

Annoter la musique : de la segmentation de fichiers audio à la publication d'articles multimédia

Nicolas Donin

IRCAM
1, place Igor Stravinsky
75004, Paris, France
nicolas.donin@ircam.fr

Samuel Goldszmidt

IRCAM
1, place Igor Stravinsky
75004, Paris, France
samuel.goldszmidt@ircam.fr

RESUME

Cet article propose un état de l'art succinct des outils de manipulation et de publication réalisés par l'équipe Analyse des pratiques musicales de l'Ircam. Des outils de segmentation et comparaison de flux musicaux jusqu'aux jeux d'interaction possibles entre une partition annotée et d'autres médias, différentes IHM sont présentées dans leur relation avec les pratiques musicales, pédagogiques et musicologiques considérées.

MOTS CLES : outils de publication multimédia, écoute musicale, guide d'écoute, partition annotée, segmentation de flux temporel, sémantique située.

ABSTRACT

This paper briefly describes various tools dealing with music data manipulation and publication, produced by Ircam research group 'Analyse des pratiques musicales'. Tools for audio segmentation and labeling, for interacting with a polyrhythmic grid, for editing and publishing listening guide with annotated scores...: all these MMI are introduced in relation with the specific musical, educational, musicological practices at stake.

CATEGORIES AND SUBJECT DESCRIPTORS:

H.5.2 [User Interfaces]: Ergonomics|Graphical User Interface|User-Centered Design; H.5.4 [Hypertext/Hypermedia]: User issues; J.5 [Arts and humanities]: Music.

GENERAL TERMS: Design, Education, Musicology, Multimedia, Sound enhancement

KEYWORDS: multimedia publication tools, musical hearing, listening guide, annotated score, temporal object segmentation, situated semantics

INTRODUCTION

Cette présentation porte sur l'*annotation de documents musicaux*, en rendant compte de différentes IHM réalisées dans l'équipe Analyse des pratiques musicales dans des contextes divers : maquettage d'outils d'aide à l'écoute, projets de recherche empirique, projets pédagogiques. Ces différentes IHM ont au moins une chose en commun : elles ont été développées, de leur

définition à leur prototypage (et le cas échéant à leur validation), en collaboration avec leur usagers. Ces derniers appartiennent à des communautés d'utilisateurs ciblées, avec une pratique d'écoute musicale singulière ; d'où des publications/démonstrateurs dans lesquels l'interaction, qu'elle soit large ou restreinte, se caractérise par sa finesse.

Selon les disciplines et selon les esthétiques musicales adressées, il y a plusieurs façons bien différentes de distinguer technologies musicales et technologies sonores. On peut par exemple considérer que le domaine musical est un *sous-ensemble* du domaine sonore, spécifié au moyen de critères de culturels implémentables dans des algorithmes (reconnaissance de patterns harmoniques, de rapports pulsation/mètre, etc.). Ici, nous considérerons plutôt que le musical et le sonore sont *deux ensembles dont le recoupement est partiel et évolutif* : un même document sonore peut ou non relever d'une ingénierie musicale selon le dispositif où il s'insère ; à l'inverse, il existe des documents musicaux qui n'ont pas de rapport direct avec le son – ainsi notamment des nombreuses notations musicales rétives à la norme MIDI (par exemple la notation neumatique). Ainsi, comme nous travaillons avec des créateurs et des commentateurs de *musique contemporaine*, les documents musicaux à traiter sont *originellement multimédia*, en ce sens qu'ils impliquent par défaut des enregistrements sonores, des programmes informatiques, et des partitions (dans des notations plus ou moins conventionnelles et des formats plus ou moins en compréhension). Les outils auteurs permettant de les traiter sont donc basés sur un flux sonore et/ou d'autres flux temporels synchronisés, ainsi que des données non temporelles qui leurs sont articulées (textes, images).

SEGMENTATION DE FLUX TEMPORELS

Une interface de segmentation linéaire et manuelle

L'une des opérations les plus minimales, mais aussi les plus importantes, de l'annotation musicale est la sélection d'un son et sa qualification verbale. Cette opération peut être réalisée en éditant un fichier audio dans un programme usuel de manipulation du son, puis en le nommant. Mais elle peut aussi être faite de façon non destructive et intégrée à une chaîne éditoriale

idoine : c'est le principe de l'outil de segmentation manuelle jUST que nous développons en relation avec un groupe de praticiens experts de la segmentation & qualification d'enregistrements musicaux, à savoir les compositeurs et enseignants du groupe de recherche Musique et Informatique de Marseille.

La pratique de ces derniers consiste à analyser des enregistrements audio d'œuvres musicales selon un lexique d'une vingtaine d'« Unités sémiotiques temporelles » (UST) [1] à vocation de lexique universel. Ces unités sont successives : une œuvre musicale doit pouvoir être décrite selon un enchaînement continu d'UST. L'assignation de ces UST à une œuvre est l'objet de séances de travail collectives où différentes segmentations/qualification sont testées puis l'une d'elles validée.

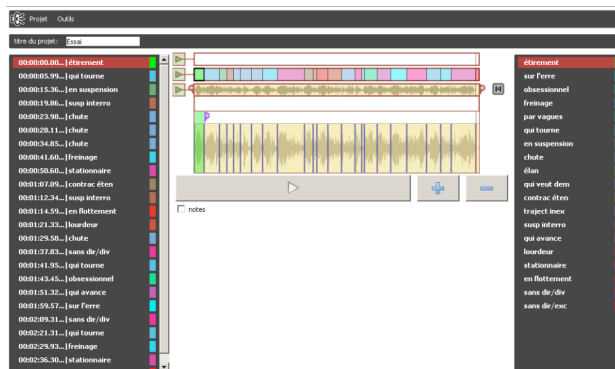


Figure 1 : Interface de jUST, outil de segmentation & qualification manuelle de fichier audio

L'outil jUST permet de segmenter un fichier son, de manière linéaire, selon un grille de lecture décrite par un fichier de définition ; déposé sur un site web, ce dernier peut être partagé par une communauté d'utilisateurs. L'outil, développé en XUL, part du composant de segmentation audio développé dans le cadre du projet RIAM SCENARIPLATFORM en 2007. Les descripteurs choisis sont de simples tags Xml associés à chaque segment. Cette liste de descripteurs, intégrée dans l'outil mais externalisée dans le cadre d'un modèle de publication de l'outil auteur Scenari utilisant le même composant de lecture audio, peut servir à marquer des segments audio selon une liste d'autorités. L'outil et sa documentation sont à : <http://recherche.ircam.fr/equipes/apm/jUST/>

Une interface d'exploration de segmentations implicites, basée sur un alignement automatique

La discrétisation manuelle de fichier audio peut également s'appuyer sur une segmentation automatique préalable. Ainsi, dans notre interface de navigation et comparaison entre différents enregistrements discographiques d'une même œuvre, l'activité de détermination & qualification de segments n'est pas un point de départ mais plutôt une finalité. Il s'agissait d'instrumenter une pratique d'écoute et d'analyse musicale particulière, la comparaison d'interprétations,

qui vise à mettre en valeur l'apport singulier de chaque interprète à un texte musical de référence – et plus spécifiquement, dans une approche musicologique, il s'agit de montrer comment chaque interprète, par le détail de sa réalisation temporelle des notes prescrites par la partition, détermine une certaine lecture de la forme de l'œuvre – une segmentation implicite [2].

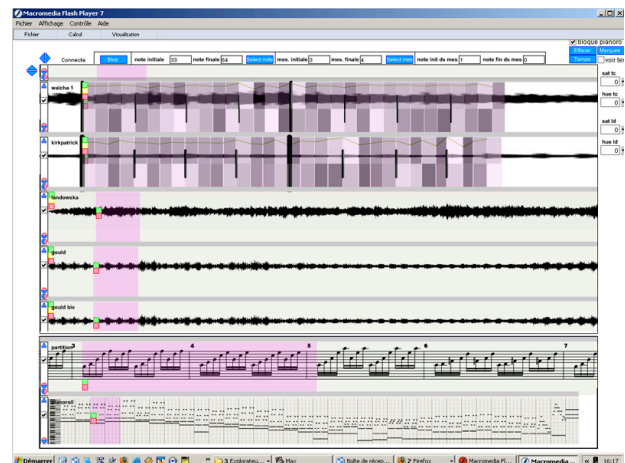


Figure 2 : Interface de navigation dans la « Bachothèque »

La "Bachothèque" rend comparable une série de versions du 1^{er} prélude du *Clavier bien tempéré* de J.-S. Bach par une représentation simultanée de certaines de leurs caractéristiques (durée entre chaque attaque, tempo moyen par mesure, intensité moyenne de chaque note). Les comparaisons s'appuient sur les résultats fournis par un algorithme d'alignement MIDI/audio produit par l'équipe Analyse/synthèse de l'Ircam, détectant automatiquement chaque attaque de note successive prédite à l'aide de l'information MIDI [4]. Une interface (figure 2) permet de naviguer entre les flux audio ainsi préparés, dans une visualisation tabulaire de chacune des interprétations (forme d'ondes) et de la partition (doublement représentée à partir du MIDI : *pianoroll* synoptique mesure par mesure, portée note par note). Une sélection depuis une représentation sélectionne aussi les passages correspondants dans toutes les autres, permettant une écoute comparée immédiate.

Au sein d'une sélection (zoomable), l'utilisateur peut appeler une visualisation de la durée et de l'intensité de chaque note (cf. lignes 1 & 2, fig. 2). Cela permet de retrouver à l'écoute – ou pas... – les infimes variations de durée détectées grâce à l'alignement des fichiers. Pour faciliter la focalisation locale de l'attention, l'utilisateur peut poser des marques sur chaque note ; il peut aussi annoter textuellement l'état d'une sélection et/ou de son calcul. Le processus de comparaison se traduit donc par un parcours exploratoire des différentes représentations des différents documents, dont des annotations liées à des *preset* gardent la trace.

PRENDRE EN COMPTE LA POLYPHONIE, OU : L'ANNOTATION VERTICALE

Les interfaces précédentes annotent les flux musicaux en se basant sur une discrétisation essentiellement horizontale. Elles n'exploitent donc pas une propriété fondamentale des musiques ainsi traitées, et plus généralement de la musique occidentale notée : la polyphonie, c'est-à-dire la cohabitation de plusieurs lignes musicales différentes sonnant simultanément. Cette dimension verticale de la musique appelle des modalités d'annotation appropriées, que nous allons mentionner à partir de deux cas contrastés.

Surlignage/démixage de couches sonores

Andrea Cera, compositeur contemporain, cherche à décomposer la singularité rythmique qu'il perçoit dans certaines œuvres de musique populaire, afin de les réinjecter dans sa propre musique. Il a développé à cet effet une pratique d'écoute intensive de courts extraits sonores comportant de nombreuses couches rythmiques simultanées, dont il réalise des transcriptions détaillées. Nous avons développé avec lui un outil permettant de "démixer verticalement" de telles boucles, puis de les rendre intelligibles pour un auditeur non expert. Ce dernier peut, *via* une interface adaptée exploitant le même outil, s'approprier le résultat en "surlignant auditivement" telle ou telle couche, grâce à une sonification supplémentaire (émission, synchrone aux impacts transcrits, d'un son similaire à celui visé). En écoutant l'une après l'autre et/ou l'une avec l'autre ces différentes couches, on se familiarise avec la façon dont Cera a exploré le son.

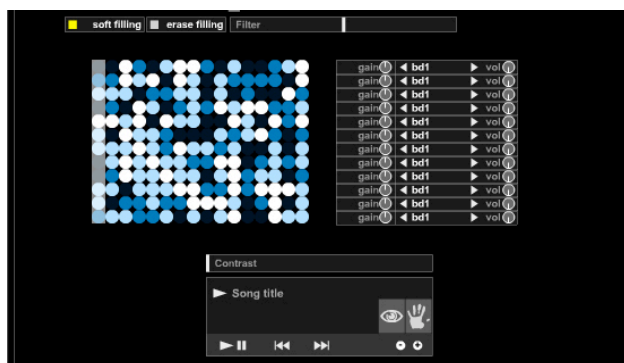


Figure 3 : Interface d'exploration auditive de boucles rythmiques *via* une transcription

Nous avons également exploité dans une autre collaboration avec un compositeur, Ph. Leroux, le fait qu'une même interface de type séquenceur puisse servir à la fois à l'analyse, à l'écoute et à la composition [6].

Annotation de partition pour guider l'audition

Dans le cas de la musique conçue sur partition, la focalisation de l'attention auditive sur différentes voix d'une polyphonie peut se faire au moyen de la lecture de la partition (cette pratique est d'ailleurs un principe de l'éducation de l'oreille). Plusieurs documents

multimédia, tant scientifiques que pédagogiques, ont été réalisés sur cette prémisse. Tous s'appuient sur l'*annotation de partition* : un série d'images (pages de partition scannées) est synchronisé à un fichier audio (enregistrement de l'œuvre correspondante), selon une échelle pertinente (par ex. à la mesure ou à la note près – ainsi un curseur peut se déplacer au rythme précis de la lecture/écoute de la partition). Sur cette base, des annotations graphiques et textuelles sont éditables à la fois spatialement (emplacement dans une image) et temporellement (apparition/disparition en fonction de ce qu'il est pertinent d'afficher à un instant donné). Dès lors, il est possible de viser des sous-ensembles de la polyphonie audio au moyen de déictiques visuels.



Figure 4 : Guide d'écoute d'une œuvre d'Anton Webern pour quatuor à cordes, version Pocket PC

La figure 4 présente un guide d'écoute de ce type, initialement réalisé en 2003 sous forme de cours en ligne dans le cadre du projet européen MusicWeb (<http://musicweb.koncon.nl>). L'annotation de la partition est animée temporellement et pointe des relations entre motifs musicaux au fil de l'écoute. Cette version pour Pocket PC a été réalisée dans le cadre du projet européen SemanticHifi (IST, FP6, voir <http://shf.ircam.fr>).

De l'annotation à sa réutilisation

Comme le suggère déjà l'exemple précédent, une annotation horizontale et/ou verticale d'un matériau musical donné peut, si elle est pertinente, être utilisée comme critère d'interaction avec ce matériau. On passe alors de la production manuelle de métadonnées à la manipulation automatisée des données. Ainsi l'outil réalisé dans le cadre de la collaboration avec Cera a-t-il également donné lieu à la conception d'un plug-in logiciel et d'une interface tactile pour le mixage en temps réel de deux fichiers sons préalablement prétranscrits selon la procédure élicitée avec Cera [5].

PUBLICATION HYPERMEDIA D'ANNOTATIONS

Le modèle éditorial de la fig. 4 apparaissant satisfaisant pour guider une première écoute et naviguer selon des critères liés à l'écriture musicale, il a notamment servi à la conception d'un logiciel d'aide à l'analyse musicale réalisé par l'Ircam en collaboration avec les Ministères

de l'Education nationale et de la Culture : ML-Annotation, un des trois outils « Musique Lab 2 » [3].

Cependant, dans beaucoup de pratiques musicales expertes de l'interaction entre lecture de partition et écoute, l'annotation graphique & textuelle à même la partition n'est pas suffisante. En particulier, l'*analyse musicale* (équivalent musicologique du commentaire littéraire) requiert une rhétorique à la fois textuelle (démonstration argumentée) et graphique (monstration par des exemples musicaux). Il faut donc pouvoir articuler la partition annotée à du texte, et, par extension, à tout autre média utile. C'est que nous avons fait pour la publication de *Parcours interactifs* basés sur des analyses pédagogiques utilisant les logiciels ML2.

Annotation pédagogique de partition annotée : décomposer et commenter des parcours analytiques

Chaque cédérom *Parcours interactif* traite d'une œuvre musicale considérée dans sa totalité et dans sa continuité. Après une page d'introduction, le lecteur/auditeur accède à un enregistrement de l'œuvre synchronisé avec une reproduction de la partition. L'interactivité est minimale, de même que le discours, car il s'agit alors de privilégier une écoute de découverte. Passant de ce mode dit de "Première écoute" au mode "Parcours interactif" (voir figure 5), l'affichage du flux précédent se trouve complété : d'une part, dans la partie inférieure de l'écran, par une *timeline* multiple comportant plusieurs couches de balises qui sont autant de parcours thématiques à travers l'œuvre ; d'autre part, dans une nouvelle colonne, par un pavé textuel commentant le parcours sélectionné depuis la *timeline* : chacun de ces textes présente les différentes occurrences de la notion qu'il commente, accessibles à chaque fois par un lien hypermédia pointant soit vers une séquence de partition annotée (réalisée au moyen de ML-Annotation), soit vers des compléments (vidéo, texte illustré).

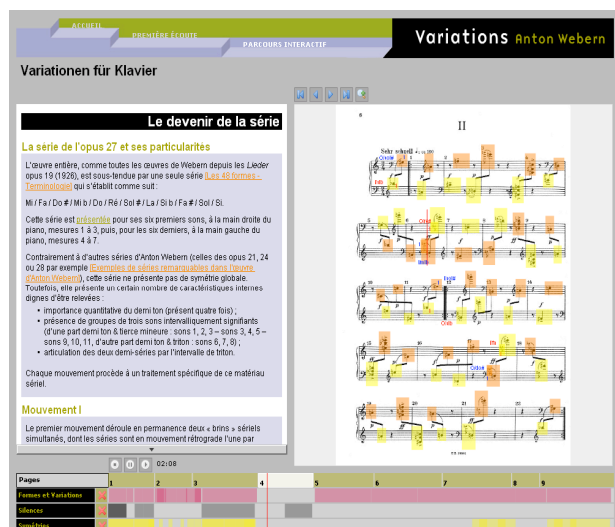


Figure 5 : Mode de navigation « Parcours interactif » dans le CD-Rom pédagogique *Un parcours interactif dans les Variations op. 27 de Webern* (Ircam/Région PACA)

Le CD-Rom fournit également à l'utilisateur les documents originaux de l'analyse (images, sons, fichiers d'annotation et logiciels correspondants), afin de lui permettre de les éditer lui-même d'une autre manière, que ce soit pour compléter ou pour critiquer l'analyse.

CONCLUSION

Toutes les IHM ici présentées a été validé relativement aux opérations et contenus spécifiquement visés dans chaque projet. Chaque dispositif est néanmoins susceptible, même s'il a d'abord thématisé la singularité d'une pratique artistique, de donner lieu à généralisation. Ainsi est-on passé, en ce qui concerne l'annotation d'image/son de partition, de la réalisation de cours d'analyse musicale en ligne au développement d'une solution éditoriale hypermédia pour un large ensemble de discours sur la musique, allant de la documentation des œuvres à la pédagogie de leur analyse. Ce parcours a mis en évidence l'intérêt de systèmes d'annotation où l'utilisateur pourra donner du sens à des éléments (par ex. définir le mode de lecture de la partition scannée, ou tagger l'information MIDI) et y entraîner des algorithmes (d'analyse audionumérique tout comme d'OCR) qui l'aideront à modéliser/tester ses prescriptions sur l'ensemble de ses matériaux.

BIBLIOGRAPHIE

1. Coll. *Les Unités sémiotiques temporelles, éléments nouveaux d'analyse musicale*. Marseille, MIM/Eska, 1996.
2. Donin, N. *Problèmes d'analyse de l'interprétation. Un essai de comparaison assistée par ordinateur d'enregistrements du premier prélude du Clavier bien tempéré*. Musurgia, Vol. XII, No 4, pp. 19-43.
3. Puig V., Guédy F., Fingerhut M., Serrière F., Bresson J. et Zeller O. *Musique Lab 2: A Three Level Approach for Music Education at School*. In *Proc. of the International Computer Music Conference (ICMC 2005)*, Barcelone, 2005.
4. Rodet, X., Escribe, J., Durigon, S. *Improving score to audio alignment: Percussion alignment and Precise Onset Estimation*. In *Proc. of the International Computer Music Conference (ICMC 2004)*, Miami, 2004, pp. 450-453.
5. Schnell, N., Schwarz, D. et Muller, R. *X-Micks - Interactive Real-Time Content Based Audio Processing*. In *Proc. of the 9th Int. Conference on Digital Audio Effects (DAFx-06)*, Montréal, 2006.
6. Theureau, J. et Goldszmidt, S. *Explicitation d'écoutes musicales singulières, analyse des activités d'écoute musicale et conception de situations d'écoute musicale*. In *XXXIIe congrès de la SELF (Saint Malo, 5, 6, 7 septembre 2007)*, Toulouse, Octares Editions, 2007, p. 79-89.