

ANIMER LA MUSIQUE SUR SUPPORT

Thomas Lucas
Puce Muse, Rungis
Institut Jean Le Rond d'Alembert
Sorbonne Université, CNRS, Paris
thomas.lucas@pucemuse.com

Serge de Laubier
Puce Muse, Rungis
sdl@pucemuse.com

Christophe d'Alessandro
Institut Jean Le Rond d'Alembert
Sorbonne Université, CNRS, Paris
christophe.dalessandro@
sorbonne-universite.fr

RÉSUMÉ

Cet article présente les principes, ainsi qu'un outil logiciel, pour « animer » de façon expressive les musiques sur support – c'est-à-dire dont le déroulement et le contenu musical sont a priori fixés. Deux étapes sont nécessaires pour cela : 1) l'analyse préalable de l'enregistrement pour la mise en place d'ancres temporelles (notes, pulsations, événements) ; 2) l'utilisation de l'analyse-synthèse et d'interfaces de contrôle pour manipuler le déroulement temporel de la pièce en temps réel. La première étape d'analyse et de marquage consiste à élaborer différentes collections d'ancres, liées à une représentation gestuelle explicite ou sous-jacente du contenu sonore. La seconde étape permet à l'utilisateur musicien de moduler le déroulement temporel et d'autres aspects du son grâce aux ancres, à l'aide d'une cellule de captation du geste et d'un module de transformation audio. Ce propos est illustré par deux exemples très différents : la *Gymnopédie n° 1* pour piano d'Erik Satie et « Accidents/harmoniques », extrait de *De Natura Sonorum* de Bernard Parmegiani. Nous discuterons leur rôle dans la conception d'une interaction expressive des différences au niveau de l'indexation du fichier audio et du geste expressif utilisé pour « animer » chacun de ces exemples.

1. INTRODUCTION ET ÉTAT DE L'ART

1.1. Animer un son

L'expression « animer un son » désigne, de façon métaphorique, l'interaction d'un utilisateur avec une application sonore interactive qui permet de diffuser des musiques enregistrées [21]. Le degré zéro de l'animation, à savoir le geste minimal, consiste à lancer et à arrêter la diffusion de l'enregistrement. On peut aussi considérer le son enregistré sur un support comme une matériau inanimé qui peut être en quelque sorte (ré)animé et réincarné lors d'une performance. Animer un son désigne alors le processus qui articule l'automation – au sens des logiciels audionumériques – et l'interaction, afin d'agir de façon expressive sur le déroulement de l'enregistrement et de varier une matière sonore a priori fixée et toujours identique. Le terme

« animer » souligne les aspects expressifs de l'interaction : le geste utilisé doit permettre d'exprimer au mieux des intentions musicales lors de la performance.

Notre travail repose sur l'hypothèse qu'il est possible d'associer des formes ou des mouvements sous-jacents au contenu d'un enregistrement sonore. À l'écoute d'une œuvre, un auditeur associe perceptivement des substituts gestuels, ou « *gesture surrogacy* » [22]. Dans la *Gymnopédie n° 1* de Satie, cette représentation découle de la partition et fait directement référence aux gestes physiques de production du son par le pianiste : il s'agit d'un substitut gestuel de deuxième ordre [22]. Dans « Accidents/harmoniques » de Bernard Parmegiani, une représentation gestuelle directe semble plus difficile à établir. Le continuum harmonique présent au début de cette œuvre électroacoustique n'évoque pas de geste instrumental et sa perception repose davantage sur les évolutions spectromorphologiques du matériau sonore : il s'agit d'un substitut gestuel éloigné [22].

Le but de cette recherche est d'élaborer un instrument sous la forme d'un outil logiciel permettant de contrôler de façon expressive le déroulement temporel d'une musique sur support. La musique est diffusée en manipulant à l'aide de capteurs une représentation gestuelle sous-jacente élaborée à partir du son perçu. La mise en œuvre d'un tel projet implique deux étapes principales.

La première étape consiste à analyser le contenu audio, afin de placer des ancres sur des éléments temporels significatifs directement liés au discours musical. Divers critères permettent de placer ces ancres. Nous examinerons ainsi en détail deux méthodes automatiques pour placer des index sur les temps ou les notes d'un morceau.

La seconde étape est celle de la définition de l'instrument, au sens des capteurs et de l'interaction. Il s'agit de définir les relations synergiques entre les pôles suivants : capteurs et analyse du geste ; analyse audio et détermination d'une collection d'index ; moteur audio et actions sur le son. Avant de décrire nos propres travaux, il est important de mentionner les travaux antérieurs visant à animer les musiques sur support fixe.

1.2. Musique concrète, *plunderphonics*, *turntablism* et *reenactment*

L'utilisation de supports fixes en tant que moyens pour créer de la musique a été désignée par l'expression « musique concrète » et formalisée par Pierre Schaeffer dans les années 1950 [20]. On parle de « *plunderphonics* » [19] – littéralement piller (*to plunder*) le son (*phonics*) – lorsque les segments audio appartenant à une œuvre musicale sur support clairement identifiable sont repris pour créer de nouvelles compositions.

Ces approches visent à créer de la musique à partir de musiques fixées sur support. Dans les années 1970, une approche portant sur l'animation de musiques sur support a été développée afin de rendre plus vivante et interactive la musique de dance par les disc jockeys. Les premiers instruments des DJs ont été la table de mixage et la platine tourne-disque, d'où le nom de *turntablism* (ou platinisme). Les mécanismes perceptifs mis en place par un utilisateur lors de son interaction avec une application audio interactive sur un enregistrement sonore ont été conceptualisés sous le nom de *reenactment* [21]. Fondé sur les théories énoncées par Francisco Varela [23], ce concept utilisé par Norbert Schnell [21] pour décrire cette façon de faire de la musique implique les dimensions perceptives et cognitives de l'interaction avec un enregistrement sonore, en rapport avec les actions évoquées par le son même. Afin d'intégrer au mieux la dimension expressive de l'interaction, le concept de *reenactment* se place au centre de notre démarche, au travers de plusieurs questions : quel serait le geste le plus approprié pour interagir de façon expressive avec un son enregistré ? S'agit-il du geste physique évoqué à l'écoute de l'enregistrement ? Comment repérer et marquer dans le flux sonore ce geste sous-jacent ou évoqué ?

1.3. Écoute active et analyse du contenu

Une démarche apparentée au *reenactment*, proposée par Masataka Goto, met en œuvre des techniques d'analyse dites de MIR (*Music Information Retrieval*) pour collecter des informations intrinsèques au discours musical, comme un changement de grille harmonique [12]. Le contenu d'un fichier audio est utilisé pour concevoir des interactions nouvelles par un processus d'écoute active – *Active Music Listening Interface* [11]. Grâce à ces applications logicielles, le rôle de l'auditeur se rapproche de celui de l'interprète, grâce aux possibilités offertes pour dérouler la lecture du fichier audio.

Cette démarche d'analyse du contenu est également utilisée pour l'analyse musicale. En s'interrogeant sur l'évolution esthétique du morceau improvisé « *Interstellar Overdrive* » de Pink Floyd au cours de différentes interprétations, Philippe Gonin [10] observe l'évolution du flux spectral pour mettre en évidence la structure musicale du morceau. L'outil d'analyse utilisé place des ancrages sur les variations brusques de tonalité, permettant ainsi d'observer la structure globale du

morceau enregistré sans passer par une transcription écrite détaillée. L'accès à une telle information pour notre étude permettrait de placer des index sur des éléments significatifs du discours musical à un niveau macro, directement à partir de l'information contenue dans l'enregistrement musical.

1.4. Méta-Piano, Méta-Instrument et synthèse vocale performative

Avec l'avènement de l'audio numérique en temps réel, les rapports entre la musique enregistrée sur divers supports et les interfaces de contrôle sont devenues particulièrement riches et ductiles. Pour les musiques notées sous forme de partition numérique, le Méta-Piano de Jean Haury [13] est une source d'inspiration pour notre démarche. Le Méta-Piano permet à un utilisateur musicien d'interpréter une pièce au piano numérique à l'aide d'un clavier MIDI réduit à quelques touches. Les quelques touches ne contrôlent pas la mélodie, comme au piano, mais le séquençage – rythme, articulation, phrasé, tempo et accentuation – des événements rythmiques sélectionnés, repérés et marqués à partir de la partition. L'utilisateur musicien propose sa propre interprétation par le contrôle de méta-paramètres du discours musical, et non par la production directe de tous les détails du son, comme pour un instrument musical ordinaire. En ce sens, il s'agit d'un méta-instrument. Le Méta-Piano permet de se concentrer entièrement sur les paramètres expressifs d'une interprétation au clavier, en utilisant l'enregistrement comme matériau et comme raccourci permettant de réduire la technique instrumentale proprement dite. Le choix de l'interface de type clavier MIDI pour régir cette application sonore permet à l'utilisateur d'exprimer en temps réel le geste physique évoqué par le son qu'il entend. Même si le morceau de musique n'est pas directement issu d'un enregistrement sonore, mais d'un fichier MIDI, l'interaction proposée avec celui-ci est directement liée au geste physique évoqué par le son. Dans le cas de cet instrument, le clavier MIDI réduit est l'interface la plus appropriée pour contrôler l'instrument de façon expressive. Qu'en est-il des musiques dont les sons qui la constituent n'évoquent pas le geste physique à l'origine de celui-ci ? Serait-il intéressant d'interagir avec leur déroulement grâce à un clavier MIDI ? Nous tenterons d'apporter quelques éléments de réponse dans notre travail.

Une méthodologie semblable à celle du Méta-Piano a été développée pour les Méta-Instruments de Serge de Laubier [4]. Le Méta-Instrument propose une interface beaucoup plus riche que le clavier MIDI, avec plusieurs dizaines de capteurs continus pour les bras, les mains et les doigts. Le module de synthèse est également riche et varié. Un mode d'utilisation, parmi d'autres, du Méta-Instrument est l'interprétation d'un texte musical enrichi de marques temporelles. Les fichiers MIDI sont joués en contrôlant non pas chaque note, mais des blocs temporels qui portent spécifiquement l'expression, et

qui sont définis lors d'une phase d'analyse préalable du contenu musical.

L'animation de la synthèse vocale – ou synthèse vocale performative – a été développée depuis le projet « *Speech conductor* » de Christophe d'Alessandro [6]. La synthèse par blocs temporels à l'aide d'ancres a été proposée dans les systèmes Vokinesis [7, 8] puis Voks [5, 24]. Dans ce cas, ce ne sont pas des fichiers MIDI qui sont joués, mais des échantillons de voix enregistrée. Grâce à un vocodeur, il est possible de changer des paramètres tels que l'intonation, la qualité vocale et la force vocale, et ainsi de jouer avec une voix synthétisée à partir de l'analyse d'une voix enregistrée. Des ancres temporelles permettent de séquencer les syllabes – avec deux points de contrôle rythmique par syllabe –, ce qui permet un jeu musical fluide avec un apprentissage rapide pour l'utilisateur musicien. Dans le cadre de la synthèse vocale, deux points de contrôle rythmique semblent nécessaires, le déclenchement et le relâchement étant également importants pour le jeu musical.

Notre système comprend donc deux grands aspects fonctionnels : la préparation, l'analyse et le marquage ou indexation du contenu de l'enregistrement, puis la définition de l'instrument en tant qu'assemblage d'interfaces pour manipuler le contenu indexé (figure 1).

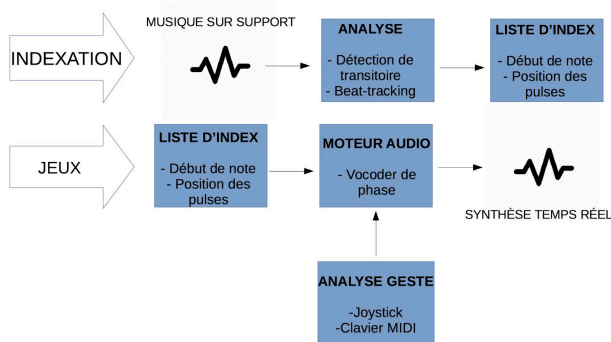


Figure 1. Principes de MonoReplay : phase d'indexation et phase de jeu.

2. ANALYSE ET INDEXATION DU CONTENU SONORE

La première étape de notre travail d'ancrage consiste à segmenter le flux audio que l'on souhaite rejouer en blocs temporels susceptibles de porter le contrôle de l'animation du fichier audio. Cette segmentation peut se faire selon plusieurs échelles temporelles : les notes jouées, les temps, la mesure ou encore les phrases musicales. Cette phase d'analyse vise à placer dans le flux temporel des points – ou ancres – afin de repérer les phases ou blocs temporels que l'on souhaite manipuler.

Pour contrôler de façon expressive le déroulement d'une musique enregistrée, il est indispensable d'obtenir préalablement des informations significatives sur les aspects temporels du discours musical pour ensuite les manipuler. Le tempo, le rythme, la position des pulsations ou encore la position temporelle des notes jouées dans un enregistrement sont des informations que

l'on récupère manuellement ou à l'aide d'algorithmes de traitement du signal. L'automatisation – ou pas – de cette étape d'indexation n'a vraisemblablement pas d'impact direct sur le résultat au moment du jeu – c'est-à-dire pour rejouer de façon expressive un enregistrement de musique –, mais permet de construire plus rapidement le répertoire musical utilisé.

2.1. Indexation de la *Gymnopédie n° 1* de Satie

La *Gymnopédie n° 1* de Satie est une célèbre pièce pour piano solo. À l'aide d'outils de traitement du signal que nous détaillons plus loin, nous élaborons trois collections d'index temporels enregistrés dans un fichier texte : 1) indexation à la mesure, avec la position temporelle du premier temps ; 2) indexation à la pulsation ; 3) indexation à l'événement, avec la position temporelle de chaque début de note ou groupe de notes jouées ensemble.

2.2. Indexation d'« Accidents/harmoniques »

« Accidents/harmoniques » est le deuxième mouvement de *De Natura Sonorum*, une pièce sur bande de Parmegiani. Ce mouvement ne possède pas de structure métrique identifiable ; c'est pourquoi nous ne cherchons pas à déterminer ici de pulsation ou de tempo. À l'aide d'outils de traitement du signal que nous détaillons plus loin, nous élaborons une collection d'index par événements : la position temporelle de chaque début d'événement sonore – qui prend le statut de note – ou de groupes d'événements sonores – c'est-à-dire des groupes de notes – possédant des transitoires d'attaque suffisamment brefs et prononcés sur le plan perceptif – est stockée dans un fichier texte. Pour cette œuvre particulière, une partition visuelle a posteriori a été élaborée [17], ce qui est précieux pour vérifier le placement des index.

2.3. Traitement du signal pour l'indexation

2.3.1. Détection des transitoires d'attaque

La recherche en indexation d'information musicale (*Music Information Retrieval*) a abouti aujourd'hui à des outils d'analyse de contenu audio extrêmement efficaces. En repérant les valeurs au dessus d'un niveau seuil sur la courbe de nouveauté, l'algorithme SuperFlux [9] permet de détecter les transitoires de sons percussifs, ainsi que les transitoires plus doux marqués par un changement de fréquence fondamentale et non par une courte tranche de bruit blanc.

Le flux spectral est la somme de la différence, bande par bande, de l'amplitude du spectre obtenue par transformée de Fourier à court terme. Les maxima observés sur la courbe représentative du flux spectral – appelée aussi courbe de nouveauté – indiquent les brusques variations spectrales dues à la présence d'un nouvel événement, que nous associons à une nouvelle note dans l'enregistrement analysé. En récupérant les

positions temporelles de ces maxima, on obtient une collection d'index où chaque événement important présent dans l'enregistrement est associé à sa position temporelle (figure 2).

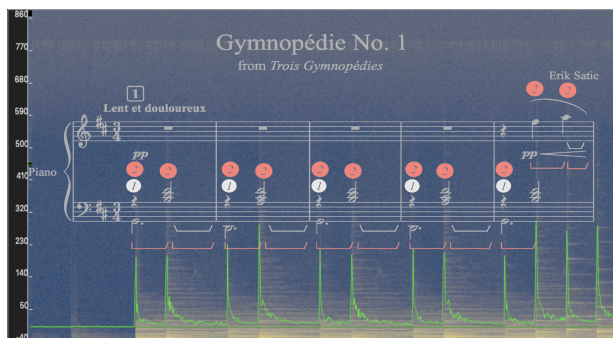


Figure 2. Détermination de la courbe de nouveauté (en vert) dans le cas de la *Gymnopédie n° 1*, avec indexation à la mesure en blanc (1) et indexation à l'événement en rouge (2).

2.3.2. Suivi de pulsation

La pulsation peut être définie comme « l'impulsion périodique induite perceptivement décrivant le mieux l'action de taper du pied sur la musique »¹. Lorsque l'on écoute une musique possédant une pulsation, il nous arrive très fréquemment d'effectuer des mouvements synchronisés avec le tempo du morceau ; c'est pourquoi nous avons choisi d'intégrer une collection d'index représentant cette pulsation lorsqu'elle existe, afin de pouvoir contrôler le déroulement d'un flux audio en tapant la pulsation. Pour constituer cette collection d'index, nous utilisons l'algorithme de détection de pulsation *Multi-feature Beat Tracking* [25]. Cet algorithme, disponible en Python dans le framework *Essentia* [1], détermine neuf fonctions de détection de transitoire d'attaque selon différentes méthodes, afin de ne garder que celle qui présente le plus de similarité avec les autres. Dans le but d'utiliser notre instrument pour interagir avec des enregistrements très différents les uns des autres, l'algorithme *Multi-feature Beat Tracking* est bien adapté, car il offre diverses méthodes de détection de transitoires, ce qui permet d'aborder la variété du paysage sonore des musiques sur support. Pour un son de corde pincée, les transitoires ne présenteront pas les mêmes propriétés spectrales que pour un son de corde frottée. On applique une fonction d'autocorrélation pour chaque fonction de détection de transitoire pour mettre en évidence les périodicités du signal. Une méthode basée sur des matrices de similarité [14] compare chaque séquence obtenue et sélectionne celle présentant le plus de similitude avec les autres comme la plus représentative de la pulsation de l'enregistrement. Cette méthode comparée à seize autres systèmes de détermination de pulsation présente les meilleurs résultats [14].

¹ « The beat (or tactus) is a perceptually induced periodic pulse that is best described by the action of "foot-tapping" to the music » [15].

Une fois la collection d'index à la pulsation déterminée, on applique à la main une correction en tenant compte de la partition pour ne garder que les premiers temps de chaque mesure. On obtient ainsi une collection supplémentaire constituée de chaque début de mesure du morceau (figure 2).

3. DÉFINITION DE L'INSTRUMENT

Au même titre que d'autres applications audio interactives [21], l'instrument *MonoReplay* s'articule autour de trois pôles : un pôle d'analyse et d'indexation, un pôle de capture et d'analyse du geste, et un pôle de synthèse sonore.

Ces trois pôles doivent fonctionner en synergie, afin d'effectuer le travail d'indexation décrit dans la partie précédente et de proposer des interactions pour agir en temps réel sur le contenu préalablement analysé à l'aide d'un geste musical enregistré par différents capteurs (figure 3).

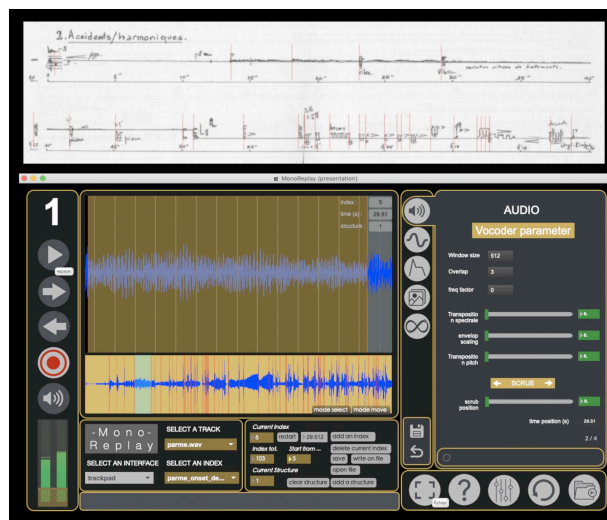


Figure 3. Interface de *MonoReplay* en phase de jeu pour un extrait d'« Accidents/Harmoniques ». En haut: partition reconstituée, avec les index placés pour *MonoReplay*. En bas : oscillogrammes avec les index de jeu.

3.1. *Gymnopédie n° 1* de Satie

À l'écoute de la *Gymnopédie n° 1*, il est relativement évident de saisir la relation entre le rendu sonore et sa causalité instrumentale : le pianiste joue les touches de son instrument et le son produit par le contact du marteau sur les cordes est enregistré. Le geste physique du pianiste entre dans notre perception au moment de l'écoute ; c'est pourquoi nous avons choisi d'utiliser comme interface un clavier MIDI pour contrôler le déroulement de cet enregistrement. En appuyant sur une touche du clavier, l'utilisateur musicien déclenche la lecture d'un segment audio du morceau. Plus il reste appuyé longtemps, plus la vitesse de lecture ralentit. Il peut passer au segment suivant en appuyant sur une autre touche du clavier et ainsi superposer la lecture de

la fin d'un segment avec la lecture du début d'un autre. L'utilisateur peut alors contrôler le rythme, le phrasé et le tempo de l'enregistrement au moment de son déroulement. Ce fonctionnement est analogue à celui du Méta-Piano dans son principe, sauf pour ce qui est de la production du son.

3.2. « Accident/harmonique » de Parmegiani

Selon Denis Smalley [22], l'écoute d'une musique enregistrée permet d'accéder à une certaine compréhension de celle-ci si nous saisissons la relation entre un corps sonore évoqué et sa cause. Dans le cas des musiques instrumentales par exemple, l'expérience d'écoute d'une performance instrumentale fixée sur support met en jeu un processus de conditionnement culturel. Ainsi, une sorte de connaissance implicite des gestes instrumentaux réalisés au moment de l'enregistrement – intrinsèquement liée à la connaissance de la musique en tant qu'activité du sujet – est mise en jeu. Dans notre étude nous souhaitons identifier les interfaces les mieux adaptées pour contrôler de façon expressive le déroulement de nos enregistrements. Nous faisons l'hypothèse que l'intégration de gestes physiques liés au substitut gestuel, virtuels lors de la création musicale et/ou évoqués lors de l'écoute des enregistrements, sont une clé pour la conception de l'interaction. Nous proposons alors à l'utilisateur deux interfaces pour interagir avec le contenu musical : un joystick permettant de contrôler la position de lecture du fichier et un clavier MIDI permettant de contrôler la vitesse de lecture et la position, comme par exemple le geste au clavier.

En effectuant un mouvement circulaire à l'aide d'un joystick, l'utilisateur musicien contrôle la position de lecture au sein d'un même segment. Pour passer au segment suivant, il suffit de faire un tour complet. Ce geste se rapproche de celui du platinière qui, en contrôlant la vitesse de rotation de ses platines ou en positionnant la tête de lecture sur une position temporelle précise du morceau, interagit de façon expressive avec le déroulement du fichier audio. Pour animer l'enregistrement d'une pièce telle que « Accidents/harmoniques », la complexité de la construction musicale et la multitude de sons utilisés rend difficile la détermination d'un seul geste expressif pour jouer au mieux l'ensemble de la pièce. La plupart des structures sonores perçues dans cette composition évoquent des substituts gestuels de troisième ordre, en suivant la typologie de Smalley [22]. Le geste physique de production du son n'est pas clairement identifiable, mais peut être néanmoins déduit ou reconstruit : lors de l'écoute, le geste évoqué peut être reconstruit en fonction de l'évolution spectromorphologique du son, davantage que par une causalité qui est de fait absente. Le choix du joystick permet d'obtenir un contrôle continu précis du déroulement sur les longues nappes du continuum harmonique entrecoupé de brefs sons percussifs au début du mouvement. Les sons percussifs délimitent la segmentation du morceau et le joystick

permet de naviguer en détail dans les textures de ce continuum.

3.3. Traitement du signal pour la transformation

À la différence d'autres applications interactives comme le Méta-Piano par exemple, ce n'est pas un fichier MIDI ou un module de synthèse qui est à la source du son, mais bien un enregistrement que l'on transforme. Ce module de transformation est enrichi de marques correspondant à la fois à la partition et au contenu sonore réel de l'enregistrement : il ne s'agit pas de synthèse mais d'analyse-synthèse, ce qui implique un processus de traitement du signal. Pour cela, un système d'analyse-synthèse est nécessaire. Même s'il ne s'agit pas de voix, ce genre de système porte le nom de vocodeur².

SuperVP et groove~ sont deux vocodeurs de phase implémentés sous la forme de *Max externals* paramétrables, permettant d'avoir un contrôle précis sur la position ou la vitesse de lecture d'un segment audio, tout en gardant de très hautes définitions au niveau des textures sonores rejouées. Le SuperVP est utilisé pour contrôler la position de déroulement du fichier audio dans le mode « *Scrub* », alors que l'objet groove~ est utilisé dans le mode « *Speed* » pour contrôler la vitesse de lecture du segment.

4. DISCUSSION ET CONCLUSION

Ce travail porte sur les problématiques liées à la question de l'animation d'une musique sur support à travers la segmentation et le contrôle du déroulement temporel, appliqués ici à deux œuvres très différentes.

En récupérant des informations sur le contenu intrinsèque du discours musical comme la pulsation, le tempo et le début des notes, on construit des collections d'index grâce auxquelles on définit des interactions permettant à un utilisateur musicien de contrôler de façon précise le déroulement de ces deux œuvres, compte tenu de leurs spécificités musicales. Le choix des interfaces utilisées s'inspire de deux points de vue complémentaires : d'un côté, on tient compte du substitut gestuel de second ordre – évoquée à l'écoute du morceau de Satie, à savoir le geste du pianiste – et, d'un autre côté, on propose une interaction inspirée du mouvement circulaire de certains platinières contrôlant la vitesse de lecture et la position d'une musique sur support discutée dans le cas de la pièce de Parmegiani. Dans le second cas, le substitut gestuel n'est pas lié directement à la causalité du son entendu, mais permet à l'auditeur-musicien de se focaliser sur les aspects spectromorphologiques de celui-ci, notamment dans le continuum harmonique situé au début du mouvement.

Les musiques écrites sous forme de partition offrent des représentations très explicites à propos du contenu musical, ce qui facilite grandement le travail d'indexation préalable des éléments significatifs du

² De *voice coder*, qui signifie codeur de voix.

discours musical, notamment à l'étape de correction où l'on vérifie que chaque index déduit par analyse algorithmique correspond à des paramètres du discours musical. Pour les musiques qui ne sont pas représentées sous forme de partition, il est plus difficile de tenir compte d'une structure explicite de la pièce au moment de l'indexation des éléments significatifs du discours musical. Il faut donc passer par une phase d'analyse préalable du contenu, voire de la structure de la pièce, phase dont la partition permet en partie l'économie. Une partition a posteriori peut ainsi être élaborée [17]. Les méthodes de représentation iconiques, par exemple celles élaborées par Pierre Couprie [3], constituent des outils efficaces pour représenter le contenu et la structure de musiques électroacoustiques sous forme graphique. En effectuant ce travail d'analyse préalable, on peut mettre en évidence les éléments saillants d'un enregistrement, effectuer une segmentation fine des structures sonores de base et proposer une représentation intégrant des éléments d'analyse comme la texture des sons, leurs enveloppes temporelles ou encore les hauteurs spectrales.

Pour l'instant, nous affichons seulement la forme d'onde et les index placés sur le fichier audio dans notre interface graphique (figure 4). Cette représentation est assez limitée en termes d'information musicale et ne permet pas d'anticiper la nature du prochain son qui sera joué. Il serait nécessaire d'intégrer une représentation plus musicale, comme par exemple une partition qui défilerait de façon synchronisée avec le son. Pour les musiques sans partition, nous pensons intégrer une représentation iconique qui défilerait elle aussi de façon synchronisée avec le son.

Les pistes de réflexion abordées dans nos propositions d'interfaces nous amènent à constater que, pour chaque son d'un enregistrement que l'on souhaite rejouer, il serait intéressant de tenir compte du substitut gestuel évoqué dans la conception de l'interaction. Cela suppose que pour un morceau de musique électronique ou électroacoustique comportant une richesse sonore aussi importante que « Accidents/harmoniques », il ne faudrait pas se limiter à un seul type de geste pour contrôler le déroulement de l'ensemble de la pièce : il faudrait au contraire un ensemble de gestes correspondant à chacun des substituts gestuels rencontrés lors de l'écoute de l'enregistrement. Les nombreux capteurs du Méta-Instrument 4 en font un outil bien adapté pour ce projet. En concevant des interactions autour des quatre types de capteurs présents sur l'instrument, on peut imaginer une multitude de mappings gestuels grâce auxquels on articulerait le passage d'un index sonore en tenant compte de la spectromorphologie des substituts gestuels évoqués par le son enregistré.

5. REMERCIEMENTS

Ce travail a bénéficié du soutien financier de la région Île-de-France et de l'Union Européenne grâce au projet SMAC (FEDER IF0011085).

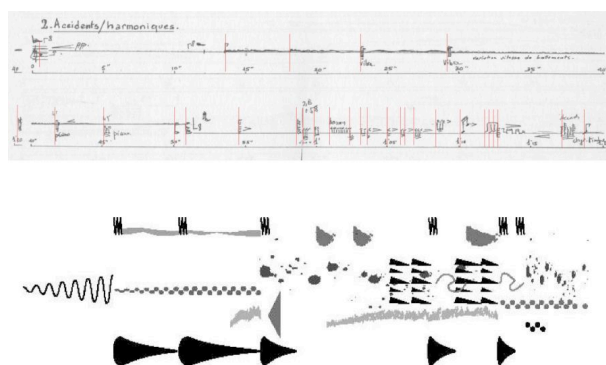


Figure 4. Deux représentations a posteriori de musique sur support : partition graphique d'« Accidents/Harmoniques » par Philippe Mion [17] (en haut) et représentation iconique de *Reflet* de Ivo Malec (0'00" à 1'00") par Pierre Couprie [3].

6. RÉFÉRENCES

- [1] Bogdanov, D., et coll. « ESSENTIA : An Open-Source Library for Sound and Music Analysis », Proceedings of the Association for Computer Machinery International Conference on Multimedia, Barcelone, Espagne, 2009.
- [2] Charles, J. F. « A Tutorial on Spectral Sound Processing using Max/MSP and Jitter », *Computer Music Journal* 32/3 (2008), p. 87-102.
- [3] Couprie, P. « Graphical Représentation : An Analytical and Publication Tool for Electroacoustic Music », *Organised Sound* 9/1 (2014), p. 109-113.
- [4] Couprie, P. « Le Méta-Instrument : genèse et évolution d'un nouvel instrument », *Musique-Images-Instruments* 17 (2018), p. 231-245.
- [5] D'Alessandro, C., et coll. « Borrowed Voices », Proceedings of the International Conference on New Interfaces for Musical Expression, Porto Alegre, Brésil, 2019.
- [6] D'Alessandro, C., et coll. « Les instruments chanteurs », *Acoustique et techniques* 89 (2019), p. 36-43.
- [7] Delalez, S., D'Alessandro, C. « Vokinesis : Syllabic Control Points for Performative Singing Synthesis », Proceedings of the International Conference on New Interfaces for Musical Expression, Université Aalborg de Copenhague, Danemark, 2017, p. 198-203.
- [8] Delalez, S., D'Alessandro, C. « Adjusting the Frame : Biphasic Performative Control of Speech Rhythm », Proceedings of the Interspeech Conference, Stockholm, Suède, 2017, p. 864-868.
- [9] Gerhard, W., Böck, S. « Maximum Filter Vibrato Suppression for Onset Detection », Proceedings of the 16th Digital Audio FX Conference, Maynooth, Irlande, 2013.

- [10] Gonin, P. « “Interstellar Overdrive” de Pink Floyd et le processus de création à travers le prisme de l'improvisation collective : du chaos à l'ordre ? », *Musimédiane* 11 (2019-2020). [www.musimediane.com/numero11/GONIN, accédé le 14/04/2020.]
- [11] Goto, M. « A Chorus Section Detection Method for Musical Audio Signals and its Application to a Music Listening Station », *Institut of Electrical and Electronics Engineers – Transactions on Audio, Speech, and Language Processing* 14/5 (2006), p. 1783-1794.
- [12] Goto, M. « Active Music Listening Interfaces based on Signal Processing », Proceedings of the International Conference on Acoustics, Speech and Signal Processing – Institute of Electrical and Electronics Engineers, Honolulu, Hawaii, USA, 2007, p. 1441-1444.
- [13] Haury, J. « La pianotechnie ou notage des partitions musicales pour une interprétation immédiate sur le métapiano », Actes des Journées d'Informatique Musicale, Grenoble, 2009, p. 9-17.
- [14] Holzapfel, A., et coll. « Selective Sampling for Beat Tracking Evaluation », *Institut of Electrical and Electronics Engineers – Transactions on Audio, Speech, and Language Processing* 20/9 (2012), p. 2539-2548.
- [15] Jehan, T. « Creating Music by Listening ? », thèse de doctorat (PhD), Massachusetts Institute of Technology, Cambridge (MA), 2005.
- [16] Mancero, D. « Composer à partir de la modélisation harmonique des soundscape-compositions : quels enjeux pour l'analyse fonctionnelle des objets sonores ? », Actes des Journées d'Informatique Musicale, Paris, 2017.
- [17] Mion, P., et coll. *L'envers d'une œuvre : De Natura Sonorum de Bernard Parmegiani*. INA-GRM/Buchet-Chastel, Paris, 1983.
- [18] Muller, M., Grosche, P. « A Mid-Level Representation for Capturing Dominant Tempo and Pulse Information in Music Recordings », Proceedings of the International Conference On Music Information Retrieval, Kobe, Japon, 2009, p. 189-194.
- [19] Oswald, J. « Plunderphonics, or Audio Piracy as a Compositional Prerogative », Proceedings of the Wired Society of Electro-Acoustic Conference, Toronto, Canada, 1985.
- [20] Schaeffer, P. *Traité des Objets Musicaux*. Seuil, Paris, 1966.
- [21] Schnell, N. « Playing (with) Sound », thèse de doctorat, Université de Graz, 2013.
- [22] Smalley, D. « Spectromorphology : Explaining Sound Shapes », *Organised Sound* 2/2 (1997), p. 107-126.
- [23] Varela, F. J., Thompson, E. T., Rosch, E. *The Embodied Mind : Cognitive Science and Human Experience*. MIT Press, Cambridge (MA), 1991.
- [24] Xiao, X., et coll. « T-Voks : The Singing and Speaking Theremin », Proceedings of the International Conference on New Interfaces for Musical Expression, Porto Alegre, Brésil, 2019, p. 110-115.
- [25] Zapata, J. R., et coll. « Multi-Feature Beat Tracking », *Institute of Electrical and Electronics Engineers – Association for Computer Machinery – Transactions on Audio, Speech, and Language Processing* 22/4 (2014), p 816-825.

Texte édité par Nathalie Hérold