques. Concernant la difficulté, les résultats sont très extrêmes sur ces exemples. Mais ce n'est pas toujours le cas. Notamment, le mode de jeu dit "rythmique" étudié en [5]³ sur les deux premières mesures du Scherzo présente des difficultés très différentes selon qu'il est joué au moyen d'une ou deux touches. En effet, pour cet exemple une difficulté nulle est obtenue avec une seule touche, alors que deux touches non génériques impliquent une difficulté de 7/8 sur l'extrait étudié.

5. CONCLUSION

Dans cet article nous avons étudié le Metapiano et le MidifilePerformer dans l'objectif de modéliser algébriquement leur fonctionnement. Il est apparu que l'expressivité de ces deux SMI réside principalement dans le traitement des suspensions. Leur fonctionnement permet d'exécuter des articulations diverses qui peuvent conduire l'interprète à jouer une exécution parfaite selon les consignes de la partition ou bien s'en éloigner à sa guise à l'intérieur des espaces d'exécution ouverts par ces SMI. Ces espaces sont bien sûr plus restreints que celui qui est accessible par un véritable piano, car un piano permettrait aussi de permuter les débuts des notes, voire de ne pas respecter du tout les notes de la partition, ce qui donnerait lieu à un nombre énorme de possibilités. En ce sens, ces SMI s'adressent aux virtuoses de par la finesse d'exécution des articulations mais aussi aux débutants voire aux personnes en situation de handicap car les gestes de sélection des touches dans la tessiture du clavier ont été éliminés [3].

Les modélisations du Metapiano et du MidifilePerformer diffèrent en somme assez peu. Si celle du Metapiano est purement temporelle, il a fallu introduire un peu d'espace dans celle du MidifilePerformer pour repérer les touches enfoncées. En effet, les touches du Metapiano sont génériques, seule la position temporelle d'une action sur une touche est utilisée. Alors que pour le MidifilePerformer, les touches enfoncées doivent être indicées pour pouvoir associer les relâchements aux enfoncements. Ces indices ajoutent un peu d'information spatiale au SMI.

Les perspectives de ce travail comprennent notamment l'indexation de la partition interactive en voix pour le jeu à plusieurs. Pour le MidifilePerformer, il faut pouvoir modifier les canaux du fichier MIDI si ceux-ci ne conviennent pas. Enfin, pour permettre de jouer une seule voix alors que le système joue les autres, il faut introduire une extraction de tempo du jeu de l'interprète pour que le système s'adapte à la vitesse d'exécution à la manière d'un suiveur de partition [7].

Il conviendrait aussi par la suite d'étudier les impossibilités de synchronisation du jeu pianistique pour les mo-

3. Il s'agit d'un mode où seuls les débuts des notes sont interactifs.

déliser et éviter ainsi les synchronisations exactes synthétiques.

Remerciements : Cette recherche s'est déroulée dans le cadre du SCRIME de l'Université de Bordeaux (Studio de Création et de Recherche en Informatique et en Musique Expérimentale) et financée par le ministère français de la culture. Les auteurs tiennent à remercier Mireille Bousquet-Mélou pour son aide précieuse dans l'énumération des espaces d'exécution.

Références

- [1] J. F. ALLEN. "Maintaining knowledge about temporal intervals". In : *Communications of the ACM* 26.11 (1983), p. 832-843.
- [2] A. BACHER. "Generalized Dyck paths of bounded height". In : <https://arxiv.org/pdf/1303.2724.pdf> (2013).
- [3] C. CADOZ. "Le geste canal de communication homme/machine : la communication "instrumentale"". In : *Revue des Sciences et Technologies de l'Information - Série TSI : Technique et Science Informatiques* 13.1 (1994), p. 31-61.
- [4] J. CHABASSIER et al. "MidifilePerformer : a case study for chronologies". In : *FARM 2021*. 2021, p. 13-22.
- [5] M. DESAINTE-CATHERINE et al. "Algebraic Framework for Design and Qualitative Evaluation of Interactive Musical Systems Temporality". In : *Sound and Music Computing*. 2022, p. 247-254.
- [6] I. A. DURAND et S. R. SCHWER. "A Tool For Reasoning about Qualitative Temporal Information : the Theory of S-languages with a Lisp Implementation". In : *Journal of Universal Computer Science* 14 (20) (2008), p. 3282-3306.
- [7] J-L GIAVITTO et al. "Time, Timelines and Temporal Scopes in the Antescofo DSL v1. 0". In : *International Computer Music Conference (ICMC)*. 2017.
- [8] J. HAURY. "Un répertoire pour un clavier de deux touches. Théorie, notation et application musicale". In : *Document numérique, URL :https://www.cairn.info/revue-document-numerique-2008-3-page-127.htm* 11 (2008), p. 127-148.
- [9] D. LEWIN. *Generalized Musical Intervals and Transformations*. Yale University Press, 1997.
- [10] G. MAZZOLA, S. GÖLER et S. MÜLLER. *The Topos of Music, Geometric logic of Concepts, Theory and Performance*. Birkhaus Verlag, 2005.
- [11] A. TARSKI. "On the Calculus of relations". In : *The journal of Symbolic logic* 6.3 (1941), p. 73-89.
- [12] N. WIENER. "A contribution to the theory of relative position". In : *Proceedings of the Cambridge Philosophical Society* 17 (1914), p. 441-449.

Actes des Journées d'Informatique Musicale 2023
24-26 mai 2023
Maison des Sciences de l'Homme Paris Nord

SESSION « COMPUTER MUSIC »

INTRODUCING COMPUTER MUSIC IN SECONDARY SCHOOL

Christopher Dobrian
University of California, Irvine
dobrian@uci.edu

RÉSUMÉ

Actuellement, l'éducation en l'informatique musicale a lieu presque exclusivement dans les universités, surtout au niveau du troisième cycle. Cet article propose que telle éducation puisse, et doive, s'introduire dans l'école secondaire. La formation en musique et en informatique peut être *combinée* de manière à ce que les deux sujets soient accessibles et engageants, même pour les élèves qui n'ont pas eu de formation traditionnelle en musique ou qui ne se croient pas doués en technologie. En rendant ces matières disponibles à un nouveau groupe d'élèves plus jeunes, on espère éviter que certains étudiants ne soient involontairement exclus de ces domaines en raison du manque d'accès à une formation précoce. Cela à son tour pourrait contribuer à accroître la diversité dans ces domaines au niveau postsecondaire. Pour que cette formation ait lieu, des ressources d'apprentissage de niveau secondaire qui combinent la musique et la programmation informatique doivent être développées, et des enseignants connaissant les deux sujets doivent être disponibles pour ces étudiants.

Computer music education currently occurs almost exclusively at the post-secondary level, mostly in graduate degree programs. This article proposes, however, that such education can and should begin in secondary school. Introductory training in music and computer programming can be *combined* in such a way as to make both subjects approachable and engaging, even for students who lack traditional music training or who might otherwise think of themselves as not technologically inclined. Making these topics accessible to a new constituency of students at a younger age may help to avoid students' being unintentionally excluded from participation in these fields due to lack of access to early training. That in turn may help to increase diversity in these fields at the post-secondary level. For such training to take place, secondary-level learning resources that combine music and computer programming must be developed, and teachers with knowledge of both topics must be available for those students.

1. INTRODUCTION

Computer music study is commonly initiated at the university level, usually in graduate school. However, it can and should be introduced at an earlier stage, in high school and even in middle school. I will explain the

reasons why, and I will propose an approach that provides knowledge and experience in both music making and computer coding to students at a young age, preparing them for better footing in those topics when they pursue postsecondary studies.

Most secondary schools provide some opportunity for musical experience (orchestra, concert/marching band, jazz band, choir), and less frequently some sort of musical training (instrumental lessons, AP theory), but almost never any discussion of computer music. By the same token, many secondary schools provide some optional opportunities to learn computer programming, but rarely is such training explicitly linked with sound and music. At the university level, some undergraduate programs do exist that focus specifically on computer music, but because most music curricula are quite full already with classical music training—instrumental lessons, chamber ensembles, large ensembles, ear training, music theory, music history, analysis, literature surveys, and so on—most music students, even those majoring in composition, receive little or no computer music training as undergraduates. Thus, most computer music education occurs during one's graduate studies.

Because music and computer programming have historically been considered separate domains, they are still taught separately at nearly every level prior to graduate school. I will explain why combining those fields at an early stage of education may be preferable.

2. HISTORY

Forty years ago, "computer music" was still largely experimental, and was taught at only a very few universities, almost exclusively in graduate-level courses that included education in digital sampling theory, the mathematics of digital signal processing and synthesis, and computer programming of digital audio. Those courses included no training in music per se; basic knowledge of music theory, notation, and composition was assumed, because most such courses were hosted in the university's music department. The musicians most interested in computer music tended to be composers who had already had some experience with analog electronic music, or less frequently electrical engineers with an interest in digital audio. Thus, computer music and composition (two largely male-dominated areas) became linked, and computer music often was treated as a

sideline subtopic within the study of composition. Alternatively, music departments might hire a digital audio researcher as a somewhat oddball lone member to represent this burgeoning field. MIDI and digital synthesizers were very new, as were personal computers, and were not particularly prevalent, so the hardware for such courses was usually a timeshared mainframe computer, and the computation was mostly non-realtime. In short, computer music equipment and knowledge was accessible only to advanced students at a research university, often in a research lab that was tangential to any academic department. Thus, the content of computer music education, and the locus of computer music activity, was established by the inherently elite nature of the field.

The definition of computer music changed somewhat in the 1990s with the greater prevalence of personal computers and speedier processors. Computers are now the basis of all music recording and production, so the term computer music includes the use of DAW software, software synthesizers, and user-friendly applications running on reasonably affordable laptop computers supplemented with affordable input control interfaces (e.g. MIDI controllers) and audio output interfaces. Thus, computer music activity has been to some extent “democratized”, and extended beyond academia to popular musicians and hobbyists. The term computer music now means different things to different people, and can have different foci in different institutions and environments. Likewise, whereas the genre of academic computer music began, and remains, closely tied to mid-twentieth century modernism, outside of academia computer music includes minimalist looping, ambient music, hip-hop, electronic dance music, and other popular styles. I personally welcome this artistic and stylistic diversity, and I hope to see that diversity increase within academia.

3. DIVERSITY

I have written previously[5] about the value of diversity in institutions, especially in academia, in the interest of promoting intellectual “hybrid vigor”. The expression of diverse viewpoints leads to healthy debate of new ideas, whereas having all like-minded people together in a room can lead to ossification of a single view, stunting growth. If only for that reason alone, I propose that diversity in computer music is inherently productive and necessary. For me, an equally strong reason for promoting diversity is the pursuit of social justice. In the United States, in many fields, implicit biases and systemic prejudices historically have privileged some people, and have tended to discourage or exclude others, due to their race and/or gender.¹

¹I recognize that some people deny that this has been the case, but they are demonstrably wrong. I will not expend the time and space here to demonstrate that.

Rectifying that wrong is of equal importance to the value gained by having diverse scholarship and aesthetics in the field. For both of those reasons—intellectual hybrid vigor and social justice—it is the responsibility of the current practitioners in the field to consciously promote greater diversity in its membership. One way of doing that is to encourage participation by as many different sorts of people as possible, namely by providing opportunity and education to all at an early age.

3.1. The pipeline problem

Youths who belong to a lower economic stratum often do not have the same access to musical instruments, music lessons, a personal computer, and after-school educational activities as their more affluent counterparts enjoy. Also, from an early age, girls—and often Black and Latino boys—are subtly and not-so-subtly steered away from mathematical, scientific, and technical topics, and also away from building and engineering (tinkering) pursuits. This early bias, especially in conjunction with lack of access to learning resources, leads to what has been termed the “pipeline problem”:² the pipeline through which potential women and minority applicants may get to graduate programs is constricted at many stages along the way, allowing only a trickle to reach the destination. This constriction of the pathway to computer music, the filtering out of certain types of people from the field, actually starts almost from birth, continues at all levels of education, and culminates in the job market. The pipeline problem is only one of the many factors that can contribute to lack of diversity in a graduate program, but it is a notable one, and is one that can potentially be addressed by providing students better access to computer music education at an earlier age.

But what can a university professor do to address this problem? University faculty are invariably extremely busy with the demands of their job: composing and/or conducting research, running computer music facilities, mentoring student creative and research projects, staying abreast of new developments in a rapidly changing field, and teaching. Even if one has the will to try to ameliorate the pipeline problem, the prospect of having to take on new work is discouraging. Reaching younger students requires extra outreach effort. My own solution has been to frame educational outreach and development as a specific research focus.

4. COMMUNITY-BASED RESEARCH

Professors in a research university are often accused of operating in an ivory tower, undertaking research and creative activity that appears to have little effect upon, or relevance to, the surrounding community. In the

²This reference to the so-called pipeline problem is not meant to ignore the equally serious “trapdoor problem” whereby systemic racism and sexism in this field causes many people to drop out or be excluded from (or avoid) the field altogether. Both problems exist and must be addressed.

University of California system, faculty research that explicitly addresses community concerns is increasingly validated and encouraged.

4.1. Outreach

One way that professors can address the pipeline problem is by leveraging the infrastructure and staff support of their own institution to offer educational outreach programs. In my school, with the help of a staff member from the Outreach office, we have instituted two-week summer academies in Digital Audio Production and in Computer Music Programming. In an effort to promote inclusivity, we offer need-based scholarships—effectively simply reducing or waiving tuition for students with demonstrated financial hardship—so that students from underprivileged areas are encouraged to participate. In addition to teaching those summer courses, I have offered weekend workshops for young student groups (from thoughtfully selected neighborhoods) so that they may be introduced to the ways in which computing and music making can be combined. Thus, the existing facilities and equipment of the university are used during times when they're otherwise unused, and the organization is facilitated by outreach staff who are happy to assist with new activities that fit perfectly within their job description and add to the university's community interaction. Once the summer and weekend courses were well established, I was able to delegate the teaching of those courses to capable graduate students, providing them with valuable teaching experience and much-needed summer employment.

4.2. Education research

Short-term outreach programs such as those are helpful for providing young students with exposure to computer music and the educational culture of the university. Their actual educational value is limited, though, because of their brevity. In order for more thorough education and deeper involvement to occur, the teaching really needs to be integrated into the students' activities on a regular and ongoing basis. This can be in the form of an after-school class (or club)—hosted either at the university or at their own school—or in some cases can even be integrated into their school's existing curriculum in math, computing, and music. Establishing such activities requires coordination with high school faculty and administrators, a task which again can be much facilitated by the university's outreach staff. In my own department at UC Irvine (a Hispanic-serving institution), a small team of three faculty members collaborate with teachers at a local high school in a low-income primarily Hispanic neighborhood, to organize additional activities in orchestral music, jazz, and music technology. In that instance, the work is *pro bono*, but again, once established, it can often be delegated to university students, who enjoy and benefit from having the teaching experience.

In most university computer music programs, pedagogy is not a primary research focus, but there's no reason it cannot be. I recently have chosen to steer my own research toward secondary-level pedagogy, with an aim toward developing and instituting a curriculum of study that will provide combined training in both music and computer programming.

As noted above, these two topics are traditionally treated completely separately until graduate school. In that way, most musicians come to programming rather late, and are often intimidated by the learning curve in a brand-new field of such complexity. Similarly, computer science and electrical engineering students casually go through their entire university curriculum of programming and/or digital signal processing without ever thinking of the potential musical applications of what they've learned. If students were to gain familiarity with both fields—music and programming—at an earlier age, combined in ways that intentionally show the potential interrelationship, those students will have a better chance of reaching a higher level of sophistication eventually.

What should be taught in a curriculum that strives to provide the basics of both music and programming, and what tools and resources should be employed? Because the field of computer music is so broad, and the hardware and software tools are so varied, and musical stylistic possibilities are so numerous, there may be many answers to that question. Even though most students today have access to computers, they are likely using different platforms—macOS, Windows, Linux, Chrome—and they may have limited access to musical instruments, keyboards, other equipment. Access is often limited by financial constraints. Thus, in my own pedagogy I am focusing on a few computer music software tools that are free, cross-platform, and easy to acquire.

5. AVAILABLE FREE TOOLS

5.1. Programming languages

- Pd (PureData)[9] is a free object-based music-oriented graphical programming language developed by Miller Puckette, creator of the ubiquitous (but not free) Max programming language. It is often cited as a free Max equivalent. It is not as fully capable as Max, but it is powerful and has a significant user base and support community.

- SuperCollider[8] is a free text-based music-oriented programming language developed by James McCartney with a significant user base and support community. As a text-based object-oriented coding language, it may be useful in teaching concepts of object-oriented programming, as is JavaScript.

- Sonic Pi[2] is a free, cross-platform high-level text programming interface designed for realtime live coding of music. It was designed mindful of a simple learning curve, for use in education.[1]

- JavaScript[6] is the predominant programming language for Web development, is free, can be composed in any text-editor, and can run in any browser. It provides access to the Web Audio API for handling a variety of audio operations and the Web MIDI API for handling MIDI data; these libraries—along with the graphical capabilities of HTML 5 (and freely available interface-building libraries such as React)—provide sufficient tools for making a browser-based music app.

5.2. DAWs

- Audacity[7] is a highly capable cross-platform sound editor, but lacks many music-specific capabilities commonly associated with DAWs, and lacks full VST support. It serves well for viewing sound characteristics in both the time domain and the frequency domain, and for teaching audio editing, mixing, and the basics of audio effects processing.

- Waveform[10] is a free cross-platform DAW with VST support and a pretty full feature set.

- Reaper[3] is essentially as powerful as a professional commercial DAW. Although not technically free, it is low-cost shareware, and the demo version continues to run with full capability (with only a brief imposed pause when loading) even after the trial period has expired.³

5.3. Synthesis

- VCV Rack[11] provides a well-conceived digital emulation of a modular analog synthesizer, retaining traditional standards such as 1V per octave for pitch, while also providing MIDI control and other benefits of digital emulation. Synthesis concepts that may be difficult to grasp in the abstract, such as low-frequency modulation, can be introduced both sonically and visually in VCV Rack before teaching them in a less visual context like Pd or a text-based context like SuperCollider or JavaScript.

6. CURRICULUM

The design of a curriculum, and a set of useful tutorials, that coherently teach both music and computer programming at an introductory level, is the subject of my new and ongoing research. There are many possible approaches to this challenge of teaching the fundamentals of music and programming simultaneously. What works best will depend on the orientation and preferences of the teacher(s), and even more importantly on the students and the context in which the teaching will occur.

In collaboration with a few of my graduate students, whose employment was funded by a grant from the University of California, I have published an online textbook of sorts, a set of lessons titled simply *Computer Music Programming*. [4] It was created for a specific

course that I teach, titled Computer Audio and Music Programming. That course is not actually intended to teach programming. It is intended for people who already know how to program in some language, and its main focus is to teach the principles of music, audio, synthesis, and processing that a programmer needs to know in order to program audio and music effectively.

The website demonstrates a modular organizational scheme consisting of many free-standing lessons, organized by topic. The approach in designing that site was that every topic has multiple lessons, and that each lesson is thought of as a “bite”—a small, concise chunk of information. Each bite should be manageably small, and if a bite has other requisite knowledge beyond its own scope, that knowledge should become its own bite, a separate lesson. The goal was to arrive at a collection of lessons—standalone-but-sometimes-interdependent educational chunks—any one of which could be useful on its own. A teacher or autodidact can then assemble those chunks in whatever order they deem appropriate for their purpose. This idea of treating information in standalone bites intended for assembly in various configurations is useful model for conceptualizing a curriculum.

Fortunately, there also already exist a great many teaching materials and videos freely available. One can take heart in the fact that much work has already been done, in the form of online textbooks, tutorials, courses, YouTube videos, etc. So, one doesn’t need to reinvent the wheel when it comes to explaining coding concepts and music concepts to beginners. There are many resources for the fundamentals of both music and coding. Where new ideas need to be developed is primarily in the meeting and merging of the two. The teacher’s task becomes one of compiling and curating existing materials, and then devising clever ways to link music theory and practice to math and coding. Here are just a few examples of “music meets coding”.

6.1. Music meets coding

- Counting is a crucial part of music composition and performance. Music is full of instructions like “play this four times, then play that two times,” or “play this chord for four measures, then play that chord for two measures,” or “play seven notes in C major as fast as you can,” or “the third time this happens, proceed to the next section.” Counting is such a well-learned procedure, used dozens of times every day in all sorts of situations, that we generally don’t think about how we do it. The process of describing exactly how we count, with the intention of teaching it to a computer, gives the student experience in formalizing a task such that a computer can carry it out. Writing a program that counts requires that the student learn about storing a number in a variable, incrementing that number, repeated processes (loops), relational operators (tests), and conditional statements.

³I advocate that everyone pay for the shareware they use. The generous policy of Cockos to leave the Reaper software fully enabled even after

the trial period expires does give users the ability to pay at whatever time they’re able.

- In most musical situation, timing is crucial. Whereas our counting program would normally run as fast as possible in a few nanoseconds, for it to be musically useful we need to introduce the idea of timed counting, which involves calculating time intervals and scheduling events with whatever scheduling mechanisms exist in the programming environment we're using. That might be **metro**, **delay**, **pipe**, **line**, etc. in Max or Pd, or it might be `setTimeout()`, `setInterval()`, `setValueAtTime()`, `linearRampToValueAtTime()`, etc. in JavaScript. All of this can be made more fun and engaging if the results are expressed sonically as well as numerically, which is now possible because we have introduced the capability for timing and rhythm.

- For music theorists and composers, particularly when discussing post-tonal music, it's common to refer to pitch classes as integers, with C=0, C#(Db)=1, D=2, etc. And for computer musicians, it's common to refer to pitches of the 12-tone equal-tempered scale with their MIDI equivalent, with middle C=60 and each semitone ± 1 from that. But why not introduce the pitch-number equivalencies immediately when discussing the fundamentals of music? That makes the pitch relationships more comprehensible to musicians and novices alike, and provides the most obvious way to manage pitches and pitch classes in a computer program. Combined with our timed counting program, we can now figure out how to play an ascending and/or descending chromatic scale, then how to alter that program to perform a whole tone scale (increment by twos) or arpeggiate a diminished seventh chord (increment by threes). Diatonic scales have an irregular pattern of intervals, usually best implemented by looking up pitches or pitch classes in a lookup table, thus introducing the concept of an array, indexing, and modulo arithmetic.

- We all have the apparently innate ability to measure the amount of time between two events, and then project that same distance mentally into the future to predict the next beat, and we can even divide that period of time into rational fractional parts with remarkable accuracy. Even though we perceive rhythm and tempo sonically, viscerally, and intuitively, conceptually rhythm is inherently linked to math, because it is conceptualized as divisions of time (as well as groupings/multiplications). In American English, the names of rhythmic values are all expressly based upon divisions of a whole unit: whole note, half note, quarter note, eighth note, etc. Rhythm and rhythmic notation in Western music have traditionally been taught through a combination of imitation and memorization of rhythms and their notated symbolic representations; however, conceptualizing rhythm as timepoints on a temporal grid, as in a drum machine or a DAW, allows one to treat rhythm mathematically and thus be better able to program rhythms in a computer.

The development of lessons that consciously integrate both music and coding simultaneously is usually best done by thinking of a very small musical concept one wants to convey, and then thinking how you would teach that to a computer step by step. From that program you

will see the sort of computer programming concepts and structures that will be necessary, and you begin to develop a set of programming techniques and their musical relatives. Conversely, one can also start with a programming concept and think of a musical use for it. For example, to demonstrate and employ the terminology of object orientation, such as *class*, *instantiation*, *properties*, and *methods*, one might choose to describe a musical note as a sound *object*. Because it is an object (in the programming sense of the word), it will have properties, such as pitch, duration, starting time, instrument, etc., and it may have methods such as `play()`, `stop()`, `transpose()`, and so on. Multiple instantiations of the note object class can be used to comprise a musical melody or phrase (or, for that matter, an entire score), which may in turn be stored as an array of note objects.

7. CMERG

To focus on this project of developing and implementing a course of study that integrates music and programming in new and effective ways, I have established at my own university a Computer Music Education Research Group (CMERG). My intention is to leverage the resources of my own institution—web hosting, physical facilities, equipment, staff, graduate students, and colleagues—and to collaborate with local secondary school educators and their high school students, trying out new ideas for education.

The goal of the group is to provide new educational opportunities, especially to students who might otherwise not receive exposure to one or both of these topics in their high school. In addition to conveying information, our goal will be to make these topics fun, engaging, and welcoming, so that all students will feel encouraged to pursue whatever interests them. Some students may develop a particular interest in programming because they are shown fun ways to use it to make music; others may develop an interest in music because they've been shown a new way to think about it, relevant to their existing interest in coding.

Realizing this goal requires compiling and curating existing educational resources, developing new lessons that integrate music and coding, collaborating with high school educators to organize classes or clubs, and then actually delivering the education. By using free, cross-platform tools and very inexpensive equipment, we can make these activities available even to students of limited financial means.

Because the effort is inherently collaborative, our first step is organizing the interested parties—university faculty and students, outreach staff, high school administrators and teachers, and eventually high school students—and setting up meetings and shared online spaces for brainstorming and planning. We welcome as many group participants as are interested in this topic, wherever they may be. Because of the ease with which it's currently possible to meet virtually and share information online, a research group like this can easily

have global scope. Thus, we welcome collaborators from other U.S. universities, and we invite our international colleagues, too. You are welcome to contact me at the email address listed in the heading of this article.

8. REFERENCES

1. Aaron, S. « Live Coding Education », *The MagPi*, Educator's Edition, v. 1, Raspberry Pi Press, Cambridge, 2016. < <https://sonic-pi.net/files/articles/Live-Coding-Education.pdf> >
2. Aaron, S. *Sonic Pi* < <https://sonic-pi.net/> >
3. Cockos. *Reaper*. < <https://www.reaper.fm/> >
4. Dobrian, C. et al. *Computer Music Programming*, online textbook developed at and hosted by the University of California, Irvine, 2019. < <https://dobrian.github.io/cmp/> >
5. Dobrian, C. and Jones, M. « Intentional Inclusion: Promoting Diversity in Graduate Study of Music Technology », *Journal SEAMUS*, 2016 v. 1 no. 1-2, Society for Electro-Acoustic Music in the United States, Beverly Hills, 2016. < <https://music.arts.uci.edu/dobrian/IntentionalInclusion.pdf> >
6. ECMA International. *JavaScript*. < <https://developer.mozilla.org/en-US/docs/Web/JavaScript> >
7. Mazzoni, D. and Dannenberg, R. *Audacity*. < <https://www.audacityteam.org/> >
8. McCartney, J. *SuperCollider*. < <https://supercollider.github.io/> >
9. Puckette, M. *Pure Data*. < <https://puredata.info/> >
10. Tracktion. *Waveform*. < <https://www.tracktion.com/products/waveform-free> >
11. VCV. *VCV Rack*. < <https://vcvrack.com/Rack> >

L'EMPREINTE DE JOHN CHOWNING DANS LA CONSTRUCTION DE L'IRCAM: DE SA RENCONTRE AVEC PIERRE BOULEZ EN 1973 A LA CREATION DE STRIA EN 1977

François-Xavier Féron

STMS – CNRS, Ircam, Sorbonne Université
francois-xavier.feron@ircam.fr

Laura Zattra

Conservatoires de Rovigo, Milan, Vicenza
laura_zattra@yahoo.it

RÉSUMÉ

John Chowning a pleinement contribué au développement de l'informatique musicale dans les années 1960 en mettant au point un algorithme de spatialisation sonore et en posant les bases de la synthèse par modulation de fréquence. Il fonde au milieu des années 1970 au sein de l'université Stanford le *Center for Computer Research in Music and Acoustics* (CCRMA) mais il participe aussi, à la même époque, à la structuration de l'Institut de recherche et coordination acoustique/musique (Ircam) que dirige Pierre Boulez. En étudiant différents fonds d'archives et en conduisant, dans le cadre du projet RAMHO (Recherche et Acoustique Musicales : une Histoire Orale), plusieurs entretiens avec Chowning, nous avons restitué son parcours professionnel entre 1972 et 1977 et mis en évidence les liens étroits qui unissent historiquement l'Ircam et le CCRMA.

1. INTRODUCTION

John Chowning est devenu une figure tutélaire de l'informatique musicale dès la fin des années 1960 grâce aux travaux pionniers sur la synthèse sonore par ordinateur qu'il développa au sein de l'université Stanford en Californie. Percussionniste aguerri et compositeur en devenir, il suit un master en Art lorsqu'il lit, au début de l'année 1964 [14], l'article séminal de Max Mathews qui avait été publié quelques mois plus tôt [13]. Avant d'entamer son doctorat en Art – qu'il accomplira sur deux années académiques entre 1964 à 1966 –, il se rend aux *Bell Telephone Laboratories* (*Bell Labs*) pour rencontrer Mathews ; il repart avec le programme Music IV sous forme de cartes perforées contenues dans une boîte à chaussure. Avec l'aide précieuse de David Poole, il s'initie à la programmation et adapte le programme pour le faire fonctionner sur l'ordinateur DEC PDP-1 connecté à un IBM 17/90 qui était alors utilisé par une poignée de chercheurs du *Stanford Artificial Intelligence Laboratory* (SAIL) [7]. Durant ses années de doctorat, Chowning parvient à synthétiser sur deux canaux des sources sonores mobiles. Il devient alors *music instructor* (1966-67) puis *assistant professor* (1967-72) au sein du département de musique de l'université où il fédère une petite équipe afin de

poursuivre les recherches sur la synthèse numérique du son en travaillant toujours avec l'équipement informatique du SAIL. Au début des années 1970, Pierre Boulez réfléchit quant à lui à la structuration d'un centre de recherche en acoustique qui sera intégré dans le futur Centre Beaubourg.

C'est donc à la même époque – à l'aube des années 1970 – que Chowning et Boulez dessinent respectivement les contours du *Center for Computer Research in Music and Acoustics* (CCRMA) et de l'Institut de recherche et coordination acoustique/musique (Ircam) qui contribueront de manière remarquable au développement de l'informatique musicale. S'il est connu que ces deux centres partagent une histoire commune [15] [1], le rôle de Chowning durant la phase de construction de l'Ircam est peu documenté. Son implication est pourtant bien réelle et elle débute au cours de l'hiver 1973 après avoir rencontré Boulez une première fois. Qui est à l'origine de cette rencontre ? Comment Chowning participe-t-il à la mise en place de l'Ircam ? Quels sont les liens entre l'institut français et son homologue californien ?

Pour répondre à ces questionnements, nous nous sommes naturellement plongés dans les nombreuses publications consacrées aux travaux de Chowning et avons étudié plusieurs fonds d'archives mais nous nous sommes surtout appuyés sur les nombreux entretiens qu'il nous a accordés en 2021 dans le cadre d'un projet de plus grande ampleur que nous présenterons brièvement dans la section 2 de cet article. Nous nous intéresserons alors au parcours professionnel de Chowning entre 1972 et 1977, en tentant de résumer dans les sections 3 à 7 les cinq années académiques durant lesquelles il participa activement à la création du CCRMA mais aussi à celle de l'Ircam.

2. LE PROJET RAMHO, ENTRE HISTOIRE ORALE ET FONDS D'ARCHIVES

2.1. Genèse et objectifs

Le projet RAMHO (Recherche et Acoustique Musicales : une Histoire Orale) a été esquissé au sein de l'équipe Analyse des Pratiques Musicales de l'Ircam par Nicolas Donin, François-Xavier Féron, Vincent Tiffon et

Laura Zattra alors que se tenait, en mai 2018, le colloque-hommage à Jean-Claude Risset (1938-2016) au cours duquel étaient réunis de nombreux pionniers de l'informatique musicale et de l'acoustique musicale. Il était question originellement de se pencher sur l'histoire de l'Ircam mais très vite le champ d'étude s'est étendu à l'acoustique musicale puis, plus généralement, à la recherche musicale, tant les frontières entre ces domaines de nature interdisciplinaire s'avéraient poreuses. La France a joué un rôle historique dans l'institutionnalisation de ces recherches en permettant la création de nombreux laboratoires, studios et instituts s'intéressant à la science des sons au prisme de la musique et des technologies. L'année 1958 marque un tournant décisif avec, d'une part, la création du Groupe de Recherches Musicales par Pierre Schaeffer (1910-1995) qui vient remplacer le Groupe de recherche en Musique Concrète (GRMC) et, d'autre part, l'organisation à Marseille, sous l'égide du CNRS, du premier congrès d'acoustique musicale par François Canac (1886-1969). Débute alors ce qu'on serait tenté de dénommer les « Trente glorieuses » de la recherche musicale, marquées par la fondation – jusqu'à la fin des années 1980 et cela sur l'ensemble du territoire – de nombreuses structures dont voici, pêle-mêle, quelques acronymes : GRM, LAM, EMAMu, GAIV, CIRM, GMEM, GMEB, LAMU, IRCAM, ACROE, LIMCA, GRAME, MIM, CICM... Comment ces structures ont-elles été mises en place ? De quels moyens humains et matériels disposaient-elles ? Comment se positionnaient-elles les unes par rapport aux autres ? Quel était l'équilibre entre recherche scientifique et production artistique ? Pour aborder de telles problématiques et révéler la nature rhizomique de ce réseau interdisciplinaire, les chercheurs associés au projet RAMHO vont à la rencontre de celles et ceux qui ont été témoins et/ou acteurs de la naissance et du développement de ces différents lieux de recherche et création.

2.2. Méthodologie et diffusion

Le projet RAMHO mêle en réalité histoire orale et recherche dans des fonds d'archives publics et privés. S'il est important dans de tels projets de définir un cahier des charges [8], nous ne donnerons ici qu'un bref aperçu du dispositif méthodologique qui fera l'objet d'un article à part entière. Considérant que la caméra pouvait être une source d'intimidation et conduire à des récits trop formatés, il a été décidé dès le début de ne pas filmer les entretiens et de privilégier ainsi l'enregistrement audio en précisant, avant chaque rencontre, qu'ils ne seraient pas diffusés de manière brute – que ce soit sous forme de fichier audio ou verbatim – faisant ainsi le choix de nous éloigner des projets classiques d'histoire orale.¹

Conduits par binôme ou trinôme, les entretiens suivent une sorte de canevas très général consistant à aborder le

parcours scolaire de l'interviewé, ses liens avec la musique, son parcours professionnel au sein des structures de recherche, le contexte institutionnel et les réseaux professionnels dans lequel il s'inscrit. Mais chaque entretien est préparé de manière individuelle à partir de recherches bibliographiques ou en archives afin de se familiariser avec le parcours de l'interviewé et de stimuler sa mémoire sur des événements spécifiques qu'il aurait pu omettre. L'interviewé est par ailleurs invité en amont à se plonger dans ses propres archives pour nous présenter éventuellement des documents inédits importants à ses yeux pour parler de son parcours.

Les entretiens ainsi semi structurés prennent la forme de longues conversations pouvant partir dans mille directions. Il arrive souvent que plusieurs entretiens soient conduits auprès d'une même personne. Les enregistrements audio sont archivés sur un serveur du laboratoire STMS au sein de l'Ircam ainsi que sur un disque dur externe. Il peut y avoir au cours des entretiens des redites, des hésitations, des contradictions, des passages « off » pour reprendre une expression courante dans le milieu journalistique... Les enregistrements ne sont donc pas éditables tels quels mais ils sont néanmoins intégralement transcrits. Les verbatim font ensuite l'objet d'un important travail d'éditionnalisation de la part des chercheurs. La version éditorialisée des entretiens, ajournée de quelques questions et commentaires visant à faciliter sa compréhension, est alors renvoyée à l'interviewé qui est invité à la relire, la corriger – en s'auto-censurant éventuellement ou en apportant de nouvelles précisions – pour finalement valider sa future publication par le biais d'une plateforme web qui sera dédiée au projet RAMHO. Celle-ci étant actuellement en phase d'élaboration, il serait prématuré d'en donner ici une description plus précise. En avril 2023, vingt personnes avaient été interviewées donnant lieu à déjà presque cent heures d'enregistrement audio.

2.3. John Chowning, un participant à part

Parmi les interviewés avec qui nous nous sommes déjà entretenus, Chowning représente un cas un peu particulier. Lors du colloque en hommage à son compagnon de route Jean-Claude Risset, il nous annonça qu'il allait revenir en France l'année suivante. Nous avons alors convenu de nous revoir à cette occasion pour parler avec lui de la création de l'Ircam. La pandémie provoquée par la COVID-19 empêcha tout voyage ce qui nous incita à organiser des visioconférences, chacun depuis son domicile (François-Xavier Féron à Pantin en France, Laura Zattra à Malo en Italie et John Chowning à Palo Alto sur la côte ouest des États-Unis) (Figure 1). Suivant ce dispositif, nous n'étions plus restreints par le temps et avons convenu tous ensemble de nous réunir à plusieurs reprises de sorte à revenir de manière chronologique sur l'ensemble de sa carrière. Finalement, neuf entretiens ont été conduits entre avril et juillet 2021

¹ Se référer par exemple à « Musique Mémoires de l'INA » (<https://entretiens.ina.fr/musiques-memoires>) ou « Music at MIT Oral History Collection » (<https://libraries.mit.edu/music-oral-history/>).

(Table 1)². Nous nous donnions rendez-vous généralement le jeudi à 17h00 pour nous qui vivons en France et en Italie, soit 08h00 pour Chowning qui vit en Californie. Nous avons collecté en tout 21h21 d'enregistrement ce qui donne, pour un entretien, une durée moyenne de 2h22. Nos rencontres étaient espacées généralement de deux semaines ce qui nous laissait du temps pour les préparer correctement en absorbant les nombreux articles et entretiens qui lui ont été consacrés.

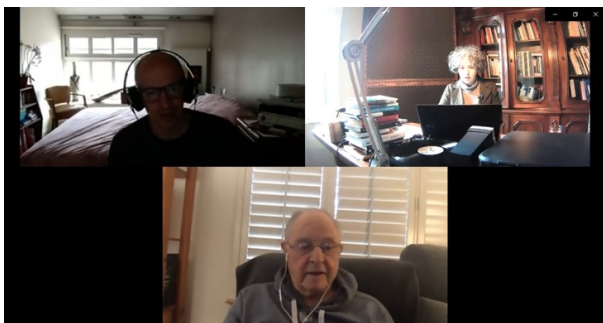


Figure 1. John Chowning lors du troisième entretien en visioconférence avec François-Xavier Féron (en haut à gauche) et Laura Zattra (en haut à droite).

Entretien 1 [JC 1]	2021-04-07	1:53:23
Entretien 2 [JC 2]	2021-04-28	2:31:40
Entretien 3 [JC 3]	2021-05-02	2:36:49
Entretien 4 [JC 4]	2021-05-13	2:47:45
Entretien 5 [JC 5]	2021-05-20	2:14:27
Entretien 6 [JC 6]	2021-06-17	2:02:56
Entretien 7 [JC 7]	2021-06-24	2:09:46
Entretien 8 [JC 8]	2021-07-22	2:20:59
Entretien 9 [JC 9]	2021-07-29	2:43:32

Table 1. Liste des entretiens menés avec John Chowning en visioconférence.

Si les sources bibliographiques ne manquent pas, les archives Chowning qui sont répartis dans différents structures et qui contiennent différents supports et contenus, représentent une source d'information tout aussi précieuse. Les enregistrements de ses œuvres, les données informatiques relatives à leur production et éventuellement les partitions sont organisées en grand partie par lui-même et stockées dans ses propres archives numériques. Certains documents papiers ou certaines photos qu'il a montrés au cours de nos entretiens étaient stockés dans son garage. L'ensemble de ces documents constitue donc ce que nous appelons l'archive privée de Chowning. Mais une grande partie de ses ressources est également conservée, depuis 2007, au sein de l'archive

SAILDART qui n'est pas en libre accès³. Ce fonds a été notamment exploré par Andrew Nelson pour reconstituer l'histoire du CCRMA dans un ouvrage monographique [15] qui est accompagné d'un site web à partir duquel il est possible d'accéder à l'ensemble des sources citées par l'auteur (<http://www.thesoundofinnovation.com/>). Le dossier associé à Chowning dans l'archive SAILDART contient entre autres des lignes de programme pour la synthèse, des calculs, mais aussi des textes ou des brouillons de lettres envoyés à des collègues externes ainsi que de simples messages émis entre collègues. Cette mine d'informations qui éclaire sur le quotidien du chercheur et sur l'ambiance qui régnait dans le laboratoire a été au cœur des investigations menées par Laura Zattra, Olivier Baudouin, Kevin Dahan et Matteo Meneghini pour reconstituer l'histoire de *Stria* [20] [23], une des premières commandes officielles passées par l'Ircam. Enfin, pour mener à bien cette recherche, d'autres archives ont été consultées, notamment celles du Stanford Daily en libre accès (<https://archives.stanforddaily.com/>), celles du Centre Pompidou et de l'Ircam mais aussi le fonds Jean-Claude Risset qui est préservé au sein du laboratoire PRISM à Marseille et qui a fait l'objet d'un précieux catalogue.

3. 1972-1973 – L'ANNEE SABBATIQUE DE CHOWNING EN EUROPE

3.1. Point sur la carrière de Chowning à la fin du premier semestre 1972

Notre récit débute au début de l'année 1972 au moment précis où Pierre Boulez contacte Jean-Claude Risset pour lui parler du futur centre de recherche acoustique de Beaubourg⁴. À des milliers de kilomètres de là, Chowning – avec qui Risset entretient déjà des liens très amicaux – est *assistant professor* à l'université Stanford. Voilà plusieurs années déjà qu'il a constitué autour de lui une petite équipe d'ingénieurs et de chercheurs travaillant sur la synthèse numérique du son en s'accaparant la nuit les systèmes informatiques du SAIL [JC 5]. Depuis 1966 il s'agit d'un PDP-10 fabriqué par la firme Digital Equipment Corporation (DEC). Chowning et Poole ont adapté pour cet ordinateur le programme Music IV issu des *Bell Labs* donnant ainsi naissance au programme Music X [1, p. 121].

En 1972, Chowning jouit déjà d'une reconnaissance importante tant le milieu de l'informatique musicale que celui de l'ingénierie sonore. Son algorithme de spatialisation sonore – capable de dessiner des trajectoires prégnantes à partir d'un dispositif de diffusion quadraphonique – avait été dévoilé dans un

² Les informations issues directement de nos entretiens sont référencées en indiquant, entre crochet, les initiales de John Chowning suivi d'un numéro indiquant l'entretien.

³ Nous avons eu l'occasion d'accéder aux fichiers stockés dans ces archives par le biais de la plateforme créée par Bruce Baumgart (<https://www.saildart.org/>) qui contient donc la plupart des documents (un peu moins d'un million de fichiers) du SAIL dont les messages envoyés sur un service de messagerie interne utilisé par les membres du laboratoire.

⁴ Risset dit avoir été contacté par Boulez dès 1970 [12, p. 12] mais c'est en réalité le 6 janvier 1972 que Boulez lui écrit depuis Londres : « Dear Mr. Risset, I had your name and address from Milton Babbitt, whom, I believe, you know very well. We are in the process of founding a centre for acoustical research in Paris to be opened in three years from now, and I would very much like to meet to discuss the possibilities of your collaboration. » [Arch. Risset – P51_004].

article paru dans le *Journal of the Audio Engineering Society* [3] et un brevet avait aussi été déposé [4]. À cette époque Chowning avait par ailleurs découvert la méthode de synthèse par modulation de fréquence qu'il présentait volontiers à différents spécialistes. Si cette découverte révolutionnaire n'avait pas encore fait l'objet de publications académiques, la firme Yamaha avait posé une option sur cette invention dès 1971 auprès de l'*Office of Technology Licensing* de Stanford que dirigeait Niels J. Reimers depuis sa création en 1969 [JC 4].

Commencée en 1970 et créée en 1971, *Sabelithe* est la première pièce de Chowning entièrement synthétisée par ordinateur. La création de *Turenas*, le 23 avril 1972, est un moment charnière dans sa carrière puisque cette deuxième pièce entièrement synthétisée par ordinateur est présentée dans le cadre de sa demande de titularisation afin de devenir *tenured professor* [JC 4]. Quelques semaines plus tard paraît le brevet pour son algorithme de spatialisation sonore alors qu'il s'apprête à entamer, en juillet 1972, une année sabbatique en Europe. Il se rend d'abord en Allemagne à l'invitation du compositeur György Ligeti qui venait de passer un semestre entier au sein du département de musique de Stanford en tant que professeur invité [JC 4].

3.2. Ligeti, une pièce centrale dans son réseau relationnel

Chowning et Martin Bresnick avait suggéré d'inviter Ligeti à Stanford pour pallier l'absence de Leland Smith alors en congés sabbatique. C'est ainsi que le compositeur hongrois séjourna en Californie de janvier à juin 1972 et qu'il assista d'ailleurs à la création de *Turenas* qui lui avait fait forte impression [JC 4]. Au mois de juillet, Ligeti invite alors Chowning à Berlin et organise un concert au cours duquel sont jouées *Sabelithe* et *Turenas*. Les deux hommes se rendent alors aux *Ferienkurse* de Darmstadt où Ligeti devait assurer des séminaires de composition durant une semaine. Chowning ne reste qu'une journée et rencontre à cette occasion pour la première fois Iannis Xenakis qui était lui aussi un des professeurs invités de cette édition [JC 2]. Parmi les élèves qui assistent au séminaire que Chowning donne le 22 juillet 1972, il nous semble important de souligner la présence de Gérard Grisey – qui posera les bases du courant spectral – et de Johannes Goebel [JC 5] – qui créera sur le label Wergo de la collection *Digital Music Digital* dédiée à la computer music et qui fondera l'*Institute for Music and Acoustics* du ZKM (Karlsruhe) ainsi que l'EMPAC (Troy). Suite à ce bref passage à Darmstadt, Chowning rejoint à Cannes (France) sa femme Elizabeth et ses deux enfants âgés respectivement de neuf et sept ans. Féru de voile depuis son plus jeune âge, il se procure un voilier et entame un long voyage en Méditerranée. Peu avant Noël la famille Chowning s'installe dans un hôtel sur l'île d'Ischia près de Naples qui deviendra leur camp de base [JC 4].

⁵ Citation originale : « Ligeti told me that when he returned at some point after he and I were in Darmstadt, he met Boulez at some point in – you know, one of their concerts or something –, and he told Boulez to pay attention to what was going on in California in computer music. So

3.3. Tournée des centres de recherche musicale et premiers contacts avec Boulez et son équipe

À Ischia, les enfants sont inscrits dans une école italienne et Chowning achève l'écriture de son fameux article sur la synthèse par modulation fréquence [5]. Il demande d'ailleurs à utiliser la machine à écrire de l'hôtel [JC 4] pour réaliser le manuscrit qu'il pense avoir envoyé dans un premier temps à Max Mathews aux *Bell Labs*. En ce début de l'année 1973, il entreprend une tournée des centres de recherche musicale suite à l'invitation faite par Lars-Gunaar Bodin – un des pionniers de la musique électronique en Suède – qui l'invite à Stockholm pour donner durant deux semaines des séminaires sur la synthèse numérique du son. Mais il se rend tout d'abord à Paris en janvier durant une ou deux semaines [JC 2]. Il revoit Xenakis à qui il pense avoir donné une copie de son manuscrit sur la synthèse par modulation de fréquence. Il visite le Groupe de Recherches Musicales (GRM) où il fait la connaissance de Bénédicte Mailliard avec qui il se liera d'amitié [JC 4]. Ce serait aussi à ce moment-là que Chowning aurait rencontré pour la première fois l'équipe mise en place par Boulez pour piloter la construction de l'Ircam qui sera finalement dissocié du futur Centre Beaubourg pour être enterré sous le parvis de l'église Saint-Merri où trône encore à ce moment-là une école.

Contrairement à ce qu'on pourrait croire, ce n'est pas Risset qui aurait introduit Chowning dans le projet Ircam mais c'est Ligeti qui aurait signalé à Boulez de s'intéresser aux recherches qui se faisaient à Stanford autour de Chowning. « Ligeti m'a raconté, qu'après notre séjour à Darmstadt, il avait rencontré Boulez lors d'un concert ou d'un autre événement. Il lui dit de prêter attention à ce qu'il se passait en Californie dans le domaine de la *computer music*. Il lui expliqua que c'était différent de tout ce qui se passait en Europe. »⁵ [JC 4].

Selon Chowning, Boulez n'était pas présent durant cette première rencontre à Paris mais Nicholas Snowman – le premier directeur artistique de l'Ircam – lui proposa de s'arrêter à Londres lorsqu'il reviendrait de Stockholm afin de discuter en personne avec Boulez. Chowning poursuit alors sa tournée en passant par la Belgique, puis la Hollande où il se rend au *Studio for Electronic Music* (STEM) de l'université d'Utrecht ; il y rencontre Gottfried Koenig ainsi que Barry Truax et Otto Laske. Il rejoint enfin Stockholm où il reste deux semaines. Sur le chemin du retour, il s'arrête donc à Londres et discute en tête-à-tête avec Boulez durant environ une heure [JC 4, 5]. « Durant la réunion, Boulez a posé quelques questions mais il a surtout écouté attentivement »⁶ [Courriel envoyé à L. Zattra le 8 juin 2019]. Suite à cette première rencontre, Boulez l'invite à se joindre à une réunion qui doit se dérouler à Paris quelques jours plus tard. Dès lors Chowning participera à de nombreuses réunions au cours desquelles seront définies les missions de l'Ircam.

he said that it is different than anything else that's going on in Europe. »

⁶ Citation originale : « At the meeting Boulez asked a few questions, but mostly listened carefully. »

3.4. Printemps-été 1973 : désenchantement et séminaire à l'abbaye Notre-Dame de Sénanque

Après cette tournée, Chowning retourne à Ischia auprès de sa famille où il apprend, au mois d'avril 1973, qu'il n'est pas titularisé à Stanford sans qu'aucune vraie raison ne lui soit donnée [JC 6]. « J'ai perdu mon poste de professeur assistant parce que je n'avais pas été titularisé comme professeur associé. J'ai donc été licencié. C'est un point de rupture dans une carrière universitaire. »⁷ [14, p. 6]. Cela peut sembler déconcertant quand on voit ce que Chowning avait accompli tant dans le domaine académique que musical et comment il était sollicité de toute part pour parler de ces recherches pionnières. Il apprendra plus tard que cela était dû à la nature conservatrice de la commission, qui ne prêtait pas attention à la recherche artistique et à la recherche en informatique appliquée aux arts. Ce n'est qu'avec le temps, grâce à une lettre de recommandation de Ligeti et à l'appui de John Pierce et Max Mathews que le même doyen de l'université Stanford (qui avait auparavant travaillé aux *Bells Labs* avec John Pierce) se rendra compte de la valeur des recherches de Chowning [JC 5].

Si cette nouvelle est un coup dur, cela ne va pas l'empêcher de poursuivre son congés sabbatique et de repartir en mer avec sa famille, longeant la côte italienne et se rendant à Malte où le rejoignent un neveu et un étudiant de Stanford. Alors que sa famille retourne à Ischia, Chowning, accompagné de ces deux jeunes matelots, navigue jusqu'en Grèce pour rejoindre Corfu [JC 4, JC 5].

Au début du mois de juillet 1973, Chowning prend un avion pour participer à un premier séminaire de réflexion organisé par l'Ircam dans l'abbaye Notre-Dame de Sénanque dans le sud de la France. Boulez, les futurs chefs de département (Jean-Claude Risset, Gerald Bennett, Luciano Berio et Vinko Globokar) et quelques personnalités extérieures (comme Milton Babbitt, Hans Peter Haller, György Ligeti ou Chowning) exposent et discutent de leurs recherches respectives dans le domaine musical. Peter Haller fait une démonstration du système de spatialisation HaLaPhon et dira à Chowning que celui-ci lui a volé la vedette en diffusant *Sabelithe* [JC 5]. Après ce séminaire qui représente une étape importante dans l'histoire de l'Ircam, Chowning réside quelques jours à Marseille chez Risset avant de rejoindre sa famille à Ischia. Il retourne enfin à Corfu pour convoyer son bateau à Ischia. L'année sabbatique touche à sa fin. En septembre 1973, la famille Chowning rentre en Californie.

⁷ Citation originale : « I lost my job because I didn't get promoted from Assistant Professor to Associate Professor. So I was let go. That's the break point in the ladder of academia. »

⁸ Citation originale : « Association with IRCAM I would propose to Boulez while I am in Europe that a time-sharing arrangement be worked out with his Institute de Recherche et Coordination Acoustique Musicale (IRCAM), where some faculty and staff of SCMP exchange

4. 1973-1974 – LES PRÉMICES DU CCRMA ET L'ACTE DE NAISSANCE PUBLIC DE L'IRCAM

4.1. Le CCRMA en gestation

En septembre 1973, alors que paraît son article sur la synthèse par modulation de fréquence [5], Chowning n'est donc plus en poste au sein de l'université Stanford qui lui confie néanmoins deux « quaters », c'est-à-dire deux charges de cours trimestriel : l'une à l'automne 1973 et l'autre à l'hiver 1974 [JC 5]. Chowning et ses fidèles compagnons de route (Andy Moorer, Loren Rush et John Grey) se lancent dans l'élaboration de deux projets de recherche complémentaires – l'un pour la National Science Foundation (NSF) et l'autre pour la National Endowment for the Arts (NEA) – avec la complicité de Leland Smith comme co-dépositaire principal en tant que professeur titulaire (pour une histoire détaillée du CCRMA se référer à [15]). Dans un document interne daté du 10 décembre 1973 concernant la demande pour la NSF, Chowning souligne que le *Stanford Computer Music Project* (SCMP) pourra se faire en collaboration avec l'Ircam : « Quand je serai en Europe, je proposerais à Boulez un partenariat avec son Institut de Recherche et Coordination Acoustique Musicale (IRCAM) de sorte à ce que des membres du corps enseignant et du personnel du SCMP et les membres du personnel de l'IRCAM puissent s'échanger leurs locaux. Pour Boulez, cela devrait l'intéresser car il aura besoin au début de ce type de 'ventilation'. Pour Stanford, cela permettrait à un membre du projet d'aller travailler à l'IRCAM pendant une période limitée et d'être ainsi dans un environnement sensiblement différent. Je suggérerais que de tels échanges puissent concerner également des étudiants diplômés de Stanford. »⁸ [Arch. SAILDART – PROP.JC,MUS 1973-12-10].

En juin 1974, Chowning et ses collègues soumettent donc à la NEA et la NSF deux projets respectivement intitulés « The computer music facility: a new musical medium » et « Computer simulation of music instrument tones in reverberant environment ». Chowning réside alors à Berlin en tant que boursier de la DAAD. Quant à l'Ircam, sa naissance avait été publiquement annoncée quelques mois plus tôt.

4.2. La conférence de presse annonçant la création de l'Ircam

Organisée par Boulez et son équipe au Théâtre de la Ville le 7 mars 1974, cette conférence de presse se déroule quelques semaines avant le décès du président

locations with some staff of IRCAM. From the point of view of Boulez it should be attractive since he will have initial need of this kind of 'ventilation'. From Stanford's position it would allow a member of the project to work at IRCAM for a limited period in a significantly different environment. I would suggest that such exchanges include Stanford graduate students as well. »

Georges Pompidou qui surviendra le 2 avril 1974. Une brochure d'une vingtaine de pages avait été distribuée [9] permettant ainsi de découvrir ce « petit Beaubourg » qui sera bâti en sous-sol sur la face nord de l'église Saint-Merri [Arch. CP – Rapport d'activité 1974]. L'Ircam est alors présenté comme un centre de recherche qui « offrira aux scientifiques de disciplines variées – acoustique, physique, psycho-acoustique, électronique, informatique, intelligence artificielle, neurophysiologie, psychologies, linguistique, sociologie – un large champ d'expérimentation, tandis que le musicien trouvera dans ce substrat de recherche une stimulation féconde pour l'intervention musicale. » [9, p. 9].

La brochure indique les fonctions de chaque personne. Pierre Boulez assure la direction générale ; Brigitte Marger est déléguée aux relations extérieures ; Nicholas Snowman, Yves Galmot et Max Mathews sont respectivement responsables artistique, administratif et scientifique ; Jean-Pierre Armand est coordinateur technique ; Vinko Globokar, Luciano Berio, Jean-Claude Risset et Gerald Bennett sont les chefs des départements « Instruments et voix », « Électro-acoustique », « Ordinateur » et « Diagonal » ; Diego Masson est responsable de l'unité mobile. Trois autres personnes forment le secrétariat. Les quatre départements sont décrits succinctement à travers les missions qu'ils rempliront alors que plusieurs pages sont consacrées au futur « Espace de projection » [Figure 2] « d'une extrême flexibilité acoustique et visuelle » [9, p. 20]. Il est mentionné que durant les deux premières années de fonctionnement – de 1976 à 1978 –, les invitations à des personnalités externes seront limitées pour mettre au point les équipements ; à partir de 1978 une unité pédagogique accueillera des étudiants et des stagiaires.

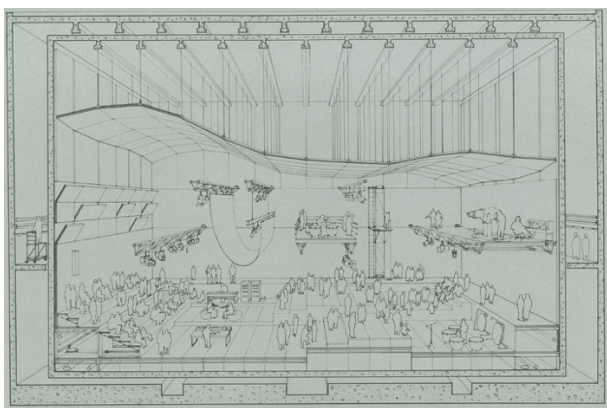


Figure 2. Croquis de l'Espace de projection tel qu'il avait été conçu originellement [9, p. 21].

Suite à cette conférence de presse qui a réuni environ un millier de personnes, la première chaîne de l'Office de radiodiffusion-télévision française (ORTF) diffuse le 23 mai 1974 un reportage sur l'institut [Arch. CP – Rapport d'activité 1974, p. 35-36]. À la même période – du 13 au 15 mai –, l'Ircam organise une de ses premières

manifestations scientifiques : un séminaire sur la recherche instrumentale réunissant des compositeurs, des instrumentistes, des acousticiens, des ingénieurs et des facteurs d'instrument [Arch. CP – Rapport d'activité 1974, p. 35-36].

4.3. Chowning de retour en Europe comme boursier du DAAD

Grâce à une bourse du DAAD – l'office allemand d'échanges universitaires – obtenue en partie avec l'appui de Ligeti, Chowning s'installe à Berlin à la fin du mois de mars et réside durant une année entière à l'Académie allemande des arts. Durant cette période il retourne à Stanford de manière ponctuelle en juillet ainsi que pour les fêtes de fin d'année. Lorsqu'il arrive à Berlin, son brevet sur la synthèse FM avait déjà été déposé [6] alors que les dossiers de demande de financement pour la NEA et la NSF étaient toujours en cours d'élaboration. C'est une année de transition, tant sur le plan personnel que professionnel, qui s'opère dans la vie de Chowning durant cette résidence. Il rencontre à Berlin des ingénieurs de Yamaha ; il continue à participer aux réunions de préparation de l'Ircam ; il est dans l'attente de ces demandes de subventions...

À l'Université Technique de Berlin, Chowning a accès à un ordinateur PDP-10 mais celui-ci est dépourvu de convertisseurs de son. Cela l'incite à étudier les processus déterministes en travaillant finalement que sur papier [19]. Il réfléchit en profondeur à la manière d'appliquer la technique FM à l'échelle d'une composition et découvre notamment qu'un rapport de fréquence entre une porteuse et une modulante s'approchant du nombre d'or produit des composantes qui sont également des puissances de la section d'or. Ce principe sera au cœur du processus compositionnel de *Stria*.

Si Chowning n'assiste pas à la conférence de presse de l'Ircam – car il n'arrive en Europe que quelques jours plus tard –, il continue à participer à plusieurs séminaires dont un organisé à Baden-Baden où le reste de l'équipe du SCMP – Grey, Moorner et Rush – avait été convié [Courriel envoyé à L. Zattra le 8 juin 2019]. Boulez sonde de manière informelle Chowning pour prendre la direction du futur département pédagogique : « Je me souviens que cela m'avait laissé perplexe parce qu'il ne s'agissait pas d'une proposition officielle. Ça s'était fait peut-être par l'intermédiaire de Snowman – je ne m'en souviens plus – mais je ne pense pas qu'il y ait une trace écrite. C'était comme si j'allais être engagé pour diriger le département pédagogique [...] Et puis cette perspective fut abandonnée lors de la deuxième réunion de Sénanque il me semble, et [cela] a beaucoup contrarié Jean-Claude [Risset]. Je pense qu'ils avaient besoin d'avoir plus de personnalités françaises. »⁹ [JC 5]. Finalement Boulez se tournera en effet vers Michel Decoust qu'il contactera en décembre 1974.

⁹ Citation originale : « I remember that puzzled me because it was not a formal offer. 'Would you be interested in the...?', and maybe it was through Snowman, I can't even remember, but I don't think anything was ever written, but it was like I would be hired to head the Education

Department [...] And then the whole subject disappeared, I think at the second Sénanque Meeting and [that] very much upset Jean-Claude [Risset]. [...] I think they needed to have more a French representation. »

La seconde retraite organisée par l'Ircam dans l'abbaye Notre-Dame de Sénanque se déroule du 8 au 13 Juillet 1974. L'équipe dirigeante est au complet et plusieurs personnalités extérieures sont à nouveau conviées : au côté de Chowning figurent ainsi John Grey, Andy Moorer, Tom Bever, Clytus Gottwald, Guido Guiducci et Jim Lawson [Arch. Ircam – « L'Ircam et le VIIe plan », juin 1974]. Dans ce même document, il est indiqué que le séminaire de Sénanque « permettra de sélectionner les projets qui seront retenus pour les cinq premières années du fonctionnement de l'Ircam 1976-1981 ».

5. 1974-1975 - LES DÉBUTS DE L'IRCAM ET L'AVÈNEMENT DU CCRMA

5.1. Les premiers événements culturels et pédagogiques de l'Ircam

L'année académique 1974-75 peut être considérée comme l'année zéro de l'Ircam. C'est dans le cadre du festival d'Automne et des Semaines musicales internationales de Paris que l'institut organise au théâtre d'Orsay, les 19, 20, 21 et 23 octobre 1974, ses premiers concerts. Chaque soirée met en valeur l'un des départements de recherche. En tant que responsable du département « Ordinateur », Risset organise une soirée sur le son numérique, avec des œuvres de John Rogers, Jonathan Harvey ainsi que sa *Computer Suite for Little Boy* et *Turenas* de Chowning. Ce dernier assiste à cette représentation et se souvient avoir été marqué par la manière dont Risset introduisait les œuvres au public et parlait des possibilités de la synthèse sonore par ordinateur. De retour à Berlin, il écrit d'ailleurs une lettre émouvante à l'attention de Max et Marge Mathews et de Rozenn Risset, lettre dans laquelle il exprime toute son admiration pour son ami [Figure 3].

Du côté de la pédagogie, l'Ircam assure du 17 au 21 mars 1975, à la demande de l'IRIA, une série de cours sur l'informatique musicale. Quelques semaines plus tard, le 3 mai 1975, est organisé un séminaire sur le thème « Linguistique et Musique ».

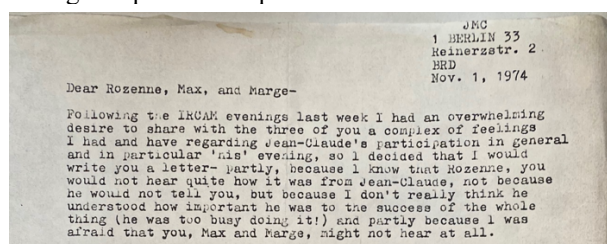


Figure 3. Début de la lettre écrite par Chowning à Rozenn Risset, Max et Marge Mathews après la série de concerts organisée par l'Ircam en octobre 1974 au Théâtre d'Orsay [Arch. Chowning].

¹⁰ Citation originale : « The researchers are trying to raise funds to establish a \$600,000 international computer music center here. According to Moorer, the group must raise \$120,000 in outside donations in order to receive an equal grant from the National Endowment for the Arts (NEA). The National Science Foundation will

5.2. Consolidation des liens entre Paris et Stanford

Au moment où se déroule la série de concerts au Théâtre d'Orsay, l'Ircam n'est qu'un immense trou. Le projet architectural originel – tel qu'il était présenté dans la première brochure [9] – dut être totalement revu pour abaisser les coûts notamment à cause de la crise économique induite par le premier choc pétrolier. Un nouveau permis de construire avait donc été déposé. Les travaux de construction du bâtiment semblent débuter au début de l'année 1975.

Les liens entre l'Ircam et l'université Stanford perdurent comme en témoigne cet extrait issu d'un article de presse paru le 5 mars 1975 dans le *Stanford Daily* au sujet du CCRMA : « Les chercheurs tentent de collecter des fonds pour créer un centre international de *computer music* d'une valeur de 600 000 \$. Selon M. Moorer, le groupe doit réunir 120 000 \$ de dons extérieurs afin de recevoir une subvention équivalente de la part de la *National Endowment for the Arts* (NEA). La *National Science Foundation* fournira le reste une fois que la NEA aura réuni les 120 000 \$. L'université approuve le projet, mais sans apporter de soutien financier, a déclaré M. Moorer. [...] Si un centre d'informatique musicale est créé ici, une étroite collaboration sera probablement mise en place avec un institut de recherche en musique et en acoustique dont l'ouverture est prévue à Paris en 1976. »¹⁰ [Arch. Stanford Daily – 5 March 1975].

C'est aussi en mars 1975 que se termine la résidence de recherche de Chowning à Berlin. Sa carrière prend alors son envol : un accord de licence autour de la synthèse par modulation de fréquence est conclu entre Yamaha et Stanford le 19 mars 1975 et, quelques semaines plus tard, la NSF fait savoir que le projet déposé par Chowning et ses collègues avait été accepté. En juin, c'est au tour de la NEA d'annoncer la même bonne nouvelle, actant ainsi la naissance du CCRMA. L'université Stanford qui, rappelons-le, n'avait pas souhaité titulariser Chowning, se trouve alors dans une situation quelque peu embarrassante et lui attribue le statut d'associé de recherche pour finalement le titulariser trois ans plus tard, en 1978.

Si l'avenir de Chowning se profile donc en Californie, ses liens avec l'Ircam ne sont pas pour autant rompus puisqu'il revient en France pour assister avec d'autres membres de l'Ircam, les 8 et 9 juillet 1975, à un important séminaire à Angers portant sur la recherche musicale en France [22]. La gronde commençait à se faire sentir, les autres centres se sentant lésés au regard des montants et des statuts alloués à l'Ircam. « Cette réunion avait surtout un but diplomatique et psychologique, *se souvient Jean Maheu* [qui était alors directeur de la musique au sein du Ministère de la culture] : essayer de décriper mais il était bien évident que l'Ircam allait avoir, avait virtuellement sa propre politique, ses propres orientations » [10].

supply the remainder once the NEA matches the \$120,000. The University approves the project, but provides no financial support, Moorer said. [...] If a computer music center is established here, it will probably work closely with an institute for research in music and acoustics scheduled to be set up in Paris in 1976. »

Quelques jours plus tard, du 30 juillet au 12 août, une douzaine de membres de l'équipe Ircam – dont les cinq chefs de département, la pédagogie ayant été créée depuis – séjourne à Stanford pour se familiariser avec les rudiments de la synthèse sonore par ordinateur [Figure 4]. Le PDP-10 du SAIL est utilisé par ces stagiaires un peu particuliers entre 5h et 9h du matin environ [JC 6] pour laisser ensuite la place aux chercheurs du laboratoire.

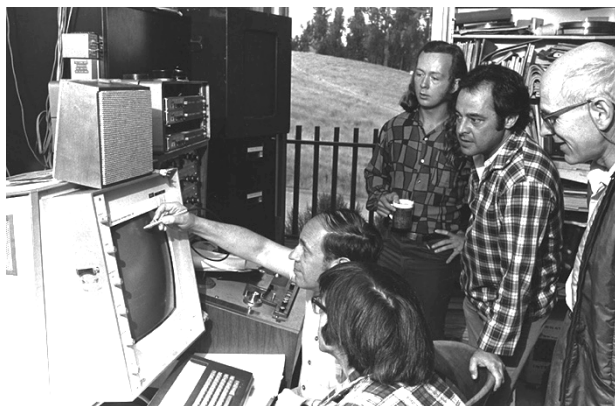


Figure 4. Pierre Boulez entouré de Andy Moorer, John Chowning, Max Mathews (debout) et Steve Martin (assis) lors du stage d'été organisé par le CCRMA pour l'équipe de l'Ircam [Arch. Chowning, photo : José Mercado].

6. 1975-1976 – À LA RECHERCHE DU TEMPS REEL

Les activités de recherche et de création au sein de l'Ircam et du CCRMA prennent leur envol durant l'année académique 1975-76. L'Ircam occupe dorénavant les locaux de la bibliothèque municipale qui faisait partie de l'ancien complexe scolaire Saint-Merri et qui surplombe alors le bâtiment souterrain en construction. L'équipement du département « Ordinateur » est mis en place au cours du second semestre 1975 [Arch. CP – Rapport d'activité 1975, p. 46] mais le PDP-10 – équipé d'un système identique à celui du CCRMA – ne sera opérationnel selon Chowning qu'à partir de 1976 [JC 5].

Vingt-deux personnes travaillent au sein de l'institut en 1975 contre dix, fin 1974 [Arch. CP – Rapport d'activité 1975, p. 46]. Parmi les nouvelles recrues figure l'ingénieur Giuseppe Di Giugno qui, à la demande de Berio, a rejoint le département « Électro-acoustique ». Di Giugno s'attèle, dès 1976, à la construction d'un processeur de sons numériques en temps réel contenant 256 oscillateurs et piloté par un PDP-11 [2, p. 3]. Ce premier processeur – la 4A – qui est opérationnel en 1976 donnera lieu à de nouveaux développements – les processeurs 4B et 4C – avant de donner naissance à la 4X au début des années 1980.

¹¹ Citation originale : « So, with the Samson box, there were hundreds of hours of music and more than a hundred compositions. I'm guessing maybe thousands of hours of music and hundreds of compositions. And so when we retired the Samson box [they were] no longer functioning. But we had all these files that had been used to run the Samson box to

Enfin, c'est en 1975 que paraît le premier livre de l'Ircam intitulé *La musique en projet* [Figure 5]. Il contient certains des discours qui avaient été donnés lors de la conférence de presse du 7 mars 1974. Marger explique dans l'avant-propos de cet ouvrage que si l'Ircam n'est pas encore véritablement né – le Centre Pompidou étant toujours en cours de construction – la production sonore et la création musicale avaient déjà leur place. Il était par conséquent « important que le public soit associé dès le départ aux réflexions et aux premières recherches qui justifient la création d'un Institut comme l'I.R.C.A.M. » [11, p. 8]. Grâce à son unité mobile, l'institut assure en effet la production de concerts mais aussi de conférences comme à Bruxelles en octobre 1975, à Grenoble en novembre 1975 ou à Metz en février 1976.



Figure 5. Couverture de livre *La musique en projet* paru fin 1975 [11].

Du côté de la Californie, c'est en 1976 que David Poole fonde la société *Foonly Computer* dans laquelle travaille Peter Samson. Le CCRMA sous-traite avec cette société afin que soit développé un outil de synthèse en temps réel. La Samson box – véritable totem du CCRMA à l'instar de la 4X pour l'Ircam – sera opérationnelle à partir du mois d'octobre 1977 [1, p. 272] et sera utilisée jusqu'en 1992. « La Samson box a permis d'avoir des centaines d'heures de musique et plus d'une centaine de œuvres. Je dirais même des milliers d'heures de musique et des centaines de compositions. Lorsque la Samson box fut mise hors service, elle ne fonctionnait plus. Mais nous étions en possession des fichiers qui avaient été utilisés pour la faire fonctionner et produire les musiques. Bill Schottstaedt, qui faisait partie de notre équipe, a donc développé un émulateur, de sorte que tout ce qui avait été synthétisé par la Samson box pouvait être recalculé avec un taux d'échantillonnage plus élevé que dans les standards audio actuels, comme 32 [bits] virgule flottante, etc. La qualité est donc très élevée. »¹¹ [JC 7].

En août 1976, alors que Boulez dirige à Bayreuth *Der Ring des Nibelungen* de Richard Wagner, mis en scène

produce the music. So, Bill Schottstaedt was on our staff, and he wrote a software emulator, so that anything had been generated on the Samson box could be recomputed and at a high sampling rate, and at current audio standard samples, like 32, that floating point sample, etc. So very high quality. »

par Patrice Chéreau, le CCRMA organise dans l'auditorium Dinkespiel son tout premier concert : « An evening of computer Music and Film ». Si au cours de cette année académique, les deux centres de recherche se concentrent essentiellement sur leur propre développement, les liens subsistent. Berio passe d'ailleurs une commande d'œuvre, au nom de l'Ircam, à Chowning.

7. 1976-1977 – DES AMÉRICAINS A PARIS

7.1. Le nouveau statut associatif de l'Ircam

Alors que le bâtiment principal du Centre Pompidou est en voie d'achèvement, Boulez connaît une fin d'année jubilatoire. Le 10 décembre 1976, il donne sa leçon inaugurale au Collège de France ; quelques jours plus tard, du 13 au 18 décembre, l'Ircam organise au Théâtre National Populaire (TNP) de Villeurbanne une session de conférences et concerts au cours de laquelle l'Ensemble Intercontemporain – qu'il avait fondé cette même année – se produit en public ; enfin, il est décidé que l'Ircam ne sera plus un département du Centre Pompidou mais un organisme associé disposant d'un budget spécifique et d'un statut juridique d'association [Arch. CP – Rapport d'activité 1975, p. 13].

Ce changement de statut permet d'échapper à certaines lourdeurs administratives et de faciliter, entre autres, l'embauche de chercheurs et techniciens étrangers dont beaucoup viennent des États-Unis comme Andy Moorer qui, après avoir participé activement à la création du CCRMA, sera employé par l'Ircam à partir de juin 1977. Ce départ est soutenu par Chowning comme en témoigne cet extrait de lettre du 25 mai 1977 à l'attention des services administratifs de l'université Stanford : « Je demande à ce que James A. Moorer puisse bénéficier d'un congé entre le 14 juin 1977 et le 1^{er} septembre 1978. Durant cette période, Moorer travaillera pour l'IRCAM à Paris en tant que conseiller scientifique. Il sera responsable du développement, sur le système PDP-10 de l'IRCAM, d'un logiciel similaire à celui du A.I. Lab. Il partagera aussi son expertise scientifique. »¹² [Arch. SAILDART – ANDY[*TXT*,*JC*]:1977-05-25].

7.2. *Stria*, une voix parmi les voix

En 1977, l'Ircam organise un nombre important de concerts dans le cadre du cycle « Passages du XX^e siècle » dont la première partie se déroule de janvier à juillet [16] et la seconde partie, de septembre à décembre [17]. Le bâtiment souterrain ouvre ses portes durant l'été 1977 mais l'installation prendra du temps et l'Espace de projection – encore en chantier se sera inauguré qu'un an plus tard. Berio conçoit pour cette inauguration partielle de l'institut un spectacle audiovisuel intitulé *La Voix des*

voies qui est diffusé au Centre Pompidou du 29 septembre au 29 octobre 1977 : « Il s'agit à la fois d'un spectacle, d'un documentaire et d'un concert. De temps en temps sur le parcours de l'exposition s'ouvrent des 'fenêtres' permettant de présenter tantôt des œuvres nouvelles (avec ou sans musiciens) commandées pour l'occasion par l'IRCAM à John Chowning, Michel Decoust, Edison Denisov, Jacob Druckman, Jean-Claude Eloy, Henri Pousseur –, tantôt d'œuvres anciennes comme *Kontakte* de Karlheinz Stockhausen » [17, p. 8].

Stria est créée publiquement dans ce contexte le 13 octobre 1977. L'idée initiale remonte, comme nous l'avons évoqué, à son séjour de recherche passé à Berlin. La composition matérielle de l'œuvre se déroule en 1977, essentiellement à partir de juillet comme l'attestent les données informatiques archivées [20] mais certaines sections ont été programmées plus tôt avec l'aide d'Andy Moorer avant que celui-ci ne s'installe à Paris en juin [JC 3]. Lorsque l'Ircam demande à Chowning le nom de son œuvre afin de l'intégrer à la brochure [17], il propose dans un premier temps *Kilo Music* puis se ravise et envoie un télégramme – en juillet ou août 1977 d'après ses souvenirs – pour signaler que l'œuvre s'intitulera finalement *Stria* [JC 3].

Chowning aurait souhaité synthétiser la pièce en temps réel avec la Samson Box, mais en raison du retard pris dans la construction de cette machine, elle fut synthétisée en temps différé. Les différentes sections composant l'œuvre ont été converties au CCRMA à partir d'un PDP-10 puis enregistrées sur autant de bandes analogiques quatre-pistes durant le mois de septembre. Elles furent en revanche assemblées à l'Ircam avec l'aide de Moorer peu de temps avant la création [23]. Pour l'anecdote, Chowning se souvient avoir déposé une impression de l'intégralité du code de cette première version à Ivanka Stoïanova alors en charge de la documentation de l'Ircam. Malheureusement ce précieux document n'a jamais été retrouvé [JC 8]¹³.

8. CONCLUSION

Le projet RAMHO dans lequel s'inscrit cette recherche vise à allier histoire orale [18] [21] et dépouillement d'archives de manière à éclairer l'histoire de l'acoustique musicale et de la recherche musicale en France au cours de la seconde moitié du XX^e siècle. Les archives sont précieuses tant pour identifier des informateurs privilégiés que pour préparer de manière personnalisée chaque entretien. En adoptant une telle approche, nous sommes parvenus à reconstituer de manière assez précise – avec la complicité de Chowning – la chronologie des événements qui ont étayé sa vie professionnelle et personnelle.

Au fil de nos entretiens, nous avons réalisé que l'Ircam et le CCRMA partageaient véritablement une histoire

¹² Citation originale : « I request that James A. Moorer be granted leave from June 14 1977 until 1 September 1978. During this period Moorer will act as scientific adviser for the IRCAM in Paris. His responsibilities at IRCAM will include the development of A.I. Lab

type software on the PDP-10 system at IRCAM in addition to providing and scientific advice. »

¹³ Les différentes versions de *Stria* existantes ont suscité des investigations d'ordre philologique qui ont permis de reconstruire l'œuvre dans son intégralité [20].

commune et que Chowning était le point d'ancrage entre ces deux centres. Après avoir rencontré Boulez à Londres pour la première fois durant l'hiver 1973, Chowning ne cessera pas de collaborer avec l'Ircam, alors même que l'institut était en phase de structuration. Il participe aux réunions et séminaires de préparation à Paris, à Baden Baden et à Sénanque. Il est joué lors de la première série de concerts que l'Ircam organise en octobre 1974 avant d'être sondé informellement par Boulez pour prendre la direction du futur département « Pédagogie ». Il organise un stage d'été de *computer music* à Stanford en 1975 à l'attention du personnel de l'Ircam. Le département « Ordinateur » que dirige Risset est pourvu du même équipement que celui du CCRMA. Enfin, Chowning figure parmi les tout premiers compositeurs à qui l'institut français commande une œuvre. L'ombre de Chowning semble planer constamment au-dessus de l'Ircam jusqu'à la création de *Stria* en octobre 1977. Mais cette connivence ne s'arrête pas là puisque Risset lui propose alors de revenir l'année prochaine en tant qu'artiste-chercheur en résidence. La suite de ce récit sera contée dans un nouvel article qui sera soumis lors du prochain appel à communication pour les Journées d'Informatique Musicale 2024 à l'occasion du 90^e anniversaire de Chowning.

9. REMERCIEMENTS

John Chowning nous a fait l'honneur de converser avec nous pour revenir sur son incroyable carrière scientifique et musicale. Nous tenons à le remercier très chaleureusement pour sa générosité, sa confiance et ses encouragements. Nous souhaitons aussi remercier le Ministère de la Culture qui a permis de recruter Justin Ratel comme vacataire de recherche. Un grand merci à lui pour avoir réalisé avec un très grand soin la transcription de nos nombreux entretiens avec Chowning.

10. REFERENCES

1. Baudouin, O. *Pionniers de la musique numérique*, Delatour France, 2012.
2. Bennett, G. « La recherche à l'IRCAM en 1977 », *Rapports IRCAM*, n° 1, 1978.
3. Chowning, J. « The Simulation of Moving Sound Sources », *Journal of the Audio Engineering Society*, vol. 19 n°1, 1971, p. 2-6 (Ré-impression dans *Computer Music Journal*, vol. 1 n°3, 1977, p. 48-52).
4. Chowning, J. « Method and apparatus for simulating location and movement of sound », United States Patent 3 665 105, filed Mar. 9, 1970, accepted May 23, 1972.
5. Chowning, J. « The Synthesis of Complex Audio Spectra by Means of Frequency Modulation », *Journal of the Audio Engineering Society*, vol. 21 n° 7, September 1973, p. 526-534.
6. Chowning, J. « Method of synthesizing a musical sound », United States Patent 4 018 121, filed May 2, 1975, accepted Apr. 19, 1977.

7. Chowning, J. « *Turenas: the realization of a dream* », *Journées d'Informatique Musicale*, Université de Saint-Etienne, France, 25-27 mai, 2011.
8. Descamps, F. « Archiver la mémoire. De l'histoire orale au patrimoine immatériel », Paris, éditions de l'EHESS, 2019.
9. *IRCAM* [Brochure de présentation], Centre Beaubourg - Ircam, 1972.
10. Kervran, P. « Pierre Boulez à tout prix – La naissance de l'Ircam », in *La Fabrique de l'histoire* [émission radiophonique], France culture, 24 décembre 2010.
11. *La musique en projet*, Paris, Gallimard, 1975.
12. Leroux, G. et Madlener, F. « À la Recherche du Temps Réel – Pierre Boulez et la fondation de l'Ircam. Entretien avec Jean-Claude Risset », *L'étincelle : Journal de la création à l'Ircam*, n° 12, 2014, p. 12-15.
13. Mathews, M. V. « The Digital Computer as a Musical Instrument », *Science, New Series*, vol. 142 n° 3592, Nov. 1, 1963, p. 553-557.
14. Means, L. « Interview with John Chowning by Loren Means », *YLEM Journal: Artists Using Science and Technology*, n° 25, 2005, p. 6-8.
15. Nelson, A. J. *The Sound of Innovation: Stanford and the Computer Music Revolution*, The MIT Press, 2015.
16. *Passage du XX^e siècle, 1^{ère} partie* [Janvier/Juillet 1977], Paris, IRCAM - Centre Georges Pompidou, 1976.
17. *Passage du XX^e siècle, 2^{ème} partie* [Septembre/Décembre 1977], Paris, IRCAM - Centre Georges Pompidou, 1977.
18. Ritchie, D. *The Oxford Handbook of Oral History*, Oxford University Press, 2011
19. Roads, C. « John Chowning on Composition », in C. Roads, ed. *Composers and the Computer*, Los Altos, William Kaufman, 1985, p. 16-25.
20. « The Reconstruction of *Stria* », *Computer Music Journal*, vol. 31 n° 3, Fall 2007.
21. Thompson, P., Bornat, J. *The voice of the past: Oral History*, Oxford University Press, 2017
22. Veitl A. *Politiques de la musique contemporaine : Le compositeur, la « recherche musicale » et l'État en France de 1958 à 1991*, Paris, L'Harmattan, 1997.
23. Zattra L. « John Chowning – *Stria* », *ANALYSES – Œuvres commentées du répertoire de l'Ircam* [En ligne], 2016. URL : <https://brahms.ircam.fr/fr/analyses/Stria/>

RECHERCHE ET ACOUSTIQUE MUSICALES A MARSEILLE ENTRE 1968 ET 1989

Baptiste Bacot
STMS – CNRS, Ircam,
Sorbonne Université
baptiste.bacot@gmail.com

Vincent Tiffon
Aix Marseille Univ, CNRS,
PRISM, Marseille, France
tiffon@prism.cnrs.fr

François-Xavier Féron
STMS – CNRS, Ircam,
Sorbonne Université
feron@ircam.fr

RÉSUMÉ

L'émergence de la recherche musicale et de l'acoustique musicale à Marseille est peu connue, bien que le caractère pionnier de l'activité phocéenne ait été déterminant pour le développement de ces travaux à l'échelle de la France. Dans cet article, nous rendons compte des premiers résultats du projet RAMHO (Recherche et acoustique musicales en France : une Histoire Orale) en étudiant l'écosystème marseillais entre 1968 et 1989. Nous traiterons ainsi trois institutions de recherche et de création, le Groupe de Musique expérimentale de Marseille (GMEM), le laboratoire Musique et Informatique de Marseille (MIM), ainsi que le Laboratoire de Mécanique et d'Acoustique (LMA) qui entretenaient tous trois, par le biais du chercheur et compositeur Jean-Claude Risset, des relations étroites. À travers les données d'une enquête fondée sur l'histoire orale et les archives, cet article s'intéresse aux liens scientifiques, institutionnels, technologiques et artistiques qui ont permis la naissance et l'établissement d'une communauté de recherche sur le son qui est aujourd'hui toujours centrale en France.

1. INTRODUCTION

Au cours de la seconde moitié du XX^e siècle, les pratiques musicales ont été totalement bouleversées par l'émergence et la démocratisation des technologies sonores et le développement des connaissances sur le son. L'acoustique musicale et la recherche sont liées par l'intégration des nouvelles technologies dans la musique [1]. Elles sont devenues des activités interdisciplinaires à part entière à partir desquelles scientifiques (chercheurs, ingénieurs ou informaticiens) et musiciens (compositeurs, chefs d'orchestre ou instrumentistes) sont amenés à collaborer étroitement au sein de nouveaux lieux. La ville de Marseille a joué un rôle pionnier dans l'émancipation de ces disciplines, en premier lieu parce qu'elle a été choisie pour l'implantation du premier laboratoire CNRS en province, le CRSIM (Centre de Recherches Scientifiques, Industrielles et Maritimes). Son directeur, François Canac (1886-1969), organise le premier colloque international d'acoustique musicale en 1958. Par ailleurs, le

conservatoire de Marseille ouvre la toute première classe d'électroacoustique en France en 1968. Si la recherche en musique et l'acoustique musicale sont bien deux champs autonomes qui ont leur finalités propres (respectivement la création artistique et la connaissance des phénomènes physiques), elles s'influençaient mutuellement par le biais d'institutions, de projets et d'acteurs scientifiques et culturels. C'est sur ces fondements que trois structures marseillaises ont contribué de façon significative au développement de la recherche sur la musique et le son au tournant du XXI^e siècle : le Laboratoire de Mécanique et d'Acoustique (LMA, descendant du CRSIM), le Groupe de Musique Expérimentale de Marseille (GMEM) et le laboratoire Musique et Informatique de Marseille (MIM).

Comment ces structures ont-elles contribué concrètement au développement de la recherche musicale et de l'acoustique musicale à Marseille, et quelles relations entretenaient-elles, non seulement entre elles, mais aussi avec les autres institutions locales et nationales chargées d'une mission pédagogique, artistique ou scientifique ? Quelles étaient les conditions d'émergence de ces approches pionnières, et que sont-elles devenues ? Nous commencerons par détailler le volet « Marseille » du projet RAMHO, sa méthode et ses données issues d'une enquête d'histoire orale et de recherches dans des fonds d'archives. Dans un second temps, l'écosystème de la recherche en acoustique et de la création musicale sera analysé dans son contexte historique à travers les figures de Marcel Frémiot et Jean-Claude Risset.

2. LE PROJET RAMHO ET SON VOLET MARSEILLAIS

2.1. Genèse et objectifs

Le projet RAMHO (Recherche et Acoustique Musicales : une Histoire Orale) a été esquissé au sein de l'équipe Analyse des Pratiques Musicales de l'Ircam par Nicolas Donin, François-Xavier Féron, Vincent Tiffon et Laura Zattra, alors que se tenait en mai 2018 le colloque-hommage à Jean-Claude Risset (1938-2016), auquel de nombreux pionniers de l'informatique et de l'acoustique musicale ont assisté. Il était initialement question de s'intéresser à l'histoire de l'Ircam, mais le champ d'étude s'est rapidement étendu à l'acoustique musicale puis à la

recherche musicale en général, tant les frontières entre ces champs interdisciplinaires s'avéraient poreuses. La France a joué un rôle historique dans l'institutionnalisation de ces recherches en permettant la création de laboratoires, studios et instituts liés à la science des sons au prisme de la musique et des technologies.

L'année 1958 marque à cet égard un tournant décisif. À Paris, Pierre Schaeffer (1910-1995) imagine de nouvelles missions pour son Groupe de Recherche en Musique Concrète (GRMC) qu'il rebaptise alors Groupe de Recherches Musicales (GRM). À Marseille, François Canac organise quant à lui, sous l'égide du CNRS, le premier congrès d'acoustique musicale. L'émancipation des recherches sur le son musical se prolonge un peu au-delà des années 1980 et sur l'ensemble du territoire, par la fondation de nombreuses structures dont voici, pêle-mêle, quelques acronymes : GRM, LAM, EMAMu, GAI, CIRM, GMEM, GMEB, LAMU, IRCAM, ACROE, LIMCA, GRAME, MIM ou CICM.

Comment ces structures ont-elles été mises en place et comment fonctionnaient-elles ? Comment se positionnaient-elles les unes par rapport aux autres et dans quelle mesure la recherche scientifique et la production artistique se rencontraient-elles dans le cadre de leur activité ? Pour tenter de répondre à ces questions et souligner le dynamisme de ce réseau de recherche interdisciplinaire, le projet RAMHO rassemble aujourd'hui des chercheurs issus de plusieurs laboratoires français et étrangers qui se chargent de réunir des données relatives à la naissance et au développement de ces structures.

2.2. Méthodologie et diffusion

Le projet RAMHO adopte les méthodes de l'histoire orale comme point de départ de ses recherches : il se fonde sur une série d'entretiens qui sont transcrits et dont les données sont croisées entre elles, ainsi qu'avec des sources écrites incluant des fonds d'archives publiques et privés. Conduits en binôme ou en trinôme par les auteurs de l'article, les entretiens suivent tous un canevas général qui vise à retracer la jeunesse de l'interlocuteur, ses liens avec la musique, son parcours et ses réseaux professionnels, le ou les contextes institutionnels qu'il a connus. La phase de préparation de l'entretien permet de prendre connaissance des travaux de la personne interrogée afin d'établir un lien de qualité avec elle, et surtout d'orienter le questionnaire sur des faits, des détails ou des interprétations jusqu'alors pas ou peu connus [19], autrement dit de vérifier la chronologie de certains événements afin de n'omettre aucun aspect des recherches considérées. L'interviewé est par ailleurs invité en amont, s'il le souhaite, à se plonger dans ses propres archives pour présenter des documents en rapport avec la ou les institutions de référence.

Concernant le système de captation, nous avons choisi d'écarter l'enregistrement vidéo : nous avons souhaité un dispositif plus léger, plus intimiste, afin de créer les conditions optimales d'un échange en toute confiance, notamment sur les sujets plus personnels, ceux qui

relèvent de la première partie des entretiens. Nous avons donc privilégié l'enregistrement audio en nous engageant à ne pas diffuser les enregistrements ni les verbatim, faisant ainsi le choix de nous éloigner d'une démarche parfois retenue pour ce type de projets consistant à publier *in extenso* l'enregistrement. Cela nous permet en retour d'intégrer le travail d'éditorialisation réalisé par l'équipe RAMHO, comme c'est le cas lors d'interviews multiples qui doivent être fusionnées pour répondre au modèle fixé préalablement. Aussi, le ou les entretiens sont ensuite structurés et éditorialisés pour éviter les redondances à la lecture. Après quoi, l'entretien transcrit est finalement soumis à l'enquête, qui peut amender ou clarifier certains passages précis, avant que le texte définitivement validé soit publié. Cette publication, accompagnée de nombreux documents d'archive, sera accessible au grand public via la plateforme web actuellement en préparation.

En avril 2023, vingt personnes avaient été interviewées donnant lieu à déjà presque cent heures d'enregistrement audio. Nous prévoyons de conduire une vingtaine d'entretiens par an, et cela jusqu'en 2027, année où sera célébré le cinquantième anniversaire de l'Ircam.

2.3. Les données du terrain marseillais

Les entretiens avec les acteurs de la recherche musicale et de l'acoustique musicale à Marseille ont été menés dans des lieux familiers des personnes interrogées, en général sur leur lieu de travail, qu'il s'agisse d'un laboratoire ou d'un studio, ou plus simplement à leur domicile. Le site du campus CNRS Joseph Aiguier a accueilli le Laboratoire de Mécanique et d'Acoustique (LMA) de 1962 à 2015. Il se situe dorénavant sur le Technopôle de Marseille Provence Château-Gombert. Les locaux vacants ont alors été investis par le laboratoire Perception, Représentations, Image, Son, Musique (PRISM). Connue de tous nos interlocuteurs, ce lieu a été privilégié pour mener plusieurs entretiens et a contribué à l'immersion dans un contexte en adéquation avec l'objet de nos discussions.

Pour aborder ce terrain marseillais, nous avons mené huit entretiens avec sept interlocuteurs entre mai 2022 et janvier 2023 (Tableau 1). Nous avons parfois eu recours à un format d'entretien hybride, dans lequel une à deux personnes participaient à la discussion en visioconférence. Nous avons ainsi collecté presque vingt heures d'enregistrements audio mais il est prévu de contacter d'autres personnes pour poursuivre nos investigations marseillaises.

Personne	Date	Lieu	Durée
Jean Kergomard [JK] ¹	31/05/2022	LMA	2:54:35
	01/06/2022	Domicile	1:49:07
Richard Kronland-Martinet [RKM]	01/06/2022	PRISM	3:28:37
Jean-Pierre Moreau [JPM]	01/06/2022	PRISM	2:07:31
Jérôme Decque [JD]	27/10/2022	Visio	2:08:55
Pascal Gobin [PG]	28/10/2022	PRISM	2:25:05
Lucien Bertolina [LB]	15/12/2022	PRISM + visio	1:55:23
Michel Redolfi [MR]	20/01/2023	Studio + visio	3:00:00
Durée totale			19:49:13

Tableau 1. Vue d'ensemble des entretiens menés à Marseille entre mai 2022 et janvier 2023.

Concernant les archives, nous avons eu accès en premier lieu à celles du CRSIM, du CRP (Centre de Recherches Physiques) et du LMA (malgré ses restructurations et ses mouvements de personnels, la continuité entre ces trois laboratoires est établie [17]) dans le fonds CNRS des Archives nationales, jusqu'à l'année 1979 incluse. Elles sont peu éclairantes dans le cadre de cet article car après une période favorable pendant laquelle François Canac dirigeait le CRSIM, la part des travaux portant sur l'acoustique musicale et la perception auditive décline fortement. On peut lire dans son compte-rendu de 1975 que le comité de direction « souhaite que l'équipe "Acoustique Musicale" ne se développe pas » (Procès-verbal de réunion du comité de direction du LMA, 6 juin 1975 [2]).

En revanche, le laboratoire PRISM abrite depuis 2018 les archives personnelles de Jean-Claude Risset. Il ne s'agit pas de documents propres aux tutelles ou aux établissements, mais bien des « archives personnelles de scientifiques » [8], qui permettent donc des études ciblées sur les recherches de Jean-Claude Risset au sein des Bell Laboratories (New Jersey, États-Unis), de l'Ircam et du LMA. Le caractère interdisciplinaire de ce fonds

d'archives permet également d'accéder aux esquisses de ses œuvres musicales, à ses carnets et ses agendas, ses brouillons de communications et articles [22].

Enfin, de manière plus ponctuelle, l'accès à des archives est également possible par le biais de certains interviewés eux-mêmes détenteurs de documents d'archives – coupures de presses, photos, affiches et programmes de concerts. C'est le cas de Jean-Pierre Moreau, l'actuel directeur du MIM, structure qui a conservé l'intégralité de ses archives, auxquelles nous avons également eu accès.

3. MARCEL FRÉMIOT, À L'INITIATIVE

3.1. Les origines du GMEM

À la fin des années 1960, la vie musicale à Marseille est le théâtre d'une mutation progressive dont l'épicentre est le conservatoire de la ville. Dirigé depuis 1963 par Pierre Barbizet (1922-1990), celui-ci fait entrer le jazz (avec Guy Longnon, en 1963) et la musique électroacoustique pour la première fois au conservatoire. Rapidement esquissée dans l'ouvrage *Les Musiques électroacoustiques* [9], cette histoire mouvementée commence en réalité à la fin de l'année 1967 quand Barbizet et Frémot évoquent l'idée d'une classe de musique expérimentale. En 1968, cette classe entre en activité et les cours sont assurés par Frémot lui-même qui s'était formé au GRM durant deux années. Seules quelques séances ont lieu avant que la classe ne soit suspendue au printemps – en raison des mouvements sociaux de Mai-68 – pour reprendre à nouveau à la rentrée de septembre. Lucien Bertolina (né en 1946), se souvient [15] que grâce à un accord passé entre le conservatoire et l'antenne Radio Provence de l'O.R.T.F., les étudiants de la classe pouvaient venir travailler dans les studios de la Radio au contact direct d'un équipement professionnel indisponible au conservatoire. Ce petit groupe, dont fait partie Lucien Bertolina, est constitué par de jeunes gens qui épouseront tous une carrière professionnelle dans la musique : Michel Redolfi, Mireille Courdeau, Georges Bœuf, Daniel Vissière, Alberto Paulin et peu après, Claude Colon. Cette situation ne dure pas : la classe de musique électroacoustique est mal perçue par les techniciens radio qui imaginent, à tort, l'activité de ce groupe comme concurrente de la leur. À la fin de l'année 1969, la classe de musique électroacoustique perd son accès aux locaux de l'O.R.T.F.² de Marseille-Provence. Toutefois, au même moment, en novembre 1969, a lieu le « Concert du groupe de musique expérimentale de Marseille » (renommé « concert des bruits » par ses détracteurs [15], une appellation qui rappelle le « concert de bruits » radiodiffusé de Pierre Schaeffer en octobre

¹ Les initiales des enquêtés serviront à faire référence aux entretiens dans le corps du texte. Les entretiens sont mentionnés dans la bibliographie, avec une cote correspondant à terme au dépôt en ligne des versions PDF (sur le serveur de l'Ircam).

² Enregistrement audio de l'intervention de Marcel Frémot dans le cadre de la journée « Musique électroacoustique : comment

enseigner une discipline en constante évolution ? », CDMC, 06 Octobre 2008 : <http://www.cdmc.asso.fr/fr/actualites/saison-cdmc/musique-electroacoustique-comment-enseigner-discipline-en-constante-evolution>

1948), première manifestation publique des jeunes musiciens, et première occurrence du nom du groupe [15]. En ce début d'année 1970, les activités pédagogiques reprennent, faute d'un meilleur endroit, dans la cuisine du logement de fonction de Pierre Barbizet, au conservatoire. Au même moment, Lucien Bertolina, Mireille Courdeau, Daniel Vissière, et Alberto Paulin, en partie en réaction à l'absence de structure d'accueil, décident de créer la SEMRA, la Société d'Enregistrement Magnétique et de Recherche Acoustique, qui s'installe rue Adolphe Thiers. Dotée de son propre matériel, la SEMRA mène différentes activités professionnelles dans le domaine de l'ingénierie sonore (enregistrements, copies, fabrication d'appareils audio), mais elle accueille également la classe de Frémot en alternance avec la cuisine de Barbizet pendant quelques mois. Ce dernier obtient finalement en 1971 des financements de la ville pour acquérir du matériel et héberger les cours au conservatoire.

En 1974, Frémot quitte brutalement le groupe. À l'été, les statuts officiels du GMEM (association loi 1901) sont déposés en préfecture. Georges Bœuf en prend la tête, épaulé par Michel Redolfi, Lucien Bertolina, Claude Colon et Alberto Paulin. Le GMEM connaît alors une période dynamique : Jacques Diennet, Patrick Portella, Frank Royon le Mée et Nadine Mistretta rejoignent le groupe, Jean-Claude Risset en devient vice-président, tandis que Michel Redolfi tisse des liens profonds aux États-Unis en particulier avec Jon Appleton, encouragé en cela par Jean-Claude Risset [16]. Redolfi introduira le Synclavier en France au début des années 1980. Durant ces quinze premières années, l'élan collectif est un véritable moteur de l'activité : « On faisait des musiques mixtes en groupe, on composait collectivement des pièces électroacoustiques, à quatre ou cinq. Le groupe était important. » [15]

En 1985, Raphaël de Vivo prend la direction du GMEM, dont le matériel se modernise : selon Jérôme Decque [13], le Yamaha DX7 fait son entrée et supplante Syter et le Synclavier II, de même que le système Pro Tools, auquel les magnétophones à bandes ne résistent pas. En 1997, le GMEM obtient le label de Centre National de Création Musicale.

3.2. L'orientation scientifique du MIM

Le MIM, créé le 6 décembre 1984 avec Pierre Barbizet comme Président, est une autre structure associative (loi 1901) portée par Marcel Frémot, après qu'il avait lui-même créé la classe d'électroacoustique du conservatoire de Marseille, puis participé à l'animation du GMEM. Les témoignages [12, 14] convergent pour mentionner l'initiative de Marcel Frémot et de Bernard Vecchione (du département de musicologie de l'Université d'Aix-Marseille I, sur le campus SHS dit « Schuman » d'Aix-en-Provence) pour créer une structure davantage orientée vers la recherche musicale que vers la production musicale. Au démarrage du MIM, le volet recherche est

ainsi principalement assumé par Bernard Vecchione, même si le principe organisationnel revendiqué est l'interdisciplinarité : la structuration initiale (cf. les premiers rapports d'activités, [1]) indique la coexistence de trois groupes : « groupe des scientifiques », « groupes des compositeurs », « groupe interdisciplinaire ». Pour autant, les compositeurs (Marcel Frémot, puis Pascal Gobin, qui intègre la structure en 1985 [PG], ainsi que Marcel Formosa) participent tout autant aux travaux de recherche que les informaticiens (Vincent Risch, Jean-François Allouis, Bernard Bel, du CNRS Marseille et Jean-Paul Allouche du CNRS Bordeaux, puis Luminy), ou les chercheurs d'horizon extra-musical (notamment le physicien et peintre Jacques Mandelbrojt). La seconde fonction prioritaire du MIM est l'accueil d'une station de travail Syter, livrée en novembre 1984 [archives MIM, JPM] au conservatoire de Marseille. Le MIM, en qualité d'association loi 1901, est en mesure de recevoir des subventions régionales et nationales, avec une mission de recherche musicale qui sort des attributions de la recherche académique du CNRS. Issue du Conservatoire de Marseille qui l'héberge dans un premier temps, cette nouvelle structure apparaît comme un intermédiaire entre le GMEM, orienté vers la production musicale, et le LMA (avec son adossement CNRS), dont la finalité est la recherche académique.

Le point culminant de la recherche musicale au MIM est le résultat d'un travail de plusieurs années (à partir de 1991) sur la création d'une méthode d'analyse auditive appelée les UST (Unités Sémiotiques Temporelles). Sous la houlette de François Delalande, parrain scientifique de ce groupe de travail interdisciplinaire, l'analyse musicale des UST s'appuie sur un catalogue de 19 « figures sonores », des segments musicaux qui possèdent une unité temporelle perçue en raison de leur organisation morphologique et cinétique intrinsèque. Xavier Hautbois décrit ainsi le travail interdisciplinaire autour des UST : « En prenant pour hypothèse que certaines figures sonores semblent produire une signification temporelle et disposer d'effets cinétiques, les membres du MIM ont recherché, dans de nombreuses œuvres musicales électroacoustiques et aussi classiques³, des effets sonores dynamiques ou statiques remarquables (comme des impressions de mouvement, d'attente ou d'immobilité). Cette démarche empirique, menée toujours de façon collective, leur a permis de constituer un véritable répertoire de figures temporelles à partir de segments d'œuvres musicales. La collecte des séquences a été réalisée en se basant sur une écoute orientée qui ne se fixe spécifiquement ni sur les hauteurs de notes, ni sur l'harmonie ou le timbre, ni sur les nuances, mais sur l'effet sonore produit par l'ensemble de ces paramètres. Le travail du MIM a été de tenter de repérer les figures sémiotiques et d'effectuer des regroupements sous le nom d'Unités Sémiotiques Temporelles. » [18]

³ Par « classique », l'auteur fait référence aux œuvres graphosphériques, en opposition aux œuvres fixées sur support électronique.

4. JEAN-CLAUDE RISSET, FIGURE TUTÉLAIRE DE L'INFORMATIQUE MUSICALE

4.1. De la faculté de sciences de Luminy à l'Ircam

Jean-Claude Risset n'est pas natif de Marseille, mais du Puy (Haute-Loire). Pour autant, il devient une figure emblématique de la recherche scientifique et musicale de la cité phocéenne. De retour de deux séjours aux Bell Laboratories (entre 1964 et 1969), fort de son expérience dans ce laboratoire interdisciplinaire, pionnier dans différents domaines et acteur majeur de l'invention d'une nouvelle discipline, Risset obtient en 1971 (après deux années de « service militaire ») un poste de Maître-assistant dans l'Unité d'Enseignement et de Recherche (UER) pluridisciplinaire de la faculté des Sciences de Luminy – Université Aix-Marseille II. C'est sous l'impulsion notamment de Daniel Kastler (fils du Prix Nobel Alfred Kastler) que cette UER pluridisciplinaire est créée, dans l'effervescence post-68, et dans une dynamique transdisciplinaire similaire à celle que Risset avait connu aux Bell Laboratories. Nous avons pu reconstituer ce moment-charnière grâce à ses agendas des années 1969-1972 [Fonds Risset/PRISM, N20_1969-NS20 1972].

Si Risset vit toujours à Paris au début des années 1970 (jusqu'à octobre 1972, [Fonds Risset/PRISM, N20_1972]), ses agendas rendent compte de ses déplacements à Marseille et de ses contacts avec des figures locales. Ainsi lit-on que Risset a écrit ou appelé Daniel Charles le 13 février et le 23 juin 1970, Daniel Kastler le 25 juillet 1970. Contact important de Jean-Claude Risset à Marseille, il sera le promoteur de l'UER interdisciplinaire de Luminy citée plus haut, conçue comme un lieu inédit en France associant la recherche scientifique, l'art et la recherche en informatique. Les contacts téléphoniques, épistolaires et physiques (à Marseille) avec Daniel Kastler seront nombreux : outre ceux mentionnés plus haut, il faut ajouter la mention au 1^{er} septembre 1971 d'un déplacement probable à Marseille, mais aussi d'une rencontre « [Daniel] Kastler, [Daniel] Charles, Marcel Frémiot à Paris » le 24 septembre 1970. Le *nota bene* du 30 septembre 1970 est également probant : « Si Marseille marche = demander Jolivet et autres personnalités qu'est-ce qu'on peut faire pour enseigner » [Fonds Risset, PRISM, N20_1970]. L'agenda du 4^e trimestre 1970 étant perdu, la fréquence de ces contacts ne peut être vérifiée. Toutefois, en 1971 [Fonds Risset/PRISM, N20_1971], Risset mentionne au 13 janvier : « après 17h Kastler, liste restreinte ? Chargé de recherche échelon 05 indice 612 ». On relève également que Jean-Claude Risset a rendez-vous avec Daniel Arfib les 19 et 26 janvier 1971 et des démarches récurrentes pour la « candidature à maître de recherche avant le 1^{er} mars avec titres et travaux en 30 ex. ».

C'est à partir du mois de septembre 1971 que Risset commence à s'intégrer pleinement dans le milieu marseillais de la recherche musicale. Toujours selon ses agendas, il a rendez-vous avec Daniel Kastler le 27 septembre 1971 et le lendemain au Conservatoire de Marseille, avec Michel Redolfi et George Bœuf, qui deviendront tous deux des amis proches. En dépit du succès pédagogique que remporte l'enseignement de la musique à Luminy, le Ministère refuse l'habilitation de cette filière [21]. Dans un article en hommage à son ami Redolfi, Risset se remémore cette période dans les termes suivants : « Le département de musique de Luminy a démarré à la rentrée 1970⁴. J'ai enseigné à des étudiants très motivés, parmi lesquels Jacques Diennet, qui deviendra bientôt l'un des compositeurs membres du GMEM. À mes cours consacrés aux musiques du XX^e siècle et à l'informatique musicale assistait un "auditeur libre" venant du conservatoire et de l'Université [département de musicologie] d'Aix-en-Provence, qui n'était autre que Michel Redolfi, musicien des grands espaces. Au-delà des démarches de la musique concrète et de la musique électronique, je présentais des éléments tout récents de la synthèse par ordinateur, que j'avais pratiquée aux États-Unis les années précédentes – un mode de production sonore qui n'était pas encore disponible en Europe. » [21]

En 1972, Jean-Claude Risset est contacté par Pierre Boulez [7] pour travailler sur la préfiguration de l'Ircam dont les travaux de construction débiteront en 1974. Entre 1975 et 1979, il est alors détaché de l'Université pendant quatre années académiques pour diriger le département « ordinateur » de l'Ircam. « Je suis parti de l'IRCAM en 1979 pour des raisons personnelles [explique-t-il]. C'était évidemment un endroit extrêmement excitant ; mais en même temps difficile à vivre, car il fallait assumer à la fois des recherches pluridisciplinaires et assurer des présentations publiques à la façon du Palais de la Découverte. L'intérêt suscité par l'IRCAM était très grand, mais la pression pour les relations publiques, la réception de visiteurs, la diffusion et la promotion occasionnaient une charge de travail et un stress considérable. Cette charge continue rendait de plus en plus difficile le véritable travail de recherche et de composition. » [22]

4.2. L'arrivée de Richard Kronland-Martinet dans l'équipe Acoustique et Informatique Musicale du LMA

Risset est promu Professeur des Universités en musicologie en 1979 juste après avoir quitté l'Ircam. Il est redevenu chercheur au LMA qui se situe sur le campus Joseph Aiguier (9^e arrondissement de Marseille). Avec l'échec de l'expérience de l'UER interdisciplinaire, Risset donne des cours de physique (sur le campus de Luminy), mais pense à intégrer le CNRS pour embrasser une carrière davantage tournée vers la recherche, en phase

⁴ Il débute ses enseignements en tant que Maître de conférences lors de la rentrée académique 1971-72.

avec sa première expérience des Bell Laboratories. Ce sera chose faite en 1985, année où il sera nommé directeur de recherche au CNRS.

Au sein du LMA, Risset anime une très petite équipe dédiée à la perception et à l'informatique musicale, intitulée AIM (Acoustique et Informatique Musicale⁵) sur le campus de Luminy (elle partage alors un ordinateur Télémécanique avec le laboratoire d'informatique fondamentale, situé au même étage que l'équipe AIM), puis au LMA sur le Campus Aiguier. L'équipe comprend Françoise Nayroles

(Ingénieure en informatique), puis une doctorante, Sylviane Sapir. Avec Daniel Arfib, qui a soutenu au LMA en 1977 une thèse intitulée « Composition assistée par ordinateur en acoustique musicale », les premiers travaux de l'équipe prolongent les recherches sur la synthèse sonore de Risset aux Bell Laboratories, via Music V et son implémentation sur l'équipement informatique du laboratoire.

L'équipe s'agrandit quelques années plus tard avec l'arrivée de Richard Kronland-Martinet. Alors qu'il était étudiant à l'Université de Marseille I (sur le campus de Luminy), il découvre au printemps 1980 l'existence de l'équipe AIM grâce à un petit prospectus ronéotypé trouvé au service de la scolarité. Il se rend au laboratoire d'informatique de Luminy où on lui signale que le bureau de Risset se situe au LMA sur le campus Joseph Aiguier. Kronland-Martinet va à sa rencontre et lui explique qu'il voit ici l'opportunité de jumeler ses deux passions – la musique et les sciences. « Il discute avec moi de façon très bienveillante. Je ne m'inscrivais pas du tout dans la musique savante, mais il a vraiment parlé avec moi... Il connaissait tout ce dont je lui parlais dans les musiques plus populaires, l'utilisation du Moog dans Pink Floyd, *et cætera*. Il me raconte ses discussions avec Moog... [...] Pour lui, toutes les musiques étaient belles, en fait. À partir du moment où tu aimais la musique, tu étais dans son clan. Ça, c'est vraiment aussi quelque chose qui montre son ouverture. Je venais donc pour discuter et finalement j'ai eu deux heures de cours de Music V ! [...] C'était un peu du chinois pour moi, parce que franchement, je rentrais dans un domaine inconnu. » [11]

Risset conseille à Kronland-Martinet de s'inscrire au DEA d'acoustique du LMA pour avoir l'opportunité de travailler sur la synthèse numérique du son. C'est ainsi que Kronland-Martinet fera ses premiers travaux en informatique musicale sur le Télémécanique (on peut raisonnablement supposer qu'il s'agit de celui qui était en usage sur le campus de Luminy). Il se souvient d'une machine de presque deux mètres de haut et qui comprenait seize touches : « On le démarrait ; il y avait des gros ventilateurs, plein de voyants qui s'allumaient, comme dans les films de science-fiction » [11]. Kronland-Martinet obtient son DEA et réalise son stage avec Daniel Arfib au cours duquel est mis au point un synthétiseur contrôlé par un mini-ordinateur PDP-11 servant à atténuer des bruits de moteur suivant les principes du contrôle

⁵ L'équipe AIM sera plus simplement renommée IM (Informatique Musicale) ultérieurement.

actif. Kronland-Martinet souhaite alors poursuivre ses recherches en thèse mais en renouant davantage avec des problématiques musicales. Risset trouve un compromis avec Bernard Nayroles, qui dirige alors le LMA, pour que Kronland-Martinet poursuive ses travaux sur l'absorption active tout en travaillant sur un sujet plus personnel relatif à la musique. Il soutiendra en juillet 1983 une thèse interdisciplinaire intitulée « Synthèse numérique en mécanique et en musique ». Il enseigne durant une année (1983-84) dans le secondaire tout en poursuivant ses recherches au LMA. Une doctorante embauchée par la Direction Générale de l'Armement travaillait alors sur la perception des sons sous-marins dans l'équipe psychoacoustique du LMA dirigé par Georges Canévet. La synthèse sonore intéressait aussi la DGA, notamment pour former les Oreilles d'or⁶. Un contrat fut donc passé entre l'équipe AIM et l'armée, ce qui permit au laboratoire s'équiper du système Syter développé par Jean-François Allouis au GRM. Kronland-Martinet a alors été embauché durant une année par la société Digilog – en charge de la commercialisation de Syter – pour travailler sur ce projet de recherche au sein du LMA. En 1986, il intègre à son tour le CNRS et se distinguera par ses recherches pionnières sur les transformées en ondelettes [20].

Alors que Kronland-Martinet s'intéresse aux méthodes d'analyse-synthèse, Risset lui suggère de rencontrer un de ses voisins qui est chercheur en physique théorique. C'est ainsi qu'il fait la connaissance d'Alex Grossman qui travaillait alors à cette époque sur la théorie des ondelettes. « Ça n'existait pas encore sous ce nom : on parlait de décomposition cycle-octave ! Il y avait eu un papier de mathématiques là-dessus... C'était purement théorique. Je commence à regarder ça et je vois le potentiel de la méthode : les fenêtres d'analyse qu'on a dans le temps-fréquence s'adaptent à la nature du phénomène. Si le phénomène était très bref, la fenêtre s'adaptait automatiquement. Et ça allait aussi dans le sens des modèles auditifs qu'on connaît comme les Gammatones. [...] J'ai alors beaucoup travaillé dans ce domaine. Ça a permis de positionner la région marseillaise sur ce sujet. J'ai fait les premières analyses en ondelettes et les premières implémentations sur Syter de la transformée en ondelettes » [11] Kronland-Martinet souligne que Syter a été un levier incroyable qui « a permis d'expérimenter la potentialité des transformées en ondelettes sur les sons ». [11]

5. UN RÉSEAU DE RECHERCHE ANCRÉ DANS LA VILLE

Ces structures – le GMEM, le MIM, le LMA et PRISM, auxquelles il faut ajouter les Universités d'Aix-Marseille I (campus Schuman) et Aix-Marseille II (campus Sciences de Luminy), les conservatoires de Marseille et d'Aix-en-Provence, la Cité de la musique de Marseille ou l'antenne de l'O.R.T.F. – forment un réseau métropolitain d'institutions qui coexistent ou se

⁶ Officier marinier spécialiste de la guerre acoustique est en charge de l'écoute des bruits à l'aide de sonars.

succèdent, et qui permet le développement des recherches sur le son en sciences et en arts en facilitant leur circulation. À la période un peu mouvementée des années 1970, durant laquelle les problèmes liés aux moyens de la recherche et les questions de structuration sont cruciaux, succède ainsi un moment de stabilité dans les années 1980. Ces différents états du paysage de la recherche à Marseille témoignent avant tout de la façon dont des acteurs comme Pierre Barbizet, Marcel Frémot ou Jean-Claude Risset ont su s'appuyer sur leurs tutelles pour susciter des initiatives et fédérer les volontés. Paradoxalement, le milieu universitaire, le CNRS et le conservatoire, malgré une rigidité administrative que l'on suppose inhérente, ont bien été pour lui des fondations sur lesquelles il a fait naître des actions de recherche musicale.

5.1. Analyse cartographique

La Carte 1 (sous la bibliographie) permet saisir en un coup d'œil la multiplication des lieux de recherche sur la musique et le son dans la ville de Marseille, où ils bourgeonnent dès la fin des années 1960. Cette expansion se poursuit durant les deux décennies suivantes avant de se stabiliser.

Une remarque s'impose d'emblée au sujet du GMEM, dont on voit que les activités se sont déroulées dans neuf lieux différents [13, 15]. Elles étaient d'abord menées dans les studios de l'O.R.T.F., allée Ray Grassi (1968-1969), puis entre la SEMRA, fondée *ad hoc*, et la cuisine du logement de fonction de Pierre Barbizet (1970-1971). Le groupe est alors accueilli au conservatoire pendant deux ans, puis s'installe au 44 rue des Dominicaines une fois son statut officialisé en 1974. Temporairement relocalisé pour un an rue Sénac de Meilhan vers 1992-1993, il intégrera par la suite la Cité de la musique, où est également domicilié, à la même période, le MIM. En 1999, le GMEM s'installe au 15 rue de Cassis, puis déménage dans des locaux conçus sur mesure sur le site de la Friche La Belle de mai en 2017. La trajectoire tumultueuse des premières années se reflète dans ces déménagements fréquents et dans l'occupation de lieux dépourvus de matériel, voire tout simplement inadaptés. Elle témoigne des efforts constants qu'ont dû fournir les quelques musiciens du Groupe pour trouver des solutions de financement et des partenariats institutionnels afin d'exercer leur activité dans de bonnes conditions. Les conceptions du travail sur le son de Schaeffer, telles qu'elles ont alors été transmises par Frémot, nécessitaient *a minima* quelques magnétophones, une table de mixage, une paire d'enceintes et un lieu pour les utiliser, ce qui représentait alors un investissement important.

Le MIM est accueilli depuis sa création dans les locaux de la Cité de la musique et n'en a jamais bougé. Le CRSIM, qui deviendra le LMA, est d'abord installé dans les bâtiments de l'actuelle préfecture des Bouches-du-Rhône, au 66 rue Saint-Sébastien. En 1962, le CRSIM prend le nom de CRP et déménage dans des locaux flambant neufs du campus Joseph Aiguier, où il restera jusqu'en 2015, année du déménagement au Technopôle.

Nous avons aussi placé sur la carte le laboratoire d'informatique du campus des sciences de Luminy : Risset y travaille au début des années 1970 et utilise régulièrement les ressources du laboratoire d'informatique, alors dirigé par Alain Colmerauer, avec qui il s'était lié d'amitié.

5.2. Le tournant informatique

L'informatique prend son essor au cours des années 1980, tant dans les structures scientifiques (LMA) que dans celles qui sont spécialisées dans la recherche et la création musicale (GMEM, MIM) : les machines font enfin leur arrivée là où elles sont nécessaires pour travailler et font moins l'objet de tensions entre les utilisateurs. On peut distinguer deux types de machines informatiques : les volumineux ordinateurs dédiés au calcul qui sont mutualisés (et qui disparaîtront progressivement) et l'informatique personnelle, qui s'installe durablement. Bernard Nayroles est conscient des difficultés matérielles rencontrées dans son équipe, dont les membres doivent se rendre au Centre de Calcul du Pharo pour réaliser certaines tâches exigeant des ressources indisponibles sur le campus Joseph Aiguier [2]. Lors du comité de direction du LMA du 3 mai 1978, il présente ses conceptions en matière d'informatique ; l'année suivante, les deux tiers du budget d'équipement du laboratoire y sont consacrés [2]. Cette décision illustre bien les tensions cristallisées par l'accès aux moyens informatiques qui traversent alors le milieu des sciences comme celui des musiques expérimentales.

Dans les années 1980, trois plateformes Syter peuvent être recensés à Marseille : au conservatoire, en 1984, ce qui déclenche la création du MIM ; au LMA, durant l'année académique, pour les besoins des recherches contractuelles du projet de Kronland-Martinet sur les méthodes de synthèse ; au GMEM enfin, où il est livré peu de temps après l'arrivée de Raphaël de Vivo à la direction, en 1986. La machine, dont les pannes fréquentes sont souvent mentionnées, est reconnue comme une innovation majeure, en même temps qu'elle est traitée avec circonspection, eu égard à son manque de fiabilité générale, sa complexité, mais surtout du fait de ses qualités sonores qui seront vite dépassées. Le GMEM et le MIM connaissent eux aussi une phase de transition entre des ordinateurs puissants mais difficiles d'accès et complexes de surcroît, et l'informatique personnelle. Toutefois, le GMEM avait déjà connu plusieurs transitions technologiques majeures qui lui avaient réussi, notamment avec l'arrivée du Synclavier I (et plus tard, du modèle II) par l'entremise de Michel Redolfi, qui donnera lieu à de nombreuses compositions, du Yamaha DX7 et des échantillonneurs Akai. Pro Tools est finalement adopté sur la recommandation de Christian Calon à la fin des années 1980 [JD]. Le logiciel permet de démultiplier les possibilités, et permet notamment au GMEM d'entamer des travaux sur la spatialisation du son avec le concours de Laurent Pottier la décennie suivante.

Pour les raisons d'encombrement et de performances évoquées, les machines ne circulent pas ou très peu. En

revanche, les figures marseillaises évoquées dans ce texte se rencontrent fréquemment : à l'université, dans les laboratoires, aux concerts ou lors d'autres événements ou encore aux rendez-vous associatifs. Il est donc facile d'imaginer que l'ordinateur alimentait alors bien des sujets de conversation et qu'il a contribué à instaurer des collaborations artistiques et des travaux de réflexion commune (comme cela a pu être le cas au MIM), et d'une manière générale, une dynamique de recherche portant non plus sur la musique ou le son, mais sur l'informatique musicale elle-même.

6. CONCLUSION

Le GMEM et le MIM, souvent opposés, apparaissent avec le recul comme deux structures complémentaires. Le premier est créé grâce au soutien d'étudiants passionnés dont les efforts contribuent à installer une dynamique durable de création électroacoustique à Marseille. Le second, lancé avec Vecchione, est à mi-chemin entre la création musicale et la recherche, au sens académique du terme. Un partage de rôle implicite se fait entre le GMEM (production) et le MIM (recherche). Après la création des UST – emblématiques de cet alliage art/science – avec les années et les dissensions internes, on assiste à une lente érosion de l'activité du MIM et à la montée en puissance du GMEM en production (davantage qu'en recherche, excepté dans les années 2000, avec les travaux sur l'espace de Laurent Pottier puis de Charles Bascou, une période située hors de notre champ historique d'étude).

Les laboratoires du CNRS ont une vie parallèle à celle des structures de recherche musicale et de création musicale, et possèdent leurs propres contraintes. D'abord très marginale au début de la période étudiée dans un laboratoire dominé par les mécaniciens, l'acoustique musicale gagne en importance au LMA à partir du début des années 1980, avec les travaux que mène l'équipe Acoustique et Informatique Musicale : ceux sur la synthèse et la perception de Jean-Claude Risset et de Daniel Arfib sur le traitement du signal, en particulier sur les effets audios et le contrôle gestuel de la synthèse. Peu de temps après, Richard Kronland-Martinet rejoint le laboratoire : il effectue d'abord des recherches sur la synthèse et le contrôle mécanique, puis devient spécialiste de l'analyse, synthèse et traitement des sons par des approches temps-fréquence et temps-échelle (ondelettes). Cette entité dédiée à l'informatique musicale finit par s'autonomiser du LMA en 2017 pour devenir PRISM, laboratoire interdisciplinaire, hébergé dans les locaux historiques du Campus CNRS Joseph Aiguier.

Nonobstant le caractère compartimenté des missions du GMEM (production), du MIM (recherche et production) et du LMA (recherche) induite par les tutelles (Ville de Marseille, ministère de la Culture, universités et CNRS), la volonté de décloisonner ces domaines est le fait de personnalités fortes (Frémiot, Risset) qui ont œuvré, chacun à leur manière, à l'interdisciplinarité (AST) indispensable au domaine de l'informatique musicale.

Prendre pour objet les recherches scientifiques et musicales à Marseille entre 1968 et 1989 permet de mettre

au jour les conditions ayant conduit à l'émergence des pratiques électroacoustiques puis informatiques de la musique en un point précisément localisé, à contrepied d'une histoire récente de la création et de la recherche musicale parfois lacunaire ou trop rapidement résumée. Les témoignages de première main sur la naissance du GMEM, du MIM et de la recherche scientifique au LMA au tournant 1970-1980 permettent de mieux comprendre les logiques d'action et le contexte dans lequel s'inscrivent les mutations esthétiques, technologiques et scientifiques dans cette période charnière. La production et le croisement de données accompagnant cette démarche révèlent à la fois certaines tensions entourant l'accès aux ressources technologiques, mais aussi les dynamiques d'émulation qui conduisent progressivement à des œuvres musicales ou scientifiques. De la classe de musique expérimentale de Marcel Frémiot aux travaux sur la perception des sons et la synthèse des scientifiques du LMA, cette étude révèle l'intrication de trajectoires humaines, technologiques et institutionnelles entourant l'émergence et l'adoption pionnière de l'informatique musicale à Marseille.

7. REMERCIEMENTS

Nous souhaitons remercier vivement les personnes interviewées pour leur accueil et leur enthousiasme.

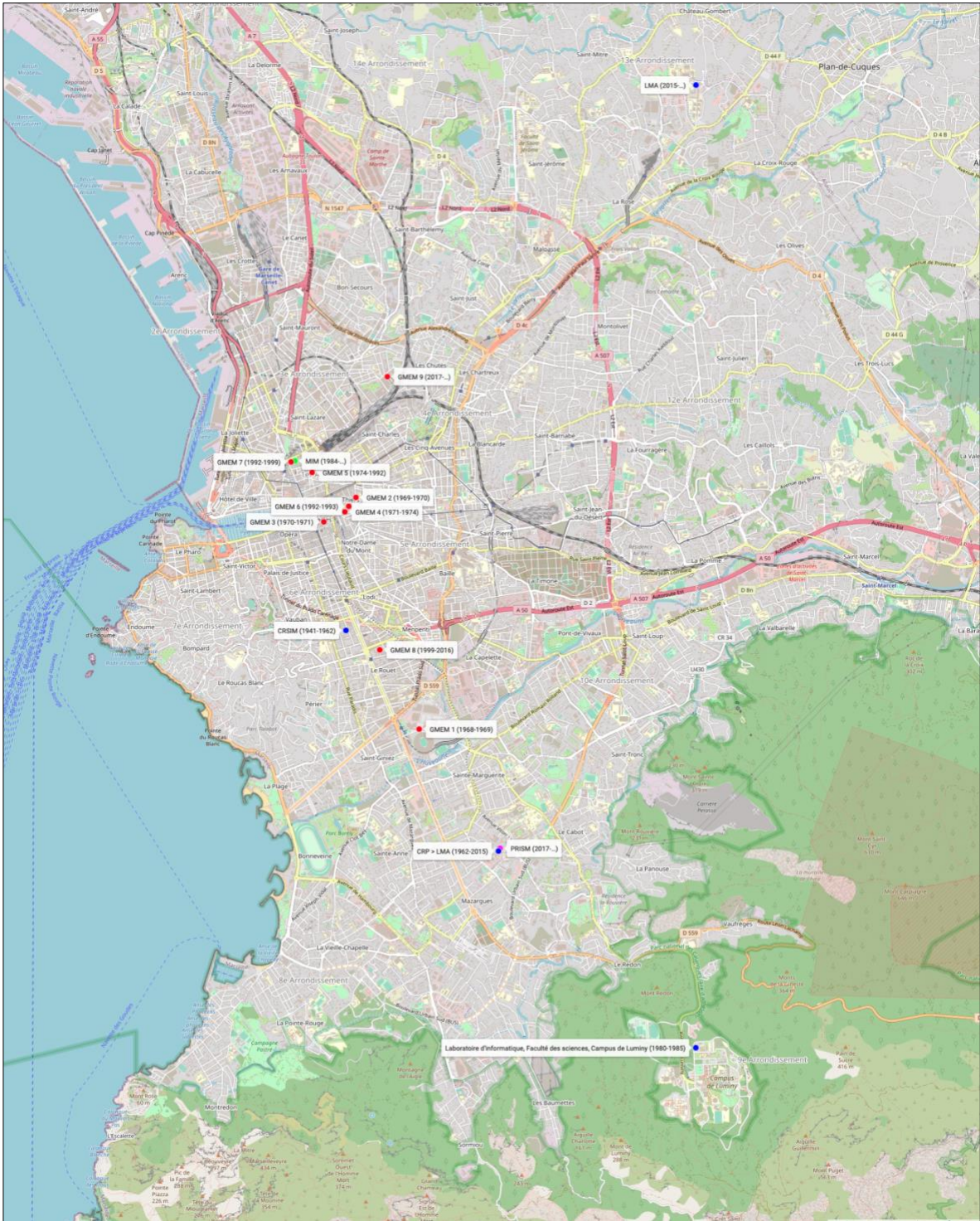
Nous remercions également le laboratoire PRISM pour les conditions matérielles de plusieurs séances d'entretiens, ainsi que l'accès aux Fonds d'archives de Jean-Claude Risset. Le processus de numérisation en cours du fonds d'archives Jean-Claude Risset (avec l'aide notamment du ministère de la Culture) a fait que nous avons eu accès aux données en rapport avec la période considérée. Ces archives pilotées par un consortium INA-Grm, Stanford, LMA et PRISM sous la responsabilité de Vincent Tiffon (PRISM), seront à termes disponibles en ligne pour la communauté des chercheurs, compositeurs et interprètes.

Cette recherche a été menée dans le cadre du projet Émergence « FOMA – Fabric of Musical Acoustics » soutenu par l'Alliance Sorbonne Université.

8. RÉFÉRENCES

1. Archives du MIM, mises à disposition par Jean-Pierre Moreau (directeur du MIM), numérisées par le laboratoire PRISM.
2. Archives Nationales, Dossiers de laboratoires du Centre national de la Recherche scientifique (1919-1986), Procès-verbaux des réunions du comité de direction du LMA, 20140644/62.
3. Archives « Fonds Jean-Claude Risset », numérisées par PRISM : agenda année 1969, cote N20_1969.
4. Archives « Fonds Jean-Claude Risset », numérisées par PRISM : agenda année 1970, cote N20_1970.
5. Archives « Fonds Jean-Claude Risset », numérisées par PRISM : agenda année 1971, cote N20_1971.
6. Archives « Fonds Jean-Claude Risset », numérisées par PRISM : agenda année 1972, cote N20_1972.

7. Archives « Fonds Jean-Claude Risset », numérisées par PRISM : correspondances, cote P51_004.
8. Charmasson, T. « Les archives des sciences en France : réalisations et projets », dans Laurent, Sébastien (dir). *Archives « secrètes », secrets d'archives ? Historiens, et archivistes face aux archives sensibles*, Paris, Éditions du CNRS, 2003.
9. Chion, M., Reibel, G. *Les Musiques électroacoustiques*, Paris, INA-GRM, 1999.
10. Delalande, F. « Qu'est-ce que la recherche musicale ? », dans *Recherche Musicale au GRM [La Revue Musicale*, quadruple numéro 394-397] sous la direction de Michel Chion et François Delalande, Richard-Masse, Paris, 1986, p. 11.
11. Entretien avec Richard Kronland-Martinet (avec Baptiste Bacot, François-Xavier Féron, Vincent Tiffon), projet RAMHO, 01/06/2022, laboratoire PRISM [cote RAMHO I 06]
12. Entretien avec Jean-Pierre Moreau (avec Vincent Tiffon, Baptiste Bacot, François-Xavier Féron), projet RAMHO, 03/06/2022, laboratoire PRISM [cote RAMHO I 06]
13. Entretien avec Jérôme Decque (avec Baptiste Bacot, Vincent Tiffon, François-Xavier Féron), projet RAMHO, 27/10/2022, Visioconférence, [cote RAMHO I 09]
14. Entretien avec Pascal Gobin (avec Vincent Tiffon, François-Xavier Féron, Baptiste Bacot), projet RAMHO, 20/10/2022, laboratoire PRISM [cote RAMHO I 10]
15. Entretien avec Lucien Bertolina (avec Vincent Tiffon, Baptiste Bacot, François-Xavier Féron), projet RAMHO, 15/12/2022, laboratoire PRISM [cote RAMHO I 12]
16. Entretien avec Michel Redolfi (avec Vincent Tiffon, François-Xavier Féron, Baptiste Bacot), projet RAMHO, 20/01/2023, Studio du compositeur/Visioconférence PRISM [cote RAMHO I 13]
17. Gazanhes, C. « Du laboratoire de la guerre sous-marine de Toulon au laboratoire de mécanique et d'acoustique de Marseille », *La revue pour l'histoire du CNRS*, en ligne : <http://journals.openedition.org/histoire-cnrs/2772>
18. Hautbois, X. « Les Unités Sémiotiques Temporelles : de la sémiotique musicale vers une sémiotique générale du temps dans les arts », *Musimédiante*, n°5, mars 2010.
19. Quinlan, M. K. « The Dynamics of interviewing », *The Oxford Handbook of Oral History*, Donald A. Ritchie (ed), Oxford university Press, 2011, pp. 23-36.
20. Kronland-Martinet R., Morlet J., Grossmann A., « Analysis of sound patterns through wavelet transforms », *Journal of Pattern Recognition and Artificial Intelligence*, Vol. 1, n° 2, 1987 pp. 273-302.
21. Risset, J.C., « Michel Redolfi, musiciens des grands espaces », disponible via <http://www.michelredolfi.info/2019/05/michel-redolfi-par-jean-claude-risset.html>
22. Castanet, P.-A., Couprie, P., Teruggi, D. et al., *Jean-Claude Risset*, « Portraits polychromes », Paris, INA-GRM, 2001
23. Tiffon, V., « Jean-Claude Risset's Paradoxical Practice of Interdisciplinarity: Making Inseparable Both Activities of Scientist and Musician, While Resolutely Distinguishing the Two Fields », in Kronland-Martinet et al. (eds.), *Perception, Representations, Image, Sound, Music*, 14th CMMR International Symposium, Marseille, France, October 14–18 2019, Springer, 2021, pp. 657-664, <https://hal.science/hal-03925283>



Carte 1. Les institutions de recherche musicale et d'acoustique musicale à Marseille [\[en ligne\]](#)

Actes des Journées d'Informatique Musicale 2023
24-26 mai 2023
Maison des Sciences de l'Homme Paris Nord

SESSION « SYMBIOSES »

L'ART DE L'ATTENTION IV : SYMBIOSES

Diane Schuh
CICM/MUSIDANSE/EDESTA
MSH Paris Nord UAR 3258
schuh.diane@gmail.com

Tania Cortés
Universidad Politécnica
de Valencia
taniapaulinacb@gmail.com

Sebastián Sarasa Molina
CRD de Pantin
CRD de Noisy
sebastiansarasa@gmail.com

RÉSUMÉ

Partants des observations et des expériences effectuées au jardin en mouvement de la MSH Paris Nord depuis novembre 2020, nous avons mené un processus de recherche-crédation aboutissant à plusieurs restitutions qui expérimentent l'adaptabilité d'un dispositif. À partir de ce processus et pour les JIM 2023, nous proposons une performance musicale et visuelle qui teste et questionne différents modèles de symbiose et leurs applications dans un dispositif d'improvisation avec traitements en temps réel qui invite à l'écoute et à l'attention pour le vivant. Ce travail fait partie de notre recherche doctorale, sous contrat EDESTA, commencé en octobre 2021 sous la direction d'Anne Sèdes et la co-direction d'Alain Bonardi. Intitulée « Symbioses, milieux, jardins en mouvement : ce que le jardinier fait à la musique », cette thèse étudie les transferts des modèles et méthodes du jardin à la composition. Notre recherche explore notamment le potentiel pédagogique et opératoire du modèle de la symbiose dans l'élaboration de dispositifs de composition et d'écoute invitant à porter attention au vivant.

1. INTRODUCTION

1.1. Jardin en mouvement : observer l'auto-organisation du vivant

La construction de notre dispositif prend pour point de départ l'écoute, l'observation et la gestion du vivant dans le jardin en mouvement de la MSH Paris Nord.

Le concept de « Jardin en Mouvement » a été forgé à la fin des années 1980 par Gilles Clément, paysagiste, botaniste et jardinier. En s'inspirant de l'observation des dynamiques du vivant dans les friches, dans les lieux abandonnés par l'humain, il invente une manière de

jardiner [3]. Il s'agit d'observer avant d'agir, « de faire le plus avec, le moins possible contre¹. » afin de s'insérer dans le « flux du vivant² ».

Très peu contraint par l'humain, le jardin en mouvement présente un type d'auto-organisation qui nous permet d'observer des morphologies émergentes et dynamiques, dépendantes de leur milieu et des modes d'interactions humains/non-humains.

Ce type de gestion, à la base du projet des paysagistes du jardin de la MSH³, a été poursuivi grâce à notre mission de gestion paysagère et d'encadrement des jardiniers pendant l'année 2021-2022⁴. Ce jardin est alors devenu notre terrain pour explorer les multiples échelles qui le constituent, et les processus qui le font évoluer. Parmi les types d'interactions observés au jardin, nous avons porté notre intérêt sur les relations de type symbiotique. D'autre part, des expériences d'écoute⁵ et d'enregistrement⁶ du milieu sonore nous ont permis de constituer une bibliothèque de sons, de définir un type d'écoute et un substrat sonore pour la composition du dispositif.

Nous nous sommes demandés comment effectuer un transfert de ces expériences dans le domaine du musical afin de composer une musique « vivante⁷ » qui invite à l'écoute et à l'attention pour le vivant.

1.2. Symbioses et systèmes complexes : une heuristique d'investigation

La symbiose peut être entendue comme une coexistence durable au sein de laquelle les partenaires différents qu'elle implique tirent des bénéfices mutuels de leurs interactions [22]. Elle participe à la dynamique et à la diversité du vivant [23]. Les différents organismes et écosystèmes modèles qui ont permis à la fois

¹ [2] p.18.

² « Pour le jardin de friches la conception intervient après : c'est une manière de sculpture qui tranche dans le substrat vivant dont elle surveille les cycles. C'est une intervention de type biologique tout entière inféodée à la maintenance d'un espace entre les bornes d'une mobilité qu'on s'est fixée. Cela revient à dire que la conception de ce jardin est perpétuellement relancée par le mouvement et le temps et qu'elle prétend également diriger ce mouvement et ce temps. » [4], p. 21 et p.29.

³ Juliette Bailly-Maître et Roman Gallais de l'agence Mutabilis.

⁴ Voir partie 2 du présent article.

⁵ <https://www.mshparisnord.fr/valorisation-partenariats/initiatives-de-valorisation/recherche-creation-jardin/>

⁶ <https://arbimon.rfcx.org/project/msh-garden-audiomoth/audiodata/recordings>

⁷ Nous nous référons ici à des concepts développés par Horacio Vaggione. Nous développons dans les parties 3 et 4 de ce texte.

l'identification⁸ et la description de la symbiose nous donnent une typologie et un spectre de caractéristiques [5] qui modèlent notre cadre théorique.

La symbiose, dans son acception mutualiste, implique des relations de réciprocité dans un réseau complexe et multi-échelle qui inclut entraides et échanges, coopérations ; mais qui n'exclut pas la compétition, la prédation et le parasitisme.

La symbiose est également entendue comme dynamique évolutive et adaptative [5]. Elle produit des attachements entre les partenaires qui permettent à la fois une adaptation au milieu et – de fait – sa transformation [14] par modification des fonctions et des morphologies.

Ainsi, cultivés dans divers domaines, les observations et les développements symbiotiques peuvent être utiles pour renouveler et vulgariser les approches de la complexité et de la diversité, telles que théorisées, entre autres, par Edgar Morin [18].

Le modèle de la symbiose nous donne des outils opératoires pour rassembler et adapter des méthodes diversifiées dans l'élaboration d'un dispositif conduit par une démarche d'investigation et non de résolution : une démarche qui se concentre sur des processus plutôt que des structures stables [12].

Comment le modèle de la symbiose, sa capacité à produire des morphologies adaptatives et l'heuristique d'investigation qu'elle demande nous inspire des manières de développer un dispositif basé sur l'improvisation qui invite à l'écoute et à l'attention pour le vivant ?

Dans un premier temps, nous décrivons l'origine et l'évolution du dispositif que nous souhaitons présenter aux JIM 2023. Nous décrivons ensuite les modèles et modélisations choisies pour la composition de ce notre système. Enfin, nous nous intéresserons au type d'écoute que nous souhaitons développer à partir d'un dispositif construit sur la notion d'émergence.

2. PERFORMANCE ET ADAPTATION

2.1. L'Art de l'attention

Le dispositif que nous avons intitulé « L'Art de l'attention IV : Symbioses » est la recomposition d'une expérience d'écoute et de création musicale fondée sur l'attention au jardin et menée depuis 2020 à la MSH Paris Nord. En partant d'une écoute attentive du milieu sonore, nous avons tenté de révéler les complexités de ce milieu

hybride - entre nature en reconquête et ville post-industrielle - en pratiquant « l'Art de l'attention ».

Anna Tsing définit « the art of noticing⁹ » comme une manière de nous rendre attentif, de mener l'enquête et d'observer le milieu afin de mettre au jour les points d'articulation, le fonctionnement d'un milieu complexe et hybride :

« [...] il faut être attentif aux manières d'exister séparément et, en même temps, observer comment ces manières se tiennent les unes les autres grâce à des coordinations sporadiques, mais cruciales¹⁰. »

Pour Tsing il s'agit de se rendre attentif à la « polyphonie¹¹ » du milieu. Écouter le jardin de la MSH nous permet d'expérimenter de multiples échelles de perception. Au-delà des classifications « Low-Fi » et « High-Fi » établies par Schafer [24], nous entraînon notre oreille à écouter les hybridations permises à la fois par la géographie, la topographie et l'écologie du milieu. En contrebas de l'Avenue George Sand et de l'Avenue de la Métallurgie¹², encerclés par les passages, accélérations et décélérations des bus 239 et 139, des voitures, au milieu des flux sonores des humains, nous pouvons entendre – en fonction des saisons – les cymbalisations de la grande sauterelle verte (*Tettigonia viridissima*), les chants du Merle noir (*Turdus merula*), du Troglodyte mignon (*Troglodytes troglodytes*), du Chardonneret élégant (*Carduelis carduelis*), du Moineau domestique (*Passer domesticus*), de l'Hirondelle rustique (*Hirundo rustica*), etc.¹³. Certaines ne font que passer et d'autres s'établissent dans le jardin. À l'Est, le grand mur de béton de la MSH réverbère les sons et les redistribue. L'acoustique est complexe et chaque point d'écoute nous révèle un monde sonore riche et changeant. Les limites entre ville et jardin s'abolissent. Nous ne pouvons pas réduire cette écoute à deux plans : un fond sonore duquel se détacheraient des signaux remarquables. Le jardin de la MSH nous permet d'expérimenter la notion de milieu dans le sens développé par Berque [1]. Les polyphonies entendues forment un espace sonore dans lequel de multiples médiations s'opèrent, entre sujets et environnement¹⁴.

C'est donc à partir des expériences d'écoutes individuelles et collectives¹⁵, et à partir des échantillons sonores collectés au jardin que nous avons proposé une expérience musicale basée sur cette notion et qui tente de « faire-jardin » dans le domaine du musical.

⁸ Depuis sa découverte en 1877 par le botaniste et mycologue Albert Bernhard Frank (1839 – 1900), de nombreuses définitions ont été proposées. De Bary (1879), par exemple, y inclut le parasitisme complet et Schwendener (1829-1919), étudie la double nature des Lichens.

⁹ [26] p. 51.

¹⁰ [26] p. 353.

¹¹ *Ibidem*

¹² Plan interactif du jardin de la MSH Paris Nord : <https://dianeschuh.fr/carte-interactive.html>

¹³ La liste des espèces végétales et animales du Jardin de la MSH Paris Nord est en cours dans le cadre de notre travail de thèse "Symbioses, milieux, jardins en mouvement : ce que le jardinier fait à la musique"

sous la direction d'Anne Sèdes et co-direction d'Alain Bonardi (Paris 8, EDESTA, CICM).

¹⁴ Ce paragraphe est une reprise d'une partie de l'article : Diane SCHUH, « L'Art de l'attention : observer, écouter et composer avec le Jardin en mouvement de la MSH Paris Nord ». « International Symposium Unheard Landscapes » organisé par FKL, École de la Nature et du Paysage, AAU CRESSON, 29 octobre 2021 à Blois, à paraître 2023.

¹⁵ Balades sonores ouvertes au public proposées dans le cadre des Rendez-vous aux jardins et des Journées Européennes du Patrimoine à la MSH Paris Nord, en 2021 et 2022 ; et expériences personnelles d'écoute et d'enregistrement depuis décembre 2020.

« L'Art de l'attention » a d'abord été une pièce électroacoustique¹⁶ puis une recomposition pour un dispositif d'improvisation pour piano préparé, saxophone et électronique en temps réel¹⁷. Enfin, la pièce a été testée pour une interaction saxophone et électronique en temps réel lors d'un concert au jardin¹⁸ (Fig. 1).



Figure 1. « L'Art de l'attention III » aux Rendez-vous aux jardins, le 3 juin 2022 à la MSH Paris Nord. Photo ©Flavie Jeannin MSHPN.

Nous lions ici heuristique d'investigation et méthodologie de recherche-crédation. Chacune des itérations du projet « L'Art de l'attention » a été l'occasion de tester des hypothèses compositionnelles qui émergent de notre recherche. À chaque restitution nous avons pu réorienter notre démarche et nos méthodologies dans des allers-retours entre recherche et création, et arts et sciences. La recherche création est ainsi entendue comme un processus continu. Les diverses recompositions de la pièce nous ont permis d'expérimenter l'évolution d'un dispositif fermé et fixé sur bande vers un dispositif ouvert qui introduit la notion de « symbiose comportementale¹⁹ » lors d'une coopération d'improvisation en temps réel et ouverte sur le monde sonore du jardin.

2.2. Adaptations et adaptabilité

Cette dernière version du dispositif – que nous proposons de jouer aux JIM23 – introduit de nouvelles notions et contraintes. Ce nouveau dispositif se base sur la manière d'introduire la notion de symbiose en tant qu'un des éléments moteurs d'un système complexe. Ce nouveau dispositif introduit aussi la composition d'un espace sonore 3D ambisonique²⁰.

La partie instrumentale, qui entrera en interaction avec la partie électronique composée par Diane Schuh, sera jouée par le saxophoniste Sebastian Sarasa.

Sebastian Sarasa est engagé dans la création contemporaine et l'interdisciplinarité. Il travaille sur des notions proches de l'écologie sonore à travers sa pratique de l'improvisation libre, notamment avec le Duo Passerelle²¹. Il collabore avec Diane Schuh depuis décembre 2021. Ensemble ils développent une relation d'improvisation saxophone et électronique en temps réel dont la dernière production a été présentée au rendez-vous aux jardins 2022 à la MSH Paris Nord.

Pour cette adaptation du dispositif nous nous associons à Tania Cortés qui viendra compléter l'expérience sensorielle par l'ajout de visuels évoluant en temps réel, en interaction avec l'improvisation musicale. Tania Cortés est une compositrice et artiste multimédia équatorienne dont le travail vise à établir une relation sonore avec d'autres disciplines, en expérimentant différents formats et médias. Pour ce dispositif elle s'inspire de la poésie de la symbiose pour construire des visuels évoluant en temps réel avec la performance musicale.

Le patch vidéo TouchDesigner²² est construit à partir de deux systèmes différents. Le premier est construit sur le principe d'émergence de morphologies à partir du bruit. Grâce à un système de feedback et un objet noise (Fig. 2), les photos que nous avons prises au jardin (notre substrat visuel) sont granulées et se révèlent petit à petit selon l'évolution de la pièce. Nous travaillons ici la notion d'émergence à partir d'une convolution entre un système statique (la photo) et une transformation dynamique (l'algorithme du patch).

Le deuxième système travaillé par Tania Cortés pour ce dispositif (Fig. 2) est construit sur la poésie de la symbiose mycorhizienne²³. À partir de l'objet « L system », construit au moyen d'un algorithme fractal, nous pouvons afficher une modélisation de la croissance des rhizomes d'une symbiose mycorhizienne.

Cette visualisation de la pousse des rhizomes formant alors un réseau se fera en interaction avec le sonore,

¹⁶ <https://lemondeautre.fr/portfolio/lart-de-lattention-sound-installation-audio-walk/>

¹⁷ Pièce jouée lors du concert pour l'ouverture du Festival EcoZone à Nanterre, Les Salons de Musique du Conservatoire de Nanterre le 13 mai 2022. Avec Henrique Cantalogo au piano préparé, Sebastian Sarasa aux saxophones et Diane Schuh à l'électronique en temps réel.

¹⁸ <https://lemondeautre.fr/portfolio/lart-de-lattention-iii-improvisation/>

¹⁹ La symbiose comportementale ou *Behavioral symbiosis* implique la coopération régulière de deux animaux sans qu'il y est attachement physique ou intégration d'un organisme dans un autre. Cette notion est à questionner lors de nos coopérations humaines, as-t-elle du sens ici ? Sur cette notion, voir [6].

²⁰ Les précédentes performances ont été diffusées sur des dispositifs de 4 et 8 haut-parleurs, en 2D.

²¹ <https://on.soundcloud.com/KAMHU>

²² <https://derivative.ca/>

²³ La mycorhize (champignon-racine en grec) ou symbiose mycorhizienne est l'association durable et mutualiste entre des champignons microscopiques et la racine d'une plante. La plante fournit des sucres issus de la photosynthèse tandis que le champignon apporte à la plante des éléments nutritifs divers comme l'azote ou le phosphore et de l'eau. Voir Selosse M.-A. [22].

renforçant alors la poésie de l'interprète-jardinier et l'interaction musicien/électronique en tant que porteuse de morphologies.

L'introduction du visuel dans la performance nous permettra aussi de questionner la représentation de la nature et de la symbiose pour ce projet. Qu'est-ce que les modélisations par algorithme disent des morphologies du vivant ? Quels types d'émergence elles décrivent et quelles poétiques elles induisent ?

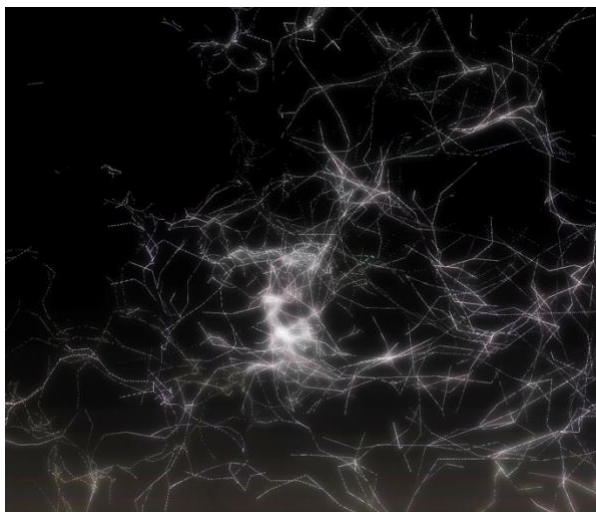


Figure 2. Deux moments de la composition visuelle interactive : à partir de l'implémentation d'un algorithme « L-Système » (rhizomes) et à partir de la granulation d'une photo (jardin). Captures d'écran du patch Touch Designer ©Tania Cortés.

L'interaction avec la vidéo introduit de nouvelles problématiques d'adaptation pour ce dispositif en temps réel. En effet, l'introduction du visuel interactif demande de fixer un cadre sur ce qui était une improvisation libre à partir d'une vingtaine de paramètres de l'électronique

(granulateurs, delays, spatialisation 8 canaux). Pour cette expérience, nous avons décidé de proposer une structure, que l'on pourrait entendre comme macro-forme musicale, au sein de laquelle l'improvisation pourra évoluer. Des jalons temporels seront définis et remarquables grâce au paramétrage des traitements et de la spatialisation. Ces jalons encadreront ce que nous pouvons nommer parties, permettant à Tania Cortés de définir l'évolution des formes du visuel en fonction de la forme musicale.

3. SYSTEMES, MODELES ET MODELISATIONS

Dans ce travail de composition à partir d'un modèle (le jardin) nous nous intéressons à divers types d'interactions du vivant que nous souhaitons modéliser dans l'environnement Max afin de contrôler les paramètres de composition de l'espace sonore ainsi que la génération de paramètres (*presets*) de traitement en temps réel.

3.1. Écosystèmes sonores et systèmes complexes adaptatifs : le patch comme « système-jardin »

Notre environnement de performance peut être entendu comme notre écosystème dans lequel la « création d'interactions voulues laisse[ant] des traces audibles²⁴ ». Il est cependant à noter que dans cette expérience de recherche-crédation nous nous éloignons du paradigme de « causalité circulaire²⁵ » et tentons d'expérimenter un système ouvert qui inclut à la fois feedbacks et circulations, mouvements périodiques induits par la modélisation mathématique de la symbiose (cf. 3.2.) et temporalités irréversibles.

3.1.1. Un système-jardin comme un Système Complexe Adaptatif

La théorie des systèmes complexes offre un cadre théorique et opératoire pour appréhender ce projet. Un Système Complexe Adaptatif (SCA) est un système constitué de plusieurs agents. Ces agents suivent des règles simples qui ne sont pas coordonnées par un processus moteur unique. De l'interaction de ces agents émergent des motifs (*patterns*). Si un ou plusieurs éléments du système sont altérés, le système réagit et/ou s'adapte [11].

Ainsi : « L'organisation [...] émerge comme le résultat de l'interaction entre les différents constituants des systèmes et son environnement. La complexité du système est le résultat des modèles d'interaction entre les éléments²⁶. »

Nous pouvons appréhender le jardin en mouvement comme un SCA et tenter une modélisation de celui-ci au moyen d'un espace composable numérique. Le jardin est

²⁴ « This is a substantial move from interactive music composing to composing musical interactions, and perhaps more precisely it should be described as a *shift from creating wanted sounds via interactive means, towards creating wanted interactions having audible traces*. In the latter case, one designs, implements and maintains a network of

connected components whose emergent behaviour in sound one calls music. » in Di Scipio Agostino [7] p.271.

²⁵ [16] p. 39-56.

²⁶ [10] p.653.

ici entendu comme matrice d'interactions, c'est-à-dire comme un ensemble de contraintes de régulation incarnées dans un espace et qui produit des morphologies en mouvement, dynamiques.

Les mécanismes biotiques (interactions du vivant sur le vivant) et abiotiques (actions du non-vivant sur le vivant) – que l'on peut transcrire en données numériques – permettant l'émergence d'un certain type de morphologies au jardin sont le moteur de notre composition. Le jardin peut être modélisé comme un système producteur de formes et de motifs récurrents dont la matrice peut être transposée dans l'espace musical pour générer des morphologies obéissant aux mêmes dynamiques

Ce dispositif conduit par ce que nous appelons un système-jardin constitue alors notre manière de « faire-jardin » dans le domaine de la composition sonore, dans le but d'un rendu sensible pour le public, d'un partage d'une écoute commune.

3.1.2. *Système-jardin et composition multi-échelle*

Cette manière de faire-jardin au niveau de l'organisation de notre espace composable rejoint la notion de « Micro-monde compositionnel²⁷ » développée par Horacio Vaggione [29]. Nous nous basons sur une approche multi-échelle de la composition qui nous permet : « de penser l'ensemble du travail de composition en termes de réseaux flexibles et ouverts où des actions de nature diverse, exercées sur des niveaux temporels parfois non-congruents, créent des processus asymétriques et modifiables à tout instant. Le cadre général capable d'intégrer cette diversité devient ainsi un lieu où se tissent des interactions entre des objets *à la fois* multiples et singuliers²⁸. »

La pensée musicale de Vaggione – issue également de la pensée de la complexité – nous sert de cadre pour organiser notre système-jardin dans le domaine du musical, en faisant une opération de translation de la notion d'agent vers celle d'objet (pensé au sein d'une composition orientée objet²⁹).

Un système-jardin est donc un environnement numérique – notre instrument – qui modélise un système complexe et opératoire pour composer le son, le visuel et leurs interactions. Certaines parties de cet instrument sont des remises à l'échelles de processus moteurs, des interprétations de modèles mathématiques [19] devenant opératoires (voire opérateurs), des mises en

correspondances d'échelles de données (*mapping*), et d'autres sont des interprétations poétiques issues d'enregistrements et de relevés faits au jardin.

Nous revenons ici plus en détail sur l'un des modules du patch conçu à partir d'une modélisation mathématique de la symbiose.

3.2. Symbioses et interactions : modèles et modélisations

La symbiose est plus qu'un état, elle est un opérateur [6] : elle est le moteur des processus de coévolution³⁰. La symbiose est un opérateur évolutif qui construit et même renforce, voire augmente la diversité [15]. Les interactions symbiotiques permettent de faire émerger une grande diversité de morphologies, de saillances, elle enrichit un (éco)système³¹. La symbiose est un modèle qu'il nous semble intéressant de tester dans le projet de composition d'un système faisant émerger de multiples discontinuités et singularités.

Une partie de notre système est ainsi mis en mouvement par la modélisation d'un système opératoire décrivant, de manière simplifiée, des interactions symbiotiques. Nous présentons dans cet article le début de notre recherche sur l'implémentation d'une modélisation mathématique de la symbiose. Nous avons le projet de complexifier cette approche dans un second temps, ainsi que de tester ce modèle avec d'autres types d'outils de contrôle dans Max.

3.2.1. *Exemple de modélisation mathématique de la symbiose*

Les mathématiques offrent diverses équations construites à partir du modèle Lotka-Volterra³². Le modèle LV est un couple d'équations différentielles non linéaires du premier ordre, développé pour rendre compte de ruptures dans « l'équilibre d'espèces en compétition, considérées comme des espèces « associées » [19] ». Ce modèle d'équation initialement développé pour rendre compte de l'impact d'une population sur une autre, dans un modèle de type proie-prédateur (Équations (1) modèle LVC, pour *Lotka-Volterra Competition*), peut aussi être utilisée pour modéliser la croissance de populations mutualistes et des interactions symbiotiques [32] (Équation (2) une adaptation possible du modèle LVS, pour *Lotka-Volterra Symbiosis*³³).

²⁷ Cf. [29] sans pagination.

²⁸ *Ibidem*.

²⁹ « C'est en pensant moi-même dans ces termes que j'ai été amené, dès 1982, à concevoir des processus de composition basés sur la définition logique d'entités actives (objets) dotées de modalités de comportement spécifiques (méthodes), déterminées de manière numérique (codes) et dont leur fonctionnalité est en même temps dépendante de ses propres méthodes et du contexte de leur utilisation. » Vaggione H. [29].

³⁰ Nous pouvons alors parler de « symbiogenèse », c'est-à-dire un processus dans lequel deux organismes différents vont former une nouvelle unité, un nouvel organisme, du fait d'interactions prolongées et à bénéfices mutuels. Les interactions peuvent ainsi amener à des synthèses.

³¹ « Symbiosis provides an additional mechanism for adaptation/complexification than available under Mendelian genetics with which Evolutionary Computation (EC) is most widely associated. [...] symbiosis appears to represent a much more effective mechanism for automatic hierarchical model building » in Heywood M. I. et Lichodziejewski P., [9] p. 51.

³² Équations proposées indépendamment par Alfred James Lotka en 1925 et Vito Volterra en 1926. Voir Lotka A. J., [13], et Volterra V. [31].

³³ « Le travail mathématique de Vito Volterra était une réponse à une situation écologique, la rupture de l'équilibre d'espèces en compétition, considérées comme des espèces « associées ». À partir des équations décrivant les variations de ces populations en compétition, un simple changement de signe suffit pour obtenir les équations décrivant les

$$\begin{cases} \frac{dx(t)}{dt} = x(t) (\alpha - \beta y(t)) \\ \frac{dy(t)}{dt} = y(t) (\delta x(t) - \gamma) \end{cases} \quad (1)$$

$$\begin{aligned} x = 0, & \quad y = \frac{\alpha x}{1 - x} \\ y = 0, & \quad y = \frac{x}{x + \beta} \end{aligned} \quad (2)$$

(1) Équation de Lotka-Volterra³⁴ où $x(t)$ = variation des proies en fonction du temps ; $y(t)$ = variation des prédateurs en fonction du temps ; alpha = taux de reproduction des proies ; beta = taux de mortalité des proies ; delta = taux de reproduction des prédateurs ; gamma = taux de mortalité des prédateurs.

(2) Modélisation mathématique d'une symbiose obligatoire [31].

Pour ce projet nous nous sommes fondés sur le concept de « cascade trophique³⁵ ». La « cascade trophique », en écologie, décrit la manière dont une interaction proie-prédateur affecte les diverses échelles d'un réseau trophique (réseaux de relations alimentaires entre organismes) et donc les morphologies d'un milieu entendu comme écosystème.

Nous nous sommes intéressés à un type de système dans lequel la modélisation d'une interactions proie-prédateur peut produire un équilibre qui devient alors bénéfique pour la formation et l'évolution d'un milieu. Nous élargissons et comprenons ici la relation de symbiose au jardin comme un équilibre entre humains et non-humains. Notre modélisation inclut ce troisième élément à la fois milieu et agent (le jardin) dont la gestion³⁶ des populations par l'humain construit les morphologies du milieu par cascade trophique. Nous considérons ici (de manière extrêmement simplifiée et pour la démonstration) les populations du jardin comme « proies », les humains comme « prédateurs » et les morphologies du jardin comme le résultat de ces interactions.

Ainsi, l'équation L-V nous sert de bloc-module pour mettre en relation diverses données qui rendent compte des interactions et des mouvements des êtres qui modèlent le jardin (le jardin étant entendu comme accumulation de micro-morphologies). Le modèle mathématique, utilisé comme expression dans notre

système, est compris comme un des éléments de la matrice de notre organisation, par la mise en réseau des interactions multiples lorsqu'il est placé dans un système multi-échelles.

En proposant un réseau de modules conduits par cette équation et mis en relation avec d'autres modules conduits par d'autres modèles génératifs, nous souhaitons ainsi proposer une interprétation du modèle qui se base sur la notion de circulation-recirculation.

En effet, comme le rappelle Perru : « En écologie, le modèle le plus simple d'un mutualisme entre deux espèces bien identifiées n'est pas spontanément mathématique. Comme de nombreux phénomènes biologiques ou écologiques, il s'inscrit d'abord dans un espace de représentation par des figures et des suites d'actions-réactions³⁷. »

Ainsi, placer ce modèle mathématique au sein d'un espace composable numérique permet l'intégration de plusieurs modes de représentation (expressions mathématique et objets mis en relation et en interactions). Pour ce dispositif, nous utilisons ce module pour effectuer une transduction d'une modélisation mathématique dans le domaine de l'espace sonore. Il s'agit ainsi de proposer une expérience sensible et sonore des phénomènes de mouvements de populations (présence/absence) par la composition de l'espace sonore.

3.2.2. Composer l'espace sonore

Dans cette première investigation expérimentale et avec le concours d'Alain Bonardi nous avons pu tester une interprétation du modèle en plan de phase de l'équation en la « mappant » sur un espace sonore en 3D, traité en 3 plans sonores au moyen de la bibliothèque HOA³⁸. Le modèle LV permet de « décrire les variations de densités de populations de deux espèces dans le mutualisme : en fonction de ces effectifs ou densités³⁹ ». Ici il nous permet d'organiser les variations et le mouvement de trois plans sonores, composant alors une situation sonore spatiale et changeante en fonction de l'évolution des paramètres (Fig. 3).

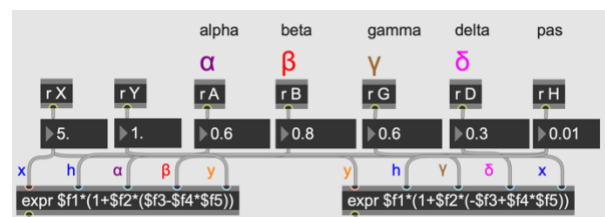


Figure 3. Implémentation de l'équation L-V dans Max : les trois plans qui composent l'espace

variations d'effectifs des populations mutualistes. » Perru Olivier [19] p. 223.

³⁴ Source Wikipedia.

³⁵ Voir le principe de « cascade trophique » dans un système qui inclut prédateurs, proies et plantes in Ripple William J. et Beschta Robert L., [21].

³⁶ par la méthode du « Jardin en mouvement ».

³⁷ [19] p. 227.

³⁸ Objets hoa.2d.map~ ; hoa.2d.projectror~ ; hoa.2d.process~ et hoa.fx.gain~ planewaves <http://hoalibrary.mshparisnord.fr/>

³⁹ [19] p. 232.

sonore sont chacun guidés par la dynamique d'une relation humains/non-humains au jardin.

Ou x = le jardin ; y = les humains ; α = taux de croissance des végétaux ; β = taux lié à la fauche des végétaux ; δ = variation de la fréquentation humaine au jardin ; γ = taux calculé en fonction des facteurs qui influencent l'absence des humains et H = le pas. Nous testons ici ce modèle avec des valeurs de références [17] qui décrivent une périodicité de deux tailles de populations⁴⁰ dont l'interaction présence/absence agit sur l'équilibre et la biodiversité d'un milieu.

Ce module rend sensible et perceptible dans l'espace sonore les mouvements induits par les fluctuations de présence/absence des humains par rapport aux non-humains. Cette expérimentation sensorielle nous permet aussi d'aborder la notion de niche et de récurrence en superposant plusieurs de ces modules et en les faisant fonctionner en parallèle.

Afin de faciliter son utilisation dans l'espace Max/MSP, Alain Bonardi a implémenté cette équation dans un module JavaScript⁴¹ qui, raccordé aux objets HOA, génère ces types de morphologies (Fig. 4.)

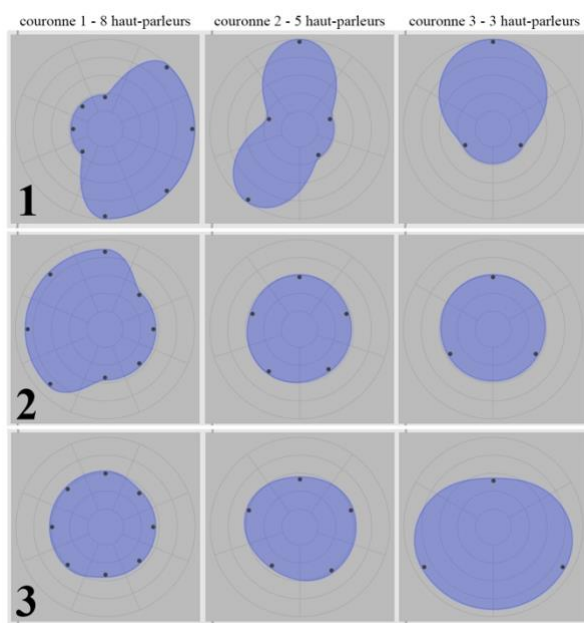


Figure 4. Représentation de l'espace sonore 3D en 3 espaces 2D superposés, commandés par le module de la Fig. 3. Le son est diffusé par trois couronnes de haut-parleurs (sur l'illustration : 8 haut-parleurs dans la colonne de gauche, 5 dans

celle du milieu et 3 dans celle de droite). L'illustration montre trois états possibles correspondant à trois temporalités, figées ici par la représentation (lignes 1 = $t1$, 2 = $t2$, 3 = $t3$).

Utiliser cette équation au sein de notre système nous permet de travailler la composition de discontinuités et leurs évolutions au niveau temporel et spatial. Nous souhaitons tester, avec cette modélisation, le potentiel de la symbiose entendue comme opérateur d'une évolution et non comme un état statique. Ce type de modélisation implique de considérer l'espace sonore comme une forme d'évolution dans le temps des matières sonores et des rythmes les uns par rapport aux autres. Le déroulement de la pièce est alors comparable à l'évolution de populations dans un environnement, en fonction de conditions de départ, c'est-à-dire une morphogenèse⁴².

Cette manière d'aborder l'espace sonore est alors complétée et complexifiée par l'utilisation de trajectoires et de transformations de l'espace sonore ambisonique au moyen des codes FAUST *décorrélation* et *trajectory* de la bibliothèque *abclib*⁴³. Ceci nous permet de rendre compte et de mapper d'autres paramètres observés au jardin. Ce système pose alors la question des échelles de données adéquates et notamment la définition même de ce qui constitue un individu au jardin dans un système de traduction basé sur le nombre⁴⁴.

Le module de base L-V et l'espace sonore qu'il transforme est intégré dans un système plus large. Il est inclus dans notre patch Max qui lui-même fait partie de notre espace composable étendu : partition, composition visuelle et performance d'improvisation électronique/saxophone. Ce dispositif teste la composition d'un espace sonore qui invite à une écoute basée sur les principes d'émergence et de diversité.

4. ÉCOUTE ET EMERGENCES

4.1. Émergences et saillances, éveiller à l'écoute des singularités

Ce dispositif ne cherche pas à reproduire le milieu sonore et visuel du jardin remis à l'échelle de la perception humaine mais à produire une expérience sensorielle et une attention sensible au vivant par le biais de l'écoute.

Composer en s'aidant de la théorie des SCA et d'un système de composition orienté objet nous permet de

⁴⁰ La prochaine étape du travail sera de définir ce que nous entendons par population au jardin, qu'est-ce que nous dénombrons et qu'est-ce que nous entendons par individu à dénombrer ici ? S'agit-il de rester au niveau poétique de l'implémentation d'une équation ou cherchons nous un « réalisme » des données ? Nous postulons que la réponse à ces questions se fait à l'écoute : quel paramétrage donne satisfaction au niveau de la recherche compositionnelle ?

⁴¹ Github avec les patches prochainement disponibles.

⁴² « Similar phenomena [emergence] can be described with rules of *morphostasis* (conservation of coherence, identity) and *morphogenesis* (dynamical behavior, change) which together capture the main

peculiarity of social and living systems: self-organisation » in Di Scipio A., [8] p.206.

⁴³ <https://github.com/alainbonardi/abclib>

⁴⁴ Lors de la constitution de notre état de l'art, nous n'avons pu mettre au jour qu'une seule utilisation de cette équation dans le champ de la composition musicale. Cette utilisation dans le domaine de l'algorithmique génétique ne montre pas les potentialités d'une modélisation multi-agents de la symbiose et reste une utilisation dans le domaine du macro temps (de la note), dans le domaine de l'imitation d'un style musical. Voir [20].

produire une musique qui rend compte du principe d'émergence en formant notamment des morphologies singulières. Nous faisons d'autre part l'hypothèse que la production de ces morphologies singulières, par émergence et donc non prévisibles, engagent alors l'auditeur et le spectateur dans une écoute impliquée, propice au développement d'un certain type d'attention : un type d'attention qui nous met à l'écoute du vivant.

Nous souhaitons ainsi développer une manière de composer qui encourage l'émergence de ces singularités : ce que Vaggione nomme « saillance⁴⁵ » et Thom « discontinuité⁴⁶ » pour l'organisation de morphologies toujours en mouvement, jamais tout à fait les mêmes.

4.2. Le substrat sonore : héritage et émergences

La poétique et l'esthétique sonore et visuelle de notre dispositif est basée sur le concept de « substrat sonore⁴⁷ ». Un substrat sonore (qui peut être spectral, structurel, timbral, morphologique) est ce qui sous-tend la composition en lui fournissant la matière première (*raw material*) qui sera mise en mouvement par le système-jardin : l'instrument et ces processus.

Ce que produit le substrat sonore, notamment l'organisation des morphologies, rejoint, par la notion d'émergence, les principes liés aux Systèmes Complexes Adaptatifs avec lesquels ils partagent des propriétés. En effet, des phénomènes d'émergence polyphoniques sont liés à la construction sous-jacente d'un substrat complexe.

Notre substrat est constitué d'enregistrements sonores faits au jardin (notre bibliothèque sonore) au moyen de dispositifs acoustiques passifs (Audiomoth⁴⁸) ainsi que du mécanisme d'organisation temporel de lecture et transformation de ces enregistrements (expliqué ci-dessus).

Les premiers essais expérimentaux ont montré que moins le substrat sonore est varié plus les émergences singulières sont rares, des morphologies identiques se réitèrent, on assiste comme en écologie à un appauvrissement du milieu, à une sorte de « monoculture ». Ce dispositif nous permet ainsi de tester les modalités de fabrication d'un substrat sonore riche : comment le paramétrer, comment l'articuler, qu'est-ce qu'il nous apprend de la notion d'héritage ?

Le paramétrage empirique et expérimental de ce système-jardin nous permet d'appréhender des caractéristiques communes à tous les systèmes complexes, et nous permettra de proposer d'autres expériences dans le champ de l'écologie sonore⁴⁹, par la recherche-crédation.

« L'Art de l'attention IV : symbiose » nous permet ainsi de tester une nouvelle évolution du dispositif, questionnant alors son adaptabilité. Ce dispositif nous permet d'expérimenter le potentiel musical d'une modélisation de la symbiose au sein d'un système de composition orienté objet. Il nous permet également d'appréhender l'espace sonore en ambisonie 3D comme espace de composition de singularités sonores mouvantes qui invitent à une certaine écoute et à une certaine attention. Ce dispositif est également l'amorce d'un questionnement sur la notion de « symbiose comportementale » dans une expérience d'improvisation en interaction avec de l'électronique en temps réel, ainsi que des questionnements sur les types de représentations du vivant que nous souhaitons développer pour ce projet.

5. RÉFÉRENCES

1. Berque A., *Médiance, De milieux en paysages*, Paris, Belin, 2000.
2. Clément G. et Lhuillier S., *Je chemine avec... Gilles Clément*, Paris, France, Éditions du Seuil, 2020.
3. Clément Gilles, « La friche apprivoisée », *Urbanisme*, n° 209, septembre 1985.
4. Clément Gilles, *Où en est l'herbe ? Réflexions sur le Jardin Planétaire*, Actes Sud, 2006.
5. Daida J., Grasso C., Stanhope Stephen A., Ross S. J., « Symbioticism and Complex Adaptive Systems I: Implications of Having Symbiosis Occur in Nature », dans *Evolutionary Programming*, 1996.
6. Daida, J.M., Ross, S.J., Hannan, B.C.: « Biological symbiosis as a metaphor for computational hybridization. » In: *Proceedings of the International Conference on Genetic Algorithms*, pp. 328–335. Morgan Kaufmann, San Francisco (1995)
7. Di Scipio A., « 'Sound is the interface': From interactive to ecosystemic signal processing », dans *Organised Sound*, vol. 8, 1 décembre 2003.
8. Di Scipio A., « Formal Processes of Timbre Composition, Challenging the Dualistic Paradigm of Computer Music, A study in Composition Theory (II) », [En ligne], <https://www.researchgate.net/publication/236025212_Formal_Processes_of_Timbre_Composition_Challenging_the_Dualistic_Paradigm_of_Computer_Music_A_study_in_Composition_Theory_II>, (Consulté le 14 février 2023).
9. Heywood M.I. et Lichodziejewski P., « Symbiogenesis as a Mechanism for Building Complex Adaptive Systems: A Review », Berlin, Heidelberg, Springer, 2010, (Lecture Notes in Computer Science), p. 51-60.

⁴⁵ « Les "saillances" sont des qualités (formes) locales qui émergent dans la tension du présent composé et qui peuvent donner lieu à une analyse morphologique approfondie afin de les faire rebondir en tant que porteuses de forme, en créant des classes qui accueillent et propagent leur spécificité : des détails morphologiques, en somme, qu'on peut faire ressortir et projeter dans d'autres localités, ici et là, dans le jeu de vecteurs de l'œuvre » Vaggione H., [28] p.55.

⁴⁶ « Le devoir premier de toute interprétation morphologique consiste dans la détermination des discontinuités d'une morphologie et des

parties stables de ces discontinuités. Dans cette interprétation apparaît la notion de singularité dont, en fait, la discontinuité est un cas particulier », Thom R. [25] p.91.

⁴⁷ [29] sans pagination.

⁴⁸ <https://www.openacousticdevices.info/audiomoth>

⁴⁹ Notamment l'étude des comparaisons entre écologie et composition du champ sonore, ainsi que la notion d'écosystème sonore investiguée par le compositeur Agostino Di Scipio.

10. Holden L. M., « Complex adaptive systems : concept analysis », *Journal of advanced nursing*, 52(6), 651–657. <https://doi.org/10.1111/j.1365-2648.2005.03638.x>.
11. Holland J. H., *Complexity: a very short introduction*, Oxford, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord, 2014.
12. Le Moigne J.-L. - *Partie 3 : L'intelligence de la conception*, 11 février 2019, [En ligne], <<https://www.youtube.com/watch?v=0-rJ7BchTiQ>>, (Consulté le 5 janvier 2023).
13. Lotka A. J., *Elements of Physical Biology*, Williams & Wilkins Company, 1925,
14. Margulis L. et Fester R., *Symbiosis as a Source of Evolutionary Innovation: Speciation and Morphogenesis*, MIT Press, 1991.
15. Maynard Smith J., « A Darwinian View of Symbiosis », ch. 3, pp. 26–39 (1991) in Margulis Lynn et Fester René, *Symbiosis as a Source of Evolutionary Innovation: Speciation and Morphogenesis*, MIT Press, 1991.
16. Meric R. et Solomos M., « Écosystèmes audibles et structures sonores émergentes dans la musique d'Agostino Di Scipio. Une collaboration entre philosophie de la musique et analyse musicale », dans *Musurgia*, vol. XVIII, Paris, ESKA, n° 3, 2011, p. 39-56, [En ligne], <<https://www.cairn.info/revue-musurgia-2011-3-page-39.htm>>, (Consulté le 14 février 2023).
17. Monod G., « Python, Méthode d'Euler explicite – Système d'ordre 1 »
18. Morin E., *Introduction à la pensée complexe*, Paris, France, Éditions du Seuil, 2005.
19. Perru O., « Modéliser la croissance des populations mutualistes : une question scientifique complexe », dans *Philosophia Scientiae*, vol. 15-3, Paris, Éditions Kimé, n° 3, 2011, p. 223-251, [En ligne], <<https://www.cairn.info/revue-philosophia-scientiae-2011-3-page-223.htm>>, (Consulté le 19 août 2022).
20. Polito J., Daida J. et Bersano T., « Musica ex machina: Composing 16th-century counterpoint with genetic programming and symbiosis », 10 avril 2006, p. 113-123
21. Ripple William J. et Beschta Robert L., « Trophic cascades in Yellowstone: The first 15 years after wolf reintroduction », dans *Biological Conservation*, vol. 145, n° 1, 1 janvier 2012, p. 205-213, [En ligne], <<https://www.sciencedirect.com/science/article/pii/S0006320711004046>>, (Consulté le 14 février 2023).
22. Selosse M.-A., *La symbiose. Structures et fonctions, rôle écologique et évolutif*, Vuibert, 2000.
23. Selosse M.-A., *Symbiose et évolution : à l'origine de la cellule eucaryote*, [En ligne], <<https://www.encyclopedie-environnement.org/vivant/symbiose-evolution-lorigine-de-cellule-eucaryote/>>, (Consulté le 22 août 2022).
24. Schafer, R. M., *Le Paysage sonore, Le Monde comme musique*, Marseille, Wildproject, 2010.
25. Thom R., *Paraboles et catastrophes*, Paris, Flammarion, 1983, p. 91.
26. Tsing A. L., *Le Champignon de la fin du monde, Sur la possibilité de vivre dans les ruines du capitalisme*, trad. Philippe Pignarre, Paris, La Découverte, 2017.
27. Vaggione H., « Composing with Objects, Networks, and Time Scales: An Interview with Horacio Vaggione », dans *Computer Music Journal*, vol. 24, The MIT Press, n° 3, 2000, p. 9-22, [En ligne], <<https://www.jstor.org/stable/3681734>>, (Consulté le 21 octobre 2022).
28. Vaggione H., « Composition musicale et moyens informatiques : questions d'approche », In M. Solomos, A. Soulez, H. Vaggione, *Formel - informel, musique et philosophie*, Paris : L'Harmattan, 2003.
29. Vaggione H., « Objets, représentations, opérations » *Article 05 - Ars Sonora n°2*, [En ligne], <<http://www.ars-sonora.org/html/numeros/numero02/02e.htm>>, (Consulté le 6 septembre 2022).
30. Vaggione H., Composition musicale : représentations, granularités, émergences. In: *Intellectica. Revue de l'Association pour la Recherche Cognitive*, n°48-49, 2008/1-2. Musique et Cognition.
31. Volterra V., « Fluctuations in the abundance of a species considered mathematically », *Nature*, no 118, 1926
32. Williams L., *Modeling symbiosis by a Lotka-Volterra-type system of differential equations*, Thesis, mai 2013, [En ligne], <<https://twu-ir.tdl.org/handle/11274/10452>>, (Consulté le 19 août 2022).

OBJETS CONNECTES AU SERVICE DE LA PERFORMANCE INTERMEDIA, EXEMPLE DE DISPOSITIF INTERACTIF POUR LA COCOMPOSITION

Laurent Di Biase

CICM / Laboratoire Musidanse

Université Paris 8

ldibiase@hotmail.com

RÉSUMÉ

Nous allons voir un exemple d'application de la technologie des objets connectés comme outils d'interaction, au service de l'expérimentation musicale pour la performance intermédia. Le cadre d'application évolue au sein d'un projet collaboratif, intitulé *Memomento*, projet écrit et réalisé en collaboration avec la compositrice et doctorante Cristina Papi. Les objectifs technologiques s'orientent selon le développement d'un design d'objets augmentés, choisis et détournés pour les besoins, selon leur potentiel poétique et artistique [1]. Une position particulière propose d'interroger la présence du vivant, ainsi que notre relation aux objets du quotidien. Ceci, dans les conditions de conserver une approche compositionnelle ouverte et résiliente selon le contexte de création de la performance, et de son écriture dans le temps. Ce projet a été développé dans le cadre de recherches doctorales auprès de l'équipe de recherche du CICM, au sein de la Maison des Sciences de l'Homme Paris Nord (MSH Paris Nord).

1. INTRODUCTION

L'Internet des Objets offre des outils permettant d'explorer le champ de l'interactivité, applicable aux domaines de la création sonore en réseau, tel l'Internet des Objets Audio [2]. La miniaturisation des systèmes embarqués répond à un besoin croissant dans le domaine de la domotique, afin de satisfaire à l'intégration des technologies de capteurs dans notre environnement en réseau pervasif. De plus, nous bénéficions des efforts constants de la communauté open-source, dans le but de démocratiser davantage les outils de prototypage libre. L'Internet des objets audio, (Internet of Audio Things, IoAuT), se développe majoritairement dans le contexte des villes et maisons intelligentes. Les objets connectés réagissent à leur environnement permettant de rendre perméable la frontière entre réel et virtuel via un système en réseau, sur un modèle d'interaction avec les ressources potentielles d'Internet. Et dans notre cas, plus particulièrement par l'utilisation du média sonore, où le contexte demande le traitement et la transmission entre

de l'information et des données audio. En cela, le contexte de la création voit émerger dès lors un champ de recherche récent, l'Internet des Objets Musicaux (Internet of Musical Things, IoMusT [3]), qui se positionne à l'intersection de l'Internet des objets, des nouvelles interfaces d'expression musicale, de la musique ubiquitaire, des interactions humain-ordinateur, de l'intelligence artificielle et de l'art participatif. Ces technologies se retrouvent aisément accessibles depuis les plateformes open-source et les outils de programmation libres. Nous avons tenté dans le cadre d'un projet de création d'en exploiter le potentiel artistique.

Ainsi, *Memomento*¹ est un projet de performance intermédia de type « work in progress », qui questionne les relations entre la mémoire [4] et les objets, dans une médiation à travers le geste et le son. L'idée motrice est donc d'intégrer dans le processus de création sonore des objets interactifs qui participent à la vie de la performance en direct. Afin que leur utilisation contribue à questionner la charge mnésique des expériences sociales et personnelles. Pour y parvenir, il s'agit pour nous d'expérimenter un dispositif hybride conçu comme un réseau de liens entre la mémoire, les objets, les gestes et le son, dans le contexte de la performance concert. Bien que ce projet, tel qu'il ait été envisagé, demande le développement de divers outils technologiques, nous nous concentrerons, dans cet article, sur ce qui a été réalisé concernant précisément les objets connectés, conçus à l'aide de microcontrôleurs Arduino [5, 6] et de capteurs. Ainsi que sur la communication de données en temps réel entre les objets et des logiciels distants sur ordinateurs, à travers une connexion en réseau WIFI.

L'enjeu principal réside dans la mise en œuvre des scénarios d'interactions avec des objets augmentés et connectés, en cohérence avec l'approche artistique du projet et leur faisabilité. Des objets capables de générer

¹ Présentation au public réalisée le 4 juin 2021 à la MSH Paris Nord

un réseau de connexion entre machines et de servir d'interface de contrôle pour interagir en temps réel avec le contexte et ses occupants. Pour ainsi permettre aux performeurs, dans des jeux de relations avec l'environnement scénique, de contrôler à distance des partitions musicales, et de sonifier des données en temps réel [7], via des patches Pure Data ou Max/MSP. A partir de cela, il a été question de définir quelle technique utiliser, et quelle méthode employer pour optimiser les interactions, soit en fonction d'objets usuels du quotidien, soit en fonction de la présence d'animaux dans l'espace scénique, comme agents participants à la composition.

2. MISE EN OEUVRE DU DISPOSITIF

2.1. Objectif et choix technologique

L'objectif du dispositif [8] est de permettre aux deux performeurs d'interagir avec des êtres vivants et des objets qui peuplent leur environnement direct, pour créer des sons. La scénographie s'organise autour de plusieurs objets augmentés et connectés, incluant la présence d'animaux, avec lesquels nous avons choisi de jouer, en utilisant leurs mouvements et déplacements. Pour ce faire, j'ai choisi d'expérimenter la plateforme Arduino MKR1010², associée à divers types de capteurs environnementaux et inertiels, pour la détection du taux d'humidité, de la température ainsi que des mouvements et déplacements dans l'espace. Cette plateforme comporte toutes les caractéristiques techniques recherchées. Par sa petite taille, elle s'intègre facilement au design. Sa faible consommation sur batterie, et sa compatibilité avec tout type de capteurs, la rend autonome et adaptable. Sa particularité, il est possible avec cette plateforme de communiquer en WIFI et d'instancier des serveurs, clients et de transmettre des paquets UDP en créant, et/ ou, en se connectant à un réseau WIFI. Cela avec le support d'une librairie Arduino intitulé WiFi-Nina³, dédiées aux cartes utilisant le module Wifi NINA. Les librairies standard qui constituent l'IDE Arduino lui sont aussi compatibles.

2.2. Sélection des objets

Des objets ont été sélectionnés en fonction des souvenirs personnels que nous gardons autour de la thématique de l'eau [9], de la mémoire, et de leur potentiel performatif. Des modèles d'interactions possibles ont ensuite été définis en fonction de leur faisabilité et des ressources disponibles. L'idée repose sur l'utilisation des données environnementales recueillies grâce aux capteurs embarqués sur des objets afin de contrôler des paramètres de jeux musicaux en temps réel sur des patches de programmation. Leur autonomie en tant qu'objet/interface et le décalage qui existe entre leur nature première et leur fonction

² <https://docs.arduino.cc/hardware/mkr-wifi-1010>

³ <https://github.com/arduino-libraries/WiFiNINA>

détournée, favorise la présence scénique et le potentiel artistique pour une forme d'usage poétique de l'informatique au service de la performance. Il sera ainsi question, par la suite, d'évoquer les différents objets réalisés pour la performance et de préciser de quel genre de modifications, ils ont fait l'objet.

2.3. Principe du dispositif réalisé

Le principe employé se décrit de la manière suivante. Cinq cartes Arduino avec leurs capteurs ont été respectivement intégrées à différents objets du quotidien par l'utilisation de petits boîtiers externes. Deux de ces cartes ont été configurées pour créer deux réseaux WIFI locaux distincts de type Point d'Accès, afin de permettre à chacun des performeurs de bénéficier de leur propre réseau.

Les cartes Arduino embarquées sur les objets sont munies de capteurs permettant de recueillir des données inhérentes à l'utilisation de ces objets, par exemple, un capteur de température installé sur une bouilloire. Ces données sont alors transmises en temps réel au format OSC via un réseau WIFI auquel se connecte l'ordinateur au travers du protocole UDP. Ces dernières sont réceptionnées directement dans un patch de programmation sur Pure Data ou Max/Msp afin de les utiliser en tant que valeurs dans des paramètres de composition. (Figure 1)

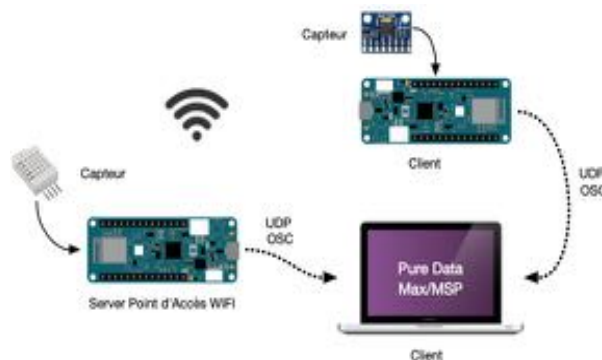


Figure 1. Schéma de principe du dispositif en réseau

3. LES DIFFERENTS OBJETS

Nous avons donc utilisé dans cette performance un dispositif composé de cinq objets interactifs, que je vais présenter ci-dessous. Il est question de proposer ici une description de principe, suivant la démarche employée pour l'intégration des technologies d'interactions et d'appropriation des objets, selon le modèle que nous avons défini dans la co-écriture du projet.

3.1. La Tasse

Les gestes inhérents à l'utilisation d'une tasse deviennent des éléments de contrôle pour générer et moduler du son. Afin d'utiliser des données provenant de l'utilisation de cette tasse, un module de capteurs de

type MPU6050⁴ est intégré dans un boîtier et accroché à la tasse (Figure 2). Ce module embarque des capteurs permettant le relevé des données accélémétriques, des données gyroscopiques, ainsi que la température.



Figure 2. La tasse équipée de son boîtier

En complément, un capteur capacitif⁵ permet de ressentir la présence de la main du performeur. A proximité ou en contact avec l'objet, le capteur fait varier des données. Un simple câble d'acier monobrin est utilisé à cet effet.

3.2. La Bouilloire

Nous avons choisi l'utilisation d'une bouilloire, destinée à faire chauffer de l'eau pour la préparation d'un thé durant la performance, d'où la tasse précédemment. Un capteur de température et d'humidité de type DHT22⁶ est utilisé, et placé à proximité de l'embouchure. (Figure 3)



Figure 3. La bouilloire équipée de son boîtier et du capteur

⁴ <https://joy-it.net/en/products/SEN-MPU6050>

⁵ <https://www.instructables.com/Capacitive-Sensing-for-Dummies/>

⁶ <https://joy-it.net/files/files/Produkte/SEN-DHT22/SEN-DHT22-Manual-20201215.pdf>

Les variations de température sont alors capturées et transmises sous forme de données et ainsi faire varier des paramètres de composition selon le déroulé de la pièce.

3.3. La Ceinture

L'espace scénique durant la performance est abordé comme un support d'enregistrement de l'instant. En cela nous avons souhaité le sonder avec l'utilisation de cette objet, utilisé à la manière d'une tête de lecture. La ceinture est conçue de façon à être portée à la taille par l'un des performeurs, dans l'action d'une interprétation chorégraphique [10]. L'électronique est fixée sur une plaque en bois et incorporée dans une housse de smartphone semblable à un ceinturon. Un module de mesure de distance LiDAR TF-Luna⁷, équipé d'un laser sensible jusqu'à 8 mètres, est utilisé comme un faisceau de lecture des distances géographiques. (Figure 4)

Un second capteur est utilisé, le module MPU6050, qui intègre de même le dispositif de la ceinture, pour permettre de capturer les données relatives aux déplacements et aux mouvements du performeur.



Figure 4. La ceinture et son matériel électronique dans une housse à face transparente

3.4. Le Plateau

Ce plateau est conçu comme une surface sensible, destinée à accueillir et à offrir une espace d'évolution pour un groupe d'escargots. Deux capteurs de distance de type IR Sharp GP2Y0A41SK0F⁸ sont disposés sur l'un des bords du plateau afin de mesurer les déplacements de ces gastéropodes. (Figure 5)

⁷ <http://en.benewake.com/product/detail/5e1c1fd04d839408076b6255.html>

⁸ https://global.sharp/products/device/lineup/data/pdf/datasheet/gp2y0a41sk_e.pdf



Figure 5. Le groupe d'escargots au repos sur leur plateau sensitif

3.5. L'Aquarium

Un poisson fait aussi partie des animaux qui composent la performance. De l'espèce des combattants, celui-ci évolue à l'intérieur de son aquarium augmenté. Les données en trois dimensions de ses déplacements dans l'eau sont recueillies à l'aide de trois plaques d'aluminium faisant offices de capteurs capacitifs⁹. Les plaques sont emboîtées sur trois des faces de l'aquarium, coté droit, en dessous et sur la face arrière où vient se rajouter le boîtier intégrant l'électronique. (Figure 6)



Figure 6. Vue de l'aquarium où réside le poisson

4. REALISATION DE L'ELECTRONIQUE ET DE LA PROGRAMMATION

Nous allons voir ici un exemple de la méthode adoptée pour la réalisation des objets interactifs en prenant l'exemple de la tasse. Cette étape concerne particulièrement la production d'un dispositif tant pour

la partie système électronique que dans la partie programmation de l'interface.

4.1. L'exemple de la Tasse

Cet objet intègre le module MPU6050, composé d'un capteur de température, ainsi que deux capteurs micromécaniques, un accéléromètre qui mesure les accélérations dans l'espace et un gyroscope qui mesure les rotations. Et il intègre également un câble monobrin dénudé servant de capteur capacitif passif. Ici le module s'alimente directement depuis la carte Arduino en 5V et utilise l'interface I2C pour communiquer, en reliant les ports SDA et SCL. Une résistance de 680KOhm est utilisée pour câbler le capteur capacitif sur deux des entrées digitales. L'ensemble est alimenté à l'aide d'une batterie Li-Po de 3,7V de 1100mA. (Figure 7)

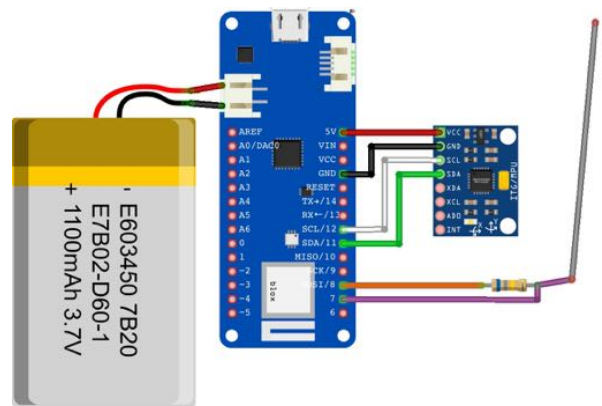


Figure 7. Schéma du dispositif de la tasse

En ce qui concerne la partie programme, les besoins étaient de créer un serveur en tant que borne de point d'accès WIFI, d'interfacer les capteurs avec le microcontrôleur, et de configurer la transmission de messages OSC portant les différentes valeurs de données des capteurs via UDP¹⁰. Ce protocole a été choisi, car il permet d'échanger des quantités de données rapidement entre un serveur et un client.

4.1.1. La programmation dans l'IDE Arduino

L'IDE Arduino fonctionne avec le langage C++. Afin de configurer la carte MKR1010, plusieurs bibliothèques ont dû être implémentées. La bibliothèque Wifi-NINA permet d'instancier le serveur et de se connecter à un réseau WIFI. Les bibliothèques SPI et Wire permettent de contrôler les appareils et de communiquer entre eux. Les bibliothèques Capacitive Sensor, Adafruit MPU6050, Adafruit Sensor, permettent de contrôler les capteurs. Et la bibliothèque OSCBundle permet de gérer la communication en OSC via UDP, de plusieurs messages différents avec les patches des ordinateurs.

⁹ <https://makezine.com/projects/a-touchless-3d-tracking-interface/>
<https://www.instructables.com/DIY-3D-Controller/>

¹⁰ https://fr.wikipedia.org/wiki/User_Datagram_Protocol

Pour configurer le serveur en point d'accès afin de transmettre des données en OSC, je me suis inspiré des travaux de Federico Peliti et de son projet Wifi Access Point for OSC¹¹, disponible sur la plateforme Github, également inspiré des travaux menés par Tom Igoe¹² et Adrian Freed¹³. Pour les objets considérés comme client, j'ai utilisé le croquis dédié provenant de la librairie Wifi-NINA. Il est donc nécessaire en premier lieu de déclarer les librairies, les informations de connexion, ainsi que les différents capteurs selon leurs raccordements aux entrées du microcontrôleur. J'ai ensuite implémenté les parties concernant la programmation des capteurs selon leurs spécificités et modifier la programmation de l'OSC pour permettre l'envoi de plusieurs paquets de données en même temps, dans le cas de l'utilisation de plusieurs capteurs en parallèles sur un même objet.

```
#include <SPI.h>
#include <WiFiNINA.h>
#include <WiFiUdp.h>
#include <OSCBundle.h>
#include <CapacitiveSensor.h>
#include <Adafruit_MPU6050.h>
#include <Adafruit_Sensor.h>
#include <Wire.h>

Adafruit_MPU6050 mpu;
Adafruit_Sensor *mpu_temp;

int status = WL_IDLE_STATUS;

char ssid[] = "NomDuReseau";
char pass[] = "MotDePasse";

IPAddress outIp(192, 168, 4, 2);
unsigned int localPort = 9999;
const unsigned int outPort = 8007;

WiFiServer server(80);
WiFiUDP Udp;

CapacitiveSensor cs_7_8 = CapacitiveSensor(7,8);
```

Ensuite, il s'agit de configurer le serveur en point d'accès wifi, ainsi que les paramètres concernant le module MPU6050. Pour synthétiser ici, j'ai conservé uniquement le capteur de température.

```
void setup() {
  Serial.begin(9600);
  while (!Serial) {
    ;
  }
  Serial.println("OSC Access Point");

  if (WiFi.status() == WL_NO_MODULE) {
    Serial.println("Communication with WiFi module failed!");
    while (true);
  }
}
```

```
Serial.print("Creating access point named: ");
Serial.println(ssid);

status = WiFi.beginAP(ssid, pass);
if (status != WL_AP_LISTENING) {
  Serial.println("Creating access point failed");
  while (true);
}
delay(10000);

server.begin();
printWiFiStatus();

Serial.println("\nStarting connection to server...");
Udp.begin(localPort);

if (!mpu.begin()) {
  Serial.println("Failed to find MPU6050 chip");
  while (1) {
    delay(10);
  }
}
```

```
mpu.setFilterBandwidth(MPU6050_BAND_21_HZ);

Serial.println("MPU6050 Found!");
mpu_temp = mpu.getTemperatureSensor();
mpu_temp->printSensorDetails();
```

Puis, l'étape permettant de recueillir les données provenant des capteurs et de les afficher.

```
void loop() {
  sensors_event_t temp;
  mpu_temp->getEvent(&temp);

  Serial.print("\t\tTemperature ");
  Serial.print(temp.temperature);
  Serial.println(" deg C");

  long start = millis();
  long touch = cs_7_8.capacitiveSensor(30);

  Serial.print(millis() - start);
  Serial.print("\t");

  Serial.print(touch);
  Serial.print("\t");
}
```

Cela nous permet par la suite de configurer la communication OSC et de définir les paquets à envoyer.

```
OSCBundle bndl;

bndl.add("/MugTouch").add((float)touch);
bndl.add("/MugAccel").add((float)temp.temperature);

Udp.beginPacket(outIp, outPort);
bndl.send(Udp);
Udp.endPacket();
bndl.empty();

delay(20);
}
```

4.1.2. La programmation des patches

Concernant la partie réception des données, l'ordinateur doit être connecté au même réseau wifi

¹¹ <https://github.com/AhUhhm/Arduino-WiFi-Access-Point-for-OSC>

¹² <https://github.com/tigoe>

¹³ <https://github.com/adrianfreed>

partagé avec les microcontrôleurs, ainsi que l'adresse IP. Nous allons voir ici comment configurer un patch sous Pure Data pour utiliser les valeurs provenant des données des capteurs. Le principe reste identique dans le cas d'un patch sous Max/MSP. (Figure 8)

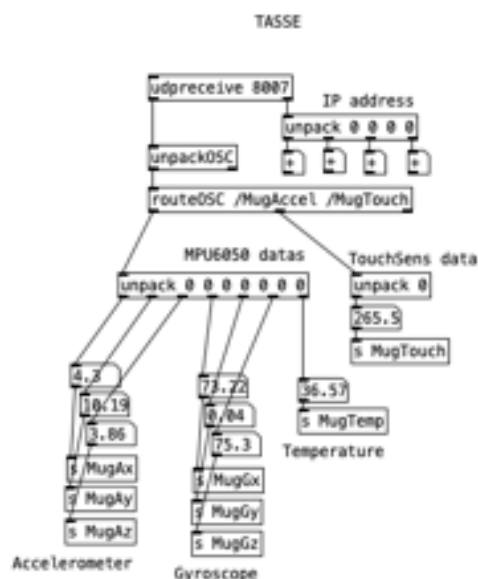


Figure 8. Réception des données OSC de la tasse dans un patch Pure Data et leur routage avec l'objet *send*

Les données, une fois réceptionnées dans Pure Data, sont transmises via le couple d'objets *send* – *receive* à la matrice de mixage des sources audio, afin d'entrer dans la chaîne de traitement. Toujours dans le cas de la tasse, les données d'inclinaisons de l'objet déclenchent des séquences audio en fonction du degré, contenant des sons évocateurs de l'eau avec par exemple des glitches retraités à l'aide de pitch-shifter. Jusqu'au point d'inclinaison horizontal où se déclenche des sons de bourdonnements d'abeilles. Pour cela un sous patch est alors conçu pour effectuer la tâche de détection des différents seuils, calibré en fonction des besoins avec une opération binaire en amont dans le patch. (Figure 9)

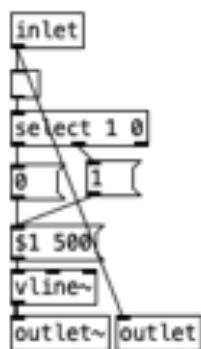


Figure 9. Détail de l'algorithme de détection de seuil, la sortie de gauche comprend les valeurs

attendues, la sortie de droite, les valeurs de déclenchement

Les traitements audio sont élaborés à l'aide de la librairie *else*¹⁴, conçue par Alexandre Torres Porres, ainsi que la librairie *faustgen2*¹⁵ développée par Pierre Guillot et Albert Graef au CICM. Le matricage des signaux est conçue pour une diffusion sonore en quadraphonie, et utilise pour ce faire les outils HOA de la librairie *ceammc*¹⁶, tenus à jour par le Centre de musique électroacoustique du Conservatoire de Moscou (CEAMMC) et par le studio ZIL-electro, pour une compatibilité avec Pure Data.

5. APPROCHE COMPOSITIONNELLE

Les degrés de collaboration dans ce projet relève du partage, avec Cristina Papi, de l'écriture de l'oeuvre performative dans l'ensemble de ses aspects, réalisation, mise en scène, situations et interactions, ainsi que de l'écriture sonore et musicale dans les grandes lignes conceptuelles. Le partage de l'espace de création entre les acteurs et agents participants, comme les animaux et les objets dans ce dispositif, crée une dynamique tripartite. De ce modèle de cocréation [11] se dessine la performance dans des jeux d'interactions entre les acteurs, les agents et les ordinateurs à travers l'usage d'objets connectés, oscillant entre actions définies et situations imprévisibles. Les animaux de part leur présence participent donc aux actions dans le déroulement de la performance, de sorte qu'ils intègrent à leur échelle le processus de création. En retour nous leur apportons de la nourriture et cela contribue à nous rappeler nos liens avec le vivant.

Ainsi, la composition s'articule autour de l'écriture d'un scénario axé sur les liens poétiques entre les objets, les gestes et le contexte, oscillant entre le réel et l'imaginaire. Afin de structurer cette approche, nous avons divisé le temps de la performance en plusieurs scènes distinctes. La production de sons est liée au dispositif instrumental, se décrivant par des événements et des situations précises. L'audio est généré en fonction de la nature des éléments déterminés, de leur apparition dans le temps et des différentes interactions établies. L'écriture compositionnelle reste ouverte au possible et se veut libre de se développer dans cette relation de coprésence avec le vivant et avec les objets, en fonction de l'instant. Une réflexion profonde a été entreprise autour des moyens permettant de formaliser, de caractériser le concept de mémoire, dans ce qui la constitue, comme flux, comme structure, mais aussi dans son altération, à travers l'oubli et l'interprétation. Par exemple avec l'utilisation de l'eau comme élément métaphorique. Sans rentrer dans les détails, nous allons voir selon quels principes s'articule le composable, ainsi que l'émergence des formes sonores dans ce projet.

¹⁴ <https://github.com/porres/pd-else>

¹⁵ <https://github.com/CICM/pd-faustgen>

¹⁶ <https://github.com/uliss/pure-data/tree/ceammc/ceammc>

5.1. Modèle d'écriture

Le modèle d'écriture s'organise principalement autour du déclenchement d'évènements selon des jeux de relations et d'interactions avec les animaux et avec les objets, en jouant avec des décalages de sens. De cette coprésence, de l'action, du geste et des mouvements tangibles prenant pieds dans cette réalité concrète, et renvoyant à la temporalité de l'œuvre, par une coprésence des durées, passée, présent et futur au moyen d'une plasticité renouvelée. Nous tentons de produire une altération du sens, de l'esthésie¹⁷, au moment même où il est en acte, vers un élargissement de l'expérience sensible. Ceci, afin de parvenir à une forme de dialogue qui s'affirme ou s'estompe selon des degrés de cohésion. Dans l'objectif de jouer avec le flou des limites potentielles entre imagination et réalité. Les dialogues se définissent selon différents schémas, soit entre les deux performeurs, soit entre un performeur, un objet, une situation, ou encore avec les animaux.

La composition s'établit donc principalement autour de jeux de dialogues à travers la présence ou l'absence des performeurs, des animaux et des objets en fonction des actions et des situations qui se créent dans le fil de la narration. Par exemple, nous faisons dialoguer la bouilloire avec le poisson, la tasse de thé avec les escargots. Ce principe se décline ainsi de suite, en se jouant des décalages entre les gestes et les sons, entre des moments précis et la mémoire personnelle ou collective. Sur scène, dans une atmosphère calme d'intérieur domestique, Cristina commence par nourrir les poissons dans leur aquarium. Pendant ce temps, Laurent prépare de l'eau chaude avec la bouilloire afin de servir deux tasses de thé. Les actions s'entremêlent entre jeux musicaux, gestes et situations du quotidien, tout en maintenant par alternance un regard sur les ordinateurs. Cristina joue avec les escargots sur leur plateau avec de la nourriture, et je l'invite ensuite à danser ensemble avec la ceinture, pour finir en cherchant à se trouver dans l'espace de part les distances qui nous séparent.

Pour exprimer les degrés de cohésion, nous avons utilisé la méthode des seuils de déclenchement, qui module les sons ou leurs traitements en fonction de la variation de l'intensité des valeurs de données générées par les interactions. En ce qui concerne les sons, nous utilisons des bruits d'animaux sauvages, baleine, dauphins, insectes, ainsi que des d'ambiances d'espaces naturels, qui sont ensuite retravaillés. Nous nous sommes inspirés de l'eau, avec l'utilisation de différents bruits enregistrés, mais aussi en tant que référence, dans le choix des traitements, comme avec de la synthèse granulaire, pour sa force évocatrice de l'élément. La synthèse soustractive est également employée pour son évocation de la perte de détails inhérents au processus de la mémoire, par phénomène de filtrage.

¹⁷ <https://fr.wiktionary.org/wiki/esth%C3%A9sie>

Ainsi dans cette intention d'explorer le concept de la mémoire, nous avons établi un répertoire d'effets sonores élémentaires [12] tel que le délai, la granulation ou l'échantillonnage, ré-échantillonnage, ainsi que des effets de perception comme la rémanence, et des effets sémantiques de décalage. Par exemple, enregistrer une séquence de la performance au hasard durant son déroulement et la rejouer à un moment précis avec de la perte d'informations, comme dans une tentative d'évoquer la forme du souvenir. Dans un jeu de tension, à l'inverse, l'idée est d'aller puiser dans l'instant, c'est-à-dire, d'inclure aussi, à un moment précis, du son d'ambiance provenant de l'environnement extérieur du lieu où se déroule la performance, créant ainsi un décalage de perception au pouvoir onirique.

6. CONCLUSION ET PERSPECTIVES

Nous avons vu dans cet article la présentation du projet collaboratif *Memomento* et de ses enjeux artistiques et techniques. Plus précisément, la description des objets interactifs qui ont servi à la réalisation d'une performance intermédia, ainsi que l'approche compositionnelle. Nous avons abordé l'intégration des technologies de l'Internet des Objets, issu de la communauté open-source, et comment celle-ci favorise dans notre projet l'utilisation de dispositifs connectés pour la production d'œuvres multimédia. Des objets capables de servir d'interfaces de contrôle, permettant d'interagir en temps réel avec des patches de programmation via des logiciels tel que Pure Data, et de jouer ainsi des partitions sonores et musicales. Pour cela, nous avons abordé les technologies utilisées et décrit le principe de mise en œuvre tant par la fabrication de ces objets, que par l'écriture de la performance, afin de répondre aux enjeux artistiques du projet.

Ce travail est le terrain d'application d'une recherche questionnant les liens entre les domaines de la performance, de l'art interactif et de la musique en réseau [13]. Dans le cadre d'une œuvre performative, il a permis d'expérimentation les interactions avec le vivant et avec des objets du quotidien, grâce à la technologie de l'Internet des Objets. L'environnement scénique y est abordé comme un terrain d'exploration artistique. La perspective de ce travail évolue désormais vers un questionnement des conditions de création en environnement extérieur et plus précisément en milieu urbain.

7. REMERCIEMENTS

De sincères remerciements sont adressés tout d'abord à ma collaboratrice, Cristina Papi, pour le développement et la réalisation en commun de ce projet, ainsi que pour nos partenaires de production et de diffusion, avec le soutien de l'EUR-ArTec ainsi que l'aide de la MSH Paris Nord.

8. REFERENCES

1. Michel de Certeau, *L'invention du quotidien, I : Arts de faire*, Folio Essai, Gallimard, 1990
2. Carlo Fischione, Luca Turchet, György Fazekas, Mathieu Lagrange, Hossein S Ghadikolaei, *The Internet of Audio Things: state-of-the-art, vision, and challenges*, IEEE Internet of Things Journal, 2020.
3. Lucas Turchet, Carlo Fischione, Georg Essl, Damian Keller, and Mathieu Barthet, "Internet of Musical Things: Vision and Challenges," IEEE Access, vol. 6, pp. 61 994–62 017, 2018.
4. Touboul, Patricia. « Ce que l'art fait à la mémoire : le renouvellement éthique de l'appropriation du temps humain », *Nouvelle revue d'esthétique*, vol. 18, no. 2, 2016, pp. 103-114.
5. Mike Cook, *Arduino music and audio projects*, Spinger Libri, 2016
6. Brent Edstrom, *Arduino for musicians: A complete guide to arduino and teensy microcontrollers*, Oxford University Press, 2016
7. Thomas Hermann, Andy Hunt, John G. Neuhoff, *The Sonification Handbook*, Logos Verlag Berlin GmbH, 2011.
8. Sandrine Baranski, *Manières de créer des sons : l'œuvre musicale versus le dispositif musical (expérimental, cybernétique ou complexe)*, Revue DEméter, Université Lille-3, 2009
9. Gaston Bachelard, *L'eau et les rêves : Essai sur l'imagination de la matière*, 1942, édition électronique Université du Québec à Chicoutimi, 2016.
10. Christopher R. Morgan, *A Motion-based Controller for Real-time Computer Music with applications for Dance Choreography and Music Composition, The Design, Construction and Programming of a Wireless accelerometer-based interface system*, Collin College, 2018
11. Alexandros Kontogeorgakopoulos, Olivia Kotsifa, *Game|lan: co-designing and co-creating an orchestra of digital musical instruments within the fab lab network*, Linux Audio Conference, CCRMA, Université de Stanford, 2019
12. Jean-Francois Augoyard et Henri Torgue, *A l'écoute de l'environnement, Répertoire des effets sonores*, Ed. Parenthèses, 1995
13. Sandrine Baranski, *La musique en réseau, une musique de la complexité ?*, Editions universitaires européennes, 2010

HEADPHONICS : MOBILE MUSIC ON THE METRO

Steve Jones

Teesside University, Royaume-Uni
stevenjones@tees.ac.uk

RÉSUMÉ

Cet article présente ma recherche concernant le paradigme de l'utilisation de l'appareil portable (smartphone ou tablette que j'appellerais désormais «mobiles») dans le contexte de la musique électronique et de la performance (dénommée «musique mobile»). Il s'agit d'explorer les dispositifs mobiles en tant qu'outils et modes d'expression artistique, et de ce fait, de placer la pratique de la performance au centre de la méthodologie. L'objectif est de montrer comment la mobilité implique des relations multiples entre un musicien, sa situation et son environnement. Je propose d'abord un aperçu du cadre théorique et pratique de la musique mobile, en m'inspirant des théories de l'oralité médiatisée, de la pratique de la marche, et des environnements urbains ou naturels comme sites de performance musicale (dénommée «headphonics»). Comme ça, je voudrais remettre en question les limites des espaces associés à la musique électronique – où elle est faite, où elle est écoutée et vécue.

1. INTRODUCTION

Imagine an electronic musician in the process of creating sound. With forefinger and thumb, the musician moves four circular icons across the x and y axis of a screen. There is a process of listening intently as each movement adds or subtracts to a variety of time delay and reverberation effects. The sound being created is a mosaic of rhythmic textures, sine tones and loops of human voices. Our musician seeks out to make constant adjustments to the sound

being created, taking spontaneous decisions about numerous variables; dynamics, textures and frequencies are in a continual state of flux. And each sequence of events will never be repeated.

If we step back from this scene, we can see our musician's surroundings are not a recording studio or a programming suite. There is no mixing desk, no racks of electronic equipment, keyboards, or near-field monitors. Instead, it is a train carriage on the Paris Métro. The musician is holding an iPad, yet to all outward appearances is simply another passenger wearing earbuds, staring at their mobile device. The physical gestures are the same as everyone else's: tapping, pinching and scrolling down a screen. No one is aware the electronic musician is inscribing this performed event as a fixed recording; it is a private, internalised experience of mobile music. The only time an audience might view or hear this music is when it is shared afterwards via YouTube or micro-blogging platform ¹.

In fact, I am the electronic musician in the above example and this article will be an autoethnographical account of an area of my research that focuses on headphone-based music performance. To describe this process, I am using the term 'headphonics' as a new model of interaction.

My practice-based research investigates the mobile paradigm in the context of electronic music, sound, and performance. It explores mobile media devices as creative tools and modes of artistic expression in everyday contexts and situations, working in dynamic and unpredictable conditions. The intention is

¹ <https://steranko.tumblr.com/>

to question the spaces commonly associated with electronic music – where it is made, where it is listened to and experienced. I consider the idea of mobile as a lens through which a new model of electronic music performance can be interrogated. In this article I hope to show how mobile music has the potential to change the experiences of making electronic music, create a novel form of music, and consequently generate a new kind of performer identity.

2. BEING MOBILE: BACKGROUND, CONTEXT

To put these questions into context, my own experiences of electronic music production have historically been situated in a recording studio environment, a place designed for recognisable, manageable, understandable and unproblematic scenarios. However, my assumptions of music making were challenged when I began attending music improvisation workshops held at the Église Saint-Merry in Paris, under the direction of Frederic Blondy. Playing with a broad mix of musicians – from amateurs to conservatoire students, hobbyists to seasoned veterans of the Paris free improvisation scene – I found it frustrating always being positioned away from the group, sat at a table obscured behind my laptop screen and an array of cables, mixing desk and loudspeakers. To counteract these vexations, I wanted to discover a new lexicon of electronic sound that did not involve a laptop, digital audio workstations or desktop computing interfaces of a QWERTY keyboard and mouse.

At this time, I was often travelling between two similar, but very different geographical sites – the UK and France. It was vital therefore to have light, portable and resilient equipment that could be easily carried. This led me to exploring the potential and limitations of generic mobile devices and commercially available apps as a music performance system, to address the gap in knowledge on how mobile music presents real challenges to the assumptions and expectations of electronic music beyond the studio.

Another significant reason driving my decision to use generic devices stems from my background rooted in electronic dance music and DJ culture. The history of modern dance culture is founded on artists experimenting and reinterpreting technologies in ways other than they were intended. From DJs misusing variable-speed turntables to mix and scratch vinyl records at dance parties, to Roland's TB-303 Bass Line accidentally becoming the definitive sound of 'acid house', dance music culture is inextricably intertwined with commercial and often overlooked products and systems, wrestling out possibilities in ways that manufacturer would never have envisioned.

Over the course of my research, I have examined a wide range of situations for mobile music making, from organising concerts for the mini festival *Mobilise*² in conjunction with IRCAM and De Montfort University (DMU), to developing a performance for groups of players carrying portable loudspeakers for the arts organisation Locus Sonus, to my somewhat precarious membership in Paris' free improvisation scene. However, in this article, I wish to focus on a group of works using earbud headphones, an approach that brought about a singular, but richly rewarding experience of making music.

2.1. FRAMING THE FIELD: MOBILE MUSIC STUDIES

Before I look specifically at examples of headphonic performance, I must consider briefly how portable auditory devices have shaped subjective experience. The precursor of this practice is the introduction of lightweight, high-fidelity headphones and the Sony Walkman. Mobile music studies can be traced back to Shuhei Hosokawa's theories on the Walkman Effect. Hosokawa describes the use of the Walkman ('*musica mobilis*', as he calls it) transforming the act of walking into a form of secret theatre: "With the appearance of this novel gadget, all passers-by are inevitably involved in the Walkman-theatre, as either actors (holders) or spectators (beholders)"

² <https://carryprinciple.wordpress.com/>

(1984: 177). Hosokawa critiques the assumption that personal stereos would lead to a disconnect among listeners, using their headphones to seclude themselves from the outside world. Instead, he proposes the affordance of a secret theatrical experience, forging new relationships between a listener, inconspicuous by-passers, and their surroundings.

Mobile media scholars such as Michael Bull similarly argued that personal stereos were employed as a way of managing the stresses of urban life. Having a musical soundtrack while negotiating everyday practices created a “cinematic experience” for the user. These changes in our listening modes, spatial relations and perceptual habits have been explored by many sound artists, among whom Janet Cardiff has been perhaps the most influential. Brandon LaBelle identified the notion of ‘headscapes’ (2015) to describe Cardiff’s audio walks and mobile mediated soundwalking, arts practices that highlighted the discrepancies fostered by wearing headphones.

Another cornerstone of this field is research exploring a smartphone’s capacity for creative musical applications, and the physical act of mobility as a performance paradigm. Mobile music as a specific genre was first identified and theorised at the Mobile Music Workshop (MMW), a series of events organised by a collective of researchers, artists and academics including Lalya Gaye, Frauke Behrendt and Atau Tanaka. The MMW group were integral in promoting the idea of mobile music as a form of new media practice that encompassed music activities, sound art and community-based projects. Mobile music was classified as a term that: ‘...covers any musical activity using portable devices that are not tethered to a specific stationary locale... thereby leveraging novel forms of musical experience.’ (Gaye et al. 2006: 22).

An important canonical work for music performance with mobile phones is *4 Hands iPhone* (2009-11) by Adam Parkinson and Atau Tanaka. Tanaka had been exploring the application of mobile technology in collaboration with Petra Gemeinboeck in a

‘multi-media mobile artwork’ *Net_Dérive* (2006), but *4 Hands iPhone* shifts the focus of the smartphone beyond simply a consumer icon to something closer to an expressive musical instrument (Tanaka, 2010). For their performance system, Parkinson and Tanaka used open-source Pure Data (Pd) ported over to the generative audio app *RjDj*. *RjDj* was promoted as a new genre of ‘reactive music’, with interactive versions of existing music recordings using a device’s internal microphone to alter the audio by triggering a set of digital signal processing (DSP) filters.

Kate Crawford’s essay, *Four Ways of Listening with an iPhone* describes *RjDj* as producing a sensation of *meta-listening*. Listening to the environment through the app’s reactive scenes creates a displacement in the real-world – ‘listening to a place in the present, but strangely modified’ (2014: 216). The in-ear style of mobile earbuds ushers in the surrounding environment with unusual closeness, while at the same time transforming perceived sounds through DSP. On a final note, Crawford considered that the option to record individual scenes with *RjDj* offers a tantalising glimpse of an *a priori* impossible opportunity; to hear another’s listening. She asks: ‘Can one make a listening listened to? Can I transmit my listening, unique as it is? That seems so impossible and yet so desirable, so necessary too’. Using the inbuilt mic of a mobile device and DSP becomes a form of digital immersion, producing an almost hallucinatory listening experience.

A decade later, the paradigms of music production made with mobile sound apps had been classified as *Situated Composition* (Thulin 2017). *Situated Composition* is an approach that entails multiple relationships between a practitioner, their situation and their surroundings. These past studies and illustrative examples provide the grounding for this article, situating mobile-mediated listening as a strategy for extending the boundaries of musical agency, and as a phenomenological approach to expanding our subjective experience of the world.

3. EXTENDED FIELD RECORDING

In 2013, RjDj's developers Reality Jockey closed its website, removing its apps from circulation. Yet the exchange of open-source materials for programming mobile apps continued; *Libpd*³ was used to build PdDroidParty⁴ for Android devices which directly inspired the iOS app PdParty⁵. A deluge of sound and music apps began to appear on Apple's App Store, taking advantage of the advanced processing and screen-based interactions of the iPad with its larger screen. Quite by chance, I stumbled upon a novel way of combining field recording and electronic music using DSP apps to process audio captured by the iPad's inbuilt microphone; its fidelity optimised for close range recording as well as capturing background noise from longer distances. Another reason for concentrating on an iPad's mic and DSP apps was the ability to overcome the often-prescribed sounds of mobile apps. Utilising a mic to capture the sounds of my immediate surroundings meant I was not confined to an app developer's embedded choice of synthesised sound.

*Xtended Field Recording: Corsica Soundscape*⁶ is an example of this fieldwork I termed as 'Xtended field recording', with the intension of going beyond field recording. It features *Samvada*⁷, a simulation of a Sitar instrument, but here I am only its accompanying drone function. Audio from the iPad's mic is passed through a comb-filter system and controlled with a simple set of slider GUI objects. Holding the device horizontally and using two or three fingers of one hand, I repurposed the app's slider controls into performance control gestures. In Figure 1, the top right-hand slider (Samvada Mic Mix) shows the dry/wet signal used to shift between the mic input and the processed signal. On the left, the tone, sustain and pluck sliders are used to alter the texture of the pitched signal, while a simple reverb adds a wider spatial element to the mix.

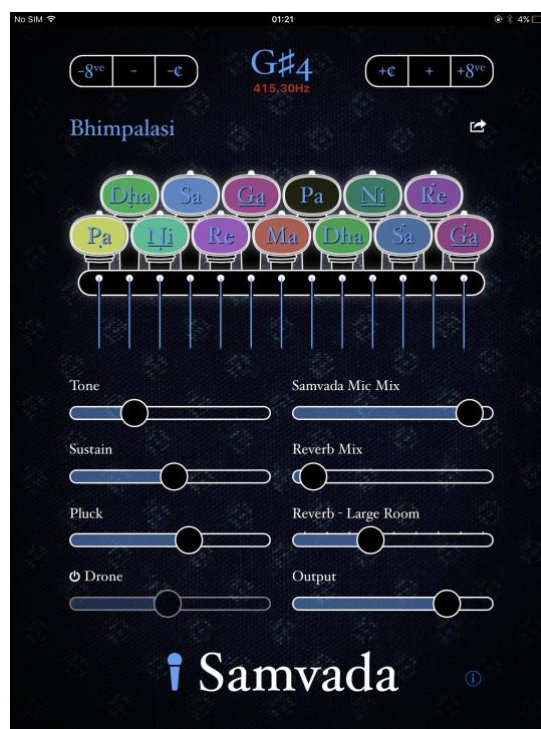


Figure 1: Samvada's slider GUI objects. Screen grab.

The key and harmony settings at the top alter the scale of the pitched filtering, modulating across harmonisations based on classical Indian raga tuning. In this way, the app is no longer used as an imitation of an acoustic instrument, but instead is employed as a multitouch signal processing system. Hidden away in the app's settings is an option to record a stereo Wav file direct to the device. In this recording, I was able to wade in the shallows of the sea while processing the audio. The ability to be so close to water while making electronic sound was a revelation. There is even a moment when a group of passers-by ask for a light for a cigarette, unaware I was playing, recording or performing. It was a lightbulb moment.

Samvada was one of the few apps that allowed audio to be recorded, otherwise I carried a separate hard disk recorder to capture the results. I began examining app interconnectivity as a method of reclaiming agency over standalone apps. *Audiobus*⁸ is an

³ <https://puredata.info/downloads/libpd>

⁴ <https://droidparty.net/>

⁵ <http://danomatika.com/code/pdparty/guide>

⁶ <https://bit.ly/2CrgcoH>

⁷ <http://iotic.com/samvada/>

⁸ <https://audiob.us/apps/>

IOS third-party app that does not generate sound itself but acts as an inter-app routing system. Audiobus uses the Inter App Audio (IAA) protocol that allows apps to announce audio input and outputs to each other. Like this, separate apps can be partitioned into input, effects and output slots. IAA allows single purpose apps to be chained together, like a guitarist patching together a series of effects pedals to create unique sounds. Now the signal stereo output could be recorded internally, thus eliminating the need for an additional audio recorder.



Figure 2: Audiobus partitioning apps into input, effects and output slots. Screen grab.

Figure 2 demonstrates how Audiobus allows individual apps to be partitioned and fed into each other. In this example, Samvada is the signal input processed through *Echopad*⁹, a multi-delay effect and sound on sound looper. Echopad's motion looper automates effects and stereo panning, with layered loops sampled on the fly represented as vinyl records (see figure 3). Patched to the output is TwistedWave Recorder, abbreviated as *TWRecorder*¹⁰. TWRecorder records either mono or stereo files as Wav or MP3 files, and I found it to be the most reliable of the iOS audio editor applications.

Although having access to one app at a time onscreen, I could move around while cradling the iPad under one arm, interacting with the

screen without monitoring it visually. The added element of freedom allowed me to experiment with mobile audio processing in places that previously would have been out of bounds. The interconnectivity of apps promised to be a way of regaining a sense of ownership when constructing a performance system out of existing systems. *Sounds of the Valley*¹¹ is an example using this system, a sound study of south-east France including church bells, crowing cockerels and flowing water. Where before I would make field recordings and take them to the studio to develop new material, now I was able to compose directly, in situ.

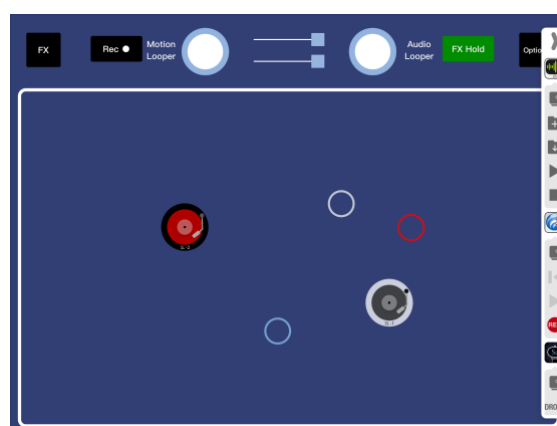


Figure 3: Echopad's looper GUIs with Audiobus controls as a vertical strip r/h side. Screen grab.

The combination of meta-listening, walking and musical creation draws heavily from Hildegard Westerkamp's practice of soundscape composition and soundwalking – listening to the environment while exposing our ears to every sound around us. Westerkamp's soundwalks can be defined as an invitation to explore a soundscape through movement: "...an embodied method of personally connecting with the soundscape through focused listening while physically moving through space" (1974: 81). Field recordist Andrea McCartney expands on this definition, proposing that soundwalking can be a practice that involves both listening *and* recording. She refers to Westerkamp's non-

⁹ <http://www.holdernessmedia.com/home/ios-applications/music/echo-pad>

¹⁰ <https://twistedwave.com/>

¹¹ https://soundcloud.com/steve_jones/larisse-bells-cockerels-and-echo-pad?in=steve_jones/sets/soundstudies-skatepark

intrusive style of field recording, suggesting the composer was learning about the Canadian soundscape ‘with an immigrant’s displaced ears’ (McCartney 2014: 220). Westerkamp herself describes the microphone as a ‘moving ear’ (1994:19). In addition to recording audio, the microphone serves as an extended method of listening, acting as a prosthesis. The difference between my work and a soundwalk is I am engaged in an augmented soundwalk, hearing electronic sound augmenting the soundscape of my environment.

4. ON THE MÉTRO

Aiming to discover out how far extended field recordings and augmented soundwalks might be taken technically and physically, I shifted my research area from rural settings to urban environments. Particularly public transport systems. From a practical point, a busy, noisy environment like the Paris Métro provides rich sonic material to excite an app’s DSP processing parameters. I was based in the city at this time and trying to devise a workable, repeatable technique for mobile music using headphones.



Figure 4: Playing Turnado on the Métro. Screen grab.

*Headphonics: Metro #02*¹² is an example of these sustained investigations, that demonstrates my preferred system of an iPad loaded with Audiobus, *Turnado*¹³, a multi-effects app, and TWRRecorder. I had also started using a GoPro camera clipped to my belt to document the onscreen interactions and show the movement and the setting of the recording (figure 4). It was now an autonomous

performance system combining real-time signal processing with audio and visual data capture. I would carry my iPad while travelling the Métro, processing sound during the entire journey. Sometimes I held the device at my side, sometimes holding the device in front of me when sitting down. I would strive to get a balance between processed audio and the sounds of my immediate environment; the rise and fall of the train’s engine, the hiss and squeal of brakes, the automated announcements of each approaching station and the familiar warning tone before the train doors close. It is this overlapping of realworld and synthesised sound that I am calling headphonics; selecting which sonic events to develop, which to omit and embracing chance proceedings as they happen.

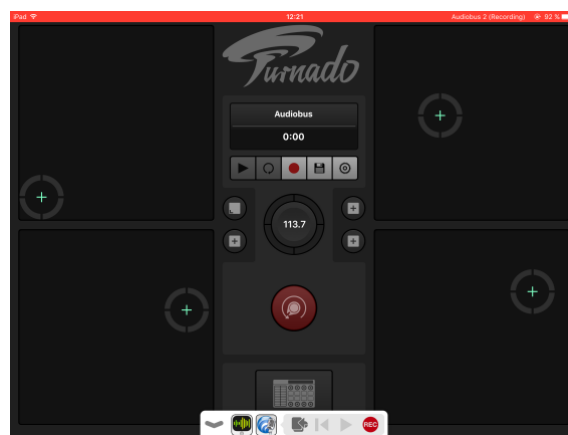


Figure 5: Turnado’s operating GUIs. Screen grab.

For a design perspective, figure 5 shows the ‘performance’ page of Turnado’s graphic user interface (GUI). Four large, circular icons are mapped to various audio effects – in this instance a combined pitch control and delay, an arpeggiated filter, a stuttering looper effect and a reverb unit. The four squares in the center provide the option to either fix the position of the icons or let them return to their starting point. The BPM counter allows the overall tempo to be set by tapping or touching the surrounding edges while red knob in the center is a reset button for all four circular icons. The overall GUI design is clean and simple,

¹² <https://youtu.be/BRk8JE0hMCU>

¹³ <https://sugar-bytes.de/turnado-ipad>

Turnado can be played almost without any visual references and the video demonstrates how rapidly the app responds instantaneously, a single gesture instantly altering a processing effect.

Gradually I learnt to develop a muscle memory with the app, remembering the positions of the processing icons like the frets of a stringed instrument or the fingerings on a piano keyboard. But note in the video how efficient thumbs are for making rapid movements across the screen's x and y axis.

In stark contrast, the interfaces of Turnado's other control pages are much smaller and more detailed, and difficult to operate in a performative manner. In the video, for example at 01:24s, we can see me struggle to change the parameters, pinching and tapping the screen. This is most likely because Turnado was originally a desktop plug-in and suggests the problems in porting established DAW designs over to iOS touchscreens. Its designer Sugar Bytes states 'Turnado's primary focus is on beat manipulation'; the app was intended as a VST plug-in using pre-existing audio files, rather than an autonomous signal processing unit. The bar above the BPM counter with its file, play and record icons shows how the app is meant to be used for playing audio samples. However, I hope this is a small demonstration of app interconnectivity as a method of reclaiming agency over blackboxed systems, with Audiobus allowing a sample playback plugin to become a powerful music instrument.

In the video we can observe my attempting to balance the indeterminacy of the situation – what sounds are available for processing, which person might enter or exit the train – while trying to consider the overall compositional process. I considered this not as improvisation, but as curating serendipity.

Quite by chance, we see a young man enter the train and immediately appeal to other passengers for spare change, food or luncheon vouchers. It was (still is?) a regular occurrence on the Métro, with homeless and obviously desperate people having to beg on the underground system. Was my sampling and manipulating his rehearsed speech a way of shutting out an uncomfortable situation? Or

does this blur the otherwise clearly demarcated boundaries between spectatorship and musicianship?

4.1 THE SECRET THEATRE OF HEADPHONICS

What is the significance of this experience of simultaneous observing and music making? How can it be a performance when unannounced to an audience, who are unable to hear the music. Pieter Verstraete argues a fundamental aspect of headphone mediated listening is its affordance of a secret theatrical experience. The user experiences listening and moving as a 'secret theatre', an experience that forges new relationships between the user, by-passers and their surroundings (Verstraete 2017:2). Verstraete's use of the term secret theatre is taken directly from Shuhei Hosokawa's investigations on the changes in listening modes, spatial relations, and perceptual habits afforded by the Sony Walkman. As we saw earlier in chapter 2.1, Hosokawa suggested that headphones allowed the mobile user to experience walking as a secret theatre, to become a secret listener as well as an actor to the outside world. This chimes with Brandon LaBelle's description of headphonic space, a space between the ears that forces the listener out-of-sync with the exterior world. Wearing headphones, LaBelle suggests, 'define[s] a very different acoustic reality to that of our physical position' (LaBelle 2015: 225). It is in this in-between, headphonic space that a user becomes a secret listener as well as an actor to the outside world.

I would also eavesdrop on other passengers' conversations; at the time I was still learning the French language. As a native English speaker, my understanding of other languages was relatively poor, and it was in public spaces that I was beginning to pick up phrases and idioms of spoken French. Theorist Michel Chion proposed this mode of listening as 'semantic listening': deciphering spoken language or learning to decode a message (1994: 28). Chion's three modes of listening is in reference to the 'audiovisual contract' in cinema, but in this context, I would argue all three modes – causal, semantic and reduced

listening – align to Michael Bull’s notion of the cinematic experience afforded by mobile auditory devices. It could further be linked to Westerkamp’s non-intrusive style of field recording, learning about a soundscape with an immigrant’s displaced ears.

There is perhaps something voyeuristic using a concealed GoPro for documentation, a practice that harks back to the tradition of Walker Evans’s still photography of New York subway passengers made between 1938 and 1941. It conflates the position of the spectator with a fellow passenger. When I first watched back the GoPro footage, I found it reminiscent of the confrontational scene shot on the Métro in Michael Haneke’s ‘Code Inconnu: Récit incomplet de divers voyages/Code Unknown’ (2000). Johan Andersson argues that the static camera should not be understood as an absence of point of view, but rather as an additional presence. It introduces the perspective of a third person, the audience as fellow passenger: “...the spectator of Code Unknown becomes personally implicated in the ethical question of whether one has a moral obligation to intervene and protect a fellow citizen but also in the refusal to look up and acknowledge” (2012: 702).

Elisabeth Weis refers to the privileged listener as the *écouteur* (the eavesdropper, from French *écouter*), equivalent to the voyeur. Weis terms *écouterism*, equating the pleasure in aural stimulation to that of voyeurism. For her, this phenomenon is central to the cinematic experience: “In every case the eavesdropper acquires some form of knowledge [...] a self-knowledge that the listener would not otherwise have recognised” (Weis 1999: 85-6). Verstraete extends this concept to mobile music situations that produce a cinema for the ears. The invisible mobile musician blurs the separate 'roles' of actor, eavesdropper, voyeur and audience.

In the Headphonics: Metro #02 video, notice also how none of the other passengers appear to be aware of my actions. Through listening on my earbuds, I am sharing the same space as everyone, yet somehow occupying another reality to the rest of my fellow passengers. To all outward appearances, there is little to

distinguish headphonics as an identifiable form of music performance. The earbuds I wore were identical to those used by most of the other passengers in the carriage. The set of physical gestures I employed to control the sound processing – tap, swipe and scroll – were like those of my immediate neighbours as they scroll through social media or play video games. This lack of obvious outward visual signs throws up a conundrum: if no one is aware a performance is taking place, does it count as a performance in its traditional sense?

4.1.1. *Headphonics as a new model of interaction*

If there are no identifiable elements to mark out headphonic performance; no culturally understood gestures associated with music making, no traditional performance venue, no one can hear the work except the performer. It raises questions about the status of performativity itself. Mobile musician Martin Koszolko cites my own research when he describes the challenge to artists using IOS devices, in their forcing the artist to redefine assumptions of what constitutes a musical instrument as well as perceptions of music practice (2019: 199). With this new model of interaction, performer and spectator are one and the same person; it starts to deconstruct the hypothetical ‘stage’.

The act of augmenting both public and private spaces with itinerant electronic sound creates a new experience of these spaces. And when music-based activities are transformed into different ways of engaging with physical places, it suggests perhaps that music activities can become closer to what Francois Bonnet describes as a ‘phenomenal investigation, rather than a music of identification’ (2016: 296). Using mobile apps allows music-based and audio recording practices to become something close to a phenomenal investigation of a place, a way of altering and remixing sound that intensifies the auditory perception of a location.

But can headphonic performance be broadened to a group setting? While devising a site-specific piece for Locus Sonus’s *Audio*

Mobility Symposium in Aix-en-Provence ¹⁴, I had intended to use a wireless headphone system utilising the Bluetooth protocol commonly found with ‘silent discos’ ¹⁵. During testing however, I discovered that DSP apps were unable to use the mic input when streamed over Bluetooth. I was limited to a set of headphones connected to a splitter, with a group of chairs for the audience. Audiobus’s developers confirmed the problem with Bluetooth is it simply does not work with apps that can record from the microphone; Apple’s iOS platform does not allow it ¹⁶. When an iOS app enables Bluetooth audio, both input and output are routed through it, disabling the internal mic and speakers. Other companies such as RØDE microphones affirm this: “Unfortunately due to issues with multiple sampling rates, both RØDE Rec and RØDE Rec LE are currently unable to support Airplay or Bluetooth output” ¹⁷. Whether it is issues with sampling rates or the need to protect listener and equipment from accidental feedback loops, this is a basic limitation of the iOS operating system. Still for now, headphonics remains a solitary, individual experience.

Yet it is an experience that can create a sense of euphoria. This euphoria seems to be due to the intense involvement in the music. But the music literally colours the visual world too. Furthermore, the outside world profoundly alters its character; it is perceived like a film.

5. CONCLUSION

I have described how the Paris Métro has been a site and stage for my research in devising a workable technique for mobile music. The intention has been to question the spaces commonly associated with electronic music – where it is made, where it is listened to and experienced. Listening on headphones to my surroundings captured by the microphone of an iPad with DSP apps heightened and expanded my perceptual range of the world. This article has considered the theoretical

aspects of the research, offering thoughts on how musicians might overcome some of the limitations of digital tools to allow new creative and compositional approaches. I hope to have drawn attention to the potential of this research and suggest that there are still new avenues to explore that entail the use of mobiles and signal processing apps to transform and remix the urban soundscape.

With headphonics, there are elements of field recording, improvisation, environmental context and temporal dislocation – sound and location are thoroughly entangled. While it might have a musical focus, it is not solely about music making. Mobile music can be as much about creating a phenomenological stance, experiencing the world through the microphone of a consumer device. This model of mobile music has the potential to change the methods and experiences of making electronic music, and consequently generate a new kind of performer identity.

6. REFERENCES

- Andersson, J. ‘Variations of an archetypal scene: the Paris Métro confrontation in Michael Haneke’s *Code Unknown*’. In *Environment and Planning D: Society and Space* 2013, volume 31, pp. 693 – 993. 2013.
- Bonnet, F. *The Order of Sounds : A Sonorous Archipelago*, translated by Robin Mackay. Falmouth : Urbanomic. 2016.
- Bull, M. *Sounding Out the City: Personal Stereos and the Management of Everyday Life*. Oxford: Berg. 2000.
- Chion, M. *Audio-Vision : Sound On Screen*. New York : Columbia University Press. 1990.
- Crawford, K. ‘Four Ways of Listening with an iPhone. From Sound and Network Listening to Biometric Data and Geolocate Tracking’. In Hjorth, L. Burgess, J. Richardson, I. (eds.) *Studying Mobile Media: Cultural Technologies, Mobile Communications and the iPhone*. New York and London: Routledge. 2012.
- Gaye, L. Holmquist, L.E. Behrendt, F. Tanaka, A. ‘Mobile Music Technology: Report on an Emerging Field’. In *Proceedings of NIME 06*, Paris, France, 2006.
- Hosokawa, S. *The Walkman Effect*. Popular Music, 4, pp.165-180. 1984. doi:10.1017/S026114300006218

¹⁴ <http://locusonus.org/>

¹⁵ <http://www.silentdisco.fr/index.php>

¹⁶ <http://www.forum.audiob.us/discussion/201/audiobus-bluetooth/pl>

¹⁷ <http://en.roderec.com/faq/compatibility-of-roderec-with-airplay-bluetooth-audio>

Koszolko, M. 'The Tactile Evolution: Electronic Music Production and Affordances of iOS Apps'. In Gullö, J-O (ed.) *Mono: Stereo: Multi*, Stockholm: Royal College of Music (KMH), p. 187-204. 2019.

LaBelle, B. *Background Noise: Perspectives on Sound Art, Second Edition*. New York, London: Bloomsbury, 2015.

McCartney, A. 'Soundwalking: Creating Moving Environmental Sound Narratives'. In Gopinath, S. Stanyek, J. (eds.) *The Oxford Handbook of Mobile Music Studies, Volume 2*. Oxford & New York: Oxford University Press. 2014.

Tanaka, A. 'Mapping Out Instruments, Affordances and Mobiles'. In *Proceedings of NIME 010*, Paris, France, 2010.

Thulin, S. 'Situating Composition in Emerging Mobile Sound Production Technologies and Practices'. In *Organised Sound 22 (1)*. Cambridge: Cambridge University Press, 2017.

Verstraete, P. 'The Secret Theatre Revisited: Eavesdropping on Locative Media Performances'. In *Journal of Sonic Studies (5)*, 2017 : <https://www.researchcatalogue.net/view/411830/411831>

Weis, E. 'Eavesdropping: An Aural Analogue of Voyeurism?'. In Brophy, Philip (ed.), *Cinesonic: The World of Sound in Film* (pp. 79-107). North Ryde: Australian Film Television & Radio School. 1999.

Westerkamp, H. 'Soundwalking'. In *Sound Heritage 3(4)*. 1974.

Westerkamp, H. 'The Soundscape on Radio'. In Augaitis, D. Lander, D (eds.) *Radio Rethink: Art, Sound & Transmission*. Banff: Walter Phillips Gallery. 1994.

Actes des Journées d'Informatique Musicale 2023
24-26 mai 2023
Maison des Sciences de l'Homme Paris Nord

**SESSION « OUTILS POUR LA COMPOSITION
ET L'IMPROVISATION »**

SYNTHÈSE CONCATÉNATIVE EN RÉALITÉ VIRTUELLE

Jonathan Bell

Prism-CNRS; XR2C2-CTEL
belljonathan50@gmail.com

RÉSUMÉ

Cette étude examine comment la métaphore d'un espace tridimensionnel peut être utilisée de manière créative pour la composition instrumentale. De nombreux outils de synthèse sonore concaténative permettent de représenter dans un espace à n dimensions de grandes quantités de sons, affichant ainsi sur une carte (2d) ou dans un espace (3d) des sons alors décorrélés de leur agencement dans le temps. Si le potentiel de ces systèmes pour créer des instruments interactifs est une évidence, leur statut en tant que partitions musicales reste à évaluer plus en détail, car la dimension temporelle y est absente. En réalité virtuelle, une telle représentation de données sonores - la partition - devient une carte (ou un monde) tridimensionnel dans lequel l'utilisateur navigue librement. L'expérimentation de ce dispositif par le biais de la composition, de la conception d'instruments et de l'improvisation a montré un potentiel de simulation d'instruments acoustiques plausibles, en utilisant des techniques d'apprentissage machines pour modéliser des instruments virtuels à partir de quantités relativement faibles de données (par exemple, 20 minutes d'audio pour modéliser une clarinette). Cette méthode offre des possibilités prometteuses pour l'exploration de fragments instrumentaux regroupés par timbre, registre, dynamique et techniques instrumentales. Que ces cartes s'identifient ou non à des partitions musicales, elles contribuent à répondre à un problème esthétique formulé par Lev Manovich : "comment fusionner base de données et narration dans une nouvelle forme".

1. INTRODUCTION

1.1. Are scores maps ?

La question soulevée dans cet article est en grande partie inspirée par une étude de Daniel Miller, *Are scores maps ? A cartographic response to Goodman* [1], dans laquelle une tension dialectique entre deux concepts (partition et carte) conduit à un questionnement sur le rôle et la fonction de la notation musicale. Miller propose que, malgré les conventions divergeantes en apparence, la structure sous-jacente des partitions est étroitement liée à celle des cartes : "*Les composantes notationnelles des partitions sont mieux comprises comme des caractéristiques de surface contingentes, renforcées par une structure représen-*

*tationnelle sous-jacente de type carte. Dans cette optique, les partitions sont considérées comme des cartes hautement codifiées, et les symboles notationnels des partitions ne constituent qu'un des multiples modes de représentation exploités par ce cadre". Les partitions peuvent donc être conçues comme un simple sous-ensemble de cartes. Le rapprochement entre ces deux notions est mieux illustré par un exemple historique : des œuvres célèbres de l'avant-garde d'après-guerre (de M. Feldman, J. Cage, E. Brown, P. Boulez, K. Stockhausen et A. Boucourechliev pour n'en citer que quelques-unes) ont pris la forme de partitions graphiques dans lesquelles l'interprète pouvait librement naviguer. Boucourechliev a nommé nombre de ses compositions "Archipels" qui évoque la même métaphore. Ces partitions présentaient typiquement des fragments de notation répartis sur toute la page, afin d'émanciper l'œuvre de la linéarité imposée par la notation traditionnelle, une idée largement discutée dans *Opera Aperta* d'U. Eco [2]. Dans la deuxième sonate pour piano de Boulez, par exemple, cette coexistence de bribes de notation conventionnelle et d'une disposition en forme de carte sur la page illustre comment les cartes permettent des cas hybrides convaincants, dans lesquels tous les symboles sur la page ne fonctionnent pas comme une carte. De même, dans les partitions plus conventionnelles/linéaires, Miller souligne que seules certaines caractéristiques sont isomorphes (ou semblables à des cartes) : "*Les partitions sont des cartes qui sont isomorphes avec les structures spatiales et temporelles des œuvres musicales qu'elles représentent, tandis que d'autres caractéristiques graphiques peuvent être purement contingentes ou accidentelles. Cela met en évidence une propriété intéressante des cartes : il suffit qu'elles soient isomorphes par rapport à un sous-ensemble de propriétés de l'espace qu'elles représentent". Les cartes, d'une manière qui rappelle les partitions, visent à guider un utilisateur ou à inciter un exécutant à agir, et cet objectif ne doit pas nécessairement obéir à des relations de correspondance strictes, systématiques et univoques. Pour Miller en effet, ce mélange d'isomorphisme et de contingence répond à la célèbre attaque de Goodman contre John Cage, formulée comme suit : "Sans stipulation d'unités significatives minimales d'angle et de distance, la p. 53 du Concert pour piano et orchestre de John Cage de 1960 n'est pas syntaxiquement différenciée". L'observation du philosophe l'a conduit à une critique virulente de l'approche de John Cage en matière de notation graphique : "Sous le système proposé, il n'y a pas de caractères joints et différenciés ou de classes de conformité, pas de**

© LA VERSION ANGLAISE DE L'ARTICLE A ÉTÉ ACCEPTÉE
À LA CONFÉRENCE TENOR 2023 À BOSTON.

notation, pas de langage, pas de partition"[3], un point de vue qui, en retour, s'est avéré largement impopulaire. Les réflexions sur les cartes et les partitions peuvent sembler ici trop théoriques. L'ère numérique, cependant, pousse les compositeurs à penser les partitions d'une manière nouvelle, dans laquelle l'hybridation entre systèmes interactifs et notation laisse plus de place à notre notion de carte comme partition. Dans "The digital score, musicianship, creativity and innovation" [3], Craig Vear propose que : *"L'objectif principal de toute partition numérique est de communiquer des idées entre musiciens à l'aide de la technologie numérique"*, plaçant ainsi la technologie au centre de la notation contemporaine, et élargissant par la même occasion l'éventail des systèmes jusqu'alors qualifiés de "partitions". Vear propose également que *"certaines partitions numériques puissent ressembler à des jeux informatiques où l'interprète prend des décisions sur ce qui se passe ensuite."* L'image du jeu vidéo, tout comme celle d'une carte que l'interprète peut parcourir, évoque d'abord la liberté, mais aussi un important changement d'orientation lorsqu'on la considère du point de vue d'un compositeur traditionnel. Son métier se rapproche alors de celui d'un concepteur d'instruments, que Vear considère toujours comme appartenant au domaine des partitions numériques. Dans un chapitre intitulé *"La nature des partitions numériques : étendre les signatures de base"*, il déclare : *"la partition peut être intégrée dans la conception d'un instrument, l'instrument peut être une partition-système d'électronique contrôlée par un logiciel génératif"*. Dans de telles circonstances, la frontière entre la partition et la conception d'un instrument numérique s'estompe, et diverses formes de systèmes interactifs interrogeant une base de données seront considérées comme une partition. Dans certains des premiers comptes rendus des utilisations artistiques de CataRT [4], Diemo Schwartz décrit ses représentations d'un corpus sonore comme des "partitions navigables" ou des "instruments de partition" : le sujet de la pièce était les "partitions navigables" ou les "instruments de partition", dans lesquels différents types d'utilisateurs joueraient le son ou la musique, en naviguant à travers la partition" [5]. La découverte de Schwartz présente alors ce paradoxe qu'elle propose une partition qui représente le temps.

1.2. Les cartes ne représentent pas le temps

Les partitions musicales sont utilisées pour transmettre des informations sur les matériaux musicaux en fonction du temps. Les cartes, en revanche, sont des représentations statiques d'une zone. Elles ne représentent pas le temps ou les changements dans le temps, et sont typiquement utilisées pour transmettre des informations sur les caractéristiques géographiques et les relations spatiales, et sont également un outil puissant de visualisation en science des données, comme nous le développerons au chapitre 5. *"De nombreux objets des nouveaux médias ne racontent pas d'histoires ; ils n'ont ni début ni fin"* [6]. Cette citation de Manovitch nous aide ici à introduire une idée stimulante pour un compositeur traditionnel : ignorer

la dimension temporelle d'une partition peut conduire à des approches innovantes de la composition. Nous allons maintenant exposer comment de telles parcelles de données, ou plus globalement des bases de données, peuvent également être considérées comme une forme d'art. Enfin, l'absence de temps inhérente à ces concepts nous incite à utiliser dans la section 1.2.2 des arguments esthétiques pour comprendre ce qui pourrait rester artistiquement intéressant dans le cadre d'un contrôle relativement souple du temps dans la composition musicale.

1.2.1. La base de donnée comme matériau artistique

Les bases de données et l'exploration de données en musique peuvent être considérées comme une forme d'art dans le sens où elles impliquent l'utilisation de la créativité et des compétences analytiques pour extraire des idées et des connaissances de grands ensembles de données sonores. Cela nécessite la capacité d'intégrer la créativité musicale en identifiant des modèles, des tendances et des relations dans les données qui pourraient ne pas être apparents autrement. Les romans et les films sont, comme la musique, la danse ou le théâtre, des formes d'art temporelles. La perspicacité de Manovitch sur la forme d'art de la base de données suggère que les nouveaux médias ne sont pas subordonnés au temps, ou narratifs, de la même manière ; cette considération peut aider un compositeur à modifier certaines idées préconçues sur la base de données. Cette considération peut aider un compositeur à modifier certains à priori sur son approche de la forme, de l'articulation ou de la narration : "Après que le roman, et par la suite le cinéma, aient privilégié la narration comme forme clé de l'expression culturelle de l'ère moderne, l'ère informatique introduit son corrélat - la base de données. De nombreux objets des nouveaux médias ne racontent pas d'histoires ; ils n'ont ni début ni fin ; en fait, ils n'ont aucun développement, thématique ou autre, qui organiserait leurs éléments en une séquence"[6]. Cette absence de narration dans l'art des nouveaux médias, que Manovitch rattache à la post-modernité, trouve un écho intéressant dans la pensée de Morton Feldman, qui voyait dans l'avant-garde européenne un désir excessif de contrôle de la temporalité.

1.2.2. Morton Feldman and the European clock makers

Dans ses écrits [7], Morton Feldman se souvient d'une discussion qu'il a eue avec Karlheinz Stockhausen : *"Il était convaincu qu'il me démontrait la réalité, que le rythme, et le placement possible des sons par rapport à celui-ci, était la seule chose à laquelle le compositeur pouvait s'accrocher de façon réaliste. [. . .] Franchement, cette approche du temps m'ennuie. Je ne suis pas un horloger. Ce qui m'intéresse, c'est de saisir le temps dans son existence non structurée. C'est-à-dire que je m'intéresse à la façon dont cette bête sauvage vit dans la jungle - et non dans un zoo"*. Feldman, souvent avec humour, insiste sur cette idée : *"Laisse les sons tranquilles, Karlheinz, ne les pousse pas - même pas un peu ?"*. Cette approche

presque passive de la composition fait écho à la pensée d'inspiration zen de John Cage, et s'inspire aussi et surtout des peintres : *"un peintre sera peut-être d'accord avec le fait qu'une couleur insiste pour avoir une certaine taille, indépendamment de ses souhaits [...] Il peut simplement lui permettre d'"être" [...] Depuis quelques années, je me rends compte que le son a une prédilection pour suggérer ses propres proportions [...] Tout désir de différenciation doit être abandonné"*. La musique de Feldman a exploré cette idée de surface de nombreuses façons, notamment par l'utilisation de lignes musicales longues, lentes et statiques, la répétition de motifs simples, l'absence de contraste (qu'il appelle différenciation) et l'utilisation de dimensions temporelles non conventionnelles (certaines œuvres, comme le deuxième quatuor à cordes, durent plus de quatre heures). Cette approche non conventionnelle de la forme se rapporte plus largement à une opposition entre les avant-gardes européenne et américaine dans les années 50-60. Les outils décrits au chapitre 5 ont incité l'auteur à penser la forme musicale comme de longues toiles temporelles statiques, comme les appellerait M. Feldman : *"Mon obsession de la surface est le sujet de ma musique. En ce sens, mes compositions ne sont pas du tout des "compositions". On pourrait les appeler des toiles temporelles dans lesquelles j'amorce plus ou moins une teinte globale de la musique"*. En écoutant presque toutes ses œuvres, on se rend compte que l'ensemble du morceau partage le plus souvent la même atmosphère du début à la fin. Une pièce avec une combinaison instrumentale reconnaissable, comme *Why patterns* pour flûte, piano et célesta ou Clarinette et Percussion, ont une forte empreinte acoustique et illustrent comment l'écoute par apprentissage automatique pourrait être utilisée pour travailler sur de tels matériaux.

2. CORPUS-BASED CONCATENATIVE SOUND SYNTHESIS (CBCS), LA SYNTHÈSE CONCATÉNATIVE AUJOURD'HUI

La synthèse sonore concaténative basée sur un corpus (CBCS) est une technique utilisée en informatique musicale qui consiste à construire un son ou un morceau de musique en concaténant (en joignant ensemble) de plus petites unités sonores, comme des phonèmes dans la synthèse vocale ou des phrases musicales dans la synthèse musicale. Il peut être utilisé pour modéliser un musicien instrumental improvisateur en créant une base de données de phrases ou de segments musicaux enregistrés qui peuvent être combinés et réarrangés en temps réel pour créer une performance musicale qui semble être improvisée.

Aujourd'hui âgé de près de 20 ans si l'on se réfère aux premières publications de CataRT [4], CBCS jouit aujourd'hui d'une popularité croissante. Diverses applications sont aujourd'hui basées sur des principes similaires (AudioStellar, Audioguide, LjudMAP ou XO). La démocratisation des outils d'analyse audio et d'apprentissage automatique tels que le package FluCoMa (pour Max, SuperCollider et Pure Data) encourage les praticiens de l'in-

formatique musicale à s'engager dans ce domaine à la croisée de la création musicale et de la science des données/apprentissage automatique.

2.1. Espace Timbre

Malgré des avancées prometteuses dans le domaine de l'apprentissage profond appliqué à la synthèse sonore [8] [9], les outils CBCS doivent leur popularité à une métaphore qui remonte aux débuts de l'informatique musicale : la notion d'espace timbrique, développée par Wessel [10] et Grey [11], selon laquelle les qualités multidimensionnelles du *timbre* peuvent être mieux comprises à l'aide de métaphores spatiales (par exemple, le timbre du cor anglais est plus proche du basson que celui de la trompette).

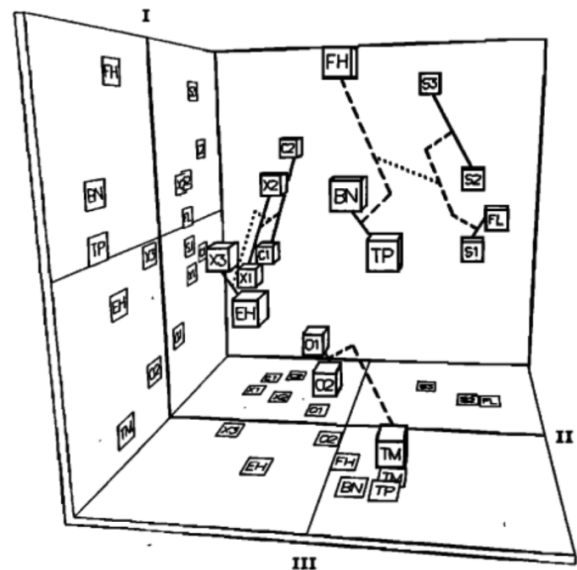


Figure 1.

L'héritage de la métaphore Timbre Space se retrouve également dans diverses itérations du projet Orchidea [12] et continue d'inspirer des générations de compositeurs et de réalisateurs en informatique musicale.¹

Les pionniers des études sur la perception du timbre tels que Grey [11], J.C. Risset, D. Wessel, [15] ou Stephen McAdams [16] [17] définissent le plus souvent le timbre en soulignant ce qu'il n'est pas. Risset et Wessel, par exemple, le définissent comme suit : *Il s'agit de l'attribut perceptif qui nous permet de distinguer les instruments d'orchestre jouant la même hauteur et ayant le même volume sonore.*

La co-variance de ces paramètres (hauteur, intensité sonore et timbre) conduit cependant Schwarz à distinguer les

1. Daniele Ghisi, co-auteur du package *bach* [13] pour Max, occupe un rôle ici puisqu'il a travaillé à la fois sur les projets *Orchidea* et *FluCoMa*. Certains objets de sa bibliothèque *dada* plus tardive [14] montrent également une influence de CataRT (l'objet *dada.catart* a été renommé plus tard *dada.cartesian*). La base *dada.base*, enfin, pourrait avoir été une source d'inspiration pour la manipulation de bases de données dans *Flucoma/Max*. Un extrait d'une de ses présentations est disponible ici : <https://youtu.be/LD0ivjyuqMA?t=3032>

notions d'espace timbrique et de CBCS : *'Notez que ce concept est similaire mais non équivalent à celui de l'espace timbrique mis en avant par Wessel et Grey [7, 24], puisque le timbre est défini comme les caractéristiques qui servent à distinguer un son d'un autre, qui restent après avoir retiré les différences de sonie et de hauteur. Notre espace sonore inclut explicitement ces différences qui sont très importantes pour l'expression musicale.* [18]

Le workflow décrit dans le chapitre 5 a donné dans la pratique une forte évidence de l'interdépendance entre le registre, le timbre et la dynamique, en particulier lorsque l'analyse est effectuée sur un seul fichier sonore d'instrument (par exemple 30 minutes de flûte solo), et découpé en échantillons courts. Le système sera alors précisément capable de trouver des similitudes entre des passages instrumentaux joués dans le même registre, la même dynamique et la même technique de jeu (par exemple, une flûte jouant des trilles rapides mezzo forte, dans un registre moyennement bas, avec de l'air).

2.2. Synthèse concaténative basée sur un corpus - État de l'art

Un large éventail de technologies peut aujourd'hui être qualifié de synthèse concaténative basée sur un corpus, dans le sens où elles permettent, par la segmentation et l'analyse, d'explorer de grandes quantités de sons. Certaines d'entre elles se présentent comme des solutions "prêtes à l'emploi", comme le récent *Audiostellar* [19], ou SC-MIR² pour SuperColider. L'AudioGuide de Hackbarth [20] offre une approche légèrement différente puisqu'il utilise la morphologie/chronologie d'un fichier son pour produire une sortie concaténée. Dans le monde de Max enfin, deux environnements apparaissent comme hautement personnalisables : MuBu [21] de l'IRCAM et le projet plus récent FluCoMa [22] financé par l'UE. CataRT est maintenant entièrement intégré dans MuBu, dont l'objectif englobe l'analyse audio multimodale ainsi que la reconnaissance automatique des mouvements et des gestes [23]. Cela rend MuBu extrêmement polyvalent, mais aussi difficile à appréhender. Les outils de traitement de données de MuBu sont principalement exposés dans le cadre du plugin pipo [24], qui peut calculer par exemple l'analyse mfcc sur un tampon audio donné³ en intégrant le plugin pipo.mfcc dans l'objet mubu.process.

FluCoMa se veut également polyvalent, mais semble

2. Une démo est disponible sur : <https://youtu.be/jxo4StjV0Cg>

3. MFCC signifie Mel-Frequency Cepstral Coefficients. Il s'agit d'une méthode d'extraction de caractéristiques couramment utilisée dans les systèmes de reconnaissance de la parole et du locuteur. Les CCFM sont utilisés pour représenter les caractéristiques spectrales d'un son sous une forme compacte, plus facile à analyser et à traiter que la forme d'onde brute. Elles sont calculées en appliquant une série de transformations au spectre de puissance d'un signal sonore, y compris une déformation à l'échelle de Mel de l'axe des fréquences, en prenant le logarithme du spectre de puissance et en appliquant une transformée en cosinus discrète (DCT) aux coefficients résultants. Les coefficients résultants, appelés MFCC, capturent les caractéristiques spectrales du son et sont généralement utilisés comme caractéristiques pour l'apprentissage de modèles d'apprentissage automatique pour des tâches telles que la reconnaissance de la parole et l'identification du locuteur.

particulièrement adapté à la réalisation de deux tâches spécifiques populaires. Avec une connaissance limitée du cadre de travail et de la théorie sous-jacente aux algorithmes qu'il utilise (comme ceux de la réduction de la dimensionnalité, de l'analyse mfcc ou de l'entraînement des réseaux neuronaux), le cadre de travail permet : 1/ de segmenter, analyser et représenter/restituer un corpus sonore 2/ d'entraîner un réseau de neurones pour contrôler un synthétiseur, d'une manière qui rappelle le Wekinator de Fiebrink [25].

Seuls les outils de segmentation, d'analyse, de représentation et de lecture (décrits en détail dans le chapitre 5) ont été utilisés ici, car ils répondent précisément aux besoins de la synthèse par corpus.

3. PRÉMICES

Certaines des premières versions du package FluCoMa fournissaient déjà des algorithmes efficaces de détection d'onset (onsetslice⁴), ce qui a encouragé l'auteur à creuser davantage dans leur environnement.

À l'époque, tout le matériel de composition généré avec le package utilisait l'objet fluid.bufstat⁵ pour exécuter une analyse statique de la hauteur (et une analyse de confiance de la hauteur) sur chaque tranche d'un fichier sonore pré-existant donné. Comme on peut l'entendre dans cet exemple d'accompagnement⁶, les tranches ont été classées par hauteur (plus l'indice est bas, plus le registre est bas), bien qu'une certaine imprécision puisse se produire en raison de la simplicité de l'analyse (la section 5 décrira des stratégies plus élaborées).

La détection d'onset combinée à l'analyse statistique de la hauteur a été utilisée pour la première fois à grande échelle dans une pièce (Chef 2.0⁷) dans lequel toutes les parties instrumentales avaient été générées avec des techniques rudimentaires comparables de recherche d'informations musicales.

Certaines expériences de diverses formes de contrôle de la RV sur de tels outils d'analyse semblaient déjà prometteuses à cette époque⁸. Le logiciel PatchXR sera abordé dans la section 5.9.

4. MOTIVATIONS

L'un des objectifs de la présente étude est de tirer parti des nombreuses capacités de visualisation, d'interaction et de données sur les gestes de mouvement disponibles dans un environnement VR. Le portage dans la VR d'une analyse réalisée dans Max/Flucoma présente des difficultés,

4. <https://learn.flucoma.org/reference/onsetslice/>

5. <https://learn.flucoma.org/reference/bufstats/>

6. https://youtu.be/UNj7_TI8SVs

7. Simulation : <https://www.youtube.com/watch?v=MkMKVm3G3W8>
Résultat : https://youtu.be/Sc4Ye_rnSO8?t=9893. Bien que cela ne soit pas pertinent pour la discussion ici, le mouvement du bras du chef d'orchestre a été mis en correspondance avec la vitesse du curseur sur les écrans des interprètes à l'aide de INScore [26] et Gesture Follower [27] dans Mubu for Max.

8. https://youtu.be/DC_BL_HGPLA

mais j'explore actuellement diverses manières possibles d'interagir avec une analyse basée sur un corpus ⁹. Après quelques essais dans lesquels les coordonnées x y z d'un monde représentaient directement des descripteurs audio tels que la hauteur du son et la centroïde, j'ai utilisé plus systématiquement l'analyse mfcc et la réduction de dimensionnalité, comme décrit dans la section suivante.

L'utilisation de l'apprentissage automatique (réduction de la dimensionnalité) dans ce dernier cas rend un monde dans lequel les coordonnées absolues de chaque point n'ont plus de lien avec l'espace des descripteurs (les sons aigus ne peuvent pas être mis en correspondance avec l'axe des y par exemple), mais offre de manière frappante des informations de regroupement convaincantes, qui sont liées aux différents styles de jeu de l'instrument analysé : par exemple, dans cet extrait basé sur des sons de flûte, l'ouverture montre une opposition claire entre deux types de gestes : 1/ les sons staccato et 2/ les matériaux de type échelle legato. Ce contraste sonore est rendu explicite par un mouvement de l'avatar qui saute d'un groupe de boutons à un autre.

5. WORKFLOW

Après de nombreuses tentatives d'écoute et de jeu avec ce système ¹⁰, j'étudie maintenant comment diversifier les métaphores VR pour exciter le moteur de synthèse basé sur le corpus, ainsi que les différentes manières dont la synthèse peut être rendue sur un orchestre de haut-parleurs équipés de RaspberryPi (voir chapitre 6).

5.1. Corpus Selection

Mes expériences se sont concentrées presque exclusivement sur les corpus d'instruments acoustiques. Les outils présentés ici peuvent générer efficacement de la musique instrumentale virtuose plausible (comme c'était également le cas dans le morceau Chef 2.0 présenté plus haut), mais des utilisations récentes ont permis d'obtenir des résultats plus satisfaisants dans des types de textures plus lentes, plus calmes, de type "Feldman". Diverses limitations du côté de la lecture (soit dans VR, soit sur un échantillonneur Pure Data pour RaspberryPi décrit dans le chapitre 6) ont imposé des restrictions dans les premières étapes sur la quantité de données qu'il pouvait gérer (moins de 5 minutes en AIFF dans VR) ou le nombre de tranches dans lesquelles l'échantillon pouvait être découpé (256 à cause de la limitation des listes dans Max). Ces deux limitations ont été surmontées par la suite (utilisation du format ogg dans VR, augmentation de la taille du buffer interne dans fluid.buf2list), ce qui a permis d'obtenir des modèles bien plus convaincants.

5.2. Analyse dans FluCoMa

L'utilisation de la synthèse concaténative pour modéliser un musicien instrumental improvisateur implique gé-

9. <https://youtu.be/cM-utI xv2Ww>

10. Extrait Cello; Extrait Flûte; Playlist pour divers instruments...

néralement plusieurs étapes :

1. Segmentation d'un grand fichier sonore : Il s'agit de diviser un grand enregistrement audio de la performance du musicien en unités ou segments plus petits.
2. Analyse : Ces segments sont ensuite organisés dans une base de données en fonction de diverses données descriptives (mfcc dans notre cas).
3. Mise à l'échelle/pré-traitement : la mise à l'échelle est appliquée pour une meilleure visualisation.
4. Réduction de la dimension : Sur la base des descripteurs mfcc, la dimensionnalité des données est réduite afin de les rendre plus gérables et plus faciles à travailler. Cette réduction peut être réalisée à l'aide de techniques telles que l'analyse en composantes principales (ACP), la décomposition en valeurs singulières (SVD) ou l'approximation et la projection uniformes des formes (UMAP, méthode préférée dans notre cas). Séquençage des proches voisins : Une fois les segments organisés et analysés, le logiciel les sélectionne et les combine en temps réel sur la base de certains paramètres ou règles d'entrée pour créer une performance musicale simulée qui sonne comme si elle était improvisée par le musicien. Nous utilisons ici un algorithme de voisinage proche, qui sélectionne les segments qui sont similaires d'une certaine manière (par exemple, en termes de hauteur, de volume ou de timbre - grâce aux similarités révélées par umap sur les mfccs dans notre cas) au segment en cours de lecture.

Nous allons maintenant décrire ces étapes plus en détail :

5.3. Slicing

Nous avons vu au chapitre 3 comment le découpage déclenche musicalement des possibilités. Dans MuBu, la détection de l'onset se fait avec pipo.onseg ou pipo.gate. FluCoMa expose cinq algorithmes de détection d'onset différents

1. fluid.ampslice : Trancheur de détritique basé sur l'amplitude
2. fluid.ampgate : Détection de porte sur un signal
3. fluid.onsetslice : Trancheur de buffer audio basé sur la différence spectrale
4. fluid.noveltyslice : Basé sur la matrice d'auto-similarité (SSM)
5. fluid.transcientslice : Implémente un algorithme de dé-clicage

Seul Onsetslice a été testé de manière approfondie. Les seuls paramètres modifiés étaient un "seuil" direct ainsi qu'un argument "longueur minimale de la tranche", déterminant la tranche la plus courte autorisée (ou la durée minimale d'une tranche) en *hopSize*. Ceci introduit une limitation commune aux outils de CBCS : le système incite

fortement l'utilisateur à choisir des échantillons courts pour de meilleurs résultats d'analyse, et plus d'interactivité, lorsqu'il contrôle la base de données avec un suiveur de geste. Aaron Einbond remarque dans l'utilisation de CataRT comment les échantillons courts convenaient le mieux à son intention : "Les échantillons courts contenant des attaques rapides et sèches, telles que des clics de touches enregistrés de près, étaient particulièrement adaptés à une impression convaincante de mouvement de la source WFS unique. L'effet obtenu est celui d'un instrument virtuel se déplaçant dans la salle de concert en même temps que les changements de son contenu timbral, réalisant ainsi la proposition initiale de Wessel." [28]

Une limitation connexe de la synthèse concaténative réside dans le fait que des échantillons courts démontrent l'efficacité de l'algorithme ; mais en même temps, ils s'éloignent de la "simulation plausible" recherchée dans la présente étude. Il faut donc trouver un équilibre entre la liberté imposée par de grands échantillons et le contrôle affiné que l'on peut obtenir avec des échantillons courts.

Une concaténation directe des tranches clique dans la plupart des cas sur le point d'édition, ce qui peut être évité par l'utilisation de rampes. Le deuxième défaut le plus notable de la concaténation concerne l'interruption des résonances du registre grave, que même une grande réverbération ne parvient pas à rendre plausible. Avoir un seuil bas et une grande "minsilicelength" résulte en des tranches équidistantes, toutes de durées identiques, comme le ferait l'objet pipo.onseg dans MuBu.

Comme nous écoutons le son dans le temps, ce paramètre responsable de la *durée des échantillons* est d'une importance primordiale.

5.4. MFCC sur chaque tranche - sur une tranche/un segment entier

Multidimensionnelle, l'analyse MFCC (Mel-Frequency Cepstral Coefficient) est une technique utilisée pour extraire des caractéristiques des signaux audio qui sont pertinentes pour la reconnaissance de la parole et de la musique. Elle consiste à calculer un ensemble de coefficients qui représentent l'enveloppe spectrale du signal audio, et peut être utilisée pour capturer les caractéristiques spectrales du style de jeu du musicien.

5.5. Analyse statistique sur chaque tranche

BufStats est utilisé pour calculer des mesures statistiques sur les données stockées dans un buffer. Un buffer est ici un type de structure de données qui contient des informations de séries temporelles, des données de descripteurs audio dans ce cas. BufStats calcule sept statistiques sur les données du buffer : moyenne, écart type, asymétrie, aplatissement, valeurs basses, moyennes et hautes. Ces statistiques fournissent des informations sur la tendance centrale des données et la façon dont elles sont distribuées autour de cette tendance. En plus de calculer des statistiques sur le buffer d'origine, BufStats peut également calculer des statistiques sur jusqu'à deux dérivés des données

d'origine, appliquer des pondérations aux données à l'aide d'un buffer de pondération, et identifier et supprimer les trames aberrantes. Ces mesures statistiques peuvent être utiles pour comparer différentes données de séries temporelles, même si les données d'origine sont de longueurs différentes, et peuvent fournir une meilleure distinction entre les points de données lorsqu'elles sont utilisées dans la formation ou l'analyse. La sortie de BufStats est un buffer comportant le même nombre de canaux que les données d'origine, chaque canal contenant les statistiques des données correspondantes dans le buffer d'origine.

5.6. Normalisation

Le package FluCoMa propose plusieurs outils de mise à l'échelle/pré-traitement, parmi lesquels la normalisation et la standardisation ont été utilisées. La normalisation et la standardisation sont des techniques utilisées pour transformer les variables afin qu'elles puissent être comparées ou combinées dans des analyses statistiques. Les deux techniques sont utilisées pour rendre les données plus comparables, mais elles fonctionnent de manière légèrement différente.

La normalisation consiste à mettre à l'échelle une variable pour qu'elle ait une moyenne de 0 et un écart-type de 1. Pour ce faire, on soustrait la moyenne de la variable de chaque point de données, puis on divise par l'écart-type. La normalisation est souvent utilisée lorsque les variables à comparer se situent sur des échelles différentes ou ont des unités de mesure différentes. Elle permet de comparer des variables qu'il serait autrement difficile de comparer directement.

La normalisation consiste à mettre à l'échelle une variable afin qu'elle ait une valeur minimale de 0 et une valeur maximale de 1. Pour ce faire, on soustrait la valeur minimale de la variable de chaque point de données, puis on divise par l'étendue (c'est-à-dire la différence entre les valeurs maximale et minimale). La normalisation est souvent utilisée lorsque les variables comparées ont une distribution asymétrique, ou lorsque les variables ne sont pas normalement distribuées. Elle permet de comparer des variables qui seraient autrement difficiles à comparer directement en raison de l'asymétrie de leur distribution.

La normalisation met à l'échelle une variable pour qu'elle ait une moyenne de 0 et un écart-type de 1, tandis que la normalisation met à l'échelle une variable pour qu'elle ait une valeur minimale de 0 et une valeur maximale de 1. La normalisation ccaling a été jugée plus facile à utiliser à la fois en 2-D (dans FluCoMa, l'objet fluid.plotter), ainsi que dans le monde VR 3D dans lequel l'origine correspond à un coin du monde. L'objet fluid.normalize possède un attribut "@max" (1 par défaut), qui correspond alors directement aux dimensions du monde VR.

5.7. Réduction de la dimensionnalité

La réduction de la dimensionnalité est une technique utilisée en apprentissage automatique pour réduire le nombre

de caractéristiques (dimensions) dans un ensemble de données. L'objectif de la réduction de la dimensionnalité est de simplifier les données sans perdre trop d'informations. Divers algorithmes de réduction de la dimensionnalité sont présentés dans une étude FluCoMa précoce[29], sans mention, curieusement, de l'UMAP, qui a ensuite été privilégié.

SOM est l'un des algorithmes de réduction de la dimensionnalité les plus populaires. Il est mis en œuvre dans la bibliothèque `ml.starmlstar` pour Max, une bibliothèque simple et pratique pour l'apprentissage automatique, l'une des nombreuses structures et algorithmes d'apprentissage automatique célèbres dans la communauté NIME. [30] [31] [32] [33].

SOM (Self-Organizing Map) et UMAP (Uniform Manifold Approximation and Projection) sont deux techniques de réduction de la dimensionnalité. SOM est un type de réseau neuronal qui est formé par apprentissage non supervisé. Il se compose d'une grille de neurones, dont chacun est associé à un ensemble de poids. Le SOM est formé en lui présentant des données d'entrée et en ajustant les poids des neurones de manière à ce que les modèles d'entrée similaires soient mis en correspondance avec les neurones voisins sur la grille. La carte résultante est une représentation à faible dimension des données d'entrée qui préserve la structure topologique des données d'origine. L'UMAP, quant à elle, est une technique non linéaire de réduction de la dimensionnalité qui repose sur les principes de l'analyse topologique des données. Elle utilise une combinaison de techniques telles que les k -voisins les plus proches, la construction de graphes pondérés et l'intégration à faible dimension pour produire une représentation à faible dimension des données d'entrée. Contrairement au SOM, qui est limité à une structure de grille fixe, l'UMAP peut produire une représentation continue et flexible des données. SOM et UMAP peuvent tous deux être utiles pour visualiser des données hautement dimensionnelles et pour découvrir des modèles et des relations dans les données. Toutefois, l'UMAP présente certains avantages par rapport au SOM, notamment la possibilité de traiter plus efficacement de grands ensembles de données et de produire des résultats plus interprétables.

L'UMAP (Uniform Manifold Approximation and Projection) peut être utilisé pour visualiser des données à haute dimension dans un espace à plus faible dimension. Appliqué à des données sonores analysées à l'aide de MFCC (Mel-Frequency Cepstral Coefficients), l'UMAP réduit la dimensionnalité des données et crée une représentation visuelle du son dans un espace à deux ou trois dimensions. Les CCFM, là encore, sont une technique d'extraction de caractéristiques couramment utilisée dans le traitement de la parole et de l'audio. Elles consistent à décomposer un signal sonore en un ensemble de bandes de fréquences et à représenter le spectre de puissance de chaque bande par un ensemble de coefficients. Les coefficients MFCC résultants capturent des caractéristiques spectrales importantes du signal sonore (bien que difficilement interprétables par l'utilisateur novice), telles que la fréquence et l'ampli-

tude des pics spectraux. En appliquant l'UMAP aux coefficients MFCC d'un signal sonore, il est possible de créer une représentation visuelle du son qui préserve les relations entre les différents coefficients MFCC (voir Fig. 2). Cela peut être utile pour des tâches telles que l'exploration de la structure d'un ensemble de données sonores, l'identification de modèles ou de tendances dans les données et la comparaison de différents sons.



Figure 2. La réduction de la dimensionnalité des MFCC permet de révéler les similitudes spectrales. UMAP produit des coordonnées en 2d ou 3d.

L'UMAP est donc utilisé en premier lieu pour ses capacités de regroupement, ce qui facilite la classification. Elle permet d'identifier des modèles ou des tendances qui ne sont pas forcément évidents à partir des données brutes. Plus important encore, les dimensions non linéaires proposées par UMAP (que ce soit en 2d dans Max ou en 3 dimensions dans PatchXR, et lorsqu'elles sont comparées à des analyses linéaires dans lesquelles, par exemple, x , y et z correspondent à la hauteur, à l'intensité sonore et à la centroïde) ont donné lieu à des regroupements bien plus "intelligents" que les types de représentations plus conventionnels et conformes aux paramètres.

5.8. Requêtes de voisinage

La fonction de recherche de voisinage est légèrement différente à chaque fois, mais elle est basée dans FluCoMa sur les arbres K-d et l'algorithme `knn`. Dans MuBu, l'objet `mubu.knn`, ainsi que l'objet `ml.kdtree` de `ml.star`, donnent des résultats très comparables à ceux obtenus avec `fluid.kdtree`.

Les arbres K-d (abréviation de "k-dimensional trees") et les k -voisins les plus proches (k -NN) sont deux algorithmes apparentés, mais dont les objectifs sont différents.

Un arbre k -d est une structure de données utilisée pour stocker et interroger efficacement un ensemble de points

dans un espace à k dimensions. Il fonctionne en partitionnant les points dans un arbre binaire, chaque nœud de l'arbre représentant un hyperplan qui divise l'espace en deux moitiés. Les points sont partitionnés de manière récursive dans les sous-arbres gauche et droit en fonction du côté de l'hyperplan sur lequel ils se trouvent. En organisant les points de cette manière, il est possible de trouver rapidement les plus proches voisins d'un point donné en ne recherchant qu'un sous-ensemble de l'arbre plutôt que l'ensemble des points.

D'autre part, l'algorithme k -NN est un algorithme d'apprentissage automatique utilisé pour la classification ou la régression. Étant donné un ensemble de points étiquetés et un nouveau point non étiqueté, l'algorithme k -NN détermine les k points de l'ensemble qui sont les plus proches du nouveau point, puis utilise les étiquettes de ces points pour prédire l'étiquette du nouveau point. La valeur de k est un hyperparamètre choisi par l'utilisateur, qui détermine le nombre de voisins pris en compte lors de la prédiction.

En résumé, un arbre k -d est une structure de données utilisée pour stocker et interroger efficacement un ensemble de points dans un espace à k dimensions, tandis que l'algorithme k -NN est un algorithme d'apprentissage automatique utilisé pour la classification ou la régression. Ces deux algorithmes sont souvent utilisés dans des applications telles que la reconnaissance des formes, la classification des images et l'exploration des données.

Alors que CataRT ou Audiostellar sont typiquement utilisés pour la génération de textures électroniques/design sonore, j'ai le plus souvent utilisé FluCoMa pour générer des instruments monophoniques (un interprète joue un instrument à la fois), dans lesquels l'avatar reproduit ce que knn ferait avec un instrument automatique : il privilégiera dans son choix l'échantillon qu'il peut atteindre à portée de main, plutôt que de sauter une grande distance entre 2 éléments (voir Fig. 3).

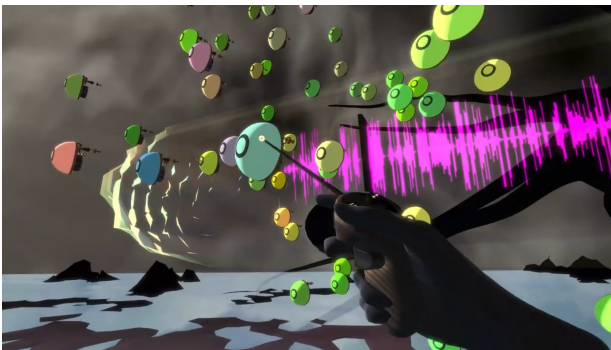


Figure 3. Une interface VR dans laquelle chaque bouton du monde correspond à une tranche du fichier sonore. L'apprentissage automatique permet de rapprocher les sons qui partagent des caractéristiques spectrales communes.

5.9. PatchXR

PatchXR [34] est une station de travail audio numérique ludique permettant de faire de la musique en VR. Sa métaphore de base correspond à ce que l'équipe FluCoMa appelle CCE (environnements de codage créatifs) dans la mesure où il fonctionne à bien des égards comme Max ou Pure Data.

L'une des raisons d'utiliser la VR pour explorer un ensemble de données 3D est qu'elle permet aux utilisateurs d'interagir avec les données d'une manière plus naturelle et immersive, en l'utilisant comme un outil de visualisation et d'analyse des données. Les utilisateurs peuvent se déplacer et explorer les données sous différents angles, ce qui peut les aider à mieux comprendre les relations entre les différents points de données et à identifier des modèles. Les utilisateurs obtiennent un sens plus intuitif des données et comprennent mieux comment elles sont structurées et comment les différents points de données sont liés les uns aux autres.

La structure d'un fichier `.patch` (un monde patchXR) suit la syntaxe d'un fichier `.maxpat` (pour Max) ou `.pd` (pour pure data) dans le sens où il déclare d'abord les objets utilisés, puis les connexions entre eux. Cette structure a rendu relativement trivial le fait de générer une routine javascript prenant en entrée un dictionnaire (fichier json) avec les coordonnées 3D de chaque segment, et en sortie un nouveau fichier `.patch` (un monde accessible en VR, voir le workflow général sur la Fig. 4). Une étude plus poussée du patching dans PatchXR a permis d'implémenter cette acrtographie de manière plus subtile (sans représenter chaque point par un bouton, comme dans d'autres exemples plus récents.¹¹

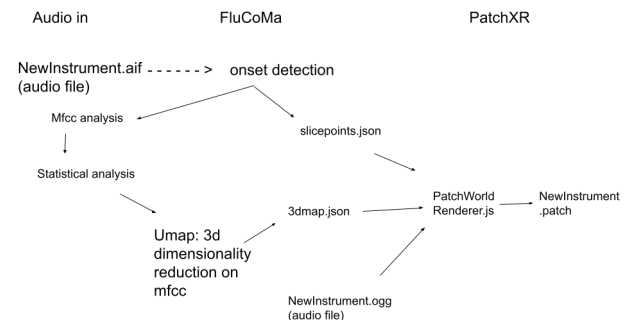


Figure 4. Flux de travail général : d'un fichier audio d'entrée à sa représentation `.patch` 3d dans PatchXR.

5.9.1. Interaction et communication OSC

PatchXR expose un large éventail de blocs (un bloc correspond à un objet dans Max ou Pure Data), ce qui permet d'accéder facilement aux données gestuelles telles que :

- La position/distance entre les mains/contrôleurs et une référence.

¹¹ . https://youtu.be/jQsZT_Tlecs?t=8

- Les angles de rotation (x y z) des contrôleurs des deux mains
- Contrôleurs 2-d de type écran tactile, où l'utilisateur déplace la position xy d'un sélecteur sur un plan en le saisissant manuellement.
- 2-d lazer-like controllers, où l'utilisateur déplace la position xy d'un sélecteur à distance, comme s'il utilisait un pointeur lazer vers un écran distant.
- 2-d pads, qui permettent d'accéder à la vélocité à laquelle le pad est frappé
- contrôleurs de type theremine 3-d, où l'utilisateur déplace la position xyz d'un sélecteur sur un plan en le saisissant manuellement.
- 1-d sliders, knobs, boutons...

L'un des défis actuels consiste à diversifier les modes d'interrogation du corpus. Les mappings un à un des résultats UMAP tels que ceux décrits dans le chapitre 5.9 privilégient les simulations pour solistes, ou duo en mode multijoueur, dans lesquelles les boutons de l'interface se font face, afin d'inciter les joueurs à se faire face (voir <https://youtu.be/LP1g79BdIpY>).

Une simulation pour un plus grand nombre d'instruments, en particulier lorsqu'on joue seul, incite à utiliser des types de contrôle de plus haut niveau sur les automates, le plus important étant la simple capacité de concaténation automatique : jouer l'échantillon suivant dès que le précédent s'est arrêté.

6. FUTURS TRAVAUX : L'ORCHESTRE RASPBERRY PI VS NMP EN MULTIJOUEUR

Dans le cadre d'une résidence d'artiste à l'UCA (Université côte d'azur), un des questionnements majeurs interroge comment les outils présentés ci-dessus (ceux qui concernent le domaine de la recherche d'information musicale - MIR) peuvent servir à contrôler une plateforme immersive composée d'un orchestre de 64 modules *Pré*. [35]

Au moment de la rédaction de ce rapport, les résultats les plus satisfaisants ont été obtenus en envoyant des messages à chaque RaspberriPi indépendamment, selon son adresse IP (statique) spécifique, avec une syntaxe simple d'une liste de 2 nombres entiers correspondant à : 1/ quel buffer consulter 2/ quelle tranche de ce buffer jouer. Poursuivant l'élaboration d'illustrations de l'espace timbrique, les modules *Pré*, avec les différentes acoustiques que sa mobilité permet, favoriseront une densité contrastée d'événements en fonction de l'acoustique de l'espace dans lequel se déroule l'écoute.

L'utilisation de Raspberry Pi dans notre cas impose des contraintes réseau (local) qui peuvent potentiellement entrer en conflit avec les possibilité récentes de multi-joueur dans PatchXR qui se déroulent en distanciel.¹²

12. <https://youtu.be/LP1g79BdIpY>

7. CONCLUSIONS

Après avoir examiné si les partitions sont des cartes ou non, nous avons proposé un workflow pour la synthèse concaténative basée sur un corpus (CBCS), en faisant valoir que les outils d'apprentissage automatique pour la visualisation des données offrent des informations révélatrices et exploitables sur la qualité timbrale du matériel qui est analysé. Du point de vue d'un compositeur, la disparition de l'axe temporel x incite à envisager la composition non pas dans un sens narratif (compris selon les réflexions sur les nouveaux médias et la notion d'art basé sur les données développée par [6]), mais plutôt, comme des "toiles temporelles", suivant l'esthétique de Morton Feldman.

Les outils de "machine listening" présentés ici (Flu-CoMa, MuBu) aident à construire des instruments intelligents avec des quantités de données relativement petites; nous avons insisté sur le contrôle de la durée des échantillons qui semble crucial en CBCS. Un équilibre doit être trouvé entre 1/ l'analyse d'échantillons de courte durée qui sont plus faciles à traiter et à catégoriser et 2/ les échantillons longs qui sonnent plus naturels dans le contexte des simulations instrumentales.

8. REMERCIEMENTS

Je suis reconnaissant pour le soutien de l'UCA/CTEL, dont le programme de recherche de résidence d'artiste a permis de réaliser ces expériences, et pour le soutien de PRISM-CNRS.

9. REFERENCES

- [1] D. Miller, "Are scores maps? a cartographic response to goodman," in *Proceedings of the International Conference on Technologies for Music Notation and Representation – TENOR'17*, H. L. Palma, M. Solomon, E. Tucci, and C. Lage, Eds. A Coru na, Spain : Universidade da Coru na, 2017, pp. 57–67.
- [2] U. Eco, P. Eco, A. Cancogni, and D. Robey, *The Open Work*. Harvard University Press, 1989. [Online]. Available : <https://books.google.fr/books?id=7jroM0M8TuwC>
- [3] C. Vear, *The Digital Score : Musicianship, Creativity and Innovation*. Routledge, 2019. [Online]. Available : <https://books.google.fr/books?id=oSblwQEACAAJ>
- [4] D. Schwarz, G. Beller, B. Verbrugghe, and S. Britton, "Real-Time Corpus-Based Concatenative Synthesis with CataRT," in *9th International Conference on Digital Audio Effects (DAFx)*, Montreal, Canada, Sep. 2006, pp. 279–282, cote interne IRCAM : Schwarz06c. [Online]. Available : <https://hal.archives-ouvertes.fr/hal-01161358>
- [5] D. Schwarz, R. Cahen, and S. Britton, "Principles and Applications of Interactive Corpus-Based

- Concatenative Synthesis,” in *Journées d’Informatique Musicale (JIM)*, Albi, France, Mar. 2008, pp. 1–1, cote interne IRCAM : Schwarz08a. [Online]. Available : <https://hal.archives-ouvertes.fr/hal-01161401>
- [6] L. Manovich, “Database as symbolic form,” *Convergence : The International Journal of Research into New Media Technologies*, vol. 5, pp. 80 – 99, 1999.
- [7] M. Feldman, B. Friedman, and F. O’Hara, *Give My Regards to Eighth Street : Collected Writings of Morton Feldman*, ser. Exact Change. Exact Change, 2000. [Online]. Available : <https://books.google.fr/books?id=hfgHAQAAMAAJ>
- [8] J.-P. Briot, G. Hadjeres, and F.-D. Pachet, *Deep Learning Techniques for Music Generation – A Survey*, Aug. 2019. [Online]. Available : <https://hal.sorbonne-universite.fr/hal-01660772>
- [9] P. Esling, A. Chemla-Romeu-Santos, and A. Bitton, “Generative timbre spaces with variational audio synthesis,” *CoRR*, vol. abs/1805.08501, 2018. [Online]. Available : <http://arxiv.org/abs/1805.08501>
- [10] D. L. Wessel, “Timbre space as a musical control structure,” *Computer Music Journal*, vol. 3, no. 2, pp. 45–52, 1979. [Online]. Available : <http://www.jstor.org/stable/3680283>
- [11] K. Fitz, M. Burk, and M. McKinney, “Multidimensional perceptual scaling of musical timbre by hearing-impaired listeners,” *The Journal of the Acoustical Society of America*, vol. 125, p. 2633, 05 2009.
- [12] C.-E. Cella, “Orchidea : a comprehensive framework for target-based computer-assisted dynamic orchestration,” *Journal of New Music Research*, vol. 0, no. 0, pp. 1–29, 2022. [Online]. Available : <https://doi.org/10.1080/09298215.2022.2150650>
- [13] A. Agostini and D. Ghisi, “A Max Library for Musical Notation and Computer-Aided Composition,” *Computer Music Journal*, vol. 39, no. 2, pp. 11–27, 06 2015. [Online]. Available : https://doi.org/10.1162/COMJ_a_00296
- [14] D. Ghisi and C. Agon, “dada : Non-standard user interfaces for computer-aided composition in max,” in *Proceedings of the International Conference on Technologies for Music Notation and Representation – TENOR’18*, S. Bhagwati and J. Bresson, Eds. Montreal, Canada : Concordia University, 2018, pp. 147–156.
- [15] J.-C. Risset and D. Wessel, “Exploration of timbre by analysis and synthesis,” *Psychology of Music*, pp. 113–169, 1999.
- [16] S. Mcadams, S. Winsberg, S. Donnadiou, G. De Soete, and J. Krimphoff, “Perceptual scaling of synthesized musical timbres : Common dimensions, specificities, and latent subject classes,” *Psychological research*, vol. 58, pp. 177–92, 02 1995.
- [17] A. Caclin, S. Mcadams, B. Smith, and S. Winsberg, “Acoustic correlates of timbre space dimensions : A confirmatory study using synthetic tones,” *The Journal of the Acoustical Society of America*, vol. 118, pp. 471–82, 08 2005.
- [18] D. Schwarz, “The Sound Space as Musical Instrument : Playing Corpus-Based Concatenative Synthesis,” in *New Interfaces for Musical Expression (NIME)*, Ann Arbour, United States, May 2012, pp. 250–253, cote interne IRCAM : Schwarz12a. [Online]. Available : <https://hal.archives-ouvertes.fr/hal-01161442>
- [19] L. Garber, T. Ciccola, and J. C. Amusatogui, “Audiostellar, an open source corpus-based musical instrument for latent sound structure discovery and sonic experimentation,” 12 2020.
- [20] B. Hackbarth, N. Schnell, P. Esling, and D. Schwarz, “Composing Morphology : Concatenative Synthesis as an Intuitive Medium for Prescribing Sound in Time,” *Contemporary Music Review*, vol. 32, no. 1, pp. 49–59, 2013. [Online]. Available : <https://hal.archives-ouvertes.fr/hal-01577895>
- [21] N. Schnell, A. Roebel, D. Schwarz, G. Peeters, and R. Borghesi, “Mubu and friends -assembling tools for content based real-time interactive audio processing in max/msp,” *Proceedings of the International Computer Music Conference (ICMC 2009)*, 01 2009.
- [22] P. A. Tremblay, G. Roma, and O. Green, “Enabling Programmatic Data Mining as Musicking : The Fluid Corpus Manipulation Toolkit,” *Computer Music Journal*, vol. 45, no. 2, pp. 9–23, 06 2021. [Online]. Available : https://doi.org/10.1162/comj_a_00600
- [23] F. Bevilacqua and R. Müller, “A gesture follower for performing arts,” 05 2005.
- [24] N. Schnell, D. Schwarz, J. Larralde, and R. Borghesi, “Pipo, a plugin interface for afferent data stream processing operators,” in *International Society for Music Information Retrieval Conference*, 2017.
- [25] R. Fiebrink and P. Cook, “The wekinator : A system for real-time, interactive machine learning in music,” *Proceedings of The Eleventh International Society for Music Information Retrieval Conference (ISMIR 2010)*, 01 2010.
- [26] D. Fober, Y. Orlarey, and S. Letz, “INScore - An Environment for the Design of Live Music Scores,” in *Linux Audio Conference*, Stanford, United States, 2012, pp. 47–54. [Online]. Available : <https://hal.archives-ouvertes.fr/hal-02158817>
- [27] F. Bevilacqua, B. Zamborlin, A. Sypniewski, N. Schnell, F. Guédy, and N. Rasamimanana, “Continuous realtime gesture following and recognition,” vol. 5394, 02 2009, pp. 73–84.
- [28] A. Einbond and D. Schwarz, “Spatializing timbre with corpus-based concatenative synthesis,” 06 2010.

- [29] G. Roma, O. Green, and P. A. Tremblay, “Adaptive mapping of sound collections for data-driven musical interfaces,” in *New Interfaces for Musical Expression*, 2019.
- [30] M. M. Wanderley and M. Battier, “Trends in gestural control of music,” 2000.
- [31] F. Bevilacqua, R. Müller, and N. Schnell, “MnM : a Max/MSP mapping toolbox,” in *New Interfaces for Musical Expression*, Vancouver, France, May 2005, pp. 85–88, cote interne IRCAM : Bevilacqua05a. [Online]. Available : <https://hal.archives-ouvertes.fr/hal-01161330>
- [32] R. Fiebrink and B. Caramiaux, “The machine learning algorithm as creative musical tool,” *ArXiv*, vol. abs/1611.00379, 2016.
- [33] B. Caramiaux and A. Tanaka, “Machine learning of musical gestures,” in *New Interfaces for Musical Expression*, 2013.
- [34] V. Bauer and T. Bouchara, “First steps towards augmented reality interactive electronic music production,” in *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2021, pp. 90–93.
- [35] « *PrÉ* » : *connected polyphonic immersion*. Zenodo, Jul. 2022. [Online]. Available : <https://doi.org/10.5281/zenodo.6806324>

SIMONE : UN INSTRUMENT DISTRIBUÉ POUR L'ÉTUDE DES INTERACTIONS IMPROVISÉES COLLECTIVES

Aliénor Golvet Benjamin Matuszewski Frédéric Bevilacqua

STMS IRCAM-CNRS-Sorbonne Université, Paris

[prenom].[nom]@ircam.fr

RÉSUMÉ

Nous présentons *Simone* [simɔ̃n], un système Web pour l'improvisation collective en réseau basé sur des entrées audio (microphone). Inspirés par plusieurs exemples de systèmes de performance musicale en réseau, notre objectif était de concevoir un instrument distribué basé sur 3 concepts : l'improvisation collective, l'utilisation de la voix dans le processus de synthèse sonore, et l'échange d'information par le réseau. *Simone* est développé avec des technologies web et est utilisable avec un équipement matériel léger, portable et adapté à plusieurs situations.

Cet article se concentre sur une présentation du design du système ainsi que de son implémentation technique. Nous présentons aussi une première utilisation de *Simone* via une étude réalisée avec 4 participants ayant improvisé avec le système. Les retours collectés ont notamment mené à un certain nombre de modifications dans le design de *Simone*. L'objectif ultérieur de ce travail est d'étudier la manière dont des utilisateur-ices peuvent s'approprier ce système et les dynamiques d'improvisation qui peuvent en résulter.

Simone est disponible au téléchargement sous licence BSD-3-Clause à l'adresse suivante : <https://github.com/ircam-ismm/simone>

1. INTRODUCTION

L'essor de la Web Audio API [1] depuis 2011 a accéléré l'utilisation de systèmes musicaux basés sur des navigateurs Web et l'émergence de nouvelles expériences musicales collectives. Ces travaux explorent notamment les diverses possibilités offertes par les réseaux : performances musicales à distance, participation du public, intégration de capteurs, systèmes de réalité virtuelle [17].

Cependant, les questions de la conception instrumentale de systèmes musicaux en réseau et de l'apprentissage et de l'appropriation collective de tels systèmes par les musicien-es restent largement inexplorées. En effet, si la communauté NIME (New Interfaces for Musical Expression) s'est depuis plusieurs années intéressée aux questions de conception et d'appropriation d'instruments [10, 20], la majorité des travaux dans ce domaine concerne des pratiques où les musicien-nes utilisent leurs propres interfaces sans forcément être connectés par un réseau.

Dans ce papier, nous présentons *Simone*, un système Web pour l'improvisation collective en réseau basé sur des entrées audio (microphone). *Simone* est un nouvel *instrument distribué* proposant une interaction collective qui se décline en plusieurs variations dont chacune correspond à une topologie d'interaction particulière. Dans chaque variation, les musicien-nes occupent des rôles variés et échangent différents types d'information. Si cette contribution se limite à la présentation technique et à une première mise en pratique de *Simone*, l'une de nos motivations à plus long terme est d'étudier l'appropriation de ce système.

La conception de *Simone* a été influencée par des exemples antérieurs de systèmes de performance musicale en réseau et par une esthétique musicale proche du collage sonore, de l'improvisation bruitiste et de la musique concrète. Une des originalités de *Simone* est d'être conçu pour être utilisé via un ensemble hétérogène d'appareils (ordinateurs, tablette, téléphones portables, nano-ordinateur type raspberry pi) permettant ainsi de le déployer facilement dans tout type d'espace grâce à des moyens techniques légers et portables.

Nous commençons par présenter divers travaux précédents de performances et d'installations musicales en réseau. Nous décrivons en particulier certains systèmes qui, comme le nôtre, utilisent des microphones. Dans une deuxième partie, consacrée à *Simone*, nous introduisons les différents objectifs ayant guidé ce travail avant de présenter les éléments de design et d'implémentation, ainsi que les différentes variations du système. Enfin, dans une dernière partie nous présentons des éléments issus d'une étude pilote que nous avons réalisée à l'Ircam durant laquelle nous avons demandé à quatre participants d'improviser avec les différents systèmes. Une version simplifiée de *Simone* a également fait l'objet d'une présentation publique lors des portes ouvertes de l'Ircam en janvier 2023.

2. EXEMPLES HISTORIQUES ET INFLUENCES

2.1. Instruments musicaux en réseau

Probablement parce qu'une analogie immédiate peut être opérée entre un réseau informatique et un ensemble de musicien-nes – dont les nœuds seraient les musicien-nes et les communications se feraient par les médiums acoustique et/ou visuel – les possibilités du réseau ont rapidement été

explorées par les artistes dans diverses modalités [4].

Les performances de musique en réseau ont ainsi été décrites comme explorant de nouveaux moyens de distribuer les ressources cognitives et l'information au sein d'un groupe, la modification des interactions et du jeu individuel en fonction de l'agentivité des autres musicien-nes et du réseau lui-même, la disparition de la séparation entre musicien-nes et public, ou encore la mise en oeuvre de processus de création collective plutôt que la production d'une oeuvre fixe [8].

L'exemple précurseur et paradigmatique de cet usage du réseau est le groupe *The Hub*, fondé en 1986 des suites de la *League of Automatic Music Composers* [6, 15]. Bien que chaque musicien du groupe contrôlait un système de sa propre confection, leurs performances étaient fondées sur l'échange d'informations via un réseau local selon un protocole défini pour chaque pièce. Ainsi par exemple dans la pièce *Waxlips* (1991), chaque musicien reçoit des notes individuelles de la part des autres membres du groupe via le réseau. Ce musicien doit d'abord jouer la note reçue puis lui appliquer une transformation arbitraire avant de l'envoyer à un autre membre du groupe ¹.

Les expérimentations de *The Hub* serviront de modèle à de nombreux ensembles de musique tels que les *laptop orchestra* dans lesquels l'infrastructure en réseau prend une importance esthétique plus ou moins importante [4, 16, 18].

En suivant la distinction faite par Rohrer [12] nous pouvons distinguer deux types de structures de partage d'information au sein d'un système de musique en réseau : soit l'accès simultané à un état commun et partagé, soit le partage et la circulation d'objets et d'informations au sein du réseau.

Dans la première catégorie nous pouvons par exemple évoquer l'*Emupo* [3], une interface pour l'improvisation musicale collective développée en Max/MSP. Destinée à être jouée de manière autonome ou pour compléter le jeu d'instrumentistes, l'*Emupo* peut-être contrôlée par plusieurs utilisateur-ices à la fois, chacun-e prenant la main sur différents paramètres de production d'un même son donnant alors lieu à une interaction "intra-instrumentale".

Dans la seconde catégorie, l'ensemble *powerbooks unplugged* [13] propose des performances de groupe basées sur l'accès à un espace de partage d'informations où circulent extraits de code, commentaires des musicien-nes et messages provoquant la production d'événements sonores sur les hauts parleurs de n'importe quel ordinateur du réseau. L'exemple de *powerbooks unplugged* met aussi en évidence un autre des avantages offerts par l'aspect décentralisé de la production sonore : la possibilité de choisir librement la manière dont le son est spatialisé ainsi que le nombre de musicien-nes participant-es.

1. On trouvera un enregistrement de cette pièce ici : <https://artifactrecordings.bandcamp.com/track/waxlips>

2.2. Utilisation du microphone

Divers travaux d'installations artistiques et d'instruments en réseau utilisant le microphone ont servi d'inspiration pour ce travail. Bien que l'usage de microphones et de la voix ait été exploré de nombreuses manières dans la communauté NIME (on en trouvera une recension et une taxonomie dans [9]), nous nous concentrons dans cette section uniquement sur les systèmes en réseau.

Voice Networks (2003) [19] de Gil Weinberg est une installation musicale collaborative proposant à des utilisateur-ices non-experts de participer à une expérience de création musicale collective et de mettre en évidence l'aspect social du jeu en groupe. L'installation est composée de quatre postes disposés en carré, face à face, autour d'un écran. Chaque poste est composé d'un microphone, d'une surface tactile et de haut-parleurs. Les participants peuvent enregistrer des boucles sonores avec leur micro et leur appliquer des transformations sonores à l'aide de la surface tactile qui contrôle à un patch Max/MSP. Le réseau est utilisé pour échanger les boucles sonores ainsi créées entre les participant-es.

Talking Drum (1995) [2] de Chris Brown est une installation interactive composée d'un ensemble de "stations" (composées d'un microphone, de haut-parleurs et d'un ordinateur) disposées librement dans un même espace. Chaque ordinateur génère un rythme à partir d'un algorithme qui analyse ce qui est capté par le microphone. Bien que les différentes stations n'échangent pas d'informations, elles sont toutes synchronisées sur une même horloge. L'installation *Talking Drum* crée ainsi une expérience d'improvisation entre des duos humain/machine au sein d'un espace acoustique partagé entre les musicien-nes et le public.

Auracle (2004) [5] conçu par Max Neuhaus est un synthétiseur collaboratif accessible sur internet ². Le système analyse la voix captée par le microphone de chaque utilisateur-ice et en extrait plusieurs types de données de différents niveaux (descripteurs audio : RMS, f_0 ; classifications de "gestes" vocaux par analyse en composantes principales) pour contrôler un synthétiseur. *Auracle* est conçu comme un système qui répond à l'activité des utilisateur-ices et cherche à encourager une communication et un dialogue non verbal à travers une expérience musicale accessible via une interface destinée à un public non expert.

3. SIMONE

Inspirés par les travaux mentionnés ci-dessus, nous avons donc développé un instrument distribué pour l'improvisation collective, proposant un ensemble de variations autour d'un scénario d'interaction.

2. <http://auracle.org/>

Notre objectif principal est de concevoir un système de création musicale collective qui favorise l'improvisation, l'utilisation de la voix dans le processus de synthèse sonore, et s'appuie sur l'échange d'information par le réseau. Le système est conçu comme un terrain d'expérimentation de divers paradigmes d'interaction collective et de leur appropriation par des musicien·nes. Il doit donc être accessible rapidement tout en possédant la profondeur et la flexibilité nécessaire pour qu'ils/elles ne se sentent pas limité·es dans un contexte d'improvisation et de création.

Un objectif secondaire est de concevoir un dispositif qui nécessite un équipement matériel minimal et suffisamment flexible pour s'adapter à différentes situations et disponible sur plusieurs types d'appareils différents : ordinateurs, téléphones portables, tablettes, nano-ordinateurs. *Simone* peut ainsi être utilisé à la fois comme installation sonore collaborative, comme instrument de musique collectif ou encore comme instrument de musique distribué contrôlable par une seule personne avec la possibilité de disposer les sources sonores dans l'espace comme on l'entend.

3.1. Éléments de design

Ce système d'improvisation collective est co-localisé et permet l'écoute mutuelle et les modalités d'interactions habituelles en improvisation (p. ex. gestes, regards). Outre l'utilisation d'interfaces classiques d'interaction humain-machine (souris, clavier, écran), qui offrent des contrôles sur la synthèse sonore, nous cherchons également à explorer l'utilisation de la voix dans la construction collective d'un espace sonore. L'utilisation conjointe du réseau et du microphone permet d'imaginer de nouvelles relations entre la voix et le son produit par l'instrument distribué et diverses manières de reconfigurer ou de déstabiliser des dynamiques d'improvisation collectives.

Par le fait que la voix enregistrée est complètement altérée par une technique de mosaicing (décrite ci-dessous), *Simone* permet une expérience accessible aux utilisateur·ices qui pourraient être inhibé·es par le fait d'utiliser leur voix dans un contexte d'improvisation collective. En effet, l'expérience de *Voice Networks* nous montre que "bien que la voix soit probablement le moyen de communication le plus intuitif et répandu dans la vie de tous les jours, certains participant·es étaient inhibé·es par le fait de l'utiliser dans le cadre d'une installation publique par pudeur ou tout simplement à cause du trac"³ [19]. Contrairement à d'autres gestes instrumentaux, l'usage de la voix engage un certain nombre de caractéristiques personnelles (âge, maîtrise du chant, santé physique, sexe, accent, etc...) qu'il peut être gênant de dévoiler.

3. Regarding the choice of the voice as an intuitive and malleable gateway for creative and collaborative interaction, it was interesting to observe that although the voice is probably the most intuitive and prevalent means of communication in everyday life, some participants were inhibited to use it in a public installation setting due to its "committing" or "revealing" nature or perhaps due to common "stage fright."

3.2. Implémentation

Principes généraux

Simone est décliné en plusieurs variations qui diffèrent quant à leurs interfaces et quant aux topologies d'interactions mises en oeuvre, mais qui partagent néanmoins un ensemble de caractéristiques communes :

1. Le microphone est utilisé comme moyen d'enregistrement afin de commander la synthèse sonore.
2. La synthèse sonore est basée sur le principe du mosaicing audio [21, 7]. Il s'agit d'une technique de synthèse dans laquelle un son appelé son *modèle* est découpé en segments et analysé. Chaque segment est ensuite associé à un segment qui lui est le plus proche dans un son dit *générateur* (en fonction d'une distance calculée sur la base de descripteurs audio pré-définis), le résultat final étant la concaténation des segments issus du son *générateur*. Ainsi, le son synthétisé partage à la fois la structure et l'évolution temporelle du son *modèle* et le timbre du son *générateur*.
3. Le système utilise un réseau local pour partager des informations entre les agents (c.à.d. utilisateur·ices et terminaux). Les informations transmises (horloge synchronisée, fichiers audio, données d'analyse) dépendent du scénario considéré (cf. Figure 1).

Aspects technologiques

L'application est entièrement développée avec le framework *soundworks* [11] basé sur les technologies Web, et est donc entièrement accessible depuis un navigateur Web. La synthèse sonore est réalisée grâce à un système d'analyse/synthèse par mosaicing développé avec l'API WebAudio [1]. Le fichier son *modèle* enregistré est analysé grain par grain en temps réel. Pour chaque grain, on calcule un vecteur de MFCC (Mel Frequency Cepstral Coefficients), et on cherche ensuite dans le son *générateur* le grain qui a les MFCC les plus semblables via une recherche dans un arbre à n dimensions.

Nous avons choisi d'utiliser des MFCC à 12 bandes de fréquences comme descripteurs audio car ceux-ci sont particulièrement adaptés à l'analyse et la représentation de la voix. Il s'agit en outre d'un descripteur de dimension suffisamment haute pour encoder plusieurs dimensions perceptives des sons, tout en étant suffisamment basse pour pouvoir trouver rapidement le grain le plus semblable à un autre.

Par ailleurs l'énergie RMS de chaque grain est calculée et est utilisée pour contrôler un gain lors de la resynthèse du son. Ainsi, un grain *modèle* de faible énergie donnera un son synthétisé de faible volume. Cet élément a été ajouté suite au retours d'utilisateurs qui semblaient confus par le fréquent manque de lien entre la dynamique de volume du son *modèle* vue sur la forme d'onde affichée et le résultat sonore.

Outre les clients accessibles depuis le navigateur Web d'un ordinateur ou d'un téléphone portable, nous avons également développé des clients exécutés dans des nano-ordinateurs de type *Raspberry Pi 4* (combinant les avan-

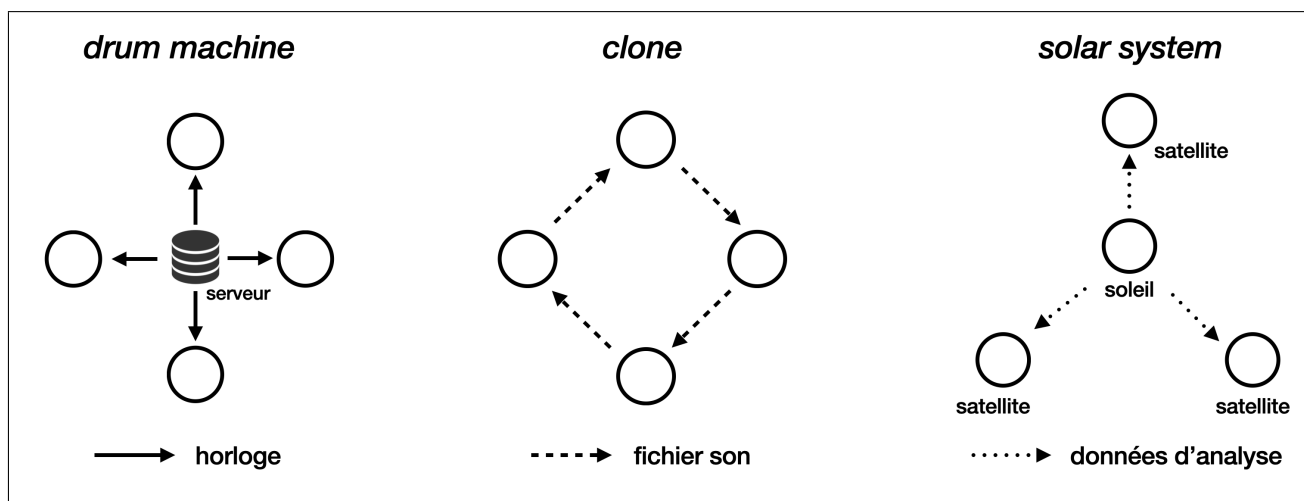


Figure 1. Diagrammes des échanges d’informations dans les trois systèmes présentés. Chaque cercle représente un client connecté au serveur.

tages d’être portables et modulaires, et possédant une bonne puissance de calcul). Ce portage simple de l’application en dehors du navigateur a notamment été rendu possible grâce à l’implémentation en Rust de la Web Audio API dont les *bindings* vers *Node.js* [14] ont permis de réutiliser de larges parties du code Javascript original.

3.3. Variations

Les principes généraux évoqués ci-dessus se déclinent selon différentes variations qui mettent en jeu différents moyens d’interactions et rôles exercés par les participant-es. Ces différents systèmes sont les suivants :

Drum Machine

Dans ce système, les utilisateur-rices enregistrent un son *modèle* avec le microphone et sélectionnent un son *générateur* parmi une banque de sons. Ils/elles peuvent ensuite sélectionner une section du son *modèle* à faire boucler. La durée de cette section est basée sur une grille rythmique commune à tou-te-s les participant-es qui partagent la même horloge. Le résultat peut être envisagé comme une sorte de boîte à rythmes distribuée.

Clone

Dans ce système, il est demandé aux utilisateur-rices de commencer à enregistrer un son d’une durée approximative de 30 secondes. Ce son est ensuite envoyé à un-e autre participant-e pour lui servir de son *générateur* dans l’interface de jeu. Ainsi, chaque participant-e doit apprendre à jouer avec la voix ou le son d’une autre personne.

Solar System

Contrairement aux précédents systèmes, les participant-es sont dans une configuration asymétrique. L’un-e d’entre eux/elles occupe un rôle central et est appelé-e le *soleil*. Le *soleil* peut enregistrer un son *modèle*. Les données d’analyse de ce son sont ensuite envoyées simultanément aux autres participant-es appelé-es *satellites* pour contrôler la

synthèse sonore par mosaïcing sur leur client. En outre, les *satellites* peuvent choisir leur propre son *générateur* parmi une banque de son et faire varier divers paramètres de synthèse. Ainsi, le même son *modèle* enregistré par le *soleil* est simultanément réinterprété par différents sons *générateur* sur chaque *satellite*.

Solar System embarqué

En plus des versions collectives que nous venons de présenter, nous avons mis au point une version solo embarquée du système *Solar System*. Celui-ci peut en effet se concevoir avec un unique agent humain prenant le rôle central de *soleil* et des clients non-humains contrôlables à distance prenant le rôle de *satellites*.

Le dispositif prend alors la forme suivante : un nombre variable de Raspberry Pi sur lesquelles est installée l’application sont connectés à un même réseau local (cf. Figure 2). L’interface de contrôle est accessible dans un navigateur web. Celle-ci, similaire à celle utilisée par l’utilisateur-riche prenant le rôle de *soleil* dans le système *Solar System*, permet d’enregistrer un son *modèle* ainsi que pour chaque *satellite* connecté de sélectionner le son *générateur* et de modifier les paramètres de synthèse.

Chaque Raspberry Pi, connectée à une paire de hauts parleurs, reçoit les données d’analyse de la part du serveur et synthétise le son en temps réel. Le système devient alors un véritable instrument où l’on contrôle simultanément plusieurs sources sonores de timbres différents mais contrôlées par le même son *modèle*. En faisant varier les volumes et les paramètres individuels, en choisissant des boucles plus ou moins longues, il est possible de créer une grande variété d’espace sonores différents.

La grande portabilité du matériel permet de disposer les sources sonores ainsi que d’en choisir le nombre de manière libre et adaptée à l’espace d’écoute. Il est à noter que le client est aussi accessible dans le navigateur internet ce qui permet à un membre du public de se connecter à l’application sur son téléphone et devenir ainsi lui-même

une source sonore.

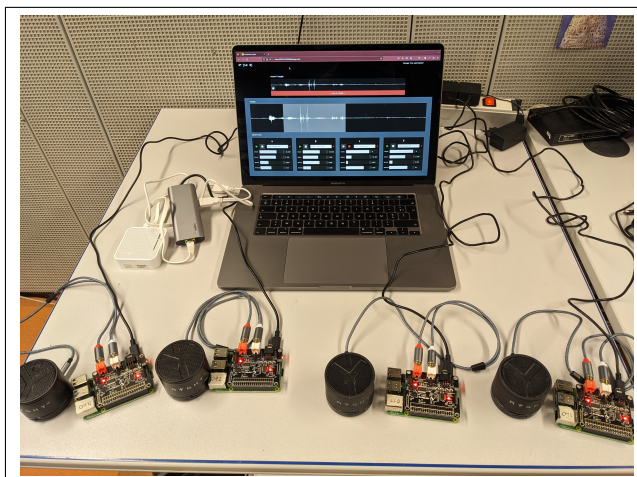


Figure 2. Le dispositif *Solar System* embarqué avec, à l'avant, quatre raspberry pi avec leurs enceintes et, à l'arrière, l'ordinateur avec l'interface de contrôle et un routeur wifi.

3.4. Interfaces

L'interface a fait l'objet de plusieurs itérations au cours des tests et réflexions. Une attention particulière a été portée sur la complexité et le nombre d'éléments de l'interface avec lesquels il est possible d'interagir, afin de faciliter l'appropriation du système dans un temps restreint et de promouvoir l'agentivité des participant-e-s dans un contexte d'improvisation collective.

Bien qu'elle diffère légèrement selon les différentes variations du système, l'interface comprend de manière générale les éléments suivants (cf. Figure 3) :

Une zone d'enregistrement qui permet de déclencher l'enregistrement avec le micro et de visualiser la forme d'onde du son enregistré, un bouton permet à l'utilisateur-ice d'utiliser cet enregistrement comme son *modèle* (zone en haut à gauche dans la Figure). En haut à droite dans la Figure, une zone avec un menu permet de choisir le son *générateur* ainsi que d'en visualiser la forme d'onde et de l'écouter.

La zone centrale montre la forme d'onde du son *modèle* actuellement utilisé. L'utilisateur-trice peut sélectionner une section à boucler en utilisant la souris. La boucle sélectionnée peut être de longueur libre ou contrainte par une grille temporelle dans le cas du système *Drum Machine*. L'affichage de la forme d'onde du son *modèle* nous semble offrir un bon équilibre entre intuition du résultat sonore et invitation à l'exploration et au jeu. En effet, si la forme d'onde permet d'observer immédiatement le niveau sonore de chaque section du son *modèle* et donc du résultat synthétisé, les MFCC représentent également d'autres dimensions du son, ce qui ne permet pas d'anticiper précisément le résultat généré, poussant ainsi l'utilisateur-ice à

l'exploration.

Enfin, dans la zone du bas de l'interface, différents sliders permettent de modifier différents paramètres de synthèse : volume, hauteur du son, période et durée des grains sonores. L'accès à ces quelques paramètres, très simples à percevoir et à contrôler, permet une plus grande variété de jeu instrumental et ouvre ainsi d'importantes possibilités de complémentarité et de dialogue entre les musicien-nes utilisateur-ices.

4. EXPÉRIENCE PILOTE

Une fois ces systèmes conçus et développés, nous avons effectué une étude préliminaire avec quatre utilisateurs experts afin d'observer quels types d'interactions et de résultat sonore pouvaient émerger. Les quatre participants de l'étude pilote étaient chercheurs en interaction humain-machine avec une expérience amateur ou professionnelle de la musique ou réalisateur en informatique musicale à l'Ircam, tous familiers de plusieurs logiciels et interfaces de création musicale numérique.

Cette étude pilote a également été l'occasion de recueillir des avis sur l'ergonomie et la simplicité d'utilisation des interfaces développées, nourrissant ainsi leur évolution vers la version présentée plus haut. Des enregistrements audio des improvisations réalisées lors de cette session sont disponibles à cette adresse : https://archive.org/details/session3_202302⁴.

4.1. Déroulé de l'expérience

L'expérience commence par une présentation générale de l'étude et une rapide explication du principe du mosaicing audio. Ensuite, pour chacun des trois systèmes, il est demandé aux participant-es d'improviser collectivement pendant 8 à 10 minutes. À la fin de chaque improvisation, une discussion a lieu afin de recueillir leurs impressions, de partager leurs expériences et de répondre à quelques questions préparées à l'avance sur la prise en main du système et des interfaces ainsi que sur les interactions qui ont eu lieu pendant l'improvisation. L'atelier se termine par une discussion générale. Nous avons choisi de présenter les systèmes dans un ordre prédéterminé (*Drum Machine*, *Clone*, *Solar System*), plutôt qu'aléatoire, pour des raisons de continuité dans l'apprentissage des interfaces.

La banque de sons *générateur* disponible propose un mélange de sons instrumentaux, de sons électroniques et de sons concrets⁵. En outre, un certain nombre d'objets et de petits instruments étaient proposés aux participants qui

4. Un extrait vidéo est également disponible à l'adresse suivante : <https://www.youtube.com/watch?v=dZPyDfiaDxI>

5. En voici le détail : un extrait du *quatuor à cordes n°5 de Bartók*, un extrait des *Variations Goldberg* de Bach, un extrait du *Lamento della ninfa* de Monteverdi, un enregistrement au microphone de contact et un enregistrement de percussions improvisées réalisés par la première autrice, un extrait d'une lecture en anglais et français par Monique Wittig, des extraits d'enregistrements de Christian Zanési et Bernard Parmegiani issus d'une banque de sample de l'INA-GRM.

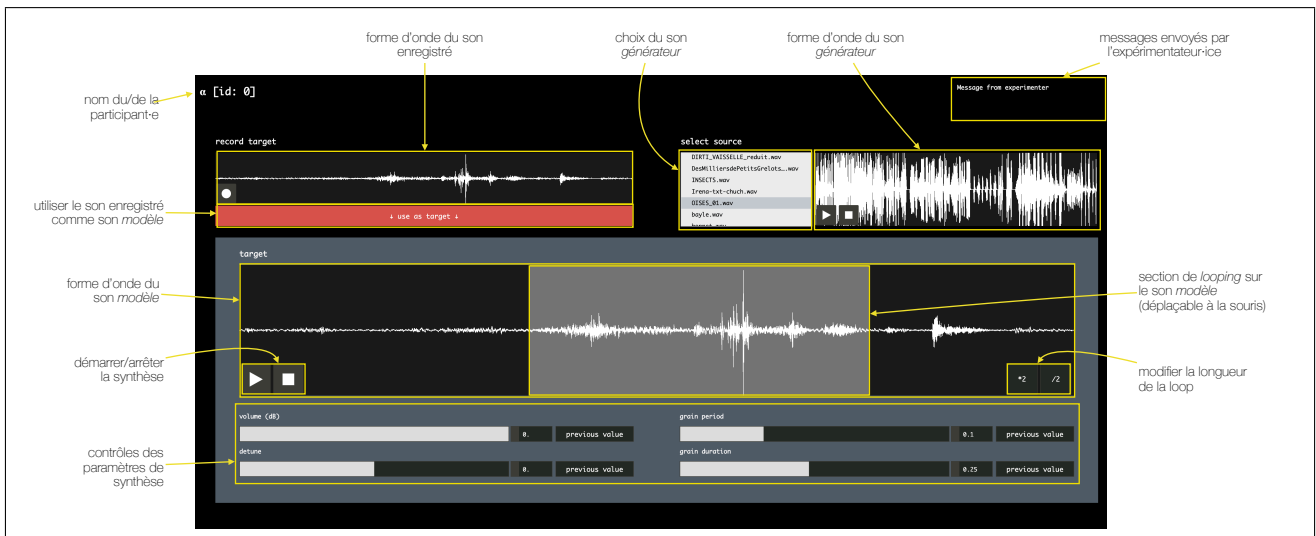


Figure 3. L'interface web du système *Drum Machine*. Les interfaces des autres systèmes sont sensiblement différentes et adaptées à chaque système.



Figure 4. Photographie de l'étude préliminaire réalisée avec quatre participants.

ne souhaitaient pas forcément utiliser leur voix : un synthétiseur pour enfant, un kalimba, un accordéon pour enfant, une grenouille en bois et un petit sanglier en céramique renfermant une bille.

4.2. Remarques générales

Les participants ont tous globalement apprécié le dispositif et les moments d'improvisation. Si la première partie de la première session peut donner l'impression d'un manque d'écoute mutuelle qui nous semble certainement dû à la période de découverte du système, les participants ont très vite proposé des stratégies pour construire l'espace sonore, en se partageant les différents registres de hauteurs, de volumes, de timbre et de rythme. Les dernières minutes de la session *Clone* nous semblent à ce titre particulièrement exemplaire : alors qu'un participant maintient un rythme plus ou moins constant avec un enregistrement de Kalimba, un autre fournit une nappe sonore et les deux participants restants interviennent de façon plus minimale avec des bruits de petites percussions répétitifs.

Les participants ont, dans l'ensemble, essayé de jouer avec les limites du système. Ils ont favorisé les boucles très courtes et ont joué avec les extrémités des sliders à leur disposition (durée des grains minimale pour donner un effet). Confrontés à la difficulté de négocier le silence et d'écouter les autres puisque le système de synthèse produit du son en continu, les participants ont rapidement appris à jouer avec le contrôle de volume.

L'utilisation du microphone à disposition a été variée et son intérêt dans la construction sonore globale a été noté : « tu peux vraiment ajuster tout ce que tu fais rythmiquement avec les boucles et savoir que tu vas enregistrer du silence. ». Cependant, les participants ont globalement moins utilisé le microphone qu'envisagé et ont préféré explorer les autres possibilités offertes par l'interface. Cela peut s'interpréter de trois façons différentes. Premièrement, les possibilités sonores du système sont larges et l'utilisation de la voix n'est strictement nécessaire qu'au début de l'improvisation : « c'est tellement riche avec très peu d'input, ça n'encourage pas de refaire des trucs avec la voix. » Deuxièmement, les participants nous ont rapporté que l'utilisation du micro était rendue difficile par l'installation technique : le type de micro utilisé – un Shure SM58 sans pied – les obligeait à utiliser une de leurs mains pour enregistrer. L'utilisation d'un micro cravate ou d'un micro sur pied pourrait régler simplement ce problème. Enfin, l'acte d'enregistrer un nouveau son avec sa voix peut être ressenti comme pouvant perturber l'espace sonore en construction : « dans le flot musical, [...] c'est une sacrée interruption. En gros, tu coupes tout ce qui se passe pour enregistrer ton truc sans que ça s'entende trop, c'est une autre sonorité... ». Ces deux derniers points mettent en évidence la manière dont des aspects tant techniques (point 2) que sociaux (point 3) sont en interrelation.

Par ailleurs, les participants ont apprécié la posture exploratoire offerte par le système : « Ouais c'est un peu au

pif. Enfin, tu fais des trucs et *oh ouais trop bien* et tout et tu gardes et si c'est pas bien ehh. Enfin du coup ça fait partie du jeu j'imagine. », tout en remarquant qu'il était parfois frustrant de ne pas pouvoir bien anticiper le type de son produit : « Je pense qu'il y a des sons que j'aurais peut-être voulu faire mais je pense que j'en ai pas été capable et je sais pas si j'en aurais été capable ». Suite à ces remarques, nous avons implémenté un gain sur la synthèse contrôlé par la RMS du son *modèle* afin de donner une appréhension plus immédiate du volume sonore produit.

4.3. Différences selon les systèmes

Les participants ont ressenti des impressions et des contraintes très différentes selon les systèmes.

Dans la session *Clone*, les participants ont trouvé le résultat sonore « plus aéré et plus subtil » et « plus facile à contrôler » qu'avec le système *Drum Machine* où « avec le choix des corpus on avait quand même des [...] présences sonores assez fortes ». L'un des participants fait ainsi remarquer que *Clone* favorise une écoute collective : « le fait que ça soit un nouveau matériau sonore et pas des matériaux qui sont dans une banque [...] incite plus à l'écoute ».

Avec le système *Solar System*, où les rôles sont asymétriques, les participants ont remarqué qu'ils avaient moins de contrôle à leur disposition sans que cela soit ressenti comme un problème : « c'était frustrant au début [mais] à la fin on sent qu'on a beaucoup plus de contraintes donc on devient un peu plus créatif ». Les participants ont aussi bien perçu l'asymétrie entre les deux rôles au sein de ce système. L'un des participants ayant le rôle de *satellite* s'est ainsi adressé au participant jouant le rôle de *soleil* en lui disant « tu es compositeur et nous on interprète ». Cependant certains participants n'ont pas eu vraiment l'impression de jouer avec le *soleil* et ont ressenti une certaine « déconnexion » en étant plus concentré sur le résultat sonore que sur ce qu'il faisait.

4.4. Modifications suite à l'expérience

Cette étude pilote a été l'occasion de faire émerger plusieurs remarques pertinentes concernant les différents systèmes, les interfaces et le déroulé de l'atelier. Ceci a donc mené à un certain nombre de modifications.

En ce qui concerne l'installation technique du système, outre la question du microphone évoquée plus haut, la question du positionnement des haut-parleurs a été évoquée par plusieurs participants. Concernant l'interface, les participants ont noté qu'il y avait quelques éléments qui « coupent la continuité ». En effet, dans cette première version du système, changer de son *générateur* menait à une interruption du son car l'analyse du fichier provoquait un blocage du processus de synthèse. Ceci a été modifié dans une version ultérieure et l'analyse est désormais effectuée dans un fil d'exécution parallèle grâce à un *Web Worker*.

Cette expérience pilote a également permis de souligner deux problèmes concernant le déroulé de l'expérience elle-même. Premièrement, elle a fait apparaître la nécessité d'un

temps d'appropriation du système en isolation par les participants avant de commencer les phases d'improvisation « pour [...] se créer une sorte de mini vocabulaire ». Deuxièmement, il est apparu nécessaire d'ajouter à l'interface un moyen de faire passer des messages aux participants pour, par exemple, débiter la fin d'une improvisation sans interférer avec l'espace sonore.

5. CONCLUSION

Nous avons présenté *Simone*, un ensemble de systèmes et d'interfaces web pour l'improvisation collective et distribuée conçu pour être utilisé par des utilisateur·ices expert·es dans un cadre d'improvisation.

À cette intention, nous avons fait tester *Simone* à un groupe de 4 utilisateurs experts lors d'une étude pilote consistant en plusieurs sessions d'improvisation collective et de discussions. Les retours collectés lors de cette expérience ont mis en évidence les points positifs et négatifs du système, et ont mené à un certain nombre de modifications dans sa conception et dans le déroulé de l'expérience.

Dans la prochaine étape de ce travail, nous allons réitérer cette expérience avec plusieurs groupes de participants·es. Il est prévu que les prochaines sessions d'expérience proposent un court moment individuel de prise en main de l'interface au début et des entretiens individuels après l'expérience, ce qui permettra de mener une analyse et une réflexion plus approfondie sur les dynamiques d'improvisation collective dans un instrument distribué tel que proposé par *Simone*. Par ailleurs, la version embarquée du système a été confiée à un artiste actif dans le milieu de l'improvisation et du spectacle vivant afin de mener une étude au plus long court sur le processus d'appropriation et de prise en main.

Références

- [1] WebAudio API Specification. <https://www.w3.org/TR/webaudio/>.
- [2] BROWN, C. *Talking Drum* : A Local Area Network Music Installation. *Leonardo Music Journal* 9 (Dec. 1999), 23–28.
- [3] CANONNE, C., MONTEIRO, R., AND RÜHL, J. L'EMUPO : Une interface logicielle pour l'improvisation collective. In *Journées d'Informatique Musicale* (Saint-Etienne, France, 2011).
- [4] COLLINS, N. *Introduction to Computer Music*. John Wiley & Sons, Hoboken, 2010.
- [5] FREEMAN, J., VARNIK, K., RAMAKRISHNAN, C., NEUHAUS, M., BURK, P., AND BIRCHFIELD, D. *Auracle* : A voice-controlled, networked sound instrument. *Organised Sound* 10, 3 (Dec. 2005), 221–231.
- [6] GRESHAM-LANCASTER, S. A Personal Reminiscence on the Roots of Computer Network Music. *Leonardo Music Journal* 27 (Dec. 2017), 71–77.

- [7] JANER, J., AND DE BOER, M. Extending voice-driven synthesis to audio mosaicing. In *Sound and Music Computing Conference (SMC '08)* (Berlin, Germany, 2008).
- [8] KIM-BOYLE, D. Network Musics : Play, Engagement and the Democratization of Performance. *Contemporary Music Review* 28, 4-5 (Aug. 2009), 363–375.
- [9] KLEINBERGER, R., SINGH, N., XIAO, X., AND VAN TROYER, A. Voice at NIME : A Taxonomy of New Interfaces for Vocal Musical Expression. In *NIME 2022* (Auckland, New Zealand, June 2022).
- [10] MAGNUSSON, T. Of Epistemic Tools : Musical instruments as cognitive extensions. *Organised Sound* 14, 2 (Aug. 2009), 168–176.
- [11] MATUSZEWSKI, B. A Web-Based Framework for Distributed Music System Research and Creation. *Journal of the Audio Engineering Society* 68, 10 (Dec. 2020), 717–726.
- [12] ROHRHUBER, J. Network music. In *The Cambridge Companion to Electronic Music*, N. Collins and J. d'Esquivan, Eds., first ed. Cambridge University Press, Dec. 2007, pp. 140–155.
- [13] ROHRHUBER, J., CAMPO, A., WIESER, R., VAN KAMPEN, J.-K., HO, E., AND HÖLZL, H. Purloined Letters and Distributed Persons. In *Music in the Global Village Conference* (Múcsarnok Budapest, 2007).
- [14] ROTTIER, O., AND MATUSZEWSKI, B. A Rust Implementation of the Web Audio API. In *Web Audio Conference* (Cannes, France, June 2022).
- [15] STONE, P. Non-Mathematical Musings on Information Theory and Networked Musical Practice. *Organised Sound* 26, 3 (Dec. 2021), 327–332.
- [16] TRUEMAN, D., COOK, P., SMALLWOOD, S., AND WANG, G. PLOrk : The Princeton Laptop Orchestra, Year 1. In *International Computer Music Conference (ICMC '06)* (New Orleans LA USA, 2006).
- [17] TURCHET, L., FISCHIONE, C., ESSL, G., KELLER, D., AND BARTHET, M. Internet of Musical Things : Vision and Challenges. *IEEE Access* 6 (2018), 61994–62017.
- [18] WANG, G., ESSL, G., AND PENTTINEN, H. Do Mobile phones Dream of Electric Orchestras? In *International Computer Music Conference (ICMC '08)* (Belfast, Ireland, 2008), p. 8.
- [19] WEINBERG, G. Voice Networks : The Human Voice as a Creative Medium for Musical Collaboration. *Leonardo Music Journal* 15 (Dec. 2005), 23–26.
- [20] ZAPPI, V., AND MCPHERSON, A. P. Dimensionality and Appropriation in Digital Musical Instrument Design. In *International Conference on New Interfaces for Musical Expression (NIME '14)* (London, UK, 2014).
- [21] ZILS, A., AND PACHET, F. Musical Mosaicing. In *COST G-6 Conference on Digital Audio Effects (DAFX-01)* (Limerick, Ireland, 2001).

UNE STATION DE TRAVAIL AUDIO-NUMÉRIQUE POUR LA PLATE-FORME WEB

Antoine Vidal-Mazuy

University Côte d'Azur, CNRS, INRIA
antoine.vidal-mazuy@etu.univ-cotedazur.fr

Michel Buffa

University Côte d'Azur, CNRS, INRIA
buffa@univ-cotedazur.fr

RÉSUMÉ

Cet article présente WAM Studio (Figure 1), une station de travail audio numérique (DAW) en ligne open source qui tire parti de plusieurs APIs et technologies standards du W3C, telles que Web Audio, WebAssembly, Web Components, Web Midi, Media Devices, etc. WAM Studio s'appuie également sur le standard Web Audio Modules (WAM), qui a été conçu pour faciliter le développement de plugins audio inter-opérables (effets, instruments virtuels, claviers virtuels de piano comme contrôleurs, etc.) sortes de "VSTs pour le Web". Les DAWs sont des logiciels riches en fonctionnalités et donc particulièrement complexes à développer en termes de conception, d'implémentation, de performances et d'ergonomie. Aujourd'hui, la majorité des DAWs en ligne sont commerciaux alors que les seuls exemples open source manquent de fonctionnalités (pas de prise en charge de plugins par exemple) et ne tirent pas parti des possibilités récentes offertes (comme WebAssembly). WAM Studio a été conçu comme un démonstrateur technologique pour promouvoir les possibilités offertes par les innovations récentes proposées par le W3C. L'article met en évidence certaines des difficultés que nous avons rencontrées (par exemple, les limitations dues aux environnements sandboxés et contraints que sont les navigateurs Web, la compensation de latence quand on ne peut pas connaître le hardware utilisé, etc.). Une démonstration en ligne, ainsi qu'un *repository* GitHub pour le code source sont disponibles.

1. INTRODUCTION

La Musique Assistée par Ordinateur (MAO) est un domaine en constante évolution qui utilise des ordinateurs pour enregistrer, éditer et produire de la musique. Les stations de travail audio numérique (Digital Audio Workstations en anglais, ou "DAWs") sont des logiciels spécialement conçus pour la MAO, permettant aux utilisateurs de créer et de manipuler du contenu audio numérique ainsi que du contenu MIDI. Les plugins audio sont des modules logiciels qui ajoutent des fonctionnalités supplémentaires aux DAWs, offrant aux utilisateurs une plus grande flexibilité et un meilleur contrôle sur leur production musicale. Le marché de la MAO est né avec l'Atari ST (1985) -premier ordinateur supportant le standard MIDI- et le DAW Cubase proposée par Steinberg (1989). Peu après, le standard VST pour les plugins audio a été proposé (1997) et depuis lors,

des milliers de plugins ont été développés, pouvant être utilisés dans les principaux DAWs natifs disponibles sur le marché.

Un DAW est un logiciel riche en fonctionnalités et donc particulièrement complexe à développer en termes de conception, d'implémentation et d'ergonomie. Il permet la création de pièces multipistes en utilisant directement des échantillons audio (par exemple en incorporant dans une piste un fichier audio ou en enregistrant à partir d'un microphone ou d'une entrée de carte son), en les mélangeant, en appliquant des effets sonores à chaque piste (par exemple, réverbération, égalisation de fréquence ou auto-tune sur les voix), mais également en utilisant des pistes avec des instruments virtuels (reproduction logicielle d'un piano, d'un violon, d'un synthétiseur, d'une batterie, etc.). La piste est ensuite enregistrée au format MIDI sous forme d'événements correspondant aux notes et à des paramètres supplémentaires (comme la vitesse avec laquelle on a appuyé sur les touches d'un clavier de piano, par exemple). Ces événements permettent à la piste d'être jouée en demandant à un instrument virtuel de synthétiser le signal. Une DAW permet donc la lecture, l'enregistrement et le mixage à la fois de pistes audio et MIDI, l'édition de ces pistes (copier/couper/coller dans une piste ou entre les pistes), la gestion des effets audio en temps réel et des instruments virtuels, le mixage et l'exportation du projet final dans un format simple (par exemple, un fichier WAVE ou MP3). Dans le monde natif, ces effets et instruments sont les "plugins audio" qui étendent les capacités des DAWs standard. Depuis 1997, un marché important pour les développeurs de plugins tiers s'est développé. Quatre DAWs dominent le marché (Logic Audio, Ableton, Pro Tools, Cubase)¹ et l'existence de plusieurs formats de plugins propriétaires complique la tâche des développeurs.

D'un autre côté, le marché de la MAO sur le Web est encore émergent, les premiers DAWs en ligne sont apparus en 2008 et exploitaient la technologie Flash. Ceux utilisant HTML5 et l'API Web Audio ne sont apparus que dans la période entre 2015 et 2016 car les technologies sont beaucoup plus récentes (les premières implémentations de l'API Web Audio dans les navigateurs datent de 2012). En 2023, plusieurs DAWs basés web sont disponibles, principalement commerciaux. Un format de plugins inter-opérables appelé Web Audio Modules (WAM) existe et est soutenu par au moins un DAW en ligne.

1. <https://tinyurl.com/s4tbjzew>

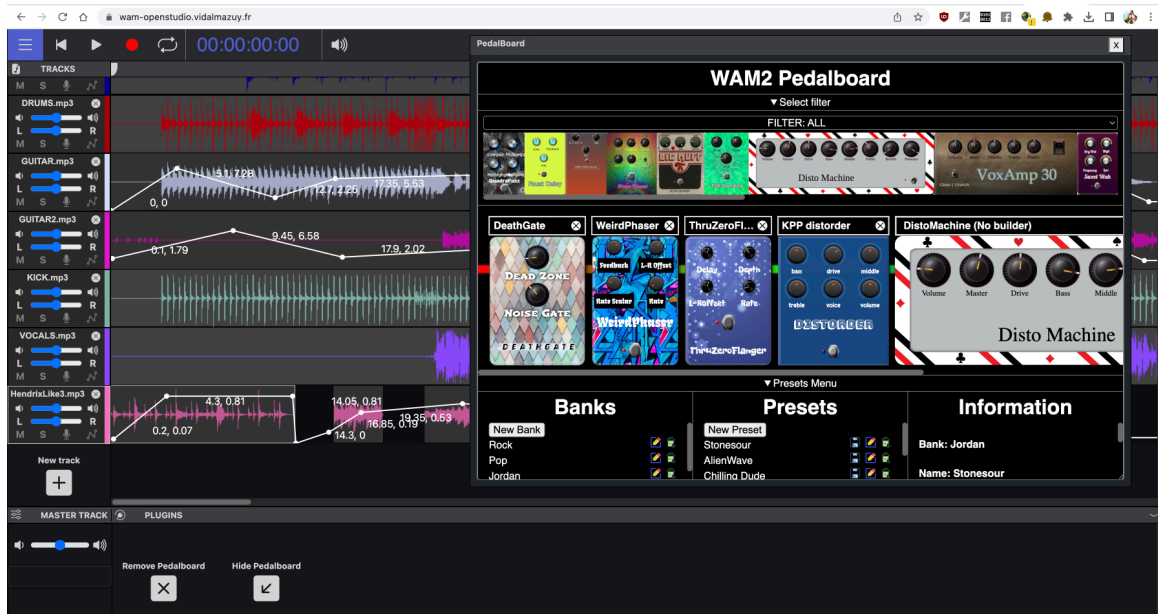


Figure 1. WAM-Studio, une station de travail audio numérique en ligne et open source. On peut voir des plugins d'effets associés à la dernière piste et des courbes d'automatisation de paramètres.

2. L'API WEB AUDIO

Depuis 2018, l'API Web Audio du W3C est une "recommandation" (une norme figée). Elle offre un ensemble de "nœuds audio" qui traitent ou produisent du son et on utilise ces nœuds depuis du code JavaScript en instanciant des classes fournies par l'API. Ces nœuds peuvent être connectés pour former un "graphe audio". Le son se déplace à travers ce graphe à la fréquence d'échantillonnage (valeur par défaut 44100Hz, cette valeur peut être modifiée) et subit des transformations [3]. Certains nœuds sont des générateurs d'ondes ou des sources sonores correspondant à une entrée de microphone ou à un fichier sonore chargé en mémoire, d'autres transforment le son. La connexion de ces nœuds dans le navigateur se fait via JavaScript et permet de concevoir une large gamme d'applications différentes impliquant le traitement audio en temps réel [6]. Les applications musicales ne sont pas les seules à nécessiter du traitement audio complexes, l'API est également conçue pour répondre à d'autres besoins, par exemple pour le développement de jeux vidéo, pour le multimédia, la visio-conférence, etc. L'API est livrée avec un ensemble limité de nœuds "standards" pour des opérations courantes telles que le contrôle de gain, le filtrage de fréquences, l'ajout de *delay*, de réverbération, pour des traitements sur la dynamique, sur la spatialisation du son 2D et 3D, etc.

En général les bibliothèques/API audio répartissent le travail entre un thread de contrôle et un thread de rendu [7, 11, 12]. L'API Web Audio ne fait pas exception et utilise également ce modèle : un thread de rendu appelé "audio thread" est chargé exclusivement du traitement du son par le graphe audio et de la livraison des échantillons sonores au système d'exploitation pour qu'ils puissent être lus par le matériel. Ce thread est soumis à des contraintes de temps

réel strictes et a une priorité élevée - s'il ne parvient pas à fournir le prochain bloc d'échantillons audio à temps (128 échantillons par défaut dans le cas de l'API Web Audio), il y aura des anomalies audibles. D'autre part, le thread de contrôle est généralement en charge d'exécuter le code JavaScript de l'interface utilisateur et d'effectuer les appels à l'API Web Audio. Il gère aussi toutes les modifications apportées au graphe audio : il permet par exemple à l'utilisateur de connecter/déconnecter des nœuds et d'ajuster leurs paramètres. Les nœuds standards sont tous des instances de sous-classes de la classe `AudioNode`. A une exception près, ils fournissent des traitements prédéfinis codés en C++ ou en Rust (Firefox) et exécutent le traitement du son dans le thread audio, mais les algorithmes utilisés ne peuvent pas être modifiés, seules les modifications de paramètres sont autorisées depuis du JavaScript. L'exception citée est l'ajout récent (2018) du nœud `AudioWorklet` qui fournit une solution pour implémenter un traitement audio personnalisé de bas niveau s'exécutant dans le thread audio [6], avec de nombreuses contraintes (pas d'opérations asynchrones, d'imports de fichiers, pas d'accès au DOM de la page HTML, etc.).

Les nœuds de l'API Web Audio peuvent être assemblés dans un graphe permettant aux développeurs de créer des effets ou instruments audio plus complexes. Voici quelques exemples construits de cette manière : effet *delay* (`DelayNode`, `BiquadFilterNode`, `GainNode`), auto-wah (`BiquadFilterNode`, `OscillatorNode`), chorus (multiples `DelayNodes` et `OscillatorNodes` pour la modulation), distorsion (`GainNode`, `WaveShaperNode`), synthétiseurs (`OscillatorNodes`), samplers (`AudioBufferNodes`), etc. Du code DSP existant dans d'autres langages tels que C/C++ ou écrit en utilisant des langages spécifiques à un domaine tels que

FAUST[13], peut être compilé en WebAssembly et exécuté dans un seul nœud AudioWorklet. Au fil des années, de nombreux effets et instruments audio de haut niveau ont ainsi été développés [4]. Cependant, il est souvent nécessaire de chaîner de tels effets et instruments audio (par exemple, le pédalier d'un guitariste est composé de pédales d'effets connectés entre elles) et lors de la composition/production musicale dans des DAWs, plusieurs effets et instruments sont en général utilisés. Ce sont des cas où les nœuds de l'API Web Audio sont de trop bas niveau, d'où la nécessité d'une unité de plus haut niveau pour représenter l'équivalent d'un plugin audio natif [1]. Pour la plate-forme Web, un tel standard de haut niveau pour les "plugins audio" et les applications "hôtes" n'existait pas avant 2015 [8]. Plusieurs initiatives ont été lancées et l'une d'entre elles est devenue un "standard communautaire" que nous détaillons dans la section suivante.

3. WEB AUDIO MODULES

En 2015, Jari Kleimola et Olivier Larkin ont proposé une norme pour des plugins Web Audio sur le Web, intitulée "Web Audio Modules" (WAM) [8]. Peu après, Jari Kleimola a participé à la création d'ampedstudio.com, l'un des premiers DAWs en ligne utilisant l'API Web Audio. En 2018, le projet WAM initial a été étendu par des chercheurs avec l'aide de développeurs de l'industrie de la MAO, ce qui a donné lieu à une norme plus polyvalente [3]. Finalement, la version 2.0 des Web Audio Modules a été publiée en 2021. Cette version vise, en plus d'établir une norme communautaire (API, SDK), à apporter plus de liberté aux développeurs (support d'outils de build, TypeScript, frameworks React, etc.), à améliorer les performances, à simplifier l'accès aux paramètres de plugin et à faciliter l'intégration dans les DAWs [5]. Nous reviendrons sur cette fonctionnalité plus tard, mais le standard WAM 2.0 utilise une conception originale pour la gestion de la communication entre les plugins et les applications hôtes qui ne repose pas sur la gestion des paramètres fournie par l'API Web Audio. La raison principale est de permettre des performances élevées dans le cas où à la fois un DAW et des plugins sont implémentés en tant qu'AudioWorklet. En effet, au moment de la conception de l'API Web Audio, les AudioWorklets n'existaient pas et certains cas d'utilisation ne pouvaient pas être pris en compte. Si le DAW est construit en utilisant des nœuds AudioWorklet pour le traitement audio, alors certaines parties du code s'exécutent dans le thread audio à haute priorité. En outre, si un plugin WAM est associé à une piste donnée dans un projet DAW, et si le plugin est lui-même construit en utilisant un nœud AudioWorklet, il dispose également de code personnalisé s'exécutant dans le thread audio. Le standard WAM a été conçu pour gérer ce cas particulier et permet une communication DAW/plugins sans quitter le thread audio. Prenons un exemple : pendant la lecture, une piste MIDI envoie des notes à un plugin d'instrument virtuel et modifie certains des paramètres de ce plugin à la fréquence d'échantillonnage (a-rate). N'oublions pas qu'un DAW peut

avoir plusieurs pistes, chacune associée à des dizaines de plugins, et que chaque plugin peut avoir des dizaines de paramètres. Le standard WAM détectera automatiquement le cas où DAW et plugins sont des AudioWorklets et optimisera en coulisse la communication (avec utilisation de mémoire partagée et de buffer tournant), sans franchir la barrière du thread audio. Pas besoin d'envoyer des événements à partir du thread de contrôle/GUI, ce qui aurait été obligatoire si la gestion des paramètres de l'API Web Audio était utilisée.

En résumé, le standard WAM simplifie la création de plugins et d'applications hôtes et permet une communication optimale entre les hôtes et les plugins.

4. ANALYSE DE L'EXISTANT : DAWs COMMERCIAUX ET OPEN SOURCE

		bandlab	ampedstudio	soundtrap	soundation	arpeggi	audiotool
Target	Desktop	X	X	X	X	X	X
	Mobile (native or responsive webapp)	X		X		X	
	Oriented to users familiar with Computer Music		X	X			
	Premium (free, with premium functionalities)	X	X	X	X		X
functionalities	Basic DAW features (recording, editing, mixing etc.)	X	X	X	X	X	X
	Audio effects and virtual instruments	X	X	X	X	X	X
	Commercials plugins support (not documented)	X	X	X	X		
	Open plugins support (WAM)		X				
	Cloud (saving project, loading, sharing, etc.)	X	X	X	X	X	X
	Collaborative (asynchronous)	X	X	X	X	X	X
	Collaborative (synchronous)			X			X
Technologies	Call in app		X	X			X
	Using AudioWorklet	X	X	?			X
	C++/wasm audio engine	X	X				
	Native mobile application	X					
Blockchain						X	
Open Source							

Figure 2. Table de comparaison des principaux DAWs commerciaux disponibles

Les premiers DAWs en ligne basés sur l'API Web Audio sont apparus entre 2015 et 2018 et sont encore disponibles aujourd'hui : Audiotools [9], Bandlab, Amed-Studio, Soundtrap [10], Soundation. Ces DAWs ont en commun le fait qu'ils sont commerciaux, fermés et ont fait l'objet de très peu de publications académiques. Ils facilitent la collaboration à distance sur des projets musicaux, notons que la pandémie de COVID-19 (2020-2022) leur a été profitable et a augmenté leur popularité. Le mode de collaboration varie (synchrone comme Google Docs ou asynchrone grâce au partage de liens), ainsi que les outils de communication fournis (chat, vidéo-conférence), mais

tous ces DAWs proposent les fonctionnalités classiques : enregistrement audio et MIDI, édition de pistes, mixage, support pour effets et instruments virtuels, etc. Certains ont été conçus principalement pour les ordinateurs de bureau, tandis que d'autres sont particulièrement adaptés aux appareils mobiles.

Ils diffèrent principalement en termes d'ergonomie : certaines de ces applications sont destinées à un public très large et ont mis l'accent sur la simplicité (BandLab, SoundTrap), tandis que d'autres ont choisi un aspect plus "professionnel" et sobre, comme AmpedStudio. De son côté, Audiotools est davantage un "studio virtuel" qu'un DAW pur et trouve son marché principalement dans le domaine de l'éducation. Les détails de mise en œuvre de ces applications sont inconnus car le code source n'est pas accessible au public. Cependant, à travers des publications académiques limitées et des présentations/interviews, on sait qu'AmpedStudio utilise un moteur C++ cross-compilé en WebAssembly et des AudioWorklets, tandis que Soundtrap a principalement utilisé les nœuds standard de l'API Web Audio. BandLab est censé avoir un noyau écrit en C++ et s'appuie également sur WebAssembly/AudioWorklet, avec peut-être une version mobile spécifique. Audiotools fonctionne également dans des nœuds AudioWorklet avec un portage du moteur audio initialement écrit en Flash. Récemment, un nouveau DAW appelé Arpeggio.io est apparu, mettant en avant l'utilisation de la technologie blockchain (pour suivre qui, quand et comment les sons et la musique sont utilisés) et des NFTs pour monétiser les créations.

Ces DAW diffèrent également dans la manière dont ils gèrent les effets et les instruments audio. Il est évident que certains prennent en charge des plugins externes : AmpedStudio prend en charge la norme WAM -les développeurs de la société ont contribué à sa création- et le site web a une boutique pour "activer" des plugins premium. Les autres DAWs en ligne commerciaux semblent utiliser un format propriétaire, tout en intégrant des portages tiers de plugins natifs (SoundTrap propose le célèbre plugin "Auto-tune" d'Antares). Propellerheads, une entreprise bien connue pour son DAW natif Reason, a également publié des versions web de son synthétiseur Europa sous forme de plugin Web (disponible maintenant dans AmpedStudio et dans Soundation, il s'agit d'un portage de Plugin Rack, un standard de plugins propre à Propellerheads dans le monde natif).

Jusqu'à présent, AmpedStudio est le seul DAW à prendre en charge une norme de plugins ouverte : Web Audio Modules. La figure 2 illustre les similitudes et les différences entre les DAW en ligne commerciaux cités.

Dans le monde de l'open source, le choix est limité : GridSound est un DAW développé depuis 2015 ² qui prend en charge l'audio et le MIDI. Bien que les effets et instruments proposés demeurent simples, GridSound est un DAW entièrement fonctionnel. Il ne prend pas en charge de plugins.

2. <https://gridsound.com/>. Voir également la présentation vidéo de GridSound à la conférence ADC 2021 : <https://www.youtube.com/watch?v=eJTtENwRxnA>

Il existe également des bibliothèques JavaScript populaires pour développer un lecteur/enregistreur multipiste, telles que wavesurfer.js ³, peaks.js ⁴ ou waveform-playlist ⁵, et des éditeurs de fichiers audio de type audacity tels que AudioMass ⁶.

Aucune de ces initiatives open source n'utilise de traitement DSP personnalisé de bas niveau (pas d'AudioWorklets), ni ne se soucie d'optimiser la communication entre DAW/plugin, ni n'utilise de plugins externes. C'est la raison principale pour laquelle nous avons développé WAM Studio : en tant que démonstrateur de ces techniques.

5. WAM STUDIO DESIGN ET IMPLÉMENTATION

Wam-Studio est un outil en ligne pour créer des projets audio que l'on peut imaginer comme de la musique multipistes. Chaque piste correspond à une "couche" différente de contenu qui peut être enregistré, édité, joué ou simplement intégré (en utilisant des fichiers audio, par exemple). Certaines pistes peuvent être utilisées pour contrôler des instruments virtuels : dans ce cas, elles ne contiennent que les notes qui doivent être jouées, avec quelques métadonnées. Des pistes peuvent être ajoutées ou supprimées, jouées isolément ou avec d'autres pistes. Elles peuvent également être "armées" pour l'enregistrement, et lorsque l'enregistrement commence, toutes les autres pistes joueront simultanément, tandis que les pistes armées enregistreront un nouveau contenu.

5.1. Les pistes

Dans un DAW, chaque piste est un conteneur de données liées à l'audio qui s'accompagne d'une représentation interactive de ces données, de fonctionnalités d'édition et de traitement ainsi que de quelques paramètres par défaut tels que le volume et le panoramique gauche/droite de la piste.

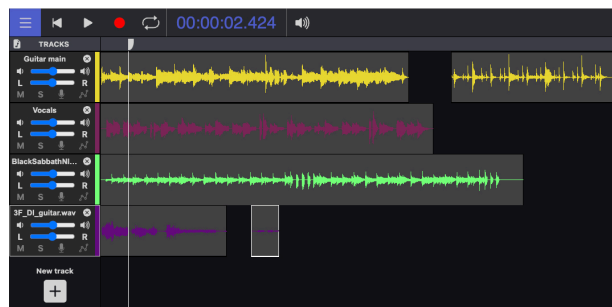


Figure 3. Pistes et régions dans la GUI du DAW Wam-Studio. Les régions peuvent être déplacées, supprimées, etc.

La figure 3 montre des pistes audio dans WAM-Studio avec l'affichage de la forme d'onde des buffers audio associés et les contrôles/paramètres par défaut des pistes,

3. <https://wavesurfer-js.org/>

4. <https://github.com/bbc/peaks.js/>

5. <https://github.com/naomiario/waveform-playlist>

6. <https://audiomass.co/>

sur le côté gauche (mute/solo, armement pour l'enregistrement, volume, panoramique stéréo, affichage des courbes d'automatisation). Comme de nombreuses pistes peuvent être affichées, scrollées pendant la lecture, zoomées et éditées, nous avons utilisé la bibliothèque pixi.js pour gérer de manière efficace le dessin et les interactions dans un canvas HTML. Cette bibliothèque utilise un rendu WebGL accéléré par GPU et offre de nombreuses fonctionnalités pour la gestion de plusieurs calques sur un seul canvas HTML5 (Figure 3). On voit sur la Figure 5 que chaque piste peut également être associée à un ensemble de plugins pour ajouter des effets audio ou générer de la musique (dans le cas de plugins instrumentaux).

La Figure 4 montre le graphe audio correspondant à la chaîne de traitement d'une piste audio. Le son va de gauche à droite : d'abord le "lecteur/enregistreur/éditeur de piste" est implémenté en tant que nœud AudioWorklet, en utilisant du code personnalisé pour le rendu d'un buffer audio, puis le signal de sortie a son gain et son panoramique stéréo ajustés, ensuite nous avons un autre nœud AudioWorklet pour le rendu du volume dans un canvas (VU-mètre), et enfin nous trouvons une chaîne de plugins WAM pour ajouter des effets audio.

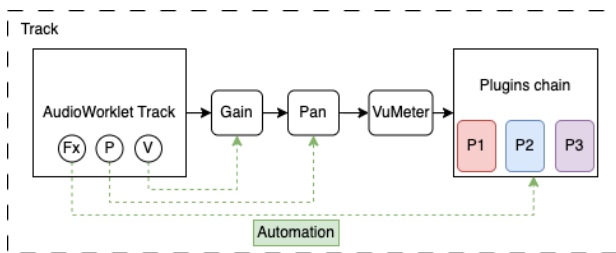


Figure 4. Graphe d'une piste audio dans Wam-Studio

Il existe deux types de pistes : **les pistes "audio"** : elles contiennent de l'audio enregistré, tel qu'une prise vocale, un enregistrement de guitare ou tout autre type de signal sonore, qui sont généralement rendus graphiquement avec une forme d'onde représentant le signal stocké (sous forme d'un ensemble d'échantillons sonores). Ces pistes peuvent être éditées, traitées et mélangées en copiant/coupant et collant des échantillons dans le buffer audio associé à la piste. Le signal de sortie des pistes audio peut être traité par une chaîne de plugins d'effet audio. **Les pistes "MIDI"** : elles contiennent des données MIDI (le pitch - la note qu'un instrument jouerait, sa vélocité et sa durée). Les données MIDI ne contiennent pas d'informations audio mais sont plutôt utilisées pour contrôler des instruments virtuels tels que des synthétiseurs, des samplers, des boîtes à rythmes, etc. Les pistes MIDI peuvent être éditées, auquel cas il est possible de modifier les événements MIDI stockés dans la piste dans un affichage de type matriciel. Les pistes MIDI sont généralement associées à un ou plusieurs plugins WAM d'instruments virtuels. Les événements MIDI peuvent cibler tous les plugins d'instruments de la piste ou certains en particulier.

Les pistes sont généralement organisées verticalement dans l'interface utilisateur et peuvent être gérées séparé-

ment en termes de volume, de panning stéréo, de plugins et d'automatisation des paramètres de ces plugins d'effets ou d'instruments (comme le montre la Figure 5).



Figure 5. Une chaîne de plugins associés à la piste sélectionnée.

Chaque sortie de piste est connectée à une "piste maître" (master track) où il est possible d'ajuster globalement le volume et le panning stéréo du mix final. Comme pour toutes les autres pistes, il est également possible d'associer une chaîne d'effets audio à la piste maître (par exemple pour ajuster la dynamique et les fréquences finales). Tous les effets audio et instruments virtuels sont des plugins WAM dans le cas de notre DAW, et tous leurs paramètres sont automatisables⁷. Cette conception offre un degré de contrôle et d'adaptabilité étendu et permet aux utilisateurs de mélanger et de manipuler le son de chaque piste avec une grande précision et sophistication, rendant ainsi plus facile la création de productions audio complexes, le tout sans quitter le navigateur Web. WAM-studio est le seul logiciel open source à proposer l'ensemble de ces fonctionnalités.

Une piste DAW est également capable de "rendre" la piste en un signal audio audible (lorsque la piste est en lecture). Lorsque l'on appuie sur le bouton de lecture du DAW, toutes les pistes sont rendues simultanément, et le signal de sortie final est ce que l'on appelle "le mix". Les DAWs disposent également d'une option de "rendu hors ligne" pour exporter le mix final sous forme de fichier, sur le disque dur local ou sur le cloud (en utilisant l'OfflineAudioContext⁸ fourni par l'API Web Audio).

5.2. Le coeur des pistes est un AudioWorklet Processor

Concentrons-nous pour le moment sur une piste de type audio. L'API Web Audio fournit le nœud AudioBufferSourceNode qui est un conteneur pour un buffer audio, et qui permet la lecture d'un buffer audio stocké en mémoire. Cette approche simple à mettre en oeuvre est néanmoins non adaptée pour un DAW devant gérer

7. A la fréquence d'échantillonnage (a-rate), du quantum audio (k-rate), personnalisée : cela dépend de l'implémentation du plugin au format WAM

8. Contrairement à un AudioContext standard, qui sert à créer les nœuds du graphe audio pour un rendu "live", un OfflineAudioContext sert lui, à générer aussi rapidement que possible un graphe audio et à envoyer le résultat dans un AudioBuffer, qui peut ensuite être converti en format WAVE par exemple.

l'automation de paramètres. Cette tâche consiste à interpoler, pendant la lecture ou l'enregistrement multipiste, les paramètres des plugins et des pistes à de hautes fréquences (à la fréquence d'échantillonnage *a-rate* ou de contrôle *k-rate*). Un paramètre automatisé peut être celui de n'importe quel plugin associé à n'importe quelle piste. Il peut y avoir potentiellement des dizaines, des centaines de paramètres à automatiser. Cela représente une quantité substantielle d'informations qui peut être échangée entre le DAW et les plugins WAM, ce qui fait de la performance un aspect crucial qui nécessite une attention particulière. En concevant notre propre lecteur audio de bas niveau (au lieu d'utiliser le nœud standard `AudioBufferSourceNode`), nous avons un contrôle total sur le comportement de lecture/enregistrement de chaque piste. Nous avons ainsi conçu le noyau de chaque piste audio comme un "AudioWorklet processor", qui permet l'exécution de code DSP personnalisé dans le thread audio à très haute priorité. Plus précisément, ce noyau est une instance d'une sous-classe de la classe `WamProcessor` fournie par le SDK des Web Audio Modules, qui hérite de la classe `AudioWorkletProcessor`. Selon l'API, chaque processeur implémente une méthode `process(input, output)` qui sera appelée à la fréquence d'échantillonnage. À l'intérieur de cette méthode, nous traitons les échantillons sonores et procédons au rendu du buffer audio, mais nous pouvons également envoyer des données aux plugins WAM. La classe `WAMProcessor` dont nous héritons fournit en effet des méthodes pour gérer efficacement la communication entre le DAW et les plugins.

Avec l'API Web Audio, les `AudioWorklets` sont composés de deux composants, reflétant la nature multi-threadée de l'environnement : le nœud `AudioWorkletNode` et le `AudioWorkletProcessor`. Le standard WAM les étend avec les classes `WamNode` et `WamProcessor`, qui s'exécutent respectivement dans le main thread du navigateur (thread de la GUI) et dans le thread audio. WAM-Studio est une application hôte WAM, et ses pistes sont donc mises en œuvre en tant que sous-classes de `WamProcessor`. Les plugins WAM associés à ces pistes peuvent également avoir leur cœur DSP implémenté en tant que sous-classes de `WamNode` avec un `WamProcessor`, mais ce n'est pas obligatoire car certaines implémentations de plugins WAM utilisent plutôt un graphe composé d'une combinaison de nœuds, dans ce cas ils étendent la classe `WAM CompositeNode`. Quelle que soit l'implémentation des plugins le code de communication de l'hôte vers le plugin est le même, et les optimisations mise en œuvre se font en coulisse (dans le WAM SDK).

Cette abstraction permet une inter-opérabilité transparente entre les hôtes et les plugins WAM, indépendamment de la mise en œuvre sous-jacente et sur tous les threads. En d'autres termes : si une piste est un `WamProcessor` et que les plugins sont également des `WamProcessor`, leur code DSP s'exécute dans le thread audio et une communication hautement optimisée peut être réalisée : les `Shared Array Buffers` peuvent être utilisés pour la communication entre DAW et plugins. Si un plugin n'est pas un `WamProcessor`, alors le code DAW qui "communique" avec lui reste

inchangé et la mise en œuvre sous-jacente sera moins optimale et impliquera le franchissement de la barrière des threads. Pour gérer plusieurs plugins en chaîne, le standard WAM utilise un objet singleton, appelé `WamEnv`, qui est attaché à la portée globale du thread audio et sert de médiateur pour les interactions entre les hôtes et les plugins. Les plugins sont structurés en `WamGroup`, regroupés dans un "WamEnv", qui encapsule le code de la version du SDK WAM utilisé. Si une nouvelle version du SDK sort, il sera ainsi possible de charger les plugins dans un environnement `WamEnv` protégé, chaque plugin utilisant sa propre version du SDK. Chaque groupe de plugins (`WamGroup`) comprend les plugins créés par un hôte spécifique et il est possible d'envoyer des événements (MIDI par exemple) à tous les plugins à la chaîne. Un effort considérable a été déployé pour réduire le nombre d'hypothèses faites par l'API WAM concernant la mise en œuvre de l'hôte, avec `WamEnv` et les instances de `WamGroup` étant les seuls objets du système WAM qui sont fournis par l'hôte. Le `WamEnv` est alloué lors de la création du DAW et les instances de `WamGroup` sont associées à chaque piste.

5.3. Gestion des chaînes de plugins

Les chaînes de plugins sont gérées à l'aide d'un plugin WAM spécial qui agit également en tant que "mini-hôte" (Figure 7). Nous l'avons appelé le WAM pedalboard [1]. Il se connecte à un serveur de plugins qui renvoie la liste des plugins disponibles sous forme de tableau JSON d'URIs (un plugin WAM peut être chargé simplement à l'aide d'une importation dynamique et de son URI, voir [5]). À partir de cette liste d'URIs, les descripteurs de plugins WAM sont récupérés, qui contiennent des méta-données sur les plugins : nom, version, fournisseur, URI d'image miniature, etc. Lorsque le plugin pedalboard est affiché dans le DAW, la chaîne de plugins actifs est vide, et les plugins peuvent être ajoutés à la chaîne de traitement, supprimés, ré-ordonnés et leurs paramètres peuvent être définis.



Figure 6. Le plugin WAM Pedalboard qui gère les chaînes de plugins associés à chaque piste. Ici chargé dans une simple page HTML hôte, en mode stand-alone.

Toute configuration peut être enregistrée en tant que preset nommé (par exemple, "son de guitare crunch 1"). Les presets peuvent être organisés en banks ("rock", "funk", "blues"). La gestion de l'organisation et de la nomination des banks et des presets relève de la responsabilité du pedalboard plugin. Les paramètres exposés par ce plugin correspondent à l'ensemble des paramètres du preset ac-

tif (c'est-à-dire la somme des paramètres des plugins du preset), et peuvent être automatisés par le DAW. L'API Web Audio Modules permet d'envoyer des événements au DAW lorsque des modifications se produisent dans la configuration du plugin (ajout ou suppression), et un menu fourni dans le DAW avec chaque piste pour sélectionner les paramètres pouvant être automatisés est automatiquement mis à jour (Figure 7).



Figure 7. Chaque piste a un menu d'automatisation qui se met à jour en fonction des plugins associés.

Tous les plugins WAMs implémentent également des méthodes getState/setState pour sérialiser/dé-sérialiser leur état. Lorsqu'un projet est enregistré, l'état de chaque piste, des buffers audio, etc. est enregistré, ainsi que l'état de la configuration de chaque plugin.

5.4. Enregistrement, gestion de la latence

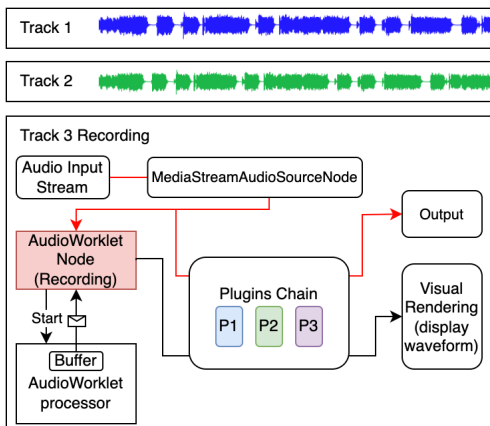


Figure 8. Schéma logique de l'enregistrement audio dans une piste de WAM-Studio

La figure 8 montre le graphe audio correspondant au processus d'enregistrement d'une piste : nous devons d'abord obtenir un flux multimédia à partir du microphone de l'utilisateur ou d'une entrée de carte son à l'aide de l'API MediaDevices. Ensuite, nous pouvons utiliser un MediaStreamSourceNode de l'API Web Audio dans le graphe audio, en passant le flux multimédia à son constructeur. C'est comme cela que nous exposons un flux audio live à l'API Web Audio. Dans un deuxième temps, nous avons besoin d'un moyen d'enregistrer ce flux dans un buffer. Cela peut être fait en utilisant l'API W3C MediaRecorder

ou en utilisant des solutions de bas niveau à l'aide d'un AudioWorklet. La solution MediaRecorder a pour avantage sa simplicité, mais tous les tests que nous avons effectués ont montré que c'est une solution peu fiable car le temps nécessaire pour allouer des buffers audio et commencer l'enregistrement est imprévisible, même lorsque nous effectuons des mesures/calibrations de latence, comme expliqué plus loin. Nous avons finalement opté pour une solution de bas niveau plus précise qui utilise un WamProcessor dont le rôle est d'enregistrer en continu les échantillons sonores dans un ring buffer contenant à peu près une seconde de son numérisé, situé dans une mémoire partagée avec un Worker JavaScript qui lui tourne dans un thread moins prioritaire et dont le rôle est de consommer les échantillons envoyés par le WamProcessor, et les copier à la fin d'un buffer qui grossit au fur et à mesure. Cette approche évite des allocations dans le thread audio, elle est également plus robuste contre un éventuel stress CPU. Le Worker envoie également sous forme de message le buffer audio courant à un code JavaScript exécuté dans le thread de la GUI. Ce dernier peut ainsi dessiner la forme d'onde du signal audio enregistré qui grandit visuellement au fur et à mesure que l'enregistrement avance. Lors qu'on stoppe l'enregistrement, la piste est prête à être jouée puisque le buffer audio a été mis à jour au fur et à mesure.

Compensation de latence : Le processus d'enregistrement avec un DAW est plus contraint et structuré que dans un simple enregistreur de mémos audio, car nous devons prendre en compte différents types de latences. La latence d'entrée est le temps entre la capture d'un signal audio par le dispositif d'entrée (carte son) et son traitement par le contexte audio. Elle dépend du système d'exploitation, de la configuration du dispositif d'entrée, du temps de traitement du système audio et de la taille du buffer utilisé par le contexte audio. La latence de sortie représente le temps entre la génération d'un son et sa perception par l'utilisateur. La somme de ces deux latences représente la latence round-trip et peut être mesurée à l'aide d'un dispositif d'enregistrement externe avec deux microphones : un sur la source du son physique (par exemple, sur le corps d'une guitare branchée sur une carte son), et un devant les haut-parleurs. Frappez le corps de la guitare, enregistrez la sortie et comparez les signaux d'entrée et de sortie. Nous avons effectué de telles mesures dans le passé[2]. On peut aussi l'estimer avec un programme émettant des sons (en général des sons de type métronome) et en enregistrant ces mêmes sons en plaçant un micro devant les haut-parleurs. Paul Adenot propose une implémentation d'un outil permettant de mesurer cette latence et aussi la latence de sortie ⁹. Le résultat est en général moins précis que lors de la mesure avec des outils externes.

Par ailleurs, l'API Web Audio expose une propriété outputLatency de l'AudioContext dont la valeur est une approximation de la latence de sortie, soit le nombre de secondes entre l'audio atteignant l'AudioDestinationNode

⁹. Voir le billet de blog de Paul Adenot <https://blog.paul.cx/post/audio-video-synchronization-with-the-web-audio-api/>

(conceptuellement, la sortie du graphe de traitement audio) et le dispositif de sortie audio. Si on sait mesurer la latence round-trip, alors on peut connaître la latence d'entrée puisque latence d'entrée + latence de sortie = latence round-trip.

Lors de l'enregistrement d'une piste de guitare, par exemple, d'autres pistes telles que des pistes de batterie et de basse peuvent être jouées simultanément. Le son de la guitare est traité en temps réel par une chaîne de plugins (simulation d'amplificateur de guitare, égaliseur, retard, etc.) et doit être entendu en temps réel par le guitariste, en synchronisation avec les autres pistes et sans latence perceptible lorsqu'il est joué (et enregistré). C'est pourquoi la latence round-trip globale doit être aussi faible que possible. Nous avons représenté la route permettant de s'entendre jouer pendant l'enregistrement avec le chemin rouge dans la Figure 8, une partie du signal traverse la chaîne de plugins et peut être entendue (*wet route*) et le signal non traité est enregistré (*dry route*, vers l'AudioWorklet). En utilisant une taille de buffer de 128 échantillons (valeur par défaut lors de la création d'un AudioContext), la latence round-trip mesurée sur MacOS est de 23 ms avec le navigateur Chrome, 15ms avec Firefox, suffisamment confortable et comparable à ce que nous pouvons obtenir avec des applications natives dans des cas d'utilisation similaires (guitare, Logic Audio, plugin de simulation d'amplificateur de guitare, taille de buffer audio de 128), même avec des guitaristes jouant très vite[2].

Néanmoins, lors de la lecture simultanée de la nouvelle piste enregistrée, un décalage de plusieurs millisecondes est visible et audible, la piste nouvellement enregistrée est "en retard" par rapport aux autres. Il est nécessaire de décaler en arrière dans le temps la nouvelle piste. Cette opération s'appelle la "compensation de latence", et la valeur de compensation est égale à *-valeur de la latence d'entrée*. Nous avons vu que cette valeur peut être calculée si on connaît la latence round-trip (puis que la latence de sortie est fournie par l'AudioContext).

Dans WAM Studio, un menu de configuration permet de calculer automatiquement la compensation de latence en approchant un micro des haut parleurs et en lançant le processus de calibration (le DAW émet des sons, enregistre le résultat et compare les temps d'émission et de réception). Pour une configuration matérielle donnée cette opération n'est à effectuer qu'une seule fois.

Le DAW AmpedStudio propose une approche similaire. Les auteurs de SoundTrap ont expliqué qu'ils utilisent plutôt des tables prédéfinies avec des entrées pour les paires OS/carte son les plus courantes, dont la latence d'entrée a été mesurée à la main et codées en dur dans le code[10]. SoundTrap utilise ensuite des heuristiques pour déterminer la configuration (vérification de l'OS, deviner le modèle de la carte son en regardant le nombre d'entrées/sorties exposées en utilisant l'API MediaDevices, etc.).

D'autres optimisations classiques ont été mises en place dans WAM-studio : dès que les pistes sont armées pour l'enregistrement, on alloue le buffer tournant, on démarre les Workers (tâches de fond JavaScript) d'enregistrement,

on pré-alloue les buffers audio et on ne commence à enregistrer que lors de l'appui sur les boutons "record" puis "play", à ce moment là, l'enregistrement à proprement parler commence. Ceci évite de perdre du temps au démarrage, avec les allocations et le lancement des Workers.

6. CONCLUSION

À ce jour, WAM-Studio est un exemple de DAW open source utilisant des fonctionnalités avancées telles que la prise en charge de plugins externes, des communications DAW/plugins à haute performance, la lecture, l'enregistrement et l'édition de pistes audio avec compensation de latence. En tant que démonstrateur, il montre les capacités de l'API Web Audio et comment le standard Web Audio Modules peut être utilisé pour faciliter le développement d'applications audio web complexes. La prise en charge des pistes MIDI est prévue à court terme ¹⁰ et le développement est actif. Le code source du DAW peut être trouvé sur GitHub ¹¹, et l'application est également disponible en ligne ¹².

7. REFERENCES

- [1] Michel Buffa, Pierre Kouyoumdjian, Quentin Beauchet, Yann Forner, and Michael Marynowic. "Making a guitar rack plugin -WebAudio Modules 2.0", *In Web Audio Conference 2022.*, Cannes, France, 2022. <https://hal.inria.fr/hal-03812948>.
- [2] Michel Buffa and Jerome Lebrun. 2017. "Real time tube guitar amplifier simulation using WebAudio" *In Proceedings of the Web Audio Conference (Queen Mary Research Online (QMRO) repository)*. <http://qmro.qmul.ac.uk/xmlui/handle/123456789/26089>. Queen Mary University of London, London, United Kingdom. <https://hal.univ-cotedazur.fr/hal-01589229>
- [3] Michel Buffa, Jerome Lebrun, Jari Kleimola, Oliver Larkin, and Stephane Letz. 2018. "Towards an open Web Audio plugin standard" *In WWW2018 - The WebConf 2018 : The Web Conference, 27th International World Wide Web Conference*. Lyon, France. <https://doi.org/10.1145/3184558.3188737>
- [4] Michel Buffa, Jerome Lebrun, Shihong Ren, Stéphane Letz, Yann Orlarey, Romain Michon, and Dominique Fober. 2020. "Emerging W3C APIs opened up commercial opportunities for computer music applications" *In The Web Conference 2020 DevTrack*. Taipei, Taiwan. <https://doi.org/10.13140/RG.2.2.16456.19202>
- [5] Michel Buffa, Shihong Ren, Owen Campbell, Jari Kleimola, Oliver Larkin, and Tom Burns. 2022. "Web Audio Modules 2.0 : An Open Web Audio Plugin

10 . Nous avons écrit des démonstrations de lecture de pistes MIDI à l'aide de WAMs sur le site de tutoriels WAM : <https://wam-examples.vidalmazuy.fr/>

11 . <https://github.com/Brotherta/wam-refont>

12 . <https://wam-openstudio.vidalmazuy.fr/>

- Standard” In *WWW '22 : The ACM Web Conference 2022*. ACM, Virtual Event, France, 364–369. <https://doi.org/10.1145/3487553.3524225>
- [6] Hongchan Choi. 2018. ”Audioworklet : the Future of Web Audio” In *Proceedings of the International Computer Music Conference*. Daegu, South Korea, 110–116.
- [7] François Déchelle, Riccardo Borghesi, Maurizio De Cecco, Enzo Maggi, Butch Rován, and Norbert Schnell. ”jMax : an environment for real-time musical applications” In *Computer Music Journal* 23, 3, 50–58. 1999.
- [8] Jari Kleimola and Oliver Larkin. ”Web Audio Modules” In *Proceedings of the Sound and Music Computing Conference 2015*.
- [9] Bojan Lazarevic and Danijela Scepanovic. ”Unrevealed Potential in Delivering Distance Courses : the Instructional Value of Audio” In *eLearning and Software for Education* (2010).
- [10] Fredrik Lind and Andrew MacPherson. ”Soundtrap : A collaborative music studio with Web Audio” In *Proceedings of the Web Audio Conference 2017*. Queen Mary University of London, London, United Kingdom.
- [11] James McCartney. ”Rethinking the computer music language : Super collider” In *Computer Music Journal* 26, 4, 61–68. 2002.
- [12] Miller Puckette. ”FTS : A real-time monitor for multi-processor music synthesis” In *Computer music journal* 15, 3, 58–67. 2002.
- [13] Shihong Ren, Stéphane Letz, Yann Orlarey, Romain Michon, Dominique Foer, et al. ”Using Faust DSL to Develop Custom, Sample Accurate DSP Code and Audio Plugins for the Web Browser”. *Journal of the Audio Engineering Society*, 68 (10), pp.703-716. 2020.

Actes des Journées d'Informatique Musicale 2023
24-26 mai 2023
Maison des Sciences de l'Homme Paris Nord

GROUPE DE TRAVAIL AFIM – MUSIQUE EN RÉSEAU

ALLÉGORIES DE L'AUDIO EN RÉSEAU

Peter Sinclair
ESAAIX (École Supérieure
d'Art d'Aix-en-Provence),
Locus Sonus Vitae
peter.sinclair@ecole-art-aix.fr

RÉSUMÉ

Locustream Open Microphone est l'un des premiers projets de cartographie sonore et à notre connaissance, le premier à être dédié à un réseau de microphones ouverts, en direct et sans traitement. Cet article propose une nouvelle approche de ce projet qui prend comme point de départ d'une réflexion plus générale sur l'art en réseau. Il porte l'attention sur la nature cyborgienne du réseau humain et machine qui caractérise l'usage mobilisé d'Internet en général, et fait appel à des analogies avec notre perception sonore pour proposer un modèle qui correspond mieux à *Locustream*, ceci dans le champ plus large de la culture en réseau. Après cet aperçu de la philosophie du projet, il fournit des détails concernant où et comment certains participants ont installé des microphones ; allant d'une île du sud de la Corée à un étang dans l'État de New York. Enfin, la dernière section traite des exemples de la diversité des créations et des usages artistiques issus du projet.

1. INTRODUCTION

Locustream Open Microphone est techniquement un ensemble d'outils – serveur de streaming, interfaces web, hardware et software de captation et encodage audio. Ce dispositif est consacré à la capture d'environnements sonores (*Soundscapes*) locaux et à leur transmission via Internet. *Locustream* est aussi un réseau de personnes qui entretiennent ces microphones et d'autres qui utilisent les sources audio qu'ils génèrent pour une diversité de réalisations [1]. Le projet existe depuis 2006, ainsi il est aussi révélateur de ce qui a changé au cours de cette période. Dans cet article, je propose d'abord une réflexion sur l'évolution de notre rapport aux réseaux et je décris les particularités de *Locustream*, avant de donner des exemples concrets d'initiatives individuelles ou collectives en lien avec ce projet.

2. L'ART EN RÉSEAU

Bien que cela puisse paraître contradictoire, puisque nous sortions depuis peu d'une période de confinement et que nous sommes encore aujourd'hui dans un esprit de

sédentarité, soit par conscience environnementale, soit à cause de restrictions budgétaires, "la mobilité" est le mot clé qui s'impose lorsque je pense à l'état actuel de l'art en réseau. J'entends par mobilité, non pas seulement le fait que nos connexions aux réseaux se sont mobilisées, mais surtout le fait, que nos vies se sont étalées dans le temps et dans l'espace, pour créer ce que l'on pourra appeler une version étendue de soi. Nos appareils (*Smartphones*), que nous transportons en permanence, participent à notre mobilité dans le sens bergsonien - ils intègrent notre faculté de diriger notre attention et par extension notre façon de percevoir le monde. Ils font partie de ce qui nous permet de « sonder » le monde et par la même il participe de notre libre arbitre.

À mesure que mon corps se déplace dans l'espace, toutes les autres images varient ; celle-ci au contraire, demeure invariable. Je dois donc bien en faire un centre, auquel je rapporterai toutes les autres images [2].

Cela crée un tel changement dans le paradigme spatial que, sans doute, rien n'est exempt de l'influence de cette évolution. Il y a quelques années, Boris Beaudé proposa qu'Internet eussent transformé nos spatialités en leur ouvrant de nouvelles dimensions géographiques [3]. Pour certains d'entre nous, la première génération du net art, il a fallu un temps pour intégrer cette approche, il y avait encore une distinction entre espaces "virtuel" et "réel". Pour la génération actuelle (nos étudiants), il est inconcevable de séparer sa vie connectée de celle non connectée, ou bien, l'idée de se déconnecter est une position de résistance, elle devient une utopie exotique et bohème. J'ai proposé récemment à un groupe d'étudiant·e·s qu'il·elle·s étaient tou·te·s des *Cyborgs* dans le sens que donne Donna Haraway à ce terme : « ... les distinctions entre naturel et artificiel, corps et esprit, autodéveloppement et création externe et tant d'autres qui permettaient d'opposer les organismes aux machines, sont devenues très vagues » [4]. Cette proposition a été reçue avec un acquiescement général, plutôt rare chez les étudiant·e·s en école d'art. Si nous avons tous des téléphones *portables* depuis longtemps, les jeunes personnes de la génération actuelle, pour la plupart, ont grandi avec un smartphone. Il·elle·s ne savent pas ce qui est une vie sans connexion et n'imaginent pas survivre

sans cela. Effectivement, l'idée qu'il·elle·s construisent leur identité propre, sur une sémiotique formellement imposée par les réseaux de communication dont chacune serait à la fois le centre et l'auteur, est plutôt une évidence. Le centre organisationnel de toute activité professionnelle et sociale – le bureau, la documentation, la bibliothèque, la mémoire, les outils audio et visuel – se tient dans la main.

Il existe au moins deux façons d'envisager ces nouvelles coutumes, cette nouvelle façon d'être au monde : soit, l'on considère qu'on va sur le réseau (comme le proposait Boris Beaudé en 2012) « Internet est un espace doté de propriétés spécifiques, qui participent des spatialités singulières qui s'y déploient. » [3]p51. Soit, on peut considérer que chacune est au centre d'un univers parfaitement enchevêtré, mais fait d'une multitude de partialités dont la configuration particulière est créée à partir de son point de vue.

3. L'ÉCOUTE

Cependant, la métaphore du « point de vue », utilisée si communément pour placer un soi au centre de l'univers, ne me semble pas bien choisie pour décrire ce contexte, et cela me permet de revenir à la question de l'espace sonore qui nous intéresse ici. Les sons que nous produisons nous-mêmes partagent indistinctement l'espace avec ceux qui émanent d'autres individus et objets, ainsi, dans l'univers sonore notre corps ne représente pas une véritable limite. On ne peut plus concevoir notre corps comme une intériorité détachée du monde. La nature de l'onde sonore et la façon dont nous la percevons signifient que le son que nous entendons, qu'il soit naturel, enregistré ou transmis (*schizophonique*, selon le vocabulaire de Murray Schafer [5]), constitue notre environnement. Pour le dire autrement, s'il y a une radio allumée dans la pièce, le son de la radio se mêle aux autres ondes de l'espace et même de l'extérieur de l'espace dans la même masse d'air vibrant, il fait partie de l'environnement sonore. Une image (sans son) exige que vous le regardiez délibérément (que vous tourniez votre regard vers elle), afin qu'elle participe à votre perception. Selon le sociologue Michael Bull, l'écoute au casque peut être si puissante et immersive, parce que la musique que vous écoutez devient la véritable bande sonore de votre vie [6]. Mais il est aussi intéressant à noter, Jean Paul Thibaud le souligne [7], qu'une autre façon de pratiquer l'écoute au casque consiste à ajuster soigneusement le niveau de votre appareil pour créer un espace sonore hybride, composé en partie du média que vous avez choisi et en partie de l'environnement naturel qui filtre à travers vos oreilles – un précurseur de la réalité augmentée, puisque cela commence avec le walkman dès les années 1980. Ainsi, l'écoute en réseau et son intégration avec nos espaces de multiples partialités, se fait aujourd'hui de façon intuitive et peut représenter une alternative à la culture du *selfie* pour envisager le modèle du « soi au centre », car construite sur une omnidirectionnalité et une acceptation de cohabitation. Je voudrais ainsi décrire notre dispositif de microphones ouverts *Locustream*, avec comme point

de référence, ce modèle de multiples mobilités qui constituent les espaces sonores.

4. LOCUSTREAM

Le dispositif *Locustream* est une ressource partagée, ouverte et en évolution permanente. La résilience et la fragilité sont deux caractéristiques qui viennent à l'esprit pour décrire ce réseau de microphones ouverts. L'infrastructure n'a pas été planifiée et ne peut pas être considérée comme définitive, car elle peut à tout moment s'effondrer comme d'autres espèces de réseau qui se sont éteintes (*MySpace* par exemple), mais dans ce cas précis, les flux audio (qui ne sont pas enregistrés) disparaîtront instantanément. Tout comme les sources sonores qu'ils reconstituent, le contenu de *Locustream* n'existe qu'en direct. Voilà pourquoi, tout en étant substantiel en ce qui concerne sa mise en œuvre (serveur, logiciel, équipements), ce projet est à l'opposé d'une sculpture monumentale ou de n'importe quel "grande œuvre" imposant. Certains microphones sont permanents, mais il s'agit d'une permanence particulière, parce que, de fait, les flux sonores sont de nature volatile et variable. Ils fluctuent également avec l'engagement des personnes qui les anime. Si certains *streamers* sont d'une constance extraordinaire, d'autres vont et viennent avec les saisons, les projets collectifs ou leurs propres projets. *Locustream* est une nébuleuse sans limites bien définies et sans intention unique – chacun·e trouvant son intention dans l'attention qu'il·elle porte au projet. Il est sans auteur proprement parlant, mais avec beaucoup de participants qui trouvent leur compte dans ces intentions et ces attentions hétéroclites. Il est non localisé et en même temps il est fortement attaché aux territoires qu'il représente.

Il est possible que certaines caractéristiques ici décrites soient aussi assimilables aux réseaux sociaux que nous connaissons tous. En effet, je le disais au début, nous sommes tous concernés par cette extension de nos espaces de captation et de représentation orchestrées par les GAFAM. Cependant, outre les différences évidentes d'envergure et de commercialité, *Locustream* réclame d'autres formes d'engagement que ces derniers. Ce sont celles que chacun·e s'invente, soit en tant que *streamer* qui prend soin d'un microphone en choisissant son emplacement et en faisant attention à son bon fonctionnement, soit en tant qu'utilisateur où l'effort artistique est largement déplacé vers l'individu écoutant. Dans tous les cas, il·elle·s participent de façon active, à ce que John Cage appelle la musique expérimentale – une musique que l'auteur découvre au moment où elle se déroule et dans une spatialité qui fait que l'expérience de chacun·e est unique.

Locustream est une œuvre ouverte, un projet de recherche participatif et un dispositif pluridisciplinaire. Les projets et aspirations qui motivent cette participation sont nombreux et divers, les exemples suivants offrent un rapide aperçu de la diversité des approches, commençant par la posé des microphones et poursuivant avec des utilisations des flux.

5. MICROPHONES OUVERTS

Les descriptions suivantes sont extraites des questionnaires que nous avons adressés aux opérateurs des microphones dans le cadre du projet « le monde aux creux de l’Oreille » [8].

5.1 Georo village on Jeju Island, Région: South Korea, Operator: Culture Space Yang (Yeonjoo Kim), Weather Report (Gangil Yi, Jiyeon Kim).

... Jeju is the largest island in South Korea. Despite its beautiful nature, Jeju has a bloody and tragic history. Many innocent Jeju people were sacrificed during the armed conflicts and suppressions following the Armed Uprising of the Jeju Chapter of the South Korean Labor Party.

... This streaming project hopes to symbolically convey the unheard voices and memories of village people through the transmission of this tranquil ecological soundscape. The stereo microphone is situated on the wall of the traditional Jeju stone house that accommodates Culture Space Yang. Usually bird calls, glass bugs, the whistling of trees, and occasionally visitor’s conversations are heard. Distant traffic and sometimes foghorns of the big ships from the nearby port are also audible. Local time is about Marseille Time plus 8 hours.

5.2 Pond Station, Location: New York State, USA, Operator: Zach Poff (Wave Farm Transmission Arts)

Two extremely sensitive hydrophones record underwater sounds of this pond. Decomposing plant matter creates a background chorus of bubbling gas. Insect, fish and frog vocalizations continue all day, combined with the traces of airborne sounds that excite the water too (like passing cars, singing birds, insects or humans). Rain on the pond creates a dense cloud of high-frequency detail, like the coals in a cooling campfire. In the winter, life in the pond slows down but there is a surprising amount of activity even when the surface is frozen.

Operator Zach Poff (Wave Farm Transmission Arts) is a media artist, educator, and maker-of-things. He often uses sound and listening to explore the web of social and ecological relationships that challenge and sustain us.

5.3 Kolkata- Chittaranjan Colony, Region: Kolkata (Calcutta) India, Operator: Sukanta Majumdar

Chittaranjan Colony, a residential area of South Kolkata is one of the biggest metropolises of India (14.1 million residents). The stereo microphone is situated under a corrugated tin shed on the roof of my flat. The houses are closely built in this part of the city and because of the tropical weather the windows and doors are open all the time. Also, this house is located beside the main

road of this neighbourhood, so human voices, sparse traffic and honking are almost always present. There is a pond on one side of the house, and a playground on the other side. Beside the playground there is a temple, where they sing in the morning and in the evening. You can hear kids playing football and cricket in the afternoon. Some beautiful old trees are still surviving around the pond. Lots of different types of birds live on those trees.

6. DÉPLOIEMENT DES STREAMS

Cette section décrit trois projets, parmi beaucoup qui existent et qui ont existé depuis le début du dispositif, il s’agit de propositions très différentes présentées avec les mots des auteurs.

6.1 New Atlantis

L’objectif de New Atlantis est d’interroger la possibilité d’utiliser les mondes virtuels partagé en réseau comme espace de création, rencontre et présentation de musiques expérimentales et arts sonores. Depuis que nous avons porté le projet en web audio, nous avons pu intégrer l’utilisation de *Locustream* qui ouvre une nouvelle voie d’exploration - notamment pour Ludmila Postel qui a fondé ses recherches doctorales sur ce projet.

3 Mai 2020 : “Je suis devant mon ordinateur, assise à mon bureau, casque vissé sur les oreilles, j’attends le feu vert des filles. Sur l’écran de droite, je vois le déroulement du streaming qu’elles envoient en ligne, ça me permet de visualiser le délai entre le son que je produis et son arrivée à leurs oreilles. Messenger sonne : “c’est bon, on commence. - Ok !”. C’est parti. Je commence à déambuler dans New Atlantis parmi des sons issus d’un flux de streaming en direct de la forêt de Kyoto et un autre de la circulation à Londres. J’entends le synthé de Crys qui augmente progressivement. Au bout de quelques minutes, Loli commence à répondre à mes chants d’oiseaux avec son synthé. Elle déclenche sa boîte à rythmes et Aurélia lance sa voix sur la piste. Je me retire quelques instants et les écoute [9].”

6.2 Deepscape, transversale

Dans le cadre de ces recherches artistiques et scientifiques, Hugo Scurto – explore les limites esthétiques de l’apprentissage profond en créant un modèle de *Soundscape* à partir des microphones ouverts de *Locustream*. Plutôt que de chercher à imiter les motifs des formes musicales, ce projet propose un générateur de *Soundscape* qui aurait puisé son inspiration dans un ensemble global des microphones ouverts. Le résultat est un mélange étonnant de profondeurs de champ sonores qui ressemble, sans jamais représenter, des espaces et des phénomènes naturels. Le son ainsi généré deviendra à son tour un nouveau flux proposé dans le cadre du projet *Locustream*.

Pour réaliser cette création, j'ai collecté à travers la planète plusieurs heures de paysages sonores via la plateforme en ligne Locustream. J'ai ensuite entraîné un réseau neuronal profond sur la base de ces données audio stéréophoniques. Le paysage computationnel résultant de cet apprentissage profond, généré en flux continu par le réseau neuronal profond, compresse et normalise les biophonies, géophonies et anthropophonies de la planète. À travers l'écoute de ce paysage, je cherche à révéler le processus de terraformation de l'intelligence artificielle (IA), dont les infrastructures informatiques, à travers les flux de médias qu'elles génèrent à l'échelle du globe, et les ressources matérielles, humaines et culturelles qu'elles exploitent intensivement, transforment nos environnements, et détournent notre attention de la planète [10].

6.3 Sound Camp Reveil

Sound Camp, Reveil consiste en un tour du monde des microphones suivant au plus près le lever du soleil et le *dawn-chorus* (chœur de l'aube) du chant des oiseaux de début mai. Le tour est mixé et proposé sous la forme d'une veille de 24h par le groupe d'artistes basées à Londres du même nom (SoundCamp). Le suivant est un extrait de mon propre carnet de notes prises à l'issue de l'expérience de l'édition 2022.

À moitié écoutant, à moitié somnolant allongée dans mon lit à 5h du matin je suis aux aguets - le voyage avance de l'est, vers chez moi et j'attends les streams que je connais (et que j'aime, si l'on peut appeler amour cet étrange attachement que nous développons pour les choses dont on s'occupe).

Parfois on entend les mouvements discrets d'une personne derrière les microphones, mais contrairement à la plupart des productions audiovisuelles où ce serait un retour à la réalité (ou à l'incrédulité) malvenus, dans ce cas, savoir que celui ou celle qui s'occupe du microphone est là et qu'il ou elle partage l'écoute avec nous me fait plaisir. Je connais beaucoup de personnes, à travers le monde qui sont derrière les micros que nous écoutons, j'en ai rencontré en personne, au cours des voyages et il y a d'autres avec qui je correspond, pour les besoins du projet. Mais, même avec ceux que je ne connais pas, je ressens une sorte de fraternité à travers l'objet et l'objectif partagé dans ce tour du monde. ... Réveil n'est pas non plus dépourvu de prises de position - le mélange de conduits de climatisation et d'oiseaux qui s'éveillent lorsque nous passons devant la bibliothèque de Scarborough est une proposition esthétique fort différent de celle du microphone que j'ai installé dans le jardin de ma mère - pourtant géographiquement assez proche. Ces différentes positions auditives représentent également des visions du monde, très ouvertes à l'interprétation, mais des positions quand même. ... Réveil n°9 touches à sa fin - je ressens un mélange de sentiments, une sorte de chaleur d'avoir partagé un moment, et en même temps un soulagement un peu comme le lendemain de Noël.

7. CONCLUSIONS

Cet article présente *Locustream Open microphone Project*, comme une forme d'art en réseau qui, par son lien à l'espace sonore et les approches que cela peut générer ; par sa forme mouvante - en évolution permanente et non intentionnelle ; par son aspect essentiellement participatif et horizontal, représente un modèle en adéquation avec les aspirations d'une génération cyborg concernées à la fois par leur position individuelle, au centre d'un monde hybride, mais également par des questions d'attention écologique et environnementale.

8. RÉFÉRENCES

1. Sinclair, Peter. « Locustream Open Microphone Project ». In *Proceedings of the 2018 International Computer Music Conference*, 271-75. Daegu Korea: The International Computer Music Association 1819 Polk Street San Francisco, California 94019 USA, 2018.
2. Bergson, Henri, Denis Forest, et Paul-Antoine Miquel. *Matière et mémoire essai sur la relation du corps à l'esprit*. Paris: Flammarion, 2012. <http://banq.pretnumerique.ca/accueil/isbn/9782081280816>.
3. Beaudé, Boris. *Internet, changer l'espace, changer la société: les logiques contemporaines de synchronisation*. Collection « Société de la connaissance ». Limoges, France: Éditions Fyp, 2012.
4. Haraway, Donna Jeanne, Laurence Allard, et Delphine Gardey. *Manifeste cyborg et autres essais: sciences, fictions, féminismes*. Essais. Paris: Exils, 2007.
5. Schafer, R. Murray. *The soundscape: our sonic environment and the tuning of the world*. Rochester, Vt. : [United States]: Destiny Books ; Distributed to the book trade in the United States by American International Distribution Corp, 1993.
6. Bull, Michael. *Sound moves: iPod culture and urban experience*. International library of sociology. London ; New York: Routledge, 2007.
7. Thibaud, Jean-Paul. « Les mobilisations de l'auditeur-baladeur : une sociabilité publique ». *Réseaux* 12, n° 65 (1994): 71-83. <https://doi.org/10.3406/reso.1994.2509>.
8. Gasc, Amandine. Sinclair, Peter. « Des mondes au creux de l'oreille - Observatoire des Sciences de l'Univers - Institut Pythéas ». Consulté le 20 février 2023. <https://www.osupytheas.fr/?Des-mondes-au-creux-de-l-oreille-2244>.
9. Postel, Ludmila. « Arts du son et jeux vidéo : la recherche-création dans New Atlantis ». Thèse de doctorat, Aix-Marseille, 2022. <https://www.theses.fr/2022AIXM0089>.
10. Scurto, Hugo. Article soumis à la publication. 2023.

JIM 2023 : NEW ATLANTIS, MÉTAVERS DU SONORE

Ludmila Postel

ESAAix, Locus Sonus Locus Vitae
ludmilapostel@gmail.com

Jonathan Tanant

Développeur indépendant
jonathan.tanant@gmail.com

RÉSUMÉ

New Atlantis un métavers de la création sonore, issue du groupe de recherche Locus Sonus Locus Vitae, de l'École Supérieure d'Art d'Aix-en-Provence. Elle s'intègre, depuis une quinzaine d'années, dans un projet de recherche aux multiples aspects — pédagogiques, techniques et artistiques — centré sur l'écoute et la création sonore en ligne. L'un des principaux objets du projet est le développement d'univers virtuels, en ligne et multijoueurs, dédié à l'expérimentation et à la création sonore. *New Atlantis* est une plateforme dédiée à l'exploration des relations entre son, image 3D et interactivité, destinée aux artistes sonores, musiciens, chercheurs et étudiants. Le nom de *New Atlantis* vient du titre d'une nouvelle utopique de 1627 du philosophe anglais Francis Bacon, considéré comme le père de l'empirisme moderne. Il y décrit une société imaginaire, sur l'île de Bensalem quelque part dans l'océan Atlantique, qui place l'éducation et le savoir au cœur de la culture, favorisant la recherche scientifique par le faire et l'observation. L'île est traversée de phénomènes et de machines sonores qui prédisent les techniques électroniques et numériques actuelles. Inspiré par l'ouvrage de Bacon, *New Atlantis* met l'accent sur la découverte, l'échange culturel, l'expérimentation, l'apprentissage et l'approfondissement des connaissances dans un environnement éducatif et créatif.

Ludmila Postel et Jonathan Tanant travaillent dans l'écosystème de *New Atlantis* depuis environ 6 ans, avec, entre autre, la production d'une thèse de recherche-crédation sur la plateforme et ces usages. Actuellement, ils formalisent les recherches menées depuis 15 ans par les différents partenaires du projet afin de valoriser ces travaux avec un outil accessible. Lors des journées d'informatiques musicales 2023, nous proposons de faire une démo de *New Atlantis*, avec un regard particulier sur

les relations avec le projet Open Microphones de Locus Sonus Locus Vitae, qui sera présentée par Peter Sinclair.

1. RÉFÉRENCES

1. Postel, L. « Penser les pratiques participatives en réseau », *Mézoaires n°1, actes du séminaire La recherche artistique participative : Nouvelles formes de recherches artistiques, l'expérimentation en réseau et le cyber activisme*, Les presses du réel, à venir courant 2023.
2. Postel, L. « Son et mondes virtuels, l'exemple du projet *New Atlantis* », Actes du colloque Rencontres Nationales de Recherches en Musique, sous la direction du Ministère de la Culture et de l'université Paris 1 Panthéon-Sorbonne, janvier 2023.
3. Postel, L. *Arts du son et jeux vidéo, la recherche création dans New Atlantis*, Thèse non-publiée, Aix-Marseille université, URM PRISM-CNRS, ESAAix, soutenue le 26 mars 2022.
4. Postel, L. « A l'écoute des jeux vidéo, Imaginer l'espace virtuel par le son », *actes du colloque Espaces Imaginés*, sous la direction du Laboratoire des Imaginaires et de l'université de Rennes 2, janvier 2021.
5. Postel, L. « *New Atlantis* », *Culture et Recherche n°141 Cinéma, audiovisuel, son*, printemps-été 2020.
6. Sinclair, P. Cahen R. Tanant, J. Gena, P. « *New Atlantis: Audio Experimentation in a Shared Online World* », *CMMR. Bridging People and Sound: 12th International Symposium, CMMR 2016*, São Paulo, Brazil, 2017.
7. Joy, J. Sinclair, P. *Locus Sonus, 10 ans d'expérimentation en art sonore*, Le Mot et le Reste, Marseille, 2015.

ART & GAMES IN METAVERSE

Frédéric Curien

EESI
f.curien@eesi.eu

Hervé Jolly

EESI
h.jolly@eesi.eu

Axel Buendia

CNAM-ENJMIN
axel.buendia@lecnam.net

Introduction

L'atelier *Art & Games* est une expérience pédagogique menée par l'École nationale des jeux et médias interactifs numériques (Cnam-Enjmin) et l'École européenne supérieure de l'image Angoulême-Poitiers (EESI). Elle réunit depuis dix ans des équipes mixtes d'étudiants et de professeurs des deux établissements en vue de créer un espace de création et de réflexion communes autour des interactions entre jeux vidéo et arts plastiques. Il a été initié à l'origine pour promouvoir une rencontre entre deux pédagogies aux méthodes et finalités différentes, parfois opposées dans leur approche du processus de création : mené en équipe pour le Cnam-Enjmin, dans une logique d'auteur individuel pour l'EESI. L'atelier a permis notamment, au cours des différentes sessions, de mettre en exergue sur le plan conceptuel le concept d'espace jouable. Cette modalité d'accès proposée au spectateur-joueur, partagée par tous les arts interactifs, a unifié les pratiques autour d'un corpus commun d'œuvres artistiques et de jeux vidéo. Par des pratiques d'hybridation hétérogène, cette approche singulière des médias interactifs a fait sortir de leur zone de confort des étudiants parfois trop captifs de leurs cursus respectifs. Ce processus de "déterritorialisation disciplinaire" s'est effectué sur plusieurs années avec le concours de chercheurs, d'artistes et de professionnels des secteurs de l'art et du jeu vidéo, comme Éric Viennot, attentifs aux évolutions technologiques et esthétiques qui traversent et impactent aujourd'hui le monde des images.

Cette année, le thème de l'atelier est le métavers, conçu comme un espace de création, un espace d'expérimentation, un réseau social, mais également un jeu. Il s'agit principalement pour les concepteurs des jeux de créer des univers, de penser des sociétés et concevoir des interactions originales à l'intérieur de ceux-ci. L'objectif premier est de créer une expérience sociale et persistante. Pour y parvenir, l'équipe, constituée de programmeurs, d'ergonomes, de chefs de projet, de graphistes et de sound designers, se doit de penser un concept du jeu qui lui soit propre et qui lui

permette de construire le dispositif adéquat incluant les multiples joueurs-spectateurs situés en réseau.

La question technologique inédite et particulièrement ardue qui en découle pour les étudiants se pose ainsi : comment fabriquer un métavers ? Comment choisir un périmètre raisonnable et créer une bulle de réalité indépendante ?

CONSIDÉRATIONS TECHNIQUES

Pour permettre un déroulement fluide de l'atelier, il était primordial de préparer un modèle utilisable par tous et toutes, faisant abstraction des complexités techniques, notamment celles liées au réseau. Les enjeux techniques et les difficultés technologiques ont donc été anticipés en avance de phase.

Le premier problème concerne la création d'applications interactives en temps réel. Pour améliorer l'efficacité des équipes, nous avons privilégié deux moteurs de jeux existants, ayant fait leurs preuves : UNITY¹ et UNREAL². Ces deux moteurs sont très couramment utilisés, possèdent une documentation conséquente et une communauté importante. L'apprentissage de ces systèmes représente également un atout supplémentaire intéressant, que ce soit pour des artistes ou bien évidemment pour des producteurs de jeux vidéo. Il aurait pu être intéressant de s'intéresser également au web et au moteur de jeu multiplateforme GODOT³, ce que nous ferons sûrement dans un avenir proche.

La différence technique entre une application interactive multi-utilisateur et un métavers réside dans le nombre potentiel d'utilisateurs bien plus important dans le cas du métavers. Les applications temps réel en réseau fonctionnent en général autour d'un seul serveur, pouvant gérer et cela est assez variable en fonction de la quantité des données synchronisées, une cinquantaine d'utilisateurs simultanés. Pour un métavers, il est important de pouvoir passer cette limite. Il convient

¹ <https://unity.com/>

² <https://www.unrealengine.com/>

³ <https://godotengine.org/features/>

alors de mettre plusieurs serveurs pour gérer toutes les connexions. Se pose alors un nouveau problème, comment synchroniser les serveurs entre eux, permettant ainsi un passage à l'échelle, avec autant d'utilisateurs et donc de serveurs que nécessaires. Une veille des technologies accessibles a été effectuée pendant plus d'une année et a permis de lister plusieurs solutions potentielles, en fonction du moteur de jeu utilisé pour réaliser l'application. Finalement, deux technologies ont été sélectionnées : SmartFox⁴ pour UNITY et Stormancer⁵ pour UNREAL. Ces deux technologies permettent aux serveurs le passage à l'échelle, grâce à un système d'autoréplication lorsque cela est nécessaire (trop d'utilisateurs), et une synchronisation automatique entre les duplicatas du serveur. Il est intéressant de noter que pendant notre veille, aucune solution open source n'était déjà disponible ou suffisamment mature.

Pour bien préparer le terrain, et mettre en évidence la faisabilité de la réalisation d'un métavers en seulement quatre jours, deux projets simples ont été réalisés en amont. Ces deux projets réalisés, ainsi qu'une documentation, ont permis aux équipes de rapidement développer leurs propres projets en seulement quatre jours. Il est cependant intéressant de noter que même si tous les projets ont abouti à une version jouable, environ la moitié avait une version serveur. Les autres projets ayant encore des bugs, souvent mineurs, au moment du rendu le dernier jour.

CONTEXTE ET HISTOIRE

Pour embarquer dans cette approche croisée des métavers, il a été nécessaire de considérer une petite histoire possible des pratiques et réalisations des cinquante dernières années qui ont pu mener aux propositions actuelles. Si l'on trouve un certain consensus dans un premier usage correspondant à la forme contemporaine de Metaverse/Métavers dans le roman de science-fiction *Snow Crash / Le Samourai virtuel*⁶, on peut observer des pratiques non-nommées, ou nommées différemment, depuis les années soixante-dix.

Parmi les premières expériences vidéoludiques en réseau notables, *Maze War*⁷ proposait déjà des notions d'avatar et d'un espace commun synchronisé multi-joueurs et joueuses. Le système réseau PLATO⁸ qui se développait aux Etats-Unis a permis d'une part l'émergence d'usages toujours contemporains (forums, email, message board, chat, etc.), et d'autre part de

formes de jeu en réseau plus avancées. Par exemple *Empire*⁹, une simulation spatiale très inspirée de *Star Trek* permettant à une trentaine de joueurs et joueuses de partager un espace commun, ou *Dnd*¹⁰, une adaptation informatique du jeu de rôle sur table *Dungeons & Dragons* qui sortait la même année, mettant simultanément en jeu une équipe d'aventuriers et aventurières. Peut-être peut-on voir ce dernier comme un prélude à *Oubliette*¹¹, *Moria*¹² et encore plus *Avatar*¹³, un Multi-User Dungeon (M.U.D.) permettant à une soixantaine de personnes de partager une aventure, avec la particularité d'une écriture nécessitant la constitution de groupes. Cela en ferait-il d'abord une aventure sociale ?

Ce côté de mise en place d'une société se retrouve dans *Habitat*¹⁴ sur Commodore 64, un environnement en ligne multi-utilisateurs et utilisatrices, avec une représentation graphique poussée pour l'époque. Ce jeu reste une référence historique pour les communautés en ligne contemporaines, non seulement pour l'accent mis sur la composante visuelle, mais aussi pour le point fort donné aux interactions entre les participants et participantes. La communauté est responsable des lois en usage et des comportements acceptables : "The essential lesson that we have abstracted from our experiences with Habitat is that a cyberspace is defined more by the interactions among the actors within it than by the technology with which it is implemented."¹⁵

Le rapport au législatif et au socialement acceptable est assez vite devenu un des enjeux des jeux de rôles massivement multi-joueurs et joueuses (MMORPG), comme les manifestations devant le château de Lord British dans *Ultima Online*¹⁶, ou l'organisation des factions de *Eve Online*¹⁷.

En France, si l'on trouve des expériences très intéressantes de mondes persistants multi-utilisateurs et utilisatrices assez tôt, comme le jeu de combat de

⁴ <https://www.smartfoxserver.com/>

⁵ <https://www.stormancer.com/product.html>

⁶ Neal Stephenson, *Snow Crash*, Bantam Books, New York, 1992.

⁷ Steve Colley, Greg Thompson, & Howard Palmer, logiciel autoédité et diffusé librement, 1973.

⁸ *Programmed Logic for Automatic Teaching Operations*, un système centralisé de temps partagé sur un serveur dédié à l'éducation dès les années soixante.

⁹ John Daleske & Silas Warner, logiciel auto édité et diffusé librement, 1974.

¹⁰ Gary Whisenhunt & Ray Wood, logiciel autoédité et diffusé librement, 1974.

¹¹ Jim Schwaiger, logiciel autoédité et diffusé librement, 1977.

¹² Kevet Duncombe & Jim Battin, logiciel autoédité et diffusé librement, 1977.

¹³ Bruce Maggs, Andrew Shapira & David Sides, logiciel auto édité et diffusé librement, 1979.

¹⁴ Randy Farmer & Chip Morningstar, Quantum Link - Fujitsu, 1985.

¹⁵ Morningstar, C. & F. R. Farmer, "The Lessons of Lucasfilm's Habitat", *The First International Conference on Cyberspace*, Austin, 1990. Une version électronique existe à

https://web.stanford.edu/class/history34q/readings/Virtua_Worlds/LucasfilmHabitat.html .

¹⁶ Richard Garriott & Raph Koster, Origin System, 1997.

¹⁷ Crowd Control Productions, 2003.

voitures post-apocalyptique *Carcrash*¹⁸, c'est l'expérience du *Deuxième monde*¹⁹ qui se pose en jalon notable. Le choix est fait d'une récréation homothétique de rues, places et architectures parisiennes, d'une liaison à une régie publicitaire et d'interactions basées notamment sur l'achat d'emplacements virtuels. Cette relation économique au terrain se retrouvera plus tard dans *Second Life*²⁰, où l'on paye pour obtenir un terrain virtuel qui est à la fois un emplacement mémoire sur le serveur, mais aussi une partie de sa puissance de calcul, ce qui implique que suivant la taille du lieu on pourra manipuler plus ou moins d'entités de construction.

À partir de ces quelques exemples anciens, il nous a paru important de proposer aux étudiants et étudiantes de se concentrer sur l'aspect social et persistant des métavers, pour essayer d'imaginer et construire ce que seraient les enjeux et pratiques des métavers à venir. Il y a peut-être une responsabilité à entreprendre dès maintenant des propositions alternatives qui mettent en avant des interactions interpersonnelles qui ne sont pas nécessairement marchandes ou violentes.

RÉSULTATS

Aspects techniques

Sur les dix projets proposés, tous proposaient une expérience interactive jouable. Seulement deux projets n'ont pas réussi à faire fonctionner la partie réseau. Cela tend à valider notre approche via des modèles de jeux préexistants.

Sur les dix projets, un seul a utilisé une technologie web indépendante de nos modèles. Le projet, bien que social et persistant, est basé sur des interactions asynchrones avec un chat et une interface de gestion commune, ce qui a rendu possible l'utilisation de cette technologie (base de données, requêtes HTTP simples). Deux projets ont été réalisés sur UNREAL, et sept sur UNITY. Le modèle alors mis en place sur UNITY était plus stable que celui sous UNREAL, ce qui a poussé les étudiants vers UNITY.

L'aspect de passage à l'échelle n'a pu être testé. Pour des raisons de politique de sécurité réseau, et malgré plusieurs semaines de délai, nous ne sommes parvenus à ouvrir le système sur l'extérieur que pendant l'atelier. Il n'a donc pas été possible de tester les expériences avec beaucoup d'utilisateurs.

Jouer à construire un espace social

Dans les productions étudiantes, on a pu observer entre autres que le sens, ou la cohérence, que l'on pense lire dans l'activation de ces espaces persistants en ligne

dépendait en bonne partie des interactions entre les joueurs et joueuses. Prenons pour exemple la proposition nommée *In Situ* par l'un des groupes²¹ : les joueurs et joueuses se retrouvent dans un monde ouvert tridimensionnel, urbain, de couleur blanche, avec la possibilité d'apposer sur le sol, les bâtiments, le mobilier ou la flore des motifs de couleur, des lettres, dessins, symboles, de les combiner ou d'en changer taille et orientation, le tout de manière persistante dans le temps.

Ces fresques, grafs, tags, inscriptions, repères, indices, signes, font certes référence à leur équivalent dans le monde réel, mais sont aussi issus de pratiques qui ont émergé en marge de jeux de tir multijoueurs en ligne et sont devenues habituelles. Dans *Counter Strike*²², alors que l'objectif exprimé est un affrontement à mort entre équipes, la possibilité d'apposer des symboles sur les murs, permettant une communication visuelle à l'intérieur du jeu, a permis la proposition artistique et militante *Velvet-Strike*²³. La gestion des motifs est améliorée et enrichie de peinture, et le groupe d'artistes performe de la peinture au milieu des autres joueurs et joueuses armes en main. Ce qui se joue alors n'est plus de combattre, mais de dénoncer de manière picturale l'idéologie militariste du jeu initial.

La persistance des motifs peints sur le serveur laisse peut-être également la possibilité d'une lecture épigraphique par les joueurs dans le futur. Déchiffrer ces inscriptions, tenter d'y lire un sens, interpréter ou sur-interpréter, compléter, laisser de nouvelles traces pour d'autres, autant d'actions qui montrent que l'aire de jeu d'*In Situ* n'est pas seulement spatiale, mais aussi temporelle.

C'est peut-être un des enjeux que nous avons pu apprendre grâce aux expériences de ce workshop : les interactions définissent le sens de l'espace en ligne, et la persistance permet que ces interactions se développent dans la dimension temporelle, et non plus simplement de manière synchrone. L'ensemble des expériences en métavers, ce que le passé nous laisse et ce que nous laissons au futur, cela nous invite peut-être à ne plus considérer uniquement le *hic et nunc*, ici et maintenant, mais *Everything Everywhere All at Once*²⁴, tout, partout, tout à la fois.

Nous finirons cet exposé sur l'utilisation des espaces sonores et musicaux dans cette expérience. Plusieurs équipes ont utilisé cette dimension pour accompagner et construire leurs espaces narratifs, par le recours à l'accompagnement musical ou simplement en renforçant le réalisme de leur environnement par l'ajout de bruitages et de sons de synthèse. Un jeu en particulier intitulé *Fruit tribal*, présente une approche intéressante

¹⁸ Serveur Minitel JEST de la revue Jeux & Stratégie, 1987.

¹⁹ Canal+ Multimedia & Cryo Interactive, 1997.

²⁰ Linden Lab, 2003.

²¹ Maud Larcher, Juliette Mourot, Silohé Malingre, Solène Cheyns, Rhea Chatziargyrou, Marc Cerutti, Camille Huynh, Johanna Birraldacci, Anthony Dannis.

²² Valve, 1999-2000.

²³ Anne-Marie Schleiner, Brody Condon, Joan Leandre, 2002, <https://anthology.rhizome.org/velvet-strike>.

²⁴ Daniel Kwan & Daniel Scheinert, A24, 2022

de la relation son-image, par l'utilisation du geste instrumental, avec un Gameplay basé principalement sur la communication et la coopération sonores entre utilisateurs. Et ce, à partir d'un orchestre d'objets-instruments, en l'occurrence des fruits exotiques, intégrés au décor que les participants pouvaient découvrir au gré de leur déambulation et récolter. Ils étaient invités de manière détournée à faire sonner de manière rythmique et mélodique ces fruits par les gestes de coupe. Cette action telle une ritournelle collective, appelant d'autres joueurs à les accompagner dans leur récolte sonore, était susceptible d'évoluer et se transformer en un véritable orchestre de joueurs-improvisateurs. Ce jeu aussi très visuel, qui a mis fortement à contribution le sound designer de l'équipe, a démontré que le sonore et le visuel peuvent trouver là dans cet espace collaboratif, un potentiel de développement tout à fait prometteur.

HERMES V2 : CREATION MUSICALE EN RESEAU

João Svidzinski
CICM-EA 1572
Université Paris 8 et IRCAM
svidzinski@gmail.com

Robin Montferme
Université Bordeaux, CNRS,
Bordeaux INP, LaBRI,
UMR 5800
robin.montferme@u-
bordeaux.fr

Joseph Larralde
Université Bordeaux, CNRS,
Bordeaux INP, LaBRI,
UMR 5800
joseph.larralde@gmail.com

RÉSUMÉ

Nous proposons la programmation, sous la forme d'atelier, de la pièce *Hermès v2*. Il s'agit d'une pièce jouée entièrement en réseau. Ce projet fait partie du groupe de travail « musique en réseau ». Cet atelier sera aussi l'occasion de discuter les aspects techniques de la pièce, à savoir le serveur et l'interface graphique avec *Open stage control*, le séquenceur Ossia Score et le moteur sonore avec Max. Après l'exécution de la pièce, nous discuterons des prochaines étapes du projet, et en particulier de la version entièrement en ligne.

1. INTRODUCTION

Hermès v2, composée par le premier auteur est la deuxième pièce d'une série d'expérimentations dont l'objectif est de proposer des solutions alternatives pour la création musicale et la réalisation de concerts en temps de crise sanitaire et de distanciation sociale. Il s'agit d'un *webconcert* collaboratif en temps réel, c'est-à-dire une pièce intégralement jouée en réseau. Les participants, à l'aide d'un dispositif mobile comme un *smartphone*, une tablette ou même un ordinateur portable, se connectent à une interface graphique via un navigateur web. Ensuite, tous les participants jouent la pièce en interaction directe avec le navigateur. L'interface graphique, quant à elle, fait office de chef-d'orchestre en donnant les indications aux participants.

Un premier prototype de cette pièce a été présenté lors des Journées d'informatique musicale 2021 en visioconférence [5]. Cette première étape a servi de terrain d'expérimentation pour la réalisation du présent projet. Ces objectifs sont organisés en trois volets : l'amélioration de l'interface utilisateur avec le logiciel *Open stage control*¹ ; la suite du développement du moteur de contrôle avec OSSIA Score² et la finalisation de la pièce en prenant en compte ses différents formules (ateliers pour grand public ou pour des artistes musiciens, restitution en ligne ou en réseau local, rendu sonore multicanal ou stéréo etc...).

Dans cet article, nous nous pencherons sur la dernière version en réseau local. Dans un premier temps, nous aborderons la manière dont l'apprentissage avec le *patching* collaboratif avec le logiciel Kiwi nous a

influencé dans la conception de cette pièce. Ensuite, nous décrirons plus précisément le développement du réseau logiciel et sa topologie. Nous concluons en présentant les perspectives.

2. APPROPRIATION DE L'EXPERIENCE AVEC LE *PATCHING* COLLABORATIF AVEC KIWI

Le développement du logiciel Kiwi a été mené dans le cadre du projet ANR MUSICOLL (2016-2018). Sa caractéristique principale est le *patching* collaboratif. Ce logiciel permet à plusieurs utilisateurs de composer ensemble en temps réel un unique processus musical sur un patch hébergé en ligne [3]. Malgré l'importance de cet outil dans la pédagogie de l'informatique musicale et dans la création en réseau [2][3], son développement a été interrompu à la fin du projet MUSICOLL. En ce moment, le logiciel n'est plus fonctionnel. En dépit de sa courte vie, notre expérience collaborative en réseau avec Kiwi nous a permis de poser les jalons de ce que nous avons réalisé par la suite avec la création de *Hermès v2* :

- Interaction en réseau par le biais d'une interface graphique
- Approche concurrentielle
- Absence de hiérarchie
- Contrôle du degré de liberté

Toutefois trois différences sont marquantes entre le *patching collaboratif* et *Hermès v2* : partage d'un même flux sonore, construction temporelle et surtout approche compositionnelle. Dans Kiwi, le moteur sonore est local, c'est-à-dire que chaque utilisateur avait son propre rendu sonore, seulement les interfaces de contrôle étaient partagées. Dans *Hermès v2*, étant donné ses qualités compositionnelles, notamment l'abondance des fonctions de contrôles aléatoires, le flux sonore est le même et diffusé vers tous les participants. *Hermès v2* est composée dans un flux temporel composé. C'est-à-dire que les parties de la pièce se succèdent les unes après les autres. Dans notre pratique avec Kiwi, il n'y avait aucun contrôle temporel. Finalement, les séances de *patching collaboratif* étaient caractérisées par une approche « commencer du zéro » [2]. L'objectif était de

¹ <http://openstagecontrol.ammd.net/> [lien vérifié le 27 février 2023].

² <https://ossia.io/> [lien vérifié le 27 février 2023].

construire un patch ensemble à partir d'un document vide. Pour cela, aucune règle n'a été déterminée au préalable. *Hermès v2* quant à elle, est déjà composée. Même si les participants ont un important degré de liberté, ses actions sont conditionnées par des instructions préétablies. Dans cette pièce nous développons donc un principe « d'interactions composées ».

Sur le plan logiciel, les deux approches ne partagent aucune procédure. Notre appropriation s'est fondée uniquement sur le plan compositionnel.

3. DEVELOPPEMENT LOGICIEL DE HERMES V2

Le développement logiciel de *Hermès v2* a eu lieu en parallèle de la composition de l'œuvre. Ainsi, une relation bilatérale a été établie : les fonctionnalités des outils techniques nous fournissaient des possibilités, alors que les critères compositionnels poussaient leur développement. Nous avons décidé de faire usage de OSSIA Score pour la gestion des événements, OSC – *open stage control* pour la création d'interface graphique et pour la gestion des flux OSC entre client et serveur et finalement un patch Max avec des objets Faust pour le moteur sonore.

3.1. Ossia Score : le chef d'orchestre

Ossia Score est un logiciel séquenceur gratuit et libre développé par le SCRIME – Studio de création et de recherche en informatique et musiques expérimentales à LaBRI – Université de Bordeaux. Cet outil est essentiel pour *Hermès v2* pour deux raisons principales : sa gestion aisée des flux OSC entre différents dispositifs et sa logique temporelle qui permet de construire des temporalités non-linéaires.

Le scénario de *Hermès v2* est divisé en quatre scènes. L'image ci-dessous montre la première (Figure 1).

La gestion temporelle de ce logiciel fonctionne comme dans la plupart des séquenceurs. Il s'agit d'un défilement temporel où les événements sont déclenchés les uns après les autres. Cependant, l'outil *trigger* (GUI jaune en bas de la Figure 1) permet d'activer des intervalles (les lignes bleues) de manière asynchrone. Ainsi, dans la première partie de la pièce, les intervalles sont activés à chaque fois qu'un bouton de l'interface graphique est pressé. Cela peut se reproduire multiples fois. En parallèle, une ligne temporelle défile ordinairement et active les statuts de configuration de SCORE, de l'interface graphique et du moteur sonore dans Max (milieu de la Figure 1). Cette stratégie permet à la fois d'avoir un contrôle temporel très précis et de donner un important degré de liberté. Dans cet exemple, est aussi évidente la facilité avec laquelle cet outil peut gérer plusieurs dispositifs, à savoir l'*Open stage control* et le patch Max. La communication s'effectue avec le protocole OSC – Open Sound Control.

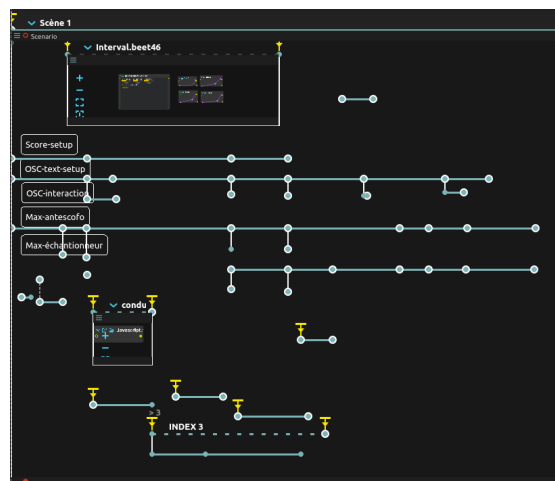


Figure 1. Première scène du scénario de *Hermès v2* dans Ossia Score.

3.2. Open stage control : interface graphique

À l'instar de Ossia Score, *Open stage control* est également gratuit et libre. Il s'agit d'un puissant outil pour la création d'interfaces graphiques simples et pour la gestion serveur et clients en OSC. La partie serveur est lancée sur le même ordinateur qui contrôle Ossia Score. Les clients sont des applications web qui peuvent être ouverts avec un navigateur (Figure 2). Le serveur fournit ainsi une adresse IP avec un port permettant à plusieurs clients de communiquer mutuellement via une même interface graphique. Cette architecture est essentielle pour *Hermès v2* car elle permet, tout d'abord, la participation entre plusieurs acteurs et aussi une approche concurrentielle : les participants peuvent agir sur l'ensemble d'éléments graphiques. Elle permet aussi une interaction sans hiérarchie, car tous les participants ont le même pouvoir sans la possibilité de tracer leurs actions.

Comme tous les clients et serveur doivent être situés sur un même réseau pour transmettre des données, sa configuration reste un aspect majeur. La stratégie dépendra de la formule d'exécution de la pièce. C'est-à-dire, si elle est donnée en réseau local (*LAN – local area network*) ou en ligne *WEB* (à distance). Nous aborderons les différences dans la prochaine section.

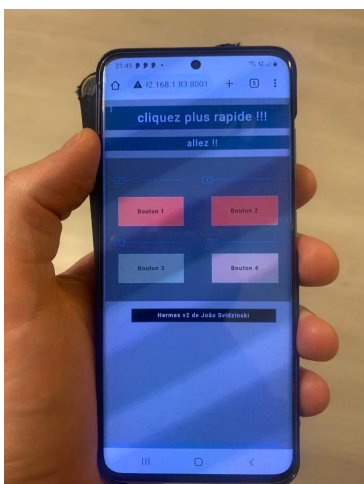


Figure 2. Interface graphique de *Hermès v2* accédée via un navigateur web dans un portable.

3.3. Max et Faust : le moteur sonore

Finalement, la branche logicielle est le patch Max avec des objets Faust. En réalité, il s'agit d'un patch très similaire avec nos pièces récentes comme *Les cris du sixième cercle* (2019) pour soprano et électronique temps réel ou *Lorenz Dream* (2020) audiovisuel. Les pièces sont composées en associant des modules de traitement numérique en Faust à des systèmes de contrôle orientée objet en JavaScript, à partir des résultats de la thèse de doctorat du compositeur. Le noyau principal de la chaîne DSP sont les objets Faust granulateur et *harmonizer* tous les deux en association avec le principe de la réinjection matricielle [4]. Différemment des pièces précédentes, dans *Hermès v2* le contrôle est réalisé avec *antescofo*³.

Deux formes d'instanciation d'événements ont été développées : des objets avec des méthodes et paramètres globaux et des *micro-événements* activés multiples fois. Pour la première, ce sont des objets qui contrôlent plusieurs modules du patch, mais surtout un échantillonneur qui, à son tour, est contrôlé par des fonctions de génération des valeurs pseudo-aléatoires. Ces objets sont instanciés avec *antescofo* par le biais des instructions envoyées par Ossia Score. Ils sont, pour la plupart, activés dans le défilement linéaire du temps. Les *micro-événements* quant à eux, sont notamment des interactions entre les utilisateurs et l'interface graphique, par exemple l'activation d'un bouton. Lorsqu'un participant appuie sur un bouton, l'application web envoie un message OSC à Ossia qui envoie par la suite un nouveau message, aussi via OSC, à Max qui finalement active un *micro-événement*, comme la lecture d'un fichier son.

Ainsi, comme pour la configuration d'*Open stage control*, le système de diffusion est différent pour chaque format. Si la pièce est donnée dans sa version locale, un rendu multicanal est généré avec la bibliothèque Hoa [1] et diffusé dans la même salle où se trouvent les

participants. Dans sa version web, la même bibliothèque est utilisée pour la production d'un rendu binaural et diffusé sur le web. Une troisième stratégie est envisageable par la transmission web d'un flux multicanal. Dans ce cas, la pièce est donnée en deux endroits différents avec une diffusion multicanale similaire. La première alternative a été utilisée dans la version jouée dans le cadre du « Symposium Arts et réseau » qui a eu lieu du 30 Juin 2022 au 2 Juillet 2022 organisé par La Métive à Notre Dame du Moulin, Moutier-d'Ahun, Creuse. La deuxième a été utilisée dans les JIM – journées d'informatique musicale 2020 [5]. Finalement la troisième, encore en développement, *webconcert* en ligne, a eu lieu 14 octobre 2022, lors d'un concert inter-dômes d'écoute spatialisée entre la MSH - Maison des sciences de l'homme Paris Nord à Saint-Denis et le SCRIME à Talence⁴.

4. TOPOLOGIE

L'aspect muable de la pièce, qui peut être donnée sous différents formats, exige également une flexibilité de configurations. La première version, donnée lors des JIM 2021, a lieu entièrement en ligne. Nous vivions encore une période tourmentée de distanciation sociale et des confinements. Cette année, les JIM ont eu lieu sous la forme de vidéoconférence. La version proposée pour les JIM 2023 reprend le même modèle que celle donnée en 2022 lors du Symposium « Art en réseau ». Il s'agit d'une version plus développée (d'après les retours d'expérience de la première version) et donnée en réseau local. Quant à la version avec la transmission web d'un flux multicanal, elle sera abordée dans la prochaine section.

4.1. Version web JIM 2021

Un premier prototype de cette pièce a été présenté lors des Journées d'informatique musicale 2021 en visioconférence [5]. À cette occasion la pièce a été donnée entièrement en ligne. Les participants, en nombre illimité, se connectaient ensemble, via un VPN, à l'interface graphique de générée par *Open stage control*. L'ensemble de données était traité par Ossia Score qui pilotait à son tour l'interface graphique et le patch Max. Celui-ci générait un rendu binaural qui était diffusé via Zoom. Un seul ordinateur concentrait les opérations : le serveur *Open stage control*, le scénario sur Ossia Score et le patch Max.

Plusieurs considérations issues de cette première expérience ont permis de nous guider pour la suite du projet. Tout d'abord, une question primordiale pour le bon déroulement de pièce a émergé : le nombre de participants. Lors des JIM 2021, nous avons décidé de ne pas limiter les participants. Entre 15 et 20 personnes y ont participé. Il s'est avéré que, avec cette quantité, les interactions de chaque participant se noyaient dans l'ensemble d'actions. Il était impossible d'avoir un retour

³ <https://antescofo-doc.ircam.fr/> [lien vérifié le 27 février 2023].

⁴ <https://sitesinsolites.cnrs.fr/visite/hermes-v2-web-concert-collaboratif-en-temps-reel-2/> [lien vérifié le 27 février 2023].

après une opération. D'une part, ce constat va à l'encontre de l'idée de l'approche concurrentielle et de l'absence de hiérarchie. Cependant, le niveau de concurrence était aussi élevé que les utilisateurs ne pouvaient pas avoir un retour de ses actions. Cela rendrait l'expérience complètement gênante. L'interface graphique était aussi peu optimisée. Il s'agissait d'une seule page statique. Seulement les instructions données par Ossia Score étaient dynamiques. L'omniprésence des interfaces graphiques, comme des boutons, qui n'avaient momentanément pas de fonction était également troublant. Le moteur sonore est resté quasiment intacte, excepté l'adaptation pour les différentes diffusions. Le scène d'Ossia Score a subi quelques améliorations qui ont optimisé son utilisation, mais qui n'ont pas changé la pièce elle-même. Par ailleurs, quelques scripts, écrits en JavaScript, ont été intégrés au scénario. Par exemple, un algorithme permet de quantifier le nombre de fois qu'un bouton a été activés par un participant. Cette donnée est ensuite utilisée comme paramètre par un autre modèle dans le séquenceur.

4.2. Version réseau local : Symposium « Art et réseau » et JIM 2023

En 2022, un grand travail d'amélioration d'interface graphique a eu lieu. *Open stage control* offre de grands avantages par rapport à l'interaction utilisateur. Par exemple, il est possible de faire apparaître et disparaître des boutons, changer les couleurs, entre autres. Ces modifications ont beaucoup contribué à l'expérience utilisateur. Tout d'abord, l'interface était plus conviviale et dynamique. Le scénario Ossia Score envoyait dynamiquement des instructions en CSS. Ainsi l'utilisateur pouvait plus facilement et directement se rendre compte des actions nécessaires. De plus, nous avons délimité le nombre maximum de participants. Désormais seulement 4 personnes peuvent participer à une même séance. Pour faciliter la reconnaissance de leurs actions, nous l'avons divisé en onglets. L'interaction de chaque participant était alors délimitée à un champ individuel. Afin de garder le caractère concurrentiel, ils peuvent quand même se déplacer entre les onglets et agir sur l'espace d'un autre participant.

Quant à la diffusion, un module de spatialisation de la bibliothèque HOA a été ajouté au patch Max de la

première version. La figure ci-dessous (Figure 3) montre la topologie de la version proposée aux JIM 2023.

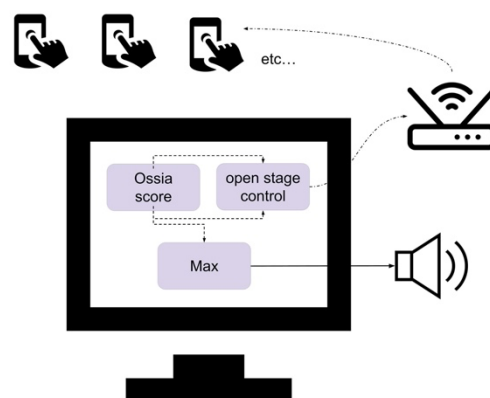


Figure 3. Schéma technique de la version réseau local de *Hermès v2*.

5. PERSPECTIVES

Dans le courant 2023, une nouvelle version sera réalisée en utilisant Quacktrip, sous forme de patch Pure Data, écrit par Miller Puckette⁵, pour le partage audio. Cet outil permettra de s'affranchir des problèmes techniques liés aux blocages institutionnels de ports. Par ailleurs, la pièce devra, dans cette nouvelle version, prendre en compte les interactions en ligne réalisées à distance pour mieux les intégrer à son propre développement. Nous viserons une amélioration du ressenti de participation lorsque plusieurs personnes jouent en même temps, notamment pour pouvoir observer et entendre mieux les conséquences de leur interaction sur la pièce jouée. Pour cela, nous envisageons modifier la méthode de serveur et interface graphique. Pour la transmission de données, nous allons intégrer le Framework Soundworks⁶ et l'interface graphique sera développée avec vue.js⁷. Ces modifications nous permettront d'avoir plus de liberté pour le développement des interfaces graphiques. De plus, comme il s'agit d'outils *open source*, nous pouvons l'adapter à plusieurs situations ultérieures comme l'hébergement de la pièce en serveur pour un usage asynchrone. Nous espérons que ces implémentations vont également nous motiver à trouver d'autres alternatives pour d'autres projets de musique en réseau.

6. MISE EN PLACE

Nous proposons un atelier aux JIM 2023 avec la pièce *Hermès v2*. Cela fera partie des activités du groupe de travail « musique en réseau ». La séance sera organisée en trois étapes. Tout d'abord, nous réaliserons une brève introduction avec une expérimentation, en guise

⁵ <http://msp.ucsd.edu/tools/quacktrip/> [lien vérifié le 27 février 2023].

⁶ <https://soundworks.dev/> [lien vérifié le 27 février 2023].

⁷ <https://vuejs.org/> [lien vérifié le 27 février 2023].

d'introduction, et finalement l'exécution de la pièce proprement dite. À la fin, un moment d'échange va clore la séance. La durée totale est d'une demi-heure (la pièce dure 10 minutes). Le nombre de participants « actifs » sera limité à quatre personnes. Quant aux « auditeurs », il n'y a aucune limite de participants. Nous enverrons ensuite un formulaire aux participants qui donnerons leurs avis par rapport à cette expérience. Ces résultats nous guiderons pour les prochaines étapes du projet, notamment la nouvelle version entièrement en ligne.

Pour notre participation, seront nécessaires un système de diffusion en huit points en cercle et un réseau local (un router wifi). Les participants accéderont à l'interface avec leurs propres dispositifs (portable, tablette etc...).

7. REMERCIEMENTS

Hermès v2 est composée en collaboration entre le laboratoire Musidanse/CICM Paris 8, le SCRIME – Studio de Création et de Recherche en Informatique et Musiques Expérimentale – de l'université de Bordeaux et la MSH Paris Nord dans le cadre du projet « Organisations musicales symbiotiques : une révision de la notion de concert de recherche-crédation faisant appel à l'informatique musicale ».

Cette recherche a été accueillie au SCRIME, plateforme de recherche de l'Université de Bordeaux (Studio de Création et de Recherche en Informatique et Musiques Expérimentales), et financée par la région et la DRAC Nouvelle Aquitaine.

8. REFERENCES

1. Colafrancesco J. *Spatialisation de sources auditives étendues - Applications musicales avec la bibliothèque HOA*, Thèse de doctorat, Université Paris 8, 2015.
2. Messina, M. Svidzinski, J. de Menezes Bezerra D. Ferreira da Costa D. « Live Patching and Remote Interaction: A Practice Based, Intercontinental Approach to Kiwi », *CMMR 2019*, The Laboratory PRISM, Oct 2019, Marseille, France. pp.696-703. ([hprints-02321033](https://hal.archives-ouvertes.fr/hal-02321033)).
3. Paris, E. Millot, J. Guillot, P. Bonardi, A. Sèdes, A. « Kiwi : vers un environnement de création musicale temps réel collaboratif (premiers livrables du projet MUSICOLL) », *Journées d'Informatique Musicale 2017*, Paris, France (2017).
4. Svidzinski, J. Bonardi, A. « Reinjection matrices with faust language : creating complex structures in real-time mixed music », *Proceedings of the International Faust Conference 2020 (IFC-20)*, 2020, Saint-Denis.
5. Svidzinski, J. « Hermès v2 - web concert collaboratif en temps réel », Journées d'Informatique Musicale 2021, AFIM, Jul 2021, Visioconférences, France <https://hal.archives-ouvertes.fr/hal-03313610>.

QUACKTRIP EN OCTOPHONIE : ETUDE DE CAS DANS *HERMES V2*

João Svidzinski
CICM-EA 1572
Université Paris 8 et IRCAM
svidzinski@gmail.com

Miller Puckette
Department of Music
University of California San Diego
msp@ucsd.edu

RÉSUMÉ

Nous proposons une démonstration et performance d'une version en ligne de la pièce *Hermès v2* avec le transfert audio en octophonie réalisé avec Quacktrip. Notre intervention aura lieu dans le cadre du Groupe de travail AFIM « art en réseau ».

1. INTRODUCTION

Quacktrip est une application informatique qui connecte des flux audios non compressés à faible latence en ligne¹. Cet outil est destiné aux musiciens souhaitant jouer ensemble à distance. Ainsi, son architecture est pensée pour le *streaming* audio stéréo entre deux ou plusieurs instrumentistes. Son efficacité a été à plusieurs fois attestée, notamment en 2020 avec la pandémie Covid-19 en période de confinement [1].

Ce projet fait partie du groupe de travail « musique en réseau » de l'association francophone informatique musicale. Cette collaboration a motivé l'incorporation de ce projet à la pièce *Hermès v2* [2].

Nous proposons donc pour les JIM une démonstration et l'exécution d'une version de *Hermès v2* avec Quacktrip en octophonie².

2. QUACKTRIP : MUSIQUE EN RESEAU

Quacktrip est une mise en œuvre, dans Pure Data, du protocole réseau Jacktrip de Chris Chafe, basé sur *jacktrip.pd* de Roman Haefeli et Johannes Schuett. Il établit une connexion, sans compression audio, point-à-point à faible latence entre deux endroits. Quacktrip est disponible sous forme de *standalones* Mac et Windows, de patchs Pure Data et Max ainsi que d'un plug-in VST. Comme il s'agit d'un outil gratuit et ouvert, il est possible de développer d'autres versions plus adaptées à des situations spécifiques. La pièce en réseau *Hermès v2* est un exemple d'œuvre qui nécessite le transfert audio en ligne d'une manière particulière. Il s'agit d'un *webconcert*, c'est-à-dire, une œuvre entièrement jouée en ligne. La pièce peut être donnée sous plusieurs formats : en réseau local ou à distance (avec des participants localisés en deux ou plusieurs endroits dont chacun

regroupant un sous-groupe des participants). Pour la dernière, nous voudrions développer un *webconcert* inter-dômes d'écoute spatialisée. Une première tentative a eu lieu en octobre 2022 lors de Fête de la science, organisé par le CNRS³. Cependant, des problèmes notamment liés à la latence et à ouverture de ports institutionnels n'ont pas permis qu'une expérience optimale de transfert audio en ligne ait lieu. Pour 2023, nous voudrions utiliser Quacktrip pour accomplir cette tâche. Cependant, les versions actuelles privilégient le transfert stéréo. Nous voudrions donc développer une version octophonie.

3. DEMONSTRATION ET PERFORMANCE

Nous proposons une intervention de trente minutes dans le cadre de la présentation du GT « musique en réseau ». Dans un premier temps nous montrerons les questions liées au développement de la version en octophonie de Quacktrip. Dans un deuxième temps, nous présenterons cet outil dans la pratique. Une version de *Hermès v2* dédiée à cette intervention sera utilisée comme exemple d'application.

Pour cette démonstration et performance, nous aurons besoin d'un système de diffusion en octophonie et une connexion internet à très haut débit.

4. REFERENCES

1. Puckette, M. « Modular software and remote workflows », *Ateliers du Forum Ircam 2021*, en ligne : <https://youtu.be/A-15Gw4F6aM> Paris, France, 2021.
2. Svidzinski, J., « *Hermès v2* - web concert collaboratif en temps réel », *Journées d'Informatique Musicale 2021*, AFIM, Jul 2021, Visioconférences, France. <https://hal.science/hal-03313610>.

¹ <http://msp.ucsd.edu/tools/quacktrip/> [lien vérifié le 27 février 2023].

² Une première version de Quacktrip en octophonie est disponible dans le lien suivant : <https://github.com/JoaoSvidzinski/Quacktrip-octo> [lien vérifié le 27 février 2023].

³ <https://www.mshparisnord.fr/event/cnrs-insolites-2022/2022-10-14/> [lien vérifié le 27 février 2023].

NETWORKED MUSIC PERFORMANCE IN PATCHXR

Jonathan Bell, Oceane Wandji
PRISM-AMU
belljonathan50@gmail.com

Shiho Yamada
XR2C2-UCA
shihomatsuda@gmail.com

RÉSUMÉ

La présente étude propose d'explorer un corpus sonore en RV, dans lequel les données audio sont découpées et analysées dans FluCoMa afin d'obtenir des collections relativement importantes d'échantillons regroupés par similarité de timbre dans une pièce en 3D. L'implémentation récente de la fonction multijoueur dans PatchXR permet d'envisager une grande variété d'interfaces de contrôle basées sur le geste pour interroger ces corpus, et dans lesquelles les interprètes peuvent interagir à distance afin de simuler une situation de musique de chambre.

1. INTRODUCTION

Le "musical metaverse" tel que l'envisage Turchet [3] pourrait apporter des réponses aux expériences de performances musicales en réseau (networked music performances, NMP) dont les prémices remontent aux débuts de l'internet [1], puis redécouvertes pendant les années covid. Le développement très récent du mode multijoueur dans PatchXR rend compte de ce constat. PatchXR [2] se définit comme une "création musicale immersive combinée à la construction d'un monde virtuel". Bien que fortement orienté vers l'EDM, sa ressemblance frappante avec Max/MSP et Pure Data (voir Fig. 1) permet diverses formes d'exploration.

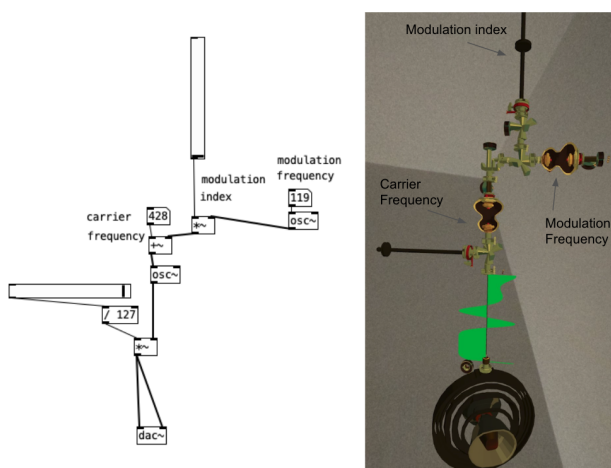


Figure 1. L'implémentation de l'algorithme de synthèse FM 1/ dans Pure Data (à gauche) 2/ dans PatchXR (à droite).

Le "musical metaverse" incite également à développer

un nouveau type de lutherie électronique afin d'interagir musicalement à distance et de manière collaborative avec un instrument numérique, dans un monde en réalité virtuelle multijoueurs (ou métavers).

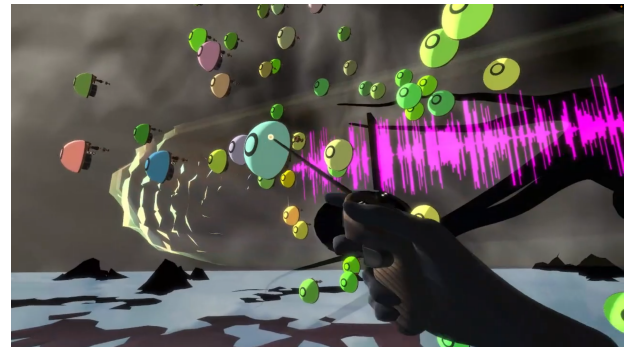


Figure 2. Une interface VR dans laquelle chaque bouton du monde correspond à un court sample (datapoint) d'un long fichier son (database). L'apprentissage automatique permet de rapprocher les sons qui partagent des caractéristiques spectrales communes.

La solution proposée ici se concentre sur l'exploration d'un corpus sonore projeté sur un espace 3D, dans lequel les utilisateurs peuvent ensuite naviguer à l'aide de leurs contrôleurs manuels (voir Fig. 2). La synthèse concaténative basée sur corpus et d'autres techniques de *music information retrieval* sont réalisées via le paquet FluCoMa [4] dans Max.

Les sessions multi-joueurs enregistrées ci-dessous témoignent du possible échange de données de contrôle à faible latence sur le piano, le saxophone, le shakuhachi, le vibraphone et la voix, alors que les interprètes étaient situés à des kilomètres les uns des autres. Dans deux premiers extraits, pour piano et saxophone par exemple, un musicien joue les boutons bleus (les échantillons de saxophone), et l'autre, les boutons jaunes (les échantillons de piano), pendant qu'un troisième bat la mesure et récite un texte :

- Was it a vision or a waking dream? (John Keats) ¹
- A beaker full of the warm sound ²

Les extraits suivants montrent différentes interfaces : 1/ l'extrait pour shakuhachi montre un instrument de type thérapie basé sur le même principe de synthèse concaténative 2/ l'extrait pour clavier rend compte des possibilités

1 . <https://youtu.be/CEhLbJeWogE?t=853>

2 . <https://youtu.be/CEhLbJeWogE?t=182>

de réalité augmentée de PatchXR ainsi de la plausible synchronisation à distance sur un rythme lent (rubato) :

— Shakuhachi ³

— Autumn Leaves ⁴

La sensation de co-présence des interprètes autour de soi est fortement renforcée par le son binaural (natif à PatchXR).

2. REMERCIEMENTS

Cette initiative a été développée dans le cadre de la résidence XR2C2 au CTEL/UCA - Université Côte d'Azur.

3. REFERENCES

- [1] Hajdu G., "Quintet.net : An Environment for Composing and Performing Music on the Internet," *Leonardo* 2005; 38 (1) : 23–30..
- [2] Andersson, T. "Immersive audio programming in a virtual reality sandbox," *journal of the audio engineering society*, march 2019.
- [3] Turchet, L. "Musical Metaverse : vision, opportunities, and challenges," *Personal and Ubiquitous Computing*, 01 2023.
- [4] Tremblay, P.A. "Enabling Programmatic Data Mining as Musicking : The Fluid Corpus Manipulation Toolkit", *Computer Music Journal* 2022; 45 (2) : 9–23..

3 . <https://youtu.be/ySKLrNlov4E>

4 . <https://youtu.be/7BKSEqnz8Jg?t=13>

Actes des Journées d'Informatique Musicale 2023
24-26 mai 2023
Maison des Sciences de l'Homme Paris Nord

SESSION « ANALYSE-SYNTHÈSE »

A MUSICOLOGICAL PIPELINE FOR SINGING VOICE STYLE ANALYSIS WITH NEURAL VOICE PROCESSING AND ALIGNMENT

Yann Teytaut
STMS (UMR 9912)

teytaut@ircam.fr

Antoine Petit
Passages XX-XXI

antoine.petit@univ-lyon2.fr

Céline Chabot-Canet
Passages XX-XXI

celine.chabot-canet@univ-lyon2.fr

Axel Roebel
STMS (UMR 9912)

roebel@ircam.fr

ABSTRACT

The study of singing style is of great interest both for expressive vocal synthesis and for the musicological analysis of vocal performances, inciting to a fruitful convergence between signal processing and musicology. However, for musicologists, these studies often come up against the absence of automatic analysis tools for voices recorded in a musical context, leading to long and tedious manual annotation work. This constraint imposes either to limit oneself to a restricted corpus, or to circumscribe one's study to experimental corpora of voices without instrumental accompaniment, thus depriving oneself of the unequalled interest that commercial recordings represent, as accomplished artistic works. This article introduces a new protocol using deep learning techniques to provide musicologists with powerful tools for the analysis of singing voices, opening up new perspectives through the automation of the different steps. We present a complete processing chain in support of musicological analysis, using neural models to isolate singing voice, predict its F0, and automatically align the syllables or notes to the audio (despite the musical accompaniment). The effectiveness of this approach is demonstrated by its practical application on two popular songs. These tools, developed in an ANR project, will soon be available to the scientific community.

Keywords: musicology, voice, singing style, popular music, voice alignment, voice processing, deep learning.

RÉSUMÉ

L'étude du style de chant revêt un fort intérêt tant pour la synthèse vocale expressive que pour l'analyse musicologique de performances vocales, incitant à une convergence fructueuse entre traitement du signal et musicologie. Cependant, pour les musicologues, ces études se heurtent souvent à l'absence d'outils d'analyse automatique de voix enregistrées en contexte musical, amenant à un long et fastidieux travail d'annotation manuelle. Cette contrainte impose soit de se limiter à un corpus restreint, soit de circonscrire son étude à des corpus expérimentaux de voix sans accompagnement instrumental, se privant alors de l'intérêt inégalé que représentent, comme œuvres artistiques abouties, les enregistrements commerciaux. Cet article introduit un protocole inédit utilisant des techniques d'apprentissage profond pour fournir aux musicologues

des outils performants d'aide à l'analyse des voix chantées, ouvrant, par l'automatisation des différentes étapes, des perspectives nouvelles. Nous présentons une chaîne de traitements complète en support à l'analyse musicologique, exploitant des modèles neuronaux pour isoler la voix chantée, prédire sa F0 et automatiquement aligner les syllabes ou notes à l'audio (malgré l'accompagnement musical). L'efficacité de cette démarche est démontrée par son application pratique sur deux chansons populaires. Ces outils, développés dans le cadre d'un projet ANR, seront bientôt disponibles pour la communauté scientifique.

Mots-clés: musicologie, voix, style de chant, musiques populaires, alignement de voix, traitement de la voix, apprentissage profond.

1. INTRODUCTION

Expressivity is one of the core elements that come into play when humans communicate with each other. Indeed, by exploiting all the prosodic resources of the spoken voice [27] – intonation, stress, paralinguistic effects, etc. – speakers are able to express a wide variety of emotions and social attitudes [43]. The singing voice, as a vector of communication, appropriates many of these codes. As a result, in the 2010s, some musicologists initiated the application of paralinguistics, phonostylistics or psycholinguistics [37, 30, 18] to vocal performance analysis [26, 6].

Between prosody, paralinguistic impregnations, rhetorical procedures and singing techniques (register, timbre, etc.), singers make use of a rich palette of vocal effects. This palette contains elements that:

- connect singers to specific groups, relating to a *generic style* through conventions allowing the identification with a “tribe” – e.g., smoothing of registers and singing formant in classical singing, yodeling and twang in country music [34], belting in musicals [25], guttural voice in extreme metal [22], etc.;
- make singers unique, relating to a *personal style* – e.g., particular vibrato, specific phrasing, intonative effects, timbre, manner of attacking, sustaining or ending notes, etc.

These characteristics can operate on a global scale (e.g., the roughness of a voice) or appear locally, as emphasis. They can be difficult to define, however, for they often strike a subtle balance between antagonistic entities –

speech and song, harmonic and noise, pure musicality and valorization of the text. This is all the more true in popular music, since its norms and conventions are looser than in classical music, due, in particular, to the absence of explicit theorization. The palette of vocal effects is potentially infinite, and even apparent mistakes can be admitted and deliberately used [8].

Although a well-defined artistic identity implies a certain amount of coherence, each performance is unique: choices made according to the needs of the song as well as the context enact a “*mouvance*” [29] characteristic of a music brought about by extemporization [4], the transience of the moment fortunately fixed by the recording, which also adds its own layer of complexity. To identify the style of an artist is thus not only to describe in detail such or such performance, but also to deduce its potentialities, its virtualities, to understand the general strategies, the processes at work, and the challenges that they support.

If the study and the interpretation of stylistic data resulting from the analysis of singing voice in a musical context can take the most diverse paths according to the disciplinary angle chosen (musicological, anthropological, philosophical, cultural, gender studies, etc.), a first stage is often essential: the description, in the most neutral and objective way possible, of the sonic materiality of the voice, as mediated by the recording and, possibly, additional studio processing (production effects).

The meticulous observation of the vocal phenomenon enables one to, for example, establish unsuspected parallels between vocal effects or techniques carrying very different meanings according to the generic, cultural and aesthetic contexts in which they appear. It is for these reasons that a growing branch of musicology is turning to the study of spectral representations of sound (sonograms), following the path traced by a few pioneers of the 1980s [11]. If we disregard identifiable production effects as such, and focus on purely vocal effects, the infinite number of observable phenomena can be abstracted to a limited number of acoustic parameters, pertaining to pitch, prosody, and quality (*i.e.*, timbre, see also [24]), which interact to create a specific *vocal delivery* [31].

The study of singing style [7], *i.e.*, the production strategies at play in a singer’s performances, and the palette of delivery effects that defines their artistic identity, is of great interest to musicologists, but the modeling of singing style also has direct applications for the synthesis of a more expressive and natural-sounding singing voice [2]. This convergence of interests has led to a close collaboration between signal processing and musicology researchers, initiating an unprecedented situation of interaction.

For years, musicologists studying singing performance have been faced with the absence of tools for automating the processes of acoustic analysis and annotation of corpora of voices recorded in a musical context, leading to long and tedious work of manual annotation (transcription) and synchronization (alignment), either by ear or through visual spectral representations like sonograms.

If the expert listening of musicologists remains indispensable to supervise both transcription and alignment, a purely manual system has its limits. It is very time-consuming, meaning the corpus must either be small, or an experimental one, comprised of voices recorded without instrumental accompaniment, which is of little interest to musicologists interested in the *actual* musical works found in commercial recordings. It can also be error-prone and, in some cases, overly subjective. Thus, relying on automatic systems for transcription [19] and alignment [40] can lead to a considerable gain in time, and help in setting a common base for the musicological community.

In this context, deep learning has found numerous applications for voice-related tasks yet, to the best of the authors’ knowledge, has not been directly dedicated to such musicological applications. In this work, taking advantage of the latest improvements in singing voice separation [10], F0 estimation [3], and voice alignment [16] algorithms based on deep learning, we introduce a complete pipeline which simplifies tedious analysis steps previously carried out by hand, and opens new perspectives through automation. With the intent to share these algorithms with the community, a companion website is under construction in the context of the ANR (French National Research Agency) project **Analysis and tRansformation of Singing style (ARS)**¹.

More specifically, our proposed contributions are:

- A complete pipeline automatically extracting voice parameters upon state-of-the-art vocals separation, and aligning the audio with lyrics and notes;
- A coupling between expert transcriptions (lyrics, notes) and audio features, allowing (computational) analyses to be performed not only on symbolic (pitch, metric position) or acoustic (F0, syllable onset and length) data, but *across* both sets of data to bridge the semiotic divide identified by [21];
- Case studies illustrating the musicological interest of our unified pipeline on two popular songs from different languages and genres;
- A website set up to give access to the tools to the scientific community².

2. RELATED WORKS

The current project relies on previous research dedicated to the analysis of singing style and its modeling for the synthesis of expressive singing, carried out within the framework of the ANR project **ChaNTeR**³, of which one can find a detailed example of application to Edith Piaf in the article [9]. As detailed in the complementary paper [2] and thesis [1], this project was aimed at incorporating expressivity, intimately related to singing style, in a concatenative synthesis system.

1. <https://ars.ircam.fr/>

2. <https://passagesxx-xxi.univ-lyon2.fr/activites/projets-anr/projet-ars-analyse-et-transformation-du-style-de-chant-1>

3. **Chant** Numérique avec contrôle Temps Réel, ANR-13-CORD-0011, 2011-2017

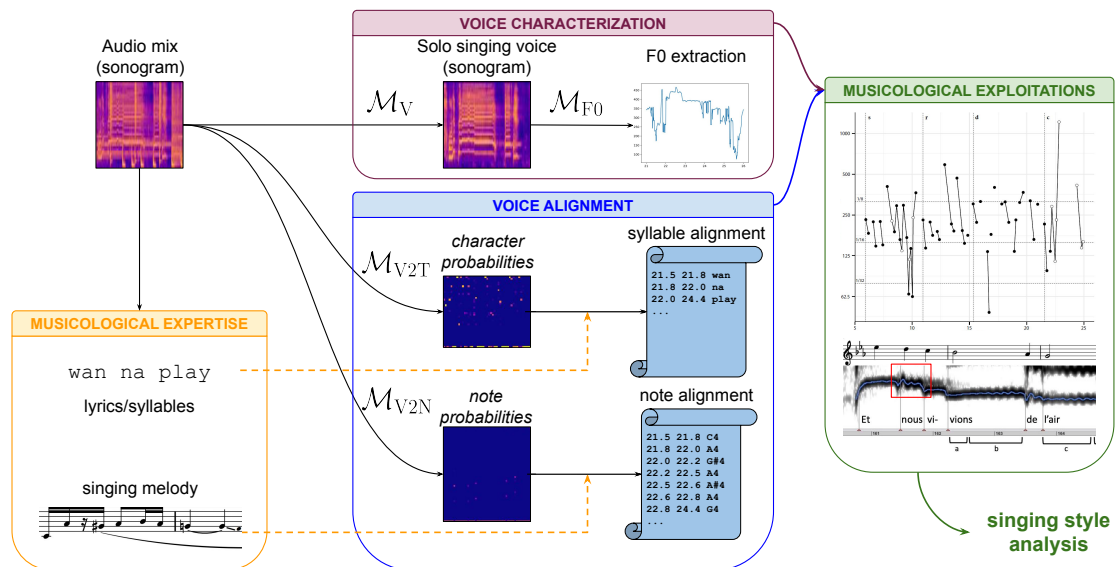


Figure 1. Overview of our complete analysis pipeline involving musicological expertise, deep learning models for the automation of voice characterization and alignment in order to help musicologists studying singing voice style.

Concretely, the synthesis of a given utterance was achieved by progressively connecting the relevant biphones (*i.e.*, two successive phonemes) of a pre-recorded singing dataset. The desired F0, derived from the score notes, was applied to each segment. Then, a musicologically-informed expressive control was developed to counteract the “unnatural” effect induced by the concatenations.

Based on extensive analyses of several performances of a representative corpus of singers illustrating several generic singing styles, a series of style-defining vocal effects were identified and described. The musicologist also established a series of local contexts to determine correlations between specific musical situations (pitch and duration of notes, position in the phrase and in the song structure, etc.) and the occurrence of the described effects. Decision trees were created to associate these contexts with each note in the score. Finally, global contexts were used to further transform the vocal parameters (*e.g.*, F0, intensity, etc.) in a consistent and expressive way.

The musicological analysis of the corpus required time-aligned lyrics and MIDI notes with respect to the audio. However, at that time, the presence of an instrumental accompaniment prevented the use of available systems, designed for *solo* singing voice. This difficulty could not be circumvented, as the desire to study iconic singers, such as Edith PIAF or Jacques BREL, required the use of commercial recordings, for which we did not have access to the vocal track independently from the full mix. As a result, the annotations and alignments were done entirely by hand. This process being very time-consuming, the choice was made to limit the corpus to 4 songs per artist studied, which offered sufficient data to obtain convincing results in the modeling of styles in a reasonable amount of time.

In this work, consequently, one of our key goals is to greatly simplify these processes and open the way for a complete and *easy-to-use* musicological analysis pipeline.

By exploiting recent deep learning models, F0 extraction is obtained automatically upon high-quality separation of the voice from the accompaniment, and lyrics and notes alignments are generated automatically even with background music present. These unprecedented possibilities open new perspectives for musicology, with the opportunity to work systematically on large corpora of recordings, considerably widening its field of investigation.

Because they contain many more instances of each effect and context, these large corpora allow for a more reliable identification of an artist’s style, and facilitate, *e.g.*, the search for similarities or voice synthesis opportunities.

3. A PIPELINE FOR SINGING STYLE ANALYSIS

This section presents our proposed pipeline. As shown in Figure 1, it consists in four main categories with highly interdependent modules, namely *voice characterization*, *musicological expertise*, *voice alignment*, and *musicological exploitations*, which are further detailed.

Note that, although the modules are introduced in their logical order of operation, the pipeline offers musicologists a more varied set of workflows: the system is flexible enough to accommodate different perspectives and practices, as well as multiple musical ontologies (*i.e.*, kinds of musical works) – see subsection 3.4.

3.1. Voice characterization

The core of any singing or singer analysis system is the voice itself. In the context of commercial recordings, the presence of background music is a major problem for state-of-the-art parameter estimation algorithms and hinders a precise description of the singers’ intonation or intensity contours. Thus, this first step aims to extract the singing voice from the background music.

3.1.1. Singing voice separation

The state of the art in source separation relies on deep learning techniques [33]. In our application, the separated vocals only serve for parameter estimation, such that a single-channel (mono), 16kHz extraction is sufficient. From the many proposed neural architectures, we chose to re-implement the network presented in [10], as it achieved state-of-the-art singing voice extraction quality with a comparatively small number of parameters.

This model, denoted as \mathcal{M}_V , has been trained using the publicly available MUSDB and CCMixer data sets, and a collection of internal data featuring solo singing voices and instrumental music including notably instruments not well covered in the public data sets. During training, the voice and music samples were randomly mixed and pitch-shifted following [12, 28]. The final model allows for efficient vocal signal separation with a very satisfying quality. It achieves an SDR of **9.2dB** for the vocals separated from the test set of the HQ version of the MUSDB [38] data set, which compares favourably with state-of-the-art performances [33]. In inference, separation is faster than real-time even when running on a CPU on a small laptop.

3.1.2. Parameter extraction

The isolated voice signal is then used to extract singing voice parameters. Currently, only the fundamental frequency (F0) and intensity contours are covered.

The F0 estimation algorithm is also a deep neural network \mathcal{M}_{F0} that has been trained using a large data set of speech signals using an analysis/resynthesis procedure that allowed the creation of a perfect annotation of the target F0 contours [3]. Despite having been trained on clean speech only, this algorithm has recently shown very good performances for pathological voice signals as well [41]. The F0 estimation has been enhanced with a voice/unvoiced algorithm forcing the estimated F0 to zero for silent parts, or segments not dominated by a single voice.

The intensity estimation is performed by means of calculating the root mean average energy of the separated voice signal over short analysis windows.

As these algorithms are limited to solo vocals, future works may focus on their robustness to polyphonic music.

3.2. Musicological expert knowledge

Upon characterization of the solo singing voice, a typical musicological objective is to correlate voice features with others, such as transcriptions of relevant information, and emphasize their relationships.

Transcription is a task aiming at predicting a symbolic sequence from another data representation. In the context of voice analysis, two transcription tasks are ubiquitous, highlighting two modalities of utmost importance: text and melody. A naive transcription may miss elements and subtleties that precisely describe the vocal performance – therefore, expert knowledge and supervision by musicologists are essential.

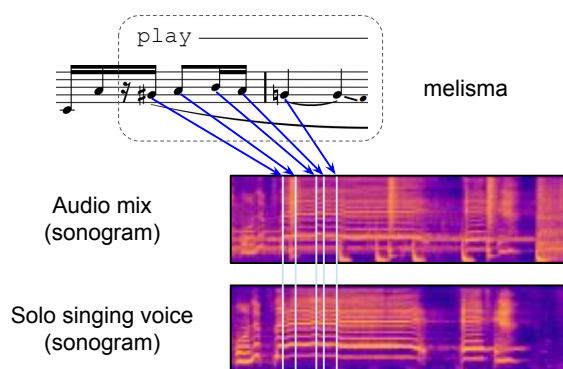


Figure 2. Illustration of the automatic analysis of a melisma: syllable-level alignment only predicts the word “play” on the full duration of this excerpt, without taking pitch variation into account – note-level alignment allows a deeper look into this gesture. See Taylor SWIFT case study – subsection 4.1.

3.2.1. Lyrics retrieval

For most vocal music today, lyrics are easily available online or via the album booklet, so that manual transcription is rarely required. However, instead of a succession of words, a *syllable*-level segmentation of the text is more relevant as singing notes are held on syllable vowels [39]. In practice, the lyrics rarely match the singing content perfectly due to additive onomatopoeia (e.g., “yeah”) or unpronounced utterances. Such local irregularities, fortunately, are not a problem for recent audio-to-text aligners, which can handle missing or additional syllables [16].

The role of the musicologists in lyrics retrieval is twofold: to ensure that the text is coherent and correctly written; and to explicitly adapt, whenever necessary, repeated syllables, missing entries, or onomatopoeia judged pertinent (i.e., conveying meaningful interpretative aspects).

3.2.2. Melodic transcription

Melodic transcription consists in determining the notes performed by the singer in, e.g., musical notation. In opposition to an F0 extraction, associated with *performance time*, a musical score is concerned with *symbolic time*.

A dedicated algorithm may be used as a first step, provided that it is robust and able to distinguish notes of the singing voice from accompaniment notes, to help in the process. In this paper, transcriptions were done entirely by musicologists, without recourse to such an algorithm.

A resulting score transcription can include one or several instances of *melisma*, i.e., multiple notes sung on the same syllable (shown in musical notation by slurs). In this study, musicologists specifically rely on note-level alignment to complement the syllable alignment in such cases. An example is depicted in Figure 2. **To the best of the authors’ knowledge, it is the first option proposed to musicologists for dealing with the automatic analysis of melisma.**

3.3. Voice-to-symbols synchronization

With the audio and symbolic sequences (syllables and notes) at disposal, an *alignment* algorithm aims to associate each element in the sequences with a time in the audio, corresponding to its onset, a key step to further study the temporal aspects in singing performances.

3.3.1. CTC-based neural alignment

The alignment models are also deep neural networks, trained to minimize the Connectionist Temporal Classification (CTC) [20], a loss function assessing sequence-to-sequence prediction that does not need aligned training data. Given a spectral representation of the audio – with a theoretical time precision δt of 16ms in our setup –, a CTC-based model outputs, for each frame, a probability distribution over an alphabet \mathcal{A} of L symbols plus a non-informative blank token. (See examples in Figure 1.) These per-frame probabilities can be used to force-align an audio with a sequence via a CTC variant of VITERBI’s decoding. The architecture of our acoustic modes (deep CNN) and the decoding module are the same as [16].

Training is done with the large collection of English songs with roughly aligned words and notes from the DALI dataset [32]. The models have the great benefit of being applicable to all languages sharing the same alphabet \mathcal{A} (although specialized, hence better, on English – see [16]) and across various musical genres irrespective of production date, meaning musicologists do *not* need to adjust parameters. This work precisely uses the same models on an American pop song from 2014 and a French chanson song from 1966. Finally, *singing voice separation is not a mandatory step for running the alignment models*, which was a major technical deadlock imposing manual alignments in [9], so that their usage goes beyond our pipeline.

This approach does not outperform the state of the art [19] but is independent from any domain knowledge [16].

3.3.2. Aligning audio with syllables and notes

Let \mathcal{M}_{V2T} denote the voice-to-text (V2T) aligner. The alphabet \mathcal{A} contains all the basic latin characters (a, b, etc.), digits (0, 1, etc.), and a space token \emptyset for separating successive syllables, hence $L = 37$. Although designed for word-level alignment, the model can also synchronize syllables, as words and syllables share the same alphabet. The only difference lies in the decoding step, as there are more spaces \emptyset between syllables than words.

Let \mathcal{M}_{V2N} denote the voice-to-note (V2N) aligner. We retrieved the F0 annotations in DALI and converted them into notes that range from C_1 to C_7 – this is particularly large for the human voice, but a manual inspection of outliers has not been pursued. The alphabet \mathcal{A} contains all 12 semitones per octave and a silence token for long pauses (>500ms in DALI annotations), hence $L = 73$. ***It is, to the best of the authors’ knowledge, the first time that an end-to-end, CTC-based model addresses note alignment*** – while note transcription was tackled [42].

	V2T	V2N
Playlist50 [16] [†]		
└ AAE (ms)	124.2 ± 58.5	232.0 ± 263.6
└ MAE (ms)	39.2 ± 14.1	105.1 ± 170.6
└ CER (%)	49.1 ± 14.2	39.7 ± 11.2
“Blank Space” (manual correction)		
└ AAE (ms)	13.5 ± 40.9	22.5 ± 63.0

[†] Metrics are averaged over all songs.

Table 1. Voice-to-text (V2T) and voice-to-note (V2N) alignment evaluations in terms of average (AAE) and median (MAE) alignment errors and character error rate (CER).

3.3.3. Alignment accuracy

We briefly assess the robustness of our models on the Playlist50 evaluation set, introduced in [16], in terms of alignment and transcription, as shown in the Table 1. Results show that it is more challenging to align notes than syllables, although overall mean errors remain below the commonly admitted 300ms perceptible threshold [13]. However, recognition of notes is better than lyrics – but it is known that a better transcription does not systematically imply a better alignment in a CTC framework [40]. Interestingly, in our “Blank Space” study, exploiting both alignments (see 4.1), the manual syllable corrections are *below* the theoretical precision δt while note alignment, although once again less stable, is very much acceptable.

3.4. Musicological exploitations

Finally, our proposed protocol allows further musicological studies while offering great flexibility.

The system only expects that symbolic data (text, notes) can be related to a recording, meaning it works equally well whether this data is transcribed from the recording or exists prior to it, as in *written* vocal music (*e.g.*, art songs), the study of which is also possible with this pipeline.

Because most modules output files directly and independently from one another, musicologists can freely decide which to use and how. One approach, which has been favored by one of the authors, is to use the data from each module to repeatedly refine the score transcription, thus enacting a sort of back and forth between symbolic and acoustic data. Also, because the data is not tied to a single working environment, it is available for computational analysis, as shown in subsection 4.1, which uses a series of scripts written by the musicologist in the R programming language.

The time markers from the alignment files can be manually corrected using visualization software, *e.g.*, Sonic Visualiser⁴ or RX⁵ – a process much less tedious than starting from scratch. As for voice parameters, F0 estimation curves require close to no manual corrections.

4. <https://www.sonicvisualiser.org>

5. <https://www.izotope.com/en/products/rx.html>

4. MUSICOLOGICAL CASE STUDIES: TAYLOR SWIFT & CHARLES AZNAVOUR

Having introduced our general pipeline, a demonstration of the whole musicological protocol is proposed via two case studies, first on Taylor SWIFT’s 2014 hit single “Blank Space”⁶ and then on Charles AZNAVOUR’s “La Bohème” from 1966⁷. In the first study (SWIFT), temporal data from the alignments is used to perform fine-grained rhythmic analyses and to investigate the structural role played by articulation and micro-rhythm. In the second one (AZNAVOUR), the fundamental frequency estimation is used to study vocal phrasing and rhetorical effects involving intonation.

4.1. Taylor SWIFT’s “Blank Space”

This song was selected for several reasons: (1) it is (at time of writing) SWIFT’s second-best charting song, having stayed seven non-consecutive weeks at the top of the Billboard Hot 100, and its analysis may contribute to a more thorough understanding of what makes a successful song; (2) it, and SWIFT’s songs more generally, have not yet been the object of much, if any, musicological attention (although see [36]); and (3) it is in standard compound AABA form, but it exhibits somewhat intricate patterning at lower levels of organization, with verse and chorus each articulating two iterations of an **srdc**⁸ structure [17] in part through shifts in vocal delivery. It thus presents a prime example of the kind of analytical work afforded to musicologists by the pipeline. The following case study focuses on the first half of the first verse (eight bars, from 5:30 to 25:30), a score transcription of which is given in Figure 3.

This excerpt can be divided into four parts: two fairly similar segments (**sr**, from 5:30 to 15:30) and a contrasting passage leading to a concluding gesture (**dc**, from 15:30 to 25:30). Following [23], we may wish to understand which criteria elicit such a segmentation, especially when it appears so self-evident.

A typical analysis, focusing mostly on pitch, would highlight the close resemblance between **s** and **r**, which share a series of two-syllables phrases⁹ on F₄ followed by a syncopated descending step-wise motion from A₄ back to F₄ (with an additional leap from G₄ to D₄ in **s**). They would then be contrasted with the series of short phrases that make up **d**, comprised exclusively of large leaps which also instantiates a hierarchy divorce [35] with the underlying harmony (C and A over a B_b chord, and F and A over a C chord), and the long melisma on “play?” that defines **c**.

6. Words and music by Taylor SWIFT, Max MARTIN et SHELLBACK. Reference recording: 1989, Big Machine, 2014.

7. Words by Jean PLANTE, music by Charles AZNAVOUR. From the operetta *Monsieur Carnaval*, 1965. Reference recording: *La Bohème*, Barclay, 1966.

8. Statement, restatement/response, departure, closure.

9. Following convention, we equate “phrase” with “breath group”. Comma-like symbols in the transcription indicate the points at which SWIFT takes a breath.

Figure 3. Taylor SWIFT, “Blank Space”, measures 3–10. [Transcription done by Antoine PETIT.]

Although such analysis is relevant, we believe that articulation and micro-rhythm play as much, if not more, of a role in shaping the form of the excerpt – dimensions that our proposed pipeline precisely allows us to investigate.

4.1.1. Application of the pipeline to the song

Syllables retrieved from a first score transcription were synchronized with the audio by inferring with the \mathcal{M}_{V2T} model. The syllable alignment was then processed with an R script to delete the end marker of syllables *not* followed by a rest in the transcription (*i.e.*, when the articulation was heard as *legato*, meaning the end of the n syllable and the onset of the $n + 1$ syllable can be taken to be identical). The same script was also used to generate an additional marker for every syllable, which is the mean of the onset and end time; this marker is meant to approximate the perceptual center (P-center) of the syllable (*i.e.*, the moment it is *heard* as beginning, as opposed to its acoustic onset – see [14] and [15] for inherent limits on P-center representation as singular time-points). The processed syllable alignment – 121 markers for 50 syllables – was then imported into RX for manual correction: the onset and (when present) end markers were corrected visually using a sonogram (with aural checking when necessary); the P-centers were systematically checked aurally by converting them to clicks using Sonic Visualiser.

The whole process was completed in roughly over an hour (*i.e.*, two markers per minute): an acceptable benchmark, given that roughly half of the markers require attentive listening upon correction, which can be expected to decrease with further experience, and a far cry from the countless hours previously required for this task.

As previously mentioned, the corrected syllable alignment was then used to revise the transcription. In particular, it allowed for a more accurate transcription of pitches when paired with the automatically-extracted F₀. The resulting string of MIDI notes was then synchronized with

	s	r	d	c
Non- <i>legato</i> notes				
└ Number	4	6	10	2
└ Proportion (%)	20	46	59	22
└ Mean duration (ms)	213	162	161	180
std. dev. (ms)	(14.1)	(71.5)	(84.8)	(175)

Table 2. Non-*legato* notes by subsection in Taylor SWIFT’s “Blank Space”, measures 3–10.

the audio by inferring with the \mathcal{M}_{V2N} model, and the note alignment was processed with another R script to delete all notes *not* part of a melisma, as well as the first note of every melisma (which had already been aligned using the syllables). The 10 remaining notes were again imported in RX for manual correction and, because we consider them to represent P-centers, aurally checked afterwards¹⁰.

4.1.2. Musicological exploitation

With this timing data at hand – onset, P-center and end of every note in the excerpt – we are now able to propose an analysis of how Taylor SWIFT uses articulation and micro-rhythm to structure her vocals.

Legato articulation, or lack thereof, can be computed by subtracting the onset time of the $n + 1$ note with the end time of the n note. Table 2 displays the number, proportion, and mean duration (with standard deviation) of non-*legato* notes by subsection, painting a vivid picture of form organized through articulation. The excerpt begins with mostly *legato* singing, interspersed with a few very homogeneous silences. SWIFT’s vocals then gradually become more jagged – mostly non-*legato*, with many overall shorter, but also much more heterogeneous, silences – before returning to the initial *legato* articulation in the concluding melisma, which is split in two by the longest silence in the excerpt, at 304ms (almost half of a beat at 96BPM). Granted, this arch-like progression seems fairly obvious upon listening (especially when it has been explicitly pointed out beforehand), but we may very well have missed it had we not been able to gather accurate timing data.

Articulation is only part of the story, however. Figure 4 maps the duration of every note (*i.e.*, the difference between the P-center of the $n + 1$ note and that of the n note, to which is subtracted the length of the intervening silence, if there is one) to its P-center, with *legato* articulation shown with connecting lines, and the internal notes of each melisma (those aligned with the \mathcal{M}_{V2N} model) shown as unfilled dots; the vertical dashed lines correspond to the beginning of the four subsections, and the horizontal ones to the projected duration of eighth, sixteenth, and thirty-second notes (the three most frequent symbolic durations in the transcription) at 96BPM.

10. Gliding intonations, such as those in **d**, are heard as a single gesture and are thus not considered as instances of melisma (this hearing is reflected in the transcription by smaller notes, approximating the initial or final pitch of the glide, followed by glissando lines).

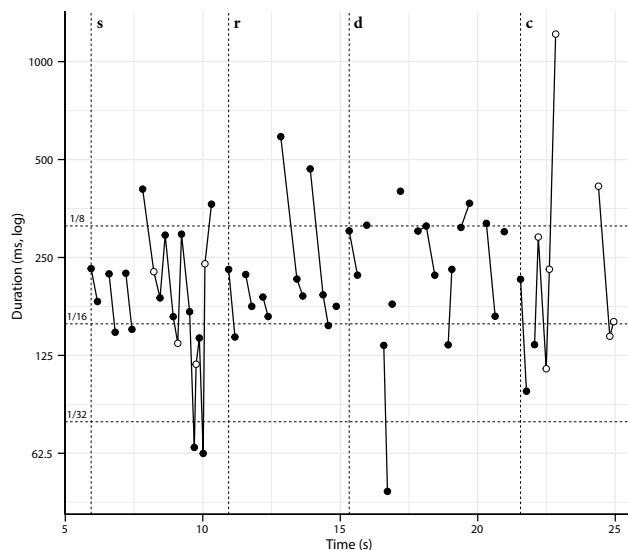


Figure 4. Note durations (log-scale) in Taylor SWIFT’s “Blank Space”, measures 3–10.

	s	r	d	c
Mean displacement (ms)	32.6	38.6	27.2	30.5
std. dev. (ms)	(38.4)	(27.7)	(9.82)	(29.8)

Table 3. Micro-rhythmic displacements by subsection in Taylor SWIFT’s “Blank Space”, measures 3–10.

Not all notes last for their projected duration. In particular, many sixteenth notes appear “uneven”, with the on-beat one being longer than the off-beat one. *Long/Short* subdivision [5] (*i.e.*, swing) is endemic to **s** and **r**, where it affects *almost* all sixteenth notes, but is absent in **d**, which prioritizes straight eighth notes.

We can thus highlight a subtle interplay between two contrasting *local* vocal styles: (1) mostly *legato*, with L/S (swung) sixteenth notes and step-wise motion (**sr** and **c**); and (2) mostly non-*legato*, with straight rhythms and large leaps filled with gliding intonations (**d**). The more jagged articulation of **r** allows SWIFT to smoothly transition from style (1) to style (2), while the leap of a major sixth coupled with L/S sixteenth notes on “wanna” at the beginning of **c** enables the reverse.

These local styles also share a number of characteristics, among which propulsive tendencies [5] in melismatic passages (*i.e.*, the notes are shorter than projected)¹¹ and lengthened notes when followed by a silence (both of these can be observed on Figure 4), a lack of vibrato (this can be observed on the automatically-extracted F0), and micro-rhythmic displacements of most notes (computed by subtracting their P-center with their projected onset at 96BPM), which consistently appear about 30ms *later* than projected, as shown in Table 3.

11. Because such propulsive tendencies are independent from the underlying pulse, which does *not* change, the last note of a melisma must last longer than projected, as “compensation”.

Thanks to the pipeline streamlining the annotation process, this analysis can easily be expanded upon, so that it encompasses the whole verse/verse-chorus unit/song, etc., up to (at least) the level of the album – thus shedding light on SWIFT’s multifaceted vocal style. The many strategies discovered during the analysis can then be compared with other artists’, and linked to, for example, the lyrics (do the L/S sixteenth notes connote out-of-breathness? do the few straight sixteenth notes constitute an early “mask-off” moment for the hysterical character portrayed by SWIFT in “Blank Space”? etc.).

4.2. Charles AZNAVOUR’s “La Bohème”

We now turn to a second example of musicological exploitation, on a fundamentally different repertoire that highlights another facet of the possible uses of the pipeline: French chanson, illustrated by an emblematic track from an equally emblematic artist: “La Bohème”, as sung by Charles AZNAVOUR in 1966. In addition to being in another language, the traditional French chanson aesthetic is markedly different from that of Anglo-Saxon pop music: the text’s primacy, and its vocal enhancement, are distinctive traits that differentiate it from almost all other genres.

While articulation and micro-rhythm in the context of a strict pulse played a central role in defining Taylor SWIFT’s style, French chanson also emphasizes other dimensions, such as paralinguistic effects involving the fundamental frequency. With this song’s nostalgic tone exploiting pathos as its main tool for seducing the listener, the rhetorical use of emphatic or euphemistic effects will be the focus of the following case study, without, however, developing all their potentialities (rhythmic placement, for example, will not be touched upon very much).

4.2.1. Application of the pipeline to the song

The first two steps in the pipeline – voice separation and F0 estimation – give particularly good results with this song. The separated voice file provides the musicologist with a working support of sufficient quality for both aural and visual analysis (including that of timbre, possibly the dimension most affected by the source separation process, but of which both the harmonic and noisy components are preserved). The quality of the source separation makes it possible to obtain an accurate fundamental frequency curve with the \mathcal{M}_{F0} model that does not require manual corrections beyond the occasional removing of the curve on silences and unvoiced consonants, whereas manual annotation as it was previously performed on a music file with accompaniment required between 30 minutes and 1 hour of work per song. Moreover, this automatically-extracted F0 is more precise than manually tracing the curve on the sonogram could ever be; it is thus an analytically reliable time saving method. The marker file resulting from the voice-to-text alignment allows the musicologist to instantaneously relate the effects heard, and observed on the sonogram, to the lyrics/syllables.

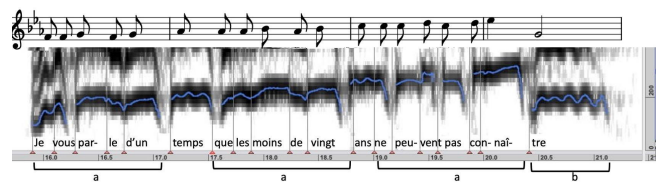


Figure 5. Verse 1, line 1 from Charles AZNAVOUR’s “La Bohème”. Score transcription, sonogram, F0 and text alignment (a = fast flow, intonational instability, and absence of vibrato; b = sustained note with vibrato on the last syllable of the phrase).

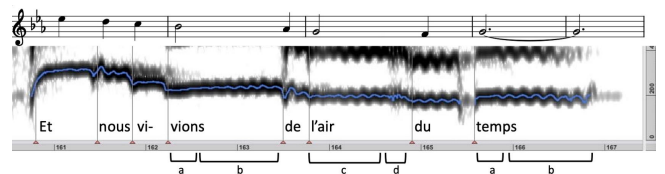


Figure 6. Chorus 3, line 4 from Charles AZNAVOUR’s “La Bohème”. Score transcription, sonogram, F0 and text alignment (a = beginning of a note without vibrato; b = arrival of the vibrato; c = vibrato on the note’s whole duration; d = rolled “r”).

4.2.2. Musicological exploitation

AZNAVOUR’s style is characterized by the coexistence of two antagonistic, yet complementary, modes of delivery: (1) a narrative everyday-speech-like phrasing defined by a fluctuating intonation, rhythmic irregularity, and a lack of vibrato; and (2) a lyrical phrasing marked by sustained notes with vibrato, stable intonation, and rhythmic precision. The first one is shown in Figure 5, with its constantly rising and falling intonation over step-wise motion paired with a fast flow of eighth notes, and is found in verses. The second one appears in verses as well, on the last syllables of phrases, but is most characteristic of choruses, where rolled “r”s further pull AZNAVOUR’s delivery away from the spoken voice and towards the singing voice. See Figure 6: the pitch remains stable throughout the notes’ duration, while vibrato, which is under very fine control, is brought in gradually on the most meaningful words – an emphatic effect. Consonants are articulated and distinct, but are somewhat temporally and dynamically euphemized, giving the choruses a resolutely vowel-like character.

While this prosody-based formal structure is typical of the French chanson genre, it takes on a particular significance with AZNAVOUR. An important axis of his performance strategy is the pairing of profound musicality (expressed through intonative and rhythmic precision) with an apparent economy of means – a sobriety associated first and foremost with his well-known veiled tone, which he knew how to make great use of in the service of pathos, evoking the tragedy of the everyday life closely related to his songs’ themes.

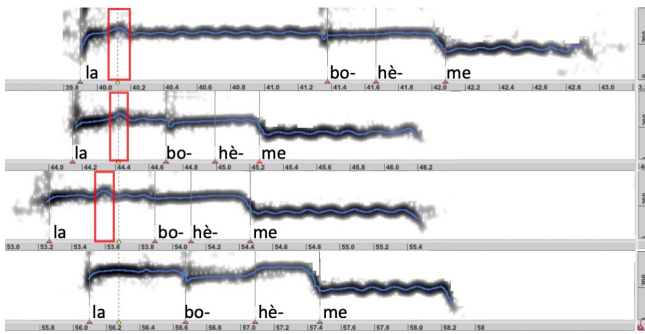


Figure 7. Synchronization of the 4 iterations of the “la bohème”-motif from the first chorus of Charles AZNAVOUR’s “La Bohème”. Sonogram, F0 curve and text alignment. The red frame highlights the position of the trill.

Another characteristic element of AZNAVOUR’s singing style is the presence of vocal ornamentation, more specifically trills and appoggiaturas. “La Bohème” contains, to varying degrees of subtlety, about 24 such effects, which are found mostly in the lyrical parts: choruses and last lines of verses. One notable use of the trill in this song is to create variations throughout the 16 iterations of the title-phrase “la bohème”, which appears 4 times per chorus on a single intonative scheme (transposed to different starting pitches) and rhythmic formula (three quarter notes followed by a half-note).

Figure 7 shows the first 4 iterations of the “la bohème”-motif (*i.e.*, those of the first chorus), synchronized in order to visualize both the distribution of the trill within the iterations and its rhythmic position. The synchronization was done not according to the onset of the first syllable, but to the rhythmic pulse of the beat as it materializes in the instrumental accompaniment, allowing the consideration of possible agogic (*i.e.*, micro-rhythmic) shifts.

5. CONCLUSION

In this paper, we introduced a complete pipeline for the musicological analysis of singing voice style. From a technical perspective, deep learning models were used for singing voice extraction from background music, voice parameter (F0) estimation, and robust automatic alignment of both syllables and notes to the audio. Not only does this pipeline greatly simplify the tedious tasks traditionally done manually by musicologists, but it also offers practical flexibility, as the two concrete case studies demonstrate: (1) text and note alignments allowed investigating articulation and micro-rhythm in a Taylor SWIFT song (American pop, 2014); and (2) F0 curves were exploited to highlight vocal phrasing and rhetorical effects involving intonation in a Charles AZNAVOUR song (French chanson, 1966). More generally, this work paves the way for future and strong collaborations between musicologists and deep learning researchers sharing a common interest in singing voice. The tools presented will be made available to the community in the form of a web interface.

6. ACKNOWLEDGEMENTS

This work has been funded by the ANR (French National Research Agency) project ANR-19-CE38-0001-01 ARS (Analysis and tRansformation of Singing style).

7. REFERENCES

- [1] Ardaillon, L. *Synthesis and expressive transformation of singing voice*. PhD thesis, Université Pierre et Marie Curie-Paris VI, 2017.
- [2] Ardaillon, L., Chabot-Canet, C., and Roebel, A. “Expressive control of singing voice synthesis using musical contexts and a parametric F0 model”. In *Proc. Interspeech 2016*, pages 1250–1254, 2016.
- [3] Ardaillon, L. and Roebel, A. “Fully-convolutional network for pitch estimation of speech signals”. In *Proc. Insterspeech 2019*, pages 2005–2009, 2019.
- [4] Caporaletti, V. *I processi improvvisativi nella musica: un approccio globale*. Libreria Musicale Italiana, 2005.
- [5] Caporaletti, V. *Swing e Groove: sui fondamenti estetici delle musiche audiotattili*. Libreria Italiana Musicale, 2014.
- [6] Chabot-Canet, C. *Léo Ferré: une voix et un phrasé emblématiques*. L’Harmattan, 2008.
- [7] Chabot-Canet, C. “L’analyse spectrale au fondement d’une rhétorique des styles interprétatifs dans la chanson française”. *Volume!*, 16(2)/17(1): 29–47, 2020.
- [8] Chabot-Canet, C. “L’interprétation de Barbara: l’expressivité de l’infime dans la performance vocale”. In Bost, S. and Douzou, C. (ed.), *Barbara en scène(s)*, pages 105–130. Presses Universitaires de Provence, 2022.
- [9] Chabot-Canet, C., Ardaillon, L., and Roebel, A. “Analyse du style vocal et modélisation pour la synthèse de chant expressif: l’exemple d’Édith Piaf”. *Volume!*, 16(2)/17(1): 63–85, 2020.
- [10] Choi, W., Kim, M., Chung, J., Lee, D., and Jung, S. “Investigating U-Nets with various intermediate blocks for spectrogram-based singing voice separation”. *arXiv preprint arXiv:1912.02591*, 2019.
- [11] Cogan, R. *New Images of Musical Sound*. Harvard University Press, 1984.
- [12] Cohen-Hadria, A., Roebel, A., and Peeters, G. “Improving singing voice separation using Deep U-Net and Wave-U-Net with data augmentation”. In *EU-SIPCO*, pages 1–5. IEEE, 2019.
- [13] Cont, A., Schwarz, D., Schnell, N., and Raphael, C. “Evaluation of real-time audio-to-score alignment”. In *ISMIR*, 2007.
- [14] Danielsen, A., Nymoen, K., Anderson, E., Câmara, G. S., Langerød, M. T., Thompson, M. R., and London, J. “Where is the beat in that note? Effects of

- attack, duration, and frequency on the perceived timing of musical and quasi-musical sounds”. *Journal of Experimental Psychology: Human Perception and Performance*, 45(3): 402–418, 2019.
- [15] Danielsen, A., Nymoén, K., Langerød, M. T., Jacobsen, E., Johansson, M., and London, J. “Sounds familiar(?): expertise with specific musical genres modulates timing perception and micro-level synchronization to auditory stimuli”. *Attention, Perception, & Psychophysics*, pages 1–17, 2021.
- [16] Doras, G., Teytaut, Y., and Roebel, A. “A linear memory CTC-based algorithm for text-to-voice alignment of very long audio recordings”. *Applied Sciences*, 13(3): 1854, 2023.
- [17] Everett, W. *The Foundations of Rock: From “Blue Suede Shoes” to “Suite: Judy Blue Eyes”*. Oxford University Press, 2009.
- [18] Fónagy, I. *La Vive Voix: essais de psychophonétique*. Payot, 1983.
- [19] Gao, X., Gupta, C., and Li, H. “Automatic lyrics transcription of polyphonic music with lyrics-chord multi-task learning”. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 30: 2280–2294, 2022.
- [20] Graves, A., Fernández, S., Gomez, F., and Schmidhuber, J. “Connectionist Temporal Classification: labelling unsegmented sequence data with recurrent neural networks”. In *Proceedings of the 23rd international conference on Machine learning*, pages 369–376, 2006.
- [21] Guillotel-Nothmann, C. “Les signes musicaux et leur étude par l’informatique: le statut épistémologique du numérique dans l’appréhension du sens et de la signification en musique”. *Revue musicale OI-CRM*, 6(2): 45–72, 2020.
- [22] Hainaut, B. *Le style black metal*. Aedam Musicae, 2017.
- [23] Hanninen, D. A. *A Theory of Music Analysis: On Segmentation and Associative Organization*. University of Rochester Press, 2012.
- [24] Heidemann, K. “A system for describing vocal timbre in popular song”. *Music Theory Online*, 22(1), 2016.
- [25] Henrich Bernardoni, N. “La voix timbrée dans les chansons: considérations physiologiques et acoustiques”. *Volume!*, 16(2)/17(1): 49–61, 2020.
- [26] Lacasse, S. “The phonographic voice: paralinguistic features and phonographic staging in popular music singing”. In Bayley, A. (ed.), *Recorded Music: Performance, Culture and Technology*, pages 225–251. Cambridge University Press, 2010.
- [27] Lacheret-Dujour, A. and Beaugendre, F. *La Prosodie du français*. CNRS, 1999.
- [28] Lancaster, E. P. and Souviraà-Labastie, N. “A frugal approach to music source separation”, 2020.
- [29] Le Vot, G. *Poétique du rock: oralité, voix et tumultes*. Minerve, 2017.
- [30] Léon, P. *Précis de phonostylistique: parole et expressivité*. Nathan Université, 1993.
- [31] Malawey, V. *A Blaze of Light in Every Word: Analyzing the Popular Singing Voice*. Oxford University Press, 2020.
- [32] Meseguer-Brocal, G., Cohen-Hadria, A., and Peeters, G. “Creating DALI, a large dataset of synchronized audio, lyrics, and notes”. *Transactions of the International Society for Music Information Retrieval*, 3(1): 55–67, 2020.
- [33] Mitsufuji, Y., Fabbro, G., Uhlich, S., Stöter, F.-R., et al. “Music Demixing Challenge 2021”. *Front. Signal Process.*, 1, 2022.
- [34] Neal, J. “The twang factor in country music”. In Fink, R., Latour, M., and Wallmark, Z. (ed.), *The Relentless Pursuit of Tone: Timbre in Popular Music*, pages 43–64. Oxford University Press, 2018.
- [35] Nobile, D. “Counterpoint in rock music: unpacking the ‘melodic-harmonic divorce’”. *Music Theory Spectrum*, 37(2): 189–203, 2015.
- [36] Petit, A. “Le travail de la pop: écouter 1989 (2014) de Taylor Swift en musicologue”. Colloque *Le Silence du mainstream*, Strasbourg, 2022. <https://www.canal2.tv/video/16316> (accessed February, 26th 2022).
- [37] Poyatos, F. *Nonverbal Communication across Disciplines*. John Benjamins Publishing, 2002.
- [38] Rafii, Z., Liutkus, A., Stöter, F.-R., Mimitakis, S., and Bittner, R. “The MUSDB18 corpus for music separation”, 2017.
- [39] Sundberg, J. *The Science of the Singing Voice*. Northern Illinois University Press, 1987.
- [40] Teytaut, Y., Bouvier, B., and Roebel, A. “A study on constraining Connectionist Temporal Classification for temporal audio alignment”. *Proc. Interspeech 2022*, pages 5015–5019, 2022.
- [41] Vaysse, R., Astésano, C., and Farinas, J. “Performance analysis of various fundamental frequency estimation algorithms in the context of pathological speech”. *J. Acoust. Soc. Am.*, 152(5): 3091–3101, 2022.
- [42] Weiß, C. and Peeters, G. “Learning multi-pitch estimation from weakly aligned score-audio pairs using a multi-label CTC loss”. In *2021 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 121–125. IEEE, 2021.
- [43] Wichmann, A. “The attitudinal effects of prosody, and how they relate to emotion”. In *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*, 2000.

ANALYSE INTERDISCIPLINAIRE DE LA SYNTHÈSE INSTRUMENTALE : ÉTUDE DE CAS AVEC *PARTIELS* DE GRISEY

Charles de Paiva Santana¹, Vincent Tiffon¹, Mylène Gioffredo^{1,3}, Micael Antunes^{1,2}, Javier Elipe Gimeno¹, Samuel Poirot¹, Sølvi Ystad¹

¹Aix-Marseille Univ, PRISM-CNRS

²NICS/IA- University of Campinas

³InCIAM (Institut Créativité et Innovation, Aix-Marseille Univ)

{depaiva, tiffon, gioffredo, elipe-gimeno, poirot, ystad}@prism.cnrs.fr
michaelant@gmail.com

RÉSUMÉ

Cet article présente les fondements d'une enquête sur la technique de composition de la synthèse instrumentale (SI), laquelle est une adaptation de la synthèse additive qui utilise des instruments orchestraux classiques pour créer de nouvelles sonorités. L'étude se concentre sur l'analyse de la SI à travers une étude de cas, *Partiels* de Gérard Grisey [6]. Des paramètres primaires et des descripteurs symboliques ont été distingués pour identifier les aspects calculables de la SI et le rôle de l'intuition dans le processus de composition. Un modèle de SI a été développé sur la base d'informations extraites de partitions musicales et des simulations ont été réalisées à l'aide d'OpenMusic et d'échantillons orchestraux. Différentes approximations de fréquences réelles dans la série harmonique ont été testées, notamment des $\frac{1}{2}$, $\frac{1}{4}$ et $\frac{1}{8}$ de ton. La précision du modèle a été évaluée en comparant les résultats avec les partitions musicales originales. Cette étude contribue à une meilleure compréhension de la SI et de son potentiel d'utilisation dans la composition de musique contemporaine. L'objectif final est de mener des expériences futures en utilisant des musiciens pour simuler la SI, ce qui permettra de mieux comprendre le potentiel de cette technique.

1. INTRODUCTION

1.1. Musique spectrale, synthèse additive, synthèse instrumentale

La "musique spectrale", dont on doit l'appellation à Hugues Dufourt [4], est née historiquement de l'alliage entre science, technologie et musique (AST). Un groupe de jeunes compositeurs, pour la plupart élèves d'Olivier Messiaen au Conservatoire de Paris (autour de l'année

1970) ont souhaité réintroduire la perception dans la composition. Par ailleurs, Gérard Grisey s'est formé à l'acoustique musicale auprès d'Emile Leipp (au LAM, aujourd'hui intégré dans l'Institut d'Alembert, Paris). Cette mise en valeur de la perception, non exclusive à ces compositeurs, s'inscrit en opposition à la supposée ou réelle hégémonie de la musique sérielle dans les années cinquante, dont les enjeux esthétiques étaient alors perçus, à tort ou à raison, comme un fétichisme du système au détriment du résultat sonore et de la perception. Des compositeurs précurseurs comme Radulescu, Tenney, Risset, des personnalités aussi différentes esthétiquement que Grisey, Murail, Dufourt, Lévinas, Tessier réunis au sein de plusieurs collectifs dont l'*Ensemble Itinéraire* (créé en 1973), auxquels s'ajoutent notamment Vivier, Dumitrescu (hors *Itinéraire*), ont ainsi créé de nouveaux processus compositionnels dont les histoires de la musique relèvent habituellement au moins quatre grands principes, largement imbriqués : 1) l'hybridation harmonie/timbre qui met en scène pour la perception une ambiguïté entre des catégories relevant de théories musicales classiques et des notions de physique acoustique ainsi que des modèles technologiques issus des manipulations électroacoustiques de studio ; 2) la notion de synthèse instrumentale¹, i.e. le transfert de procédés électroacoustiques dans le champ de l'orchestration instrumentale induisant un processus perceptif de fusion ; 3) la notion de musique liminale (initialement consensuée par G. Grisey), jouant sur les effets de seuil perceptif ; et 4) les processus de transformations continues.

Le point de départ de cette investigation est l'exploration du second domaine, la notion de SI proposée par Gérard Grisey (1946-1998). Cette technique, en référence à la synthèse additive, vise à créer des textures sonores en transcrivant les paramètres trouvés dans l'analyse temps-fréquence des sons sur une partition

¹ Cette technique, qui relève de la transcription, n'est pas le fait unique des musiciens dits spectraux, mais est partagée par d'autres compositeurs, comme par exemple F.B. Mâche [7, 3]

orchestrale [8]. La synthèse additive est une technique de création sonore par l'ajout d'ondes sinusoïdales de fréquences variées et selon des comportements dynamiques dans le temps également variables. De son côté, la SI permet l'émergence d'une nouvelle qualité orchestrale en la "synthétisant" avec des sons orchestraux [11]. Cette nouvelle qualité engendre un état qui, selon Grisey, est "un être hybride de notre perception, un son qui, sans être encore un timbre, n'est déjà pas tout à fait un accord" [9]. Les principales compositions qui utilisent le procédé de SI sont présentées dans la Table 1.

Œuvre	Année	Modèle Acoustique
Gérard Grisey, <i>Périodes</i>	1974	son de spectre trombone
Gérard Grisey, <i>Partiels</i>	1975	son de spectre théorique (<i>mi grave</i>)
Gérard Grisey, <i>Modulations</i>	1976	son de cuivres
Tristan Murail, <i>Gondwana</i>	1980	sons de cloches
Gérard. Grisey, <i>Transitoires</i>	1980-1981	note de contrebasse (corde <i>mi</i>) jouée par différents modes d'exécution (7 au total)
Tristan Murail, <i>Sillages</i>	1985	klaxons des tramways de Kyoto.
Marc-André Dalbavie, <i>Les Miroirs transparents</i>	1985	
Gérard Grisey, <i>L'Îcône Paradoxe</i>	1992 - 1994	sons vocaux (lectures des signatures de Piero de la Francesca et extraits de son traité de perspective, <i>De perspectiva pingendi</i> .)
Tristan Murail, <i>Le Partage des eaux</i>	1995	
Jonathan Harvey, <i>Speakings</i>	2008	sons vocaux
Tristan Murail, <i>Les Neiges d'antan</i>	2019	cloches des troupeaux de vaches

Table 1. Liste non exhaustive de compositions utilisant la synthèse instrumentale.

1.2. Objectifs

L'utilisation des techniques électroniques de studio a eu un impact significatif sur la musique instrumentale moderne. Notre étude vise à une meilleure compréhension de sa signification et de son impact. En outre, nous cherchons à mettre en lumière le travail et l'héritage des compositeurs qui ont participé à l'École spectrale française. En analysant des compositions telles que *Partiels* de Gérard Grisey et *Désintégrations* de Tristan Murail [9], cette étude vise à apporter un nouvel éclairage sur les compositions qui utilisent les techniques technomorphologiques [16], en particulier la SI.

Cette étude pourrait être utile aux compositeurs et aux étudiants en quête de moyens permettant d'étendre et de transformer les techniques du passé dans le but de proposer de nouvelles formes d'écriture compositionnelle. Elle peut donner un aperçu concret de la façon dont le timbre et le son peuvent servir de point de départ à un idiome compositionnel. Nous espérons également que notre travail pourra aider les musiciens à aborder les partitions musicales qui utilisent la technique de la SI. Cela peut offrir de nouvelles perspectives sur la dualité entre imagination et algorithme, ou "décision et automatisme" qui se manifestent dans la création contemporaine.

1.3. Questions de recherche et hypothèses

Comment décrire analytiquement un passage de musique écrite qui utilise la SI ? Notre première hypothèse est que les caractéristiques timbrales et structurelles des objets sonores de SI dans la musique spectrale peuvent être décrites dans le domaine symbolique par des mesures quantitatives/statistiques tels que la densité des partiels, les partiels les plus renforcés etc. Par l'analyse du signal sonore, d'autres aspects liés à la structure musicale tels que forme de l'onde ou le contenu spectral peuvent être mis en évidence à travers des descripteurs de timbre (e.g. centroid spectral, flux spectral, forme de l'enveloppe spectrale, ...). Néanmoins, même si ces descripteurs permettent de montrer des variations d'interprétation à travers différents enregistrements, les effets perceptifs chez les auditeurs sont difficiles à prédire.

Si la synthèse additive nécessite un certain nombre de paramètres, quels sont les paramètres de la SI ? À l'exception de la phase, la plupart des paramètres de synthèse additive peuvent être utilisés pour modéliser la SI. Cependant, notre seconde hypothèse est que le choix des instruments utilisés dépend largement de l'intuition du compositeur et de son savoir-faire comme de contraintes externes telles que l'instrumentation de l'œuvre commandée, les interprètes, et la salle de concert de la première exécution publique.

Dans un tel contexte où intuition, modèle acoustique et métier d'orchestrateur s'interpénètrent, est-il possible de déterminer exhaustivement les paramètres significatifs pour modéliser la SI, pourrions-nous les utiliser pour recomposer des musiques existantes (vocation pédagogique) et/ou générer algorithmiquement de

nouvelles proposition de musique spectrale (vocation artistique)? En troisième hypothèse, nous pensons pouvoir modéliser et simuler algorithmiquement la technique de la SI, à l'instar de ce que font déjà les logiciels d'orchestration assistée tels qu'Orchidea. Mais comment valider la méthodologie et les résultats analytiques ? Un modèle de SI peut être validé en faisant des simulations et en utilisant des outils analytiques ou des tests d'écoute expérimentaux pour estimer la précision des résultats simulés par rapport aux exemples originaux. Cette phase de travail est en cours de réalisation dans le cadre du projet ANTISCOPE décrit dans les "remerciements" (partie 5).

2. METHODOLOGIE

Cette étude vise à décrypter les rouages de la SI et à développer une méthodologie pour analyser les techniques de composition qui la rendent possible (Figure 1).

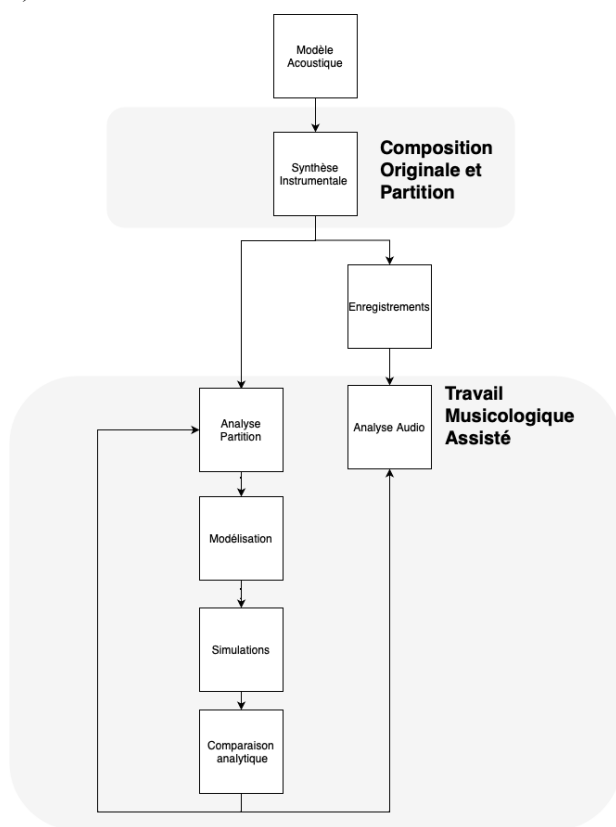


Figure 1. Représentation graphique de la méthodologie.

Sur la base des informations extraites des partitions musicales, un modèle de SI a été développé. Une fois le modèle développé, il a été utilisé pour simuler la technique à l'aide de logiciels comme *OpenMusic* et d'échantillons orchestraux trouvés dans la station de travail audio numérique *Logic Pro*. À ces différentes étapes s'ajoute l'analyse des enregistrements (Table 2) avec des descripteurs audio : l'ensemble de ce travail peut être qualifié de travail musicologique assisté par

ordinateur. Les expériences ont été menées en utilisant différentes approximations des fréquences réelles de la série harmonique, y compris des $\frac{1}{2}$, $\frac{1}{4}$ et $\frac{1}{8}$ de ton. Ces simulations ne donnent qu'un aperçu de ce qui peut être accompli à l'avenir, car l'objectif ultime est de mener des expériences en faisant appel à de vrais musiciens pour simuler les SI. Cela permettra de mieux comprendre le fonctionnement de cette technique de composition unique et la façon dont elle peut être utilisée pour créer de la musique nouvelle. La précision du modèle a ensuite été évaluée en comparant les résultats avec les partitions musicales originales.

En parallèle à ce travail de modélisation et de simulation, et forts des résultats obtenus, nous avons procédé à un travail comparatif sur la base de descripteurs audio de trois enregistrements de *Partiels* de Grisey. Les trois enregistrements sélectionnés sont les suivants :

Fig 3	Album	Interprètes	Editeur
B	Gérard Grisey, <i>Les Espaces Acoustiques</i>	Gérard Caussé (alto), Ensemble Court-Circuit / Pierre-André Valade (direction), Frankfurter Museumorchester / Sylvain Cambreling (direction)	CD Accordia 4 1999.
C	MusikFabrik Edition 13. Crossings	Nicolaou, Schöllhorn, Grisey, Mack, Ensemble Musikfabrik, Emilio Pomarico,	CD Wer WER68
D	Gérard Grisey, <i>Les Espaces acoustiques</i> ,	Garth Knox (alto), Asko Ensemble, WDR Sinfonieorchester Köln / Stephan Asbury (direction)	CD Kair 001242 2005.

Table 2. Sélection des enregistrements phonographiques de *Partiels* de Grisey.

3. RESULTATS

3.1. Que montre l'analyse de la première section de partiel ?

3.1.1 Identité de *Partiels* de Grisey

Partiels est une composition pour 18 musiciens, écrite en 1975 et créée en 1976. Elle fait partie d'un cycle de compositions appelé *Les Espaces Acoustiques*. Le compositeur a mis plus de dix ans pour composer ce cycle d'une durée d'une heure et trente minutes. La première pièce est pour instrument solo (alto) et la dernière pièce pour quatre solistes (quatre cors) et grand orchestre. Chaque section de *Partiels* est le résultat d'un procédé de composition différent. Certaines d'entre elles sont de nature technomorphologique. Alors que la section 2 explore l'harmonie et le rythme générés à partir de *sons de combinaison*, la section 1 se concentre sur la SI d'un son *mi* grave d'un spectre théorique. En tant que telle, cette analyse se concentrera uniquement sur cette première section.

La première section de *Partiels* a été étudiée en détail par François-Xavier Féron [6]. Elle peut être divisée en onze sous-sections. Chaque sous-section correspond à un chiffre de répétition dans la partition, qui a été segmentée par le compositeur lui-même.

Lorsque la deuxième pièce du cycle *Les espaces acoustiques* intitulée *Périodes* est jouée sans interruption

avec *Partiels*, elles partagent une section commune. Ainsi, la dernière page de *Périodes* sert de section initiale à *Partiels*, où l'on peut entendre la SI : cette partition est l'une des pages les plus reconnaissables de la musique spectrale.

3.1.2 Plan formel

La première page renommée de *Partiels*, écrite en notation proportionnelle et ressemblant à un sonagramme, est répétée plusieurs fois avec de légères variations de durée. Plus qu'une simple transcription d'un modèle acoustique ou d'un sonagramme, le début de *Partiels* présente un geste orchestral original qui peut être subdivisé en sous-sections ou moments correspondant à ses quatre mesures.

Le premier moment est un motif pratiquement périodique sur la contrebasse et le trombone, formant une anacrouse pour la SI. Celle-ci elle-même peut être subdivisée en trois moments correspondant à l'enveloppe d'amplitude d'un son acoustique : fondamentale + transitoires d'attaque, régime de maintien et transitoires d'extinction (Table 3, Figure 2).

Chiffre	Fund + Trans. Att.	Régime de maintien	Trans. d'extinction	Total
#1	4	9	3,75	16,75
#2	4	10	3,75	17,75s
#3	3,5	8,5	4,5	16,5
#4	3	8	4,25	15,25
#5	3,5	9,75	4,25	17,50
#6	3,25	7,75	4,5	15,5
#7	2,75	8,25	5,25	16,25
#8	2,5	7,5	5,25	15,25
#9	2	9	6	17
#10	1,5	9,5	6,75	18,75
#11	1	9	9	19

Table 3. Durées des sous-sections de la première section relevées sur la partition Ricordi [10, p. 1-14, avec les proportions attaque/maintien/extinction (durées exprimées en secondes).

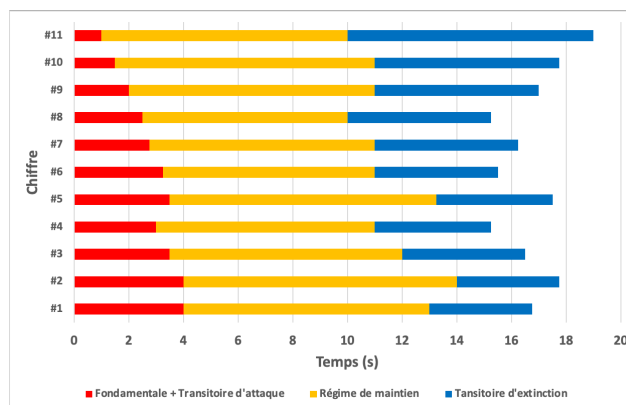
Bien que la SI ait offert une nouvelle opportunité pour la composition musicale en fournissant une expérience auditive qui n'était ni strictement tonale ni atonale et en formant la base d'une musique où le "son" était le matériau principal plutôt que les notions traditionnelles de "note", elle posait également un défi compositionnel. Une fois ce geste équilibré présenté et pouvant être répété plusieurs fois sans inconfort, comment pourrait-il être développé en un discours musical ?

Il convient de mentionner notre choix de travailler sur des valeurs discrètes pour "coller" à la partition [10]

(figure 3, ligne A) qui propose, *de facto*, presque exclusivement des valeurs discrètes. La nature du signal audio présuppose des dynamiques continues, alors que la partition fonctionne en termes de dynamiques discrètes.

Figure 2. Représentation graphique de la table 2.

La figure 3 montre l'écart interprétatif entre les trois versions enregistrées (figure 3, lignes B, C, D) par rapport aux données initiales fournies par la partition [10] (figure 3, ligne A). Les écarts sont *a minima* de deux ordres :



distorsion temporelle, par un élargissement dans la version D ; écart dynamique, par disposition spectrale différentes d'une interprétation à l'autre. Si l'élargissement temporel s'explique difficilement du fait de la précision de la partition, la question des dynamiques est plus compréhensible, du fait du caractère théorique des indications symboliques dans la partition, et des différences en termes de prise de son des trois enregistrements. Ici se joue l'adéquation entre les intentions du compositeur quant au modèles spectraux et la nécessaire interprétation des musiciens et du preneur de son, en connaissance de cause ou non des intentions du compositeurs et des connaissances en termes de physique acoustique. L'analyse détaillée de ces écarts et les conclusions qui s'y rapportent seront l'occasion d'une prochaine publication.

3.1.3 De l'harmonicité vers l'inharmonicité

La première section de *Partiels* est, en quelque sorte, une longue séquence d'objets de SI. Pour construire la forme et le discours à partir de ce matériau de base, le compositeur utilise une méthode qui consiste à transformer l'objet initial d'un état d'harmonicité et de périodicité vers l'inharmonicité et l'apériodicité.

Grisey *Partiels* Première section :
 Comparaison de la forme et de la durée entre la partition
 et trois enregistrements différents

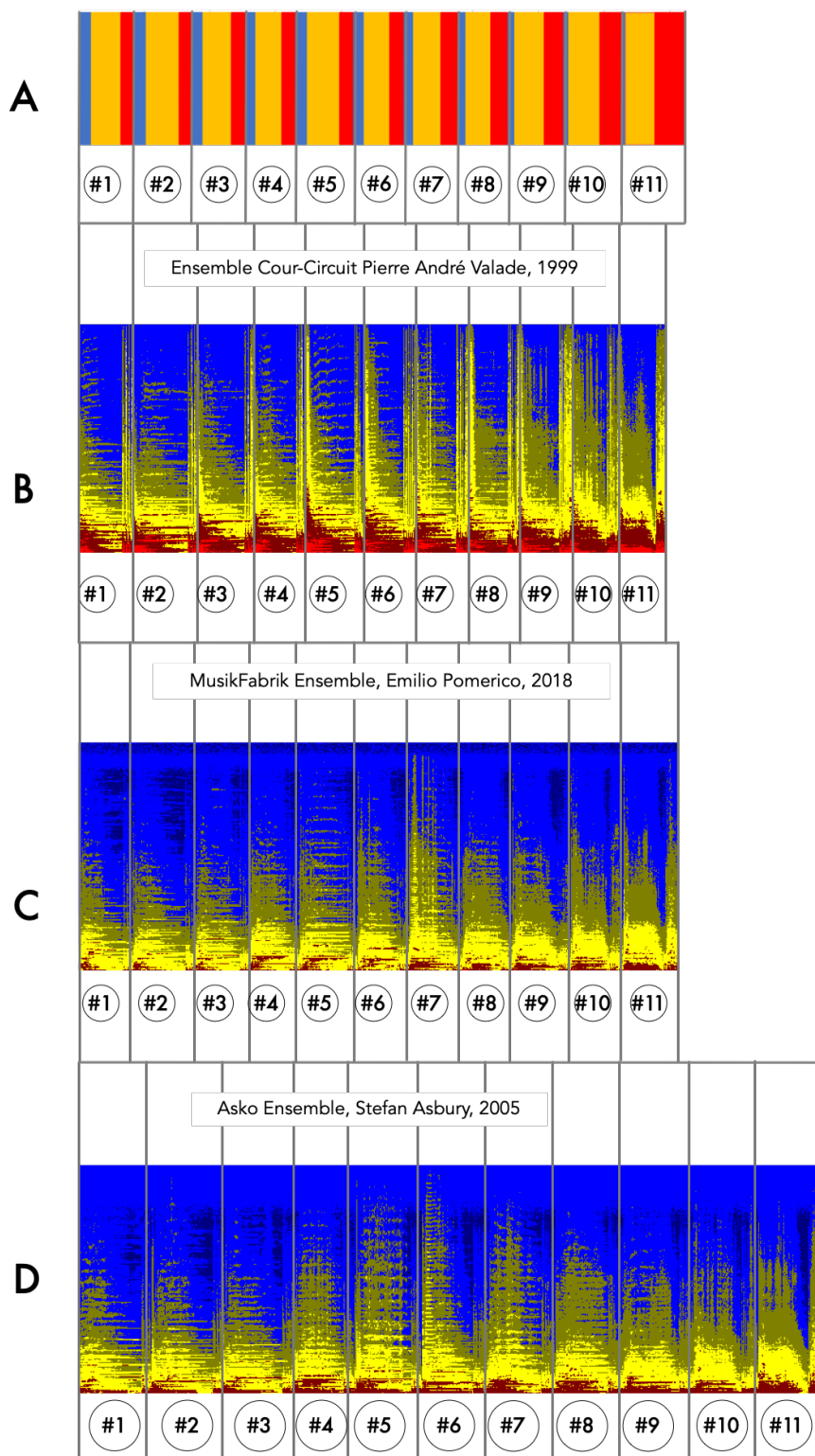


Figure 3. Représentation de l'information symbolique de la première section (A) au regard de trois sonagrammes de trois interprétations enregistrées (B-C-D) disponibles en CD (cf. Table 2).

Grisey compartimente le spectre synthétisé en plusieurs zones de fréquences différentes: zones hors formants et zones formantiques. Ces zones reçoivent un traitement orchestral différencié (en les associant à une famille d'instruments particulière, par exemple), et font l'objet de transformations. Du point de vue formel, ces dernières participent à la trajectoire de la musique de l'harmonie à l'inharmonie.

La zone hors formant est la base du spectre synthétique subissant le moins de transformations et qui maintenant une certaine identité du spectre. Dans le cas de *Partiels*, cette zone est principalement confiée à la clarinette, au trombone, au cor et aux cordes. Grisey fait encore une autre distinction entre les fréquences du spectre en dehors du formant et les "transitoires", mais à ce stade de la recherche nous n'avons pas encore travaillé avec cette distinction. Dans *Partiels*, Grisey distingue deux zones formantiques qui correspondent grosso modo aux partiels pairs et impairs confiés pour l'essentiel aux bois et aux percussions. Cette zone formantique se charge progressivement vers le grave et s'enrichit d'inharmoniques. Ces dernières, dans le cas de notre modèle, sont représentées par des nombres fractionnaires.

L'approche de Grisey pour transformer le matériau est perceptible pour l'auditeur et repose sur un ensemble d'opérations très schématiques, bien qu'elle permette également une grande liberté en termes de créativité de la part du compositeur. À chaque répétition du trombone synthétique, Grisey transpose certains des partiels supérieurs vers une ou plusieurs octaves inférieures, ce qui augmente la prévalence des battements de fréquence. Parallèlement, il demande des techniques de jeu étendues de plus en plus extrêmes. Le motif joué par la contrebasse devient apériodique, voire complètement aléatoire et bruyant. La subtilité des attaques de triple piano qui ouvrent la composition cède progressivement la place à des sons de plus en plus agressifs.

De manière évidente, les principales particularités de la SI reposent sur la combinaison de plusieurs critères : tout d'abord, le filtrage effectué dans les rangs de partiels d'une série harmonique donnée (une composition pour 18 musiciens a tout intérêt à pouvoir écrire des spectres qui vont bien au-delà du 18ème partiel du fondamental donné) ; le choix des instruments ; leur mode de jeu ou d'articulation respective (le compositeur optimisera souvent, comme on peut facilement le conjecturer, l'orchestration pour créer un spectre plus riche en exploitant les qualités timbrales de sa palette orchestrale) ; l'indétermination qui découle de la tentative de reproduire un modèle acoustique avec des composants complexes (sons orchestraux) qui résisteront à la fusion et produiront de nouvelles sonorités inattendues.

Le tableau ci-dessous (Table 4) présente les paramètres du modèle de SI qui nous servira de base pour une pratique analytique mais aussi pour une implémentation informatique qui nous permet d'exécuter des simulations d'objets de SI.

Paramètre	Type	Description
Fondamentale	N. Entier (Hz)	La fréquence fondamentale du spectre produit
Rangs de la zone "hors-formants" ²	Liste de flottants	Les harmoniques présents dans le spectre, exprimés en multiples de la fréquence fondamentale
Rangs de la zone formantique ³	Liste de flottants	Les partiels harmoniques et inharmoniques du spectre. Ils sont exprimés par des nombres décimaux
Attaques	Liste d'entiers	Les moments où les partiels commencent à être joués
Durée totale	Liste d'entiers	La durée pendant laquelle les harmoniques sont joués
Intensité initiale	Liste d'entiers	La nuance initiale avec laquelle le partiel est joué
Intensité maximale	Liste d'entiers	Les nuances la plus forte atteinte pendant la durée du partiel
Intensité minimale	Liste d'entiers	Les nuances avec laquelle la note s'éteint
Instruments de l'orchestre	String / Entier correspondant au canal MIDI	La liste des instruments de l'orchestre utilisés pour produire les partiels
Durée du transitoire d'attaque	Entier (ms)	Proportion de la durée totale attribué pour la phase initiale du spectre
Durée du régime de maintien	Entier (ms)	Proportion de la durée totale attribué pour la phase de maintien du spectre
Durée de la résonance	Entier (ms)	Proportion de la durée totale attribué pour la phase d'extinction du spectre
Modes de jeu	String / Entier (canal MIDI)	Les modes de jeu utilisés pour produire les partiels, comme la position de l'archet pour un violon (<i>Sul tasto</i> , <i>Sul ponticello</i>)
Articulation	String / Entier (canal MIDI)	Certains instruments permettent le découplage entre mode de jeu et articulation

Table 4. Liste des paramètres avec les types correspondants servant de base à une modélisation informatique de la synthèse instrumentale.

3.2. Intuition et données moins formalisables

Pour comprendre la réalité concrète de la SI et son supposé impact sur la perception, d'autres aspects non

² Voir 3.1.3

³ Voir 3.1.3

négligeables et moins facilement objectivables doivent être identifiés et intégrés dans le processus analytique. On peut citer pêle-mêle : 1) la manipulation de la forme de l'objet de synthèse (transitoires d'attaque, maintien, extinction) comme indiqué dans la Table 1 ; 2) le choix de l'instrumentation (qui relève du "métier" de compositeur, singulièrement dans l'esthétique dite "spectrale" ou "liminale") ; 3) l'attribution et le rôle des instruments (via leurs spécificités timbrales, leurs modes de jeu, etc.) dans la formation des spectres ; 4) les procédures de renforcement ou de densifications de certains partiels par l'ajout de fréquences supplémentaires (comme par exemple l'accordéon dans *Partiels*) ; 5) les stratégies de création d'un taux d'inharmonicité des spectres synthétiques instrumentaux (en tenant compte également de leur place dans l'ambitus du spectre).

Rappelons que Grisey travaille sur trois zones de partiels⁴ : deux zones de type formantique (qui conditionnent le glissement des partiels vers le grave) et une troisième hors formantique. Ce choix de processus n'est pas directement indexé par une objectivité scientifique, mais relève de la créativité propre du compositeur. De ce fait, si certaines de ces procédures sont objectivables, elles peuvent toutes ou en partie relever de la "responsabilité" du compositeur, de son "intuition", parfois conscientisée en amont ou en aval de la composition. Grisey s'exprimait ainsi pour exemplifier la technique de la SI [9, p. 107] : "L'excessive épaisseur et la redondance restent les principaux dangers de ce type d'écriture [i.e. la synthèse instrumentale]. Il nous reste sans doute beaucoup à apprendre et à expérimenter afin de suggérer un spectre synthétique par ses composantes les plus saillantes tout en supprimant l'inutile".

Aux données objectives relevant pour l'essentiel d'un "projet" compositionnel adossé à des modèles acoustiques [2] s'ajoutent ainsi des données empiriques de l'activité de "recherche musicale" des compositeurs, qu'il convient de formaliser le plus possible.

3.3. Analyse avec descripteurs Audio et Symbolique

3.4.1 Analyse Symbolique

À ce stade de notre recherche, nous avons utilisé des descripteurs symboliques afin de mesurer le degré d'écart entre les choix faits par le compositeur lors de l'écriture de la partition et la réalité acoustique de l'exécution de la pièce. Nous nous sommes concentrés sur trois analyses, en utilisant divers descripteurs symboliques⁵.

Dans le premier cas, nous avons étudié l'évolution de la densité des partiels au cours des onze premières sous-sections de la partition (Figure 4a). L'écoute de la pièce suggère un processus de complexification qui favorise le passage progressif de l'harmonicité et de l'inharmonicité. Nous avons émis l'hypothèse que, sur le plan symbolique, l'usage des partiels se densifierait au cours de la première

section. Notre analyse montre qu'effectivement la section moins dense "harmoniquement" (en termes de nombre de hauteurs jouées dans une sous-section) est la sous-section introductive (sous-section 0 - première page), avec seulement quatorze hauteurs différentes utilisées. Elle est suivie par la première sous-section qui atteint 18 partiels et la seconde avec 24.

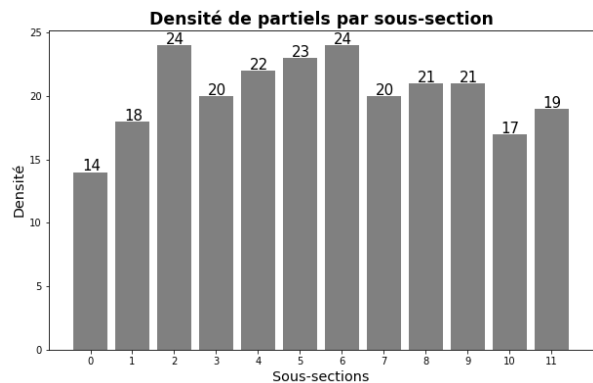


Figure 4a. Représentation graphique des données de l'analyse des descripteurs symboliques de la première section de *Partiels*.

La courbe de densité de la première section de partiels ressemble davantage à un arc qu'à une ligne toujours croissante, ce qui ne correspond peut-être pas exactement à ce que nous entendons ou pensons de la pièce. Bien que ce paramètre puisse sembler moins important, il a certainement fait partie du raisonnement et de l'arsenal du compositeur lors de la création de la pièce.

Dans toute la première sous-section, Grisey n'emploie pas plus de 24 partiels à la fois, ce qui est compréhensible étant donné l'effectif de 18 musiciens. Après avoir atteint 24 hauteurs différentes simultanées à la sous-section 6, la densité diminue progressivement, atteignant 19 hauteurs différentes au chiffre 11. Cela peut s'expliquer par le fait qu'au fur et à mesure que la pièce avance dans la première section, les instruments doivent jouer des techniques étendues ou des sons bruités, ce qui rend difficile de demander aux instrumentistes de jouer des multiphoniques, d'autant plus que certaines techniques étendues rendent déjà le son de l'instrument plus riche.

La deuxième analyse symbolique (Figure 4b) que nous avons menée a consisté à corrélérer le partiel le plus élevé par sous-section avec le nombre d'inharmoniques par section. Le suivi du plus haut rang est intéressant car même s'il est souvent joué subtilement à ce moment de la pièce, il peut marquer la perception de la pièce comme étant la plus haute hauteur à chaque sous-section, établissant avec la fondamentale, l'ambitus, et la perception spatiale de la hauteur. En lisant attentivement la partition et en écoutant la pièce, nous avons émis l'hypothèse que le nombre d'inharmoniques augmente au

⁴ cf. 3.1.3

⁵ Dans le cadre de ce papier, l'équipe n'est pas exhaustive dans les descripteurs choisis. D'autres descripteurs sont en cours d'étude : ratio

entre partiels nouveaux, partiels maintenus et partiels disparus par sous-sections ; évolution des timbres par sous-section ; sons résultants engendrés par sous-section ; répartition des intensités des partiels par sous-section...

fur et à mesure que les partiels de rang le plus élevé diminuent (car ils sont transposés vers le registre grave).

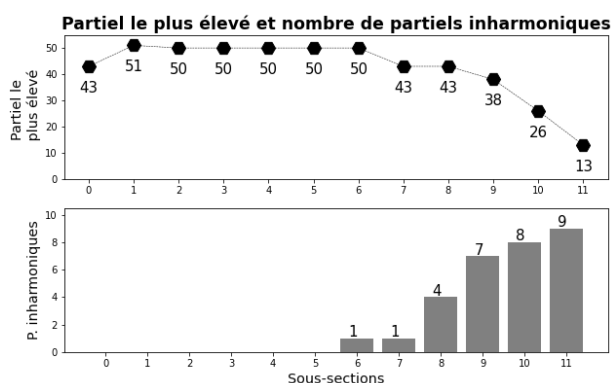


Figure 4b. Représentation graphique des données de l'analyse des descripteurs symboliques de la première section de *Partiels*.

Notre analyse montre que ces deux paramètres (à l'exclusion de ce qui est joué par l'accordéon) indiquent clairement le moment où la texture spectrale, qui était plus ou moins finement construite au début, se fracture et inverse son caractère vers l'inharmonicité. La première page de la partition, tout en étant la moins dense comme nous l'avons remarqué plus haut, correspond à une de plus grande sous-section en termes d'ambitus, celle-ci énonçant le partiels 43 et la page suivante énonçant à la fois la fondamentale et le 51° partiel. À partir de la deuxième sous-section et jusqu'à la sixième, l'ambitus reste stationnaire entre l'intervalle formé autour de la fondamentale et le 50° partiel. À partir de la septième sous-section, le partiel le plus élevé tombe au 43° rang lorsque le premier inharmonique apparaît. Dans la sous-section suivante, cette tendance s'accélère pour atteindre, dans le chiffre 11, le 13° partiel comme le plus haut, et neuf inharmoniques.

Cela indique que la première section de partiels ne représente pas seulement une progression de l'harmonicité à l'inharmonicité, mais qu'elle reflète également une intention compositionnelle claire de transformer la texture de clairsemée à très étroite, tout en déplaçant simultanément l'ambitus vers le registre inférieur. De plus, il indique que le principal outil utilisé pour créer de l'inharmonicité au niveau de la hauteur était de fournir des battements de fréquence. Dans la dernière sous-section, nous avons une combinaison d'harmoniques et d'inharmoniques, ce qui renforce cet effet par rapport au cas où seulement les inharmoniques sont présentes.

La troisième analyse symbolique (Figure 4c) montre que le renforcement du deuxième partiel par le trombone et parfois la clarinette basse, le cor ou le violoncelle, est responsable de la sonorité puissante et éclatante du registre grave de la section. Le renforcement du deuxième partiel reste constant tout au long de la section. Tandis que

l'ambitus devient plus restreint, davantage de ressources sonores sont sollicitées pour le renforcer (cinq au total).

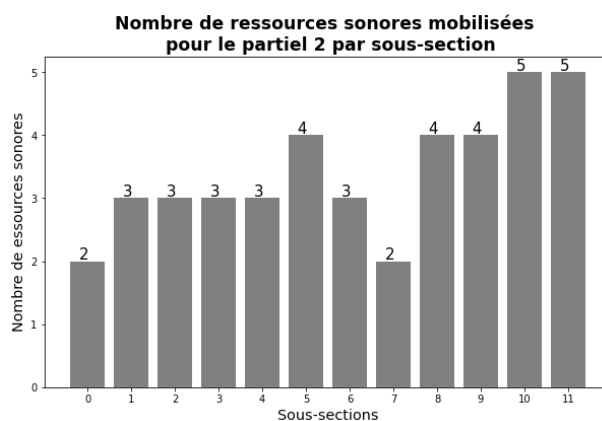


Figure 4c. Représentation graphique des données de l'analyse des descripteurs symboliques de la première section de *Partiels*.

Notre analyse des descripteurs symboliques met en évidence la nécessité de concilier des aspects variés et parfois opposés du matériau musical lorsqu'on travaille avec la SI.

3.4.2 Analyse Signal des deux enregistrements

Pour mieux comprendre la nature des résultats sonores engendrés par la SI, il est essentiel d'examiner comment la musique est interprétée et comment les prescriptions symboliques du compositeur visibles dans la partition deviennent du son. Si la SI est réellement capable de générer une sorte de perception hybride, l'analyse de la partition est nécessaire mais pas suffisante. C'est pourquoi notre approche vise à combiner des études symboliques, audio et, dans une étape future, d'écoute expérimentale.

Pour mener notre analyse de la composition de *Partiels* de Gérard Grisey, nous avons sélectionné deux enregistrements différents, deux des trois enregistrements cités plus haut (table 2 ; figure 3, lignes B et D). Ces enregistrements proposent des interprétations de l'œuvre qui présentent des solutions musicales différentes. En analysant ces deux enregistrements (et potentiellement 3 ou 4), nous cherchons à acquérir une compréhension plus profonde des choix musicaux issus des interprétations de l'œuvre de Grisey.

Nos premiers résultats quant à l'analyse comparée des signaux audios des deux enregistrements cités ci-dessus font l'objet de communications et publications séparées à venir⁶. Ces résultats tentent de répondre à l'objectif initial visant à mieux comprendre les phénomènes de fusion perceptive, ces "être[s] hybride[s]" pour notre perception, un son qui, sans être encore un timbre n'est déjà plus tout à fait un accord, sorte de mutant de la musique 'aujourd'hui, issu de croisements opérés

⁶ Cf. Micael Antunes, Mylène Gioffredo, Javier Elipe-Gimeno, Charles de Paiva, Samuel Poirrot, Vincent Tiffon and Sølvi Ystad, "On the Perceptual Fusion in Grisey's *Partiels*: a Comparative Audio-Descriptor

Analysis of Two Recordings", *Timbre 2023*, 3rd International Conference on Timbre, Thessaloniki, Greece [présentation à venir : 10-12 juillet 2023].

entre les techniques instrumentales nouvelles et les synthèses additives réalisées par ordinateur” [9, p. 107].

Afin d'éprouver la fiabilité des analyses ci-dessus, nous suggérons une nouvelle étape de simulations des processus de SI via OpenMusic, avant d'entrer dans les tests d'écoute expérimentaux.

3.4. Travail en cours : simulations via OpenMusic

L'objectif de cette étape est d'explorer la technique de composition de la SI à l'aide du logiciel OpenMusic. Plus précisément, nous souhaitons générer des spectres arbitraires à partir d'un modèle donné, tester différentes approximations et stratégies pour l'inharmonicité, et analyser les résultats à l'aide de descripteurs ou de mesures symboliques. Enfin, nous prévoyons de comparer la sortie synthétisée avec la source originale et d'utiliser cette expérience pour concevoir un protocole expérimental impliquant des tests d'écoute.

Notre implémentation de la SI dans OpenMusic a nécessité la construction de plusieurs classes, objets et fonctions. La première étape a été de concevoir une classe pour représenter un "partiel" qui comprenait tous les paramètres détaillés dans le Tableau 2. Cette classe a servi de preuve de concept et nous a aidé à comprendre les exigences de la description d'une ligne ou d'un partiel dans une texture dérivée de la SI.

Ensuite, nous avons construit une deuxième classe appelée *Spectre* qui a hérité de tous les attributs de la classe *Partiels*. La classe *Spectrum* est essentiellement une collection de classes *Partiels*, ce qui permet de les ordonner de manière synchrone. L'intégration de cette classe aux éditeurs et bibliothèques d'OM a été bénéfique pour explorer la microtonalité et d'autres aspects.

En utilisant ces objets, nous avons pu simuler le spectre dans la pièce *Dérives* de Grisey, qui était la première ébauche de la SI et qui est antérieure à *Périodes*. Notre simulation du spectre de *Dérives* a révélé qu'il correspond à une série dodécaphonique lorsqu'il est considéré comme une collection de classes de hauteurs. Cependant, lorsqu'elles sont entendues dans leurs octaves respectives, les hauteurs fusionnent pour créer une perception de son harmonique.

Pour simuler la première section de la partition de *Partiels*, nous avons développé une série de méthodes pour modéliser les processus de distorsion du spectre observé dans la partition de Grisey. Nous avons expérimenté deux types de méthodes ou de fonctions : l'une qui transpose les partiels sélectionnés vers des octaves inférieures et l'autre qui les déforme proportionnellement en multipliant la fréquence par un facteur (similaire au processus utilisé par Tristan Murail dans *Désintégrations*).

L'un des défis auxquels nous avons été confrontés dans nos simulations informatiques était la gestion d'événements continus, tels que les changements de modes de jeu et les changements progressifs dans la dynamique des sons soutenus. Malgré la mise au point de solutions de contournement pour tenir compte de la microtonalité dans le protocole MIDI, le rendu des pages

les plus simples de la notation proportionnelle, qui font appel à des techniques étendues, s'est avéré être une tâche ardue, nécessitant de nombreux ajustements et solutions de contournement.

4. CONCLUSION ET PERSPECTIVES

Cet article présente les fondements d'une enquête sur la technique de la SI, laquelle est une adaptation de la synthèse additive utilisant des instruments orchestraux classiques pour créer de nouveaux sons. L'étude s'est concentrée sur l'analyse de la SI à travers une étude de cas de *Partiels* de Gérard Grisey, et nous avons montré des analyses basées sur des descripteurs symboliques tels que la densité des partiels par sous-section, le partiel de rang le plus élevé par sous-section, le nombre d'inharmoniques par sous-section, le partiel le plus renforcé, et les ressources sonores par partiel le plus renforcé. En outre, nous avons effectué une analyse temporelle de la première section, en segmentant les sous-sections (chiffres 1 à 11). Un modèle SI a été développé sur la base des informations extraites des partitions musicales, et des simulations potentielles ont été esquissées en utilisant OpenMusic et des échantillons orchestraux. Diverses approximations de fréquences réelles dans la série harmonique sont également en train d'être testées.

Au delà de contribuer au dépassement de l'écart entre musicologie computationnelle et traditionnelle par l'intégration des outils du traitement du signal et de la science expérimentale à la pratique musicologique [15], nous envisageons plusieurs types de retombées dans le champ de l'acoustique, du design et de la sonification : le développement de nouvelles stratégies de sonification intégrant des techniques issues de la musique spectrale et validées par les tests d'écoute du projet (comme par exemple la substitution de l'axe grave vs aigu par inharmonique vs harmonique, périodicité vs apériodicité, etc.) ; les enjeux autour de la communication non verbale et substitution sensorielle, en prenant appui sur le médium sonore ; la meilleure compréhension des processus de fusion acoustique, les frontières de fusion, les liens entre texture sonore et tension. Dans le champ des SHS, nous envisageons une réflexion sur la notion de dé-coïncidence [12] avec la notion de dissonance au sens large en considérant la musique contemporaine comme laboratoire d'une pratique de la poétique de la "dé-coïncidence". Enfin, nous envisageons des applications dans le champ de la clinique (priorisation des procédés valorisant l'écart, la discontinuité, chez les sujets pour lesquels l'activité de filtrage perceptif est inopérante).

5. ACKNOWLEDGMENTS

Ces travaux sont le produit d'une équipe interdisciplinaire au sein du laboratoire PRISM (UMR 7061) à Marseille, réunissant des musicologues (Charles de Paiva Santana, Javier Elipe-Gimeno, Vincent Tiffon, Mylène Gioffredo, Micael Antunes), des chercheurs dans le champ de la perception et des descripteurs audios (Sølvi Ystad, Samuel Poirot) et une psychologue (Laetitia Petit)

autour d'un projet appelé ANTISCOPE (Analyse transdisciplinaire de l'impact des stratégies compositionnelles de la musique spectrale sur la perception) : <https://www.prism.cnrs.fr/projet/antiscope/>

Le projet ANTISCOPE est lauréat de deux AAP : Financement PRISM (Appel à Projet Interdisciplinaire interne au laboratoire) ; Financement SEPIA (Appel à Projet MITI - Mission pour les Initiatives Transverses et Interdisciplinaires - du CNRS. Par ailleurs, l'IESM (Institut d'Enseignement Supérieur de la Musique, Aix-en-Provence) est également partenaire du projet, dans la phase d'enregistrement des tests perceptifs expérimentaux.

6. REFERENCES

1. « Grisey/Murail », *Entretemps* n°8, sept. 1989.
2. Barrière, J.-B. éd, *Le Timbre, métaphore pour la composition*. Bourgois/IRCAM, Paris, 1991.
3. Donin, N. « Sonic Imprints: Instrumental Resynthesis in Contemporary Composition ». In Gianmario Borio (ed.) *Musical Listening in the Age of Technological Reproducibility*, Farnham/Aldershot: Ashgate, 2014
4. Dufourt, H. « Musique spectrale », *Conséquences*, 7-8 (1985-1986), 111-115 (1^o éd., 1979).
5. Dufourt, H. *La musique spectrale : une révolution épistémologique*. Éditions Delatour, Sampzon, 2014.
6. Féron, F.-X., « Gérard Grisey: Première section de *Partiels* (1975) », *Genesis*, 31, 2010, 77-97. <https://doi.org/10.4000/genesis.352>
7. Féron, F.-X. « Le son d'une voix (1964) by François-Bernard Mâche: the introduction of the Sona-Graph in the composer's workshop ». *Tracking the Creative Process in Music*, Oct 2019, Lisbonne, Portugal. halshs-02427028
8. Grisey, G. « À propos de la synthèse instrumentale », *Écrits* (2e édition). Éditions MF, Paris, 2019.
9. Grisey, G. « La musique : le devenir des sons », *Écrits* (2e édition). Éditions MF, Paris, 2019.
10. Grisey, G. *Partiels* (1975). Ricordi, Milan, n° 132423.
11. Hasegawa, R. « Timbre as harmony—harmony as timbre », In E. I. Dolan & A. Rehding (Eds.), *The Oxford handbook of timbre* (1st ed., 525–551). Oxford University Press, Oxford, 2021. <https://doi.org/10.1093/oxfordhb/9780190637224.013.11>
12. Julien, F.. *Dé-coïncidence. D'où viennent l'art et l'existence*, Grasset, Paris, 2017.
13. Murail T. *Désintégrations* (1982). Lemoine, Paris, n° 28227.
14. Orcalli, A. « La pensée spectrale », in *Théories de la composition musicale au xx^e siècle*, Nicolas Donin, Laurent Feneyrou (dir.), 1511-1574, 2014.
15. Schaub, S., Simurra, I., Fernandes Tavares, T. « Mixing symbolic and audio data in computer assisted music analysis a case study from J. Harvey's speakings (2008) for orchestra and live electronics », *Proceedings of Sound and Music Computing Conference*, 2013.
16. Wilson, P. N. « Vers une écologie des sons », *Entretemps*, n°8, septembre 1989.

Actes des Journées d'Informatique Musicale 2023
24-26 mai 2023
Maison des Sciences de l'Homme Paris Nord

**Articles soumis en relation
au programme artistique des JIM 2023**

HERMES V1 : CREATION AUDIOVISUELLE EN REALITE VIRTUELLE

João Svidzinski
CICM-EA 1572
Université Paris 8 et IRCAM
svidzinski@gmail.com

RÉSUMÉ

Hermès v1 est une pièce audiovisuelle entièrement en réalité virtuelle. Développée avec le moteur de jeu Unity, elle se sert des modules DSP en Faust et des méthodes de contrôle orientée objet en C# pour créer une pièce interactive et immersive. Jouée avec Oculus Quest 2, cette pièce fait écho aux deux autres pièces de cette série, *Hermès v2* et *Hermès v3* qui seront aussi programmées aux JIM 2023.

1. INTRODUCTION

La série de pièces *Hermès* a été proposée dans le cadre du projet « Organisations musicales symbiotiques : une révision de la notion de concert de recherche-crédation faisant appel à l'informatique musicale » soutenu par le conseil scientifique de la MSH PN – Maison des sciences de l'homme Paris Nord – porté par l'auteur. Son objectif est de revisiter la pratique actuelle musicale notamment par le biais de la musique en réseau [1].

Hermès v1 est le premier volet de trois pièces de ce projet. Il s'agit d'organisation d'installations sonores dans laquelle chaque auditeur peut vivre l'expérience musicale immersive en simulation 3D de manière asynchrone. Un seul auditeur à la fois peut y accéder en se servant d'un casque de réalité virtuelle. Cette pratique respecte ainsi la distanciation sociale tout en restant interactive.

Nous proposons donc pour les JIM une installation pour la diffusion en réalité virtuelle de *Hermès v1*.

2. VIRTUALISATION DE L'ESPACE DU CONCERT PAR LE BIAIS DE LA REALITE VIRTUELLE

Pour revisiter l'activité de concert dans le contexte de crise Covid, le multimédia apparaît naturellement comme l'option la plus appropriée. L'audiovisuel offre la possibilité de nouvelles et originales manières de faire de la musique. Dans ce scénario, les moteurs de jeux vidéo, comme Unity, sont en grand essor. Ces outils sont caractérisés par un grand degré d'interaction homme-machine et leur déploiement peut se faire sous divers formats comme les applications mobiles et sur le Web. Depuis les années 2000, plusieurs domaines des sciences

humaines ont recours aux jeux vidéo. La notion de *gamification* a pris une importance notable dans le contexte scientifique actuel. Pour notre recherche, il ne s'agit pas d'incorporer des notions ludiques provenant des situations de jeu, comme les bonifications ou les récompenses. Nous nous intéressons plutôt à l'interactivité entre l'auditeur/joueur et le système informatique. Ainsi, l'œuvre musicale se compose en symbiose avec l'auditeur. Au lieu de diffuser une pièce électroacoustique fixée en réalité virtuelle, ce sont les interactions entre le participant et la pièce qui composent la pièce tout étant attaché à une ligne temporelle définie.

Au niveau logiciel son, nous utilisons des modules Faust pour le traitement audio temps réel et le plugin de spatialisation Hoa¹ pour l'immersion 3D. Le moteur DSP est similaire à nos autres pièces. Il s'agit des modules de granulation et *harmonizer*. Pour la gestion temporelle et le paramétrage des objets graphiques et des modules audio nous faisons appel à une logique orientée objet fondée sur la thèse de doctorat du compositeur [2].

3. MISE EN PLACE

Pour la programmation de *Hermès v1*, nous proposons l'installation d'un poste avec un ordinateur et casque VR pour que les participants la jouent à tour de rôle. Chaque utilisateur prendra environ 5 minutes.

L'auteur est complètement autonome avec l'installation et l'approvisionnement d'équipements, à savoir le casque VR Oculus Quest 2, l'ordinateur, la carte son et le casque audio. Seulement le mobilier et prises électriques doivent être fournis par l'équipe locale.

4. REFERENCES

1. Svidzinski, J. « *Hermès v2* - web concert collaboratif en temps réel », *Journées d'Informatique Musicale 2021*, AFIM, Jul 2021, Visioconférences, France. <https://hal.science/hal-03313610>.
2. Svidzinski, J. *Modélisation orientée objet-opératoire pour l'analyse et la composition du répertoire musical numérique*. Thèse de doctorat, Université Paris VIII, 2018. Français. [\[NNT : \]](#). [\(tel-02045765\)](#)

¹ <http://hoalibrary.mshparisnord.fr/telechargements/unity/> [lien vérifié le 27 février 2023].

BOATNOA : AN AUDIOVISUAL REFLECTION ON A MODERN CITY

Óscar Bahamonde
dotzerosix

Universitat Politècnica de
València
obahsua@posgrado.upv.es

Tania Cortés
dotzerosix

Universitat Politècnica de
València
tcorbec@posgrado.upv.es

Gabriela Yáñez

Independent Sound Designer
gyanez.sound@gmail.com

ABSTRACT

L'article suivant est une description étendue d'une composition audiovisuelle immersive et d'une performance appelée *boatnoa*. Cette œuvre représente les conditions modernes de la ville dans laquelle elle est présentée. Grâce aux méthodes de collecte et aux processus numériques utilisés, les artistes commentent les activités et les attitudes omniprésentes facilitées par les dispositifs numériques et les « boîtes noires ».

Dans ce contexte, cette œuvre offre au public une expérience pour réfléchir aux sujets proposés et, par conséquent, méditer sur leurs rôles dans le temps, l'espace et la caractérisation de la ville dans laquelle ils vivent.

1. INTRODUCTION

boatnoa is a live A/V piece with real-time audio reactive visuals. As a composition, it maintains a general structure, while also allowing for improvised sections. In this work, the real-time audio reactive performance employs an ambisonic system and two parallel projections in order to generate an immersive space. Visuals react to sound, creating a space that invites the audience to reflect on the concept of 'a modern city', and what it means to inhabit it.

Audio and image are used to rethink the city (referencing its physicality, its people) and the digitality inadvertently present in it. On one hand, the evident references of the city are created from material taken from Paris itself: images, videos, soundmarks and soundscapes are components that describe its sensory characteristics. On the other hand, the methods used to gather this material speak of the ubiquitous digital activities: photogrammetry, recordings, digital photography, and sound and image manipulation.

The output of this work aims to offer two possibilities to its public: first, to allow a place and time for reflection, as the audience faces a condensed version of their urban living and different angles of it; and second, to expose the audience to uncommon creative languages, promoting a normalization process towards digital artistic works.

This paper will focus on the aesthetic representation of the mentioned concepts, as well as the description of the system behind *boatnoa*.

2. BRIEF OVERVIEW OF BLACK-BOXES IN DAILY LIVING

This work is a creative reflection on modern devices, which have almost become an extension of our bodies and our perceptions. These are so ubiquitous, and yet they are designed to keep most people ignorant in regard to their inner workings. These unknown, hidden but essential parts of our devices are known as black boxes, and are purposely designed to be only used and not understood [2] (Hertz & Parikka, 2012).

As *boatnoa* uses sound and image to bring the hidden to the audience, the artists aim to allow for some reflection on the matter: how many invisible things happen in front of us, and because of us, every day? How do these ghostly processes affect the characteristics of the city we live in?

3. CONCEPT AND AESTHETIC OUTPUT

3.1. Structural concepts

Since *boatnoa* talks about a modern city and its inhabitants, it is essential to revise the presence of devices in daily living. These ubiquitous machines are the core topic of this work; in consequence, the audiovisual work approaches their two sides: the one that is manipulated by citizens and have become part of their routines (audio recordings, photos, videos), and the other known as 'black boxes' that is unreachable and mostly unknown to people. This concept is adopted for this work and presented as purpose-built devices that shape the sound and sight of *boatnoa*.

There are other secondary concepts that describe a modern city, which also relate to technology and describe human complexity. All these are part of the discourse expressed through sound and image in *boatnoa*: intergenerational links, diversity and plurality, mobility and sense of belonging.

3.2. Aesthetic choices

Specific processing techniques (sonic and visual) were chosen in order to cover three creative needs:

- a) to symbolize the ideas in the aforementioned concepts, bringing the digital (computers) into the human (creativity),
- b) to keep a high degree of coherence in relationship to the artists particular mode of expression, and
- c) to reinforce a certain level of connection between the main topics and the audience, as they would associate certain effects to the idea of “digitalism”.

The techniques that translate to aesthetic choices in *boatnoa* are as follows:

3.2.1. Sound processes

Delay:

This time-based effect symbolizes the presence of intergenerational bonds. A modern city is a complex superimposition of age groups, where some relate to technology more than others, but all are affected by it. Digitalism has generated conflict between these groups, but it has also brought opportunities for communication and closeness.

Harmony:

A modern city cannot be described without thinking of plurality. When using effects for harmonization in *boatnoa*, the intention is to mention all this complexity that makes the city: human mobility, efforts for inclusion, and diversity are topics sensibly thought of with this work. The constant presence of digital tools has given access to information and organization related to those topics. From the artistic point of view, digital tools have become a platform for a plurality of voices which otherwise would not be heard so closely.

Granulation:

The process of granulation is employed as an analogy. Just as this effect breaks down a sound into its micro-components, *boatnoa* aims to analyze the different smaller elements that compose the modern city. The use of granulation in this work elevates the complexity of its elements by looking closely into them. Just as this process zooms in into the texture of a picture and breaks down the grains of a voice note, in the context of a city each person represents the granulated whole. As a result, granulation is a comment on communal responsibility, since a city is the expression of its inhabitants, on the ways they come together or disregard each other.

3.2.2. Visual processes

Glitch:

The visual elements of *boatnoa* are constructed from the same materials (digital captures of the city) but are used to reveal the inner workings of the human-machine relationship. Hence, the evident visual errors known as

“glitches” are purposely employed as a hint about devices and black boxes. This visual characteristic is in itself the result of working with machines and, more specifically, it is the result of aesthetically embracing errors [1] (Cascone, 2000). Glitches become small, almost magical, displays of the automatic processes that surround modern life, even when they are mostly ignored. In consequence, making these errors visual and audible become another creative possibility. Glitch is picked as a stylistic choice to express how the value of technology relies on the way people use it, even when it remains a mystery for most.

Photogrammetry:

This technique consists of recreating a 3D image parting from a series of 2D captures. By using this process, visual information from the city, taken with regular devices, brings depth into the representation of the place. To obtain these 3D models, the artists complete a series of processes which by their own digital nature bring more glitch into the final result. The models become a series of point-cloud images which are explored as a series of particles (Figure 1), each one capable of bringing in more visual effects: gravity, weight, movement, trajectories, loops, fusing of images, and more (Figure 2).



Figure 1. Original 2D visual capture (left) can become a series of point-cloud information (right).

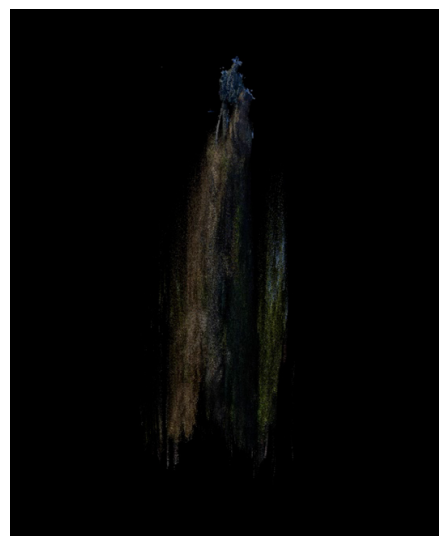


Figure 2. The point-cloud data can then be processed with effects for movement and texture.

Additionally, visuals are modulated by several sound parameters and triggers, which allows a solid and coherent link between sound and image. As a result, the audience experiences unconventional ways of perceiving their city, exploring different angles simultaneously.

Finally, the visual concept of this work is expanded by the way projections work. Two different surfaces are used for the image. The back screen represents those hidden processes on our devices, as they cannot be seen but yet are essential. The second projected surface is the frontal face of the performance. Made from gaze, this last layer represents the palpable interaction with all digital devices. Between both layers are the people, those who make technology meaningful.

4. TECHNICALITIES

The present work employs a 2D ambisonic array for sound and a two-screen projection that limits the area. Both sound and image are manipulated from the center stage, using personal computers and purpose-built devices. As mentioned before, the projection surfaces are set in front and behind the performers (Figure 3), while the speaker system surrounds them (Figure 4).

It is important to note that the chosen sound system will maintain the immersive character of the work while being adaptable. This work is built to fit different venues since its sound spatialization can be adjusted for any multichannel speaker configuration.

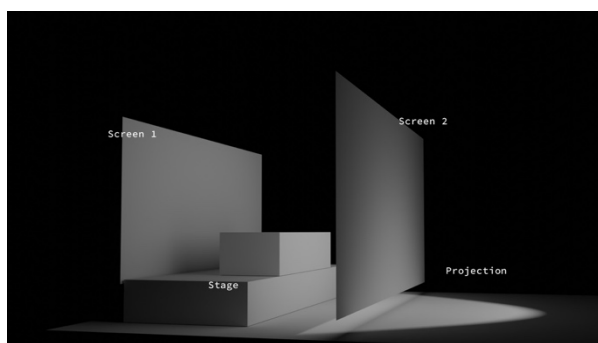


Figure 3. Stage set up.

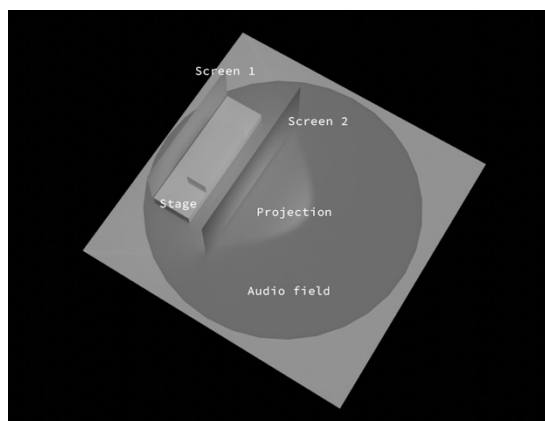


Figure 4. The ambisonic set up can be adjusted to different multichannel speaker configuration, while the visuals are projected in two surfaces.

5. CONCLUSIONS

This extended description of the audiovisual work *boatnoa*, exposes its inner workings and the thought process of the artists. The marriage of both gave way to an immersive audiovisual composition and performance that serves as a place (time and space) for an audience to reflect upon the concept of a ‘modern city’. Besides the sensory and conscious elements of a place perceived daily through sight and sound, it is essential to recognize the mostly unconscious presence of digital devices. This concept is highlighted by the methods of representation employed in *boatnoa*. Consequently, the audience is invited to recognize this now omnipresent element that influences and affects their daily urban living, in order to rethink their roles in the city with and through digital devices.

6. REFERENCES

1. Cascone, K. “The Aesthetics of Failure: ‘Post-Digital’ Tendencies in Contemporary Music.” *Computer Music Journal*, vol. 24, Cambridge, USA, 2000.
2. Hertz, G. & Parikka, J. “Zombie Media: Circuit Bending Media Archaeology into an Art Method.” *Leonardo Music Journal*, Cambridge, USA, 2012.
3. Russolo, L. “THE ART OF NOISE” [online], <http://www.artype.de/Sammlung/pdf/russolo_noise.pdf>, (Accessed February 28, 2023).
4. Thomson, P. “Atoms and errors: Towards a history and aesthetics of microsound.” *Organised Sound*, vol. 9(2), 207–218, [online], <<https://doi.org/10.1017/S1355771804000299>>, (Accessed February 28, 2023).