Traitement par le contenu des signaux musicaux

Hugues Vinet

Traitement par le contenu des signaux musicaux

Hugues Vinet, Directeur scientifique de l'Ircam

Biographie

Hugues Vinet est depuis 1994 directeur scientifique de l'Ircam, dont il dirige le département Recherche et développement. Ses domaines d'intérêt portent notamment sur le traitement du signal audionumérique, les interfaces homme-machine, l'ingénierie des connaissances musicales et l'épistémologie des relations entre recherche scientifique, développement technologique et création musicale. Il assure la coordination de projets européens (Cuidado, SemanticHIFI) et nationaux (Ecrins, Ecoute, Voxstruments, Sample Orchestrator 1&2). De formation scientifique et musicale, il a précédemment travaillé au Groupe de recherches musicales de l'Institut national de l'audiovisuel où il a animé de 1987 à 1994 les activités de recherche et développement, ayant notamment abouti à la réalisation des premières versions des logiciels GRM Tools, Acousmographe et MIDI Formers. Il est membre de diverses instances scientifiques et depuis 2006, Vice-Président Europe de l'International Computer Music Association.

Résumé

Les technologies musicales font largement appel à des fonctions de traitement sonore, qu'il s'agisse d'instruments électroniques ou logiciels, d'outils de création et de production ou d'interfaces de navigation et de lecture. Longtemps limitées à des algorithmes opérant de manière globale et indifférenciée sur tous types de sons, les recherches récentes dans ce domaine tendent à s'orienter vers des traitements par le contenu, reposant sur une analyse de leurs caractéristiques et structures internes. Ces avancées ouvrent de nombreuses perspectives nouvelles, tant en matière de manipulation intuitive et de possibilités créatives que d'amélioration qualitative de traitements existants ou d'automatisation d'opérations fastidieuses. Cet article propose une vue d'ensemble de l'état de l'art du domaine, en en précisant les approches et problématiques et en les illustrant par des exemples récents d'applications, notamment issus des recherches de l'Ircam.

Le signal musical, capté par un microphone ou issu d'un enregistrement, est le support d'un contenu sonore fortement structuré, qui résulte d'une séquence d'opérations techniques complexes combinant lutherie, composition, interprétation, prise de son, post-production, mixage et *mastering*. Si l'essentiel de cette organisation sonore est perceptible à l'écoute, sa modification l'est plus difficilement, les différentes représentations intermédiaires des contenus musicaux – partitions, données d'interprétation et enregistrements d'instruments et composantes sonores isolés – n'étant pas codées ni disponibles séparément dans le signal résultant. Ainsi, les possibilités de manipulation de tels contenus enregistrés se sont longtemps limitées à des traitements élémentaires - volume, équilibre spectral, balance, effets, réverbération artificielle ou à des algorithmes de traitement plus élaborés destinés à la création, tels que ceux des modules GRM Tools[1], imprimant leur transformation à l'intégralité du contenu du signal. De plus, l'accès aux structures temporelles de sons enregistrés se réduit généralement à la position d'un index de lecture sur un segment, et au mieux, dans les logiciels d'édition, à une représentation des amplitudes du signal au cours du temps.

Dans l'état de l'art actuel de la recherche, les traitements parmi les plus complexes applicables à tous types de sons et effectuant la modification de paramètres musicaux avec une qualité satisfaisante pour des applications professionnelles sont réalisés par une modélisation des signaux reposant sur le vocodeur de phase, ou transformée de Fourier à court terme [Allen and Rabiner 1977, Moorer 1978] et permettent des transformations telles que la transposition (modification de la hauteur sans changement de la durée) ou la compression-expansion temporelle ou *time stretching* (traitement symétrique lié à un changement de la vitesse de lecture sans altération du contenu fréquentiel). Ces traitements s'appliquent globalement à l'ensemble des éléments sonores constitutifs du signal et offrent un niveau de contrôle adapté à l'utilisateur musicien, en lien direct avec les paramètres de l'écriture (hauteurs, tempo)[2].

Traitement par le contenu et niveaux d'abstraction des informations musicales

Plus généralement, la réalisation de fonctions de traitement adaptées aux usages musicaux implique qu'elles se fondent sur un contrôle de « haut niveau » - notion que nous nous proposons de préciser dans la suite – en lien direct avec les paramètres constitutifs de la composition musicale. Il peut s'agir soit pour certaines applications créatives d'extraire certains paramètres d'un son pour les appliquer à un autre ou le plus souvent de transformer un son en fonction de critères relevant directement du vocabulaire musical.

L'analyse de cette problématique peut s'appuyer sur une typologie que nous avons proposée des représentations numériques des informations musicales, organisant

[1] http://www.inagrm.com/grmtools

[2] Les récentes améliorations de ces algorithmes font cependant intervenir une adaptation au contenu du son traité : la transposition de la voix d'un locuteur sans altération de son timbre implique l'extraction préalable et la préservation de son enveloppe spectrale, caractéristique de la réponse acoustique de son conduit vocal ; la réduction des artefacts liés au ralentissement d'un son implique la détection et un traitement différencié assurant la préservation des transitoires d'attaques [Roebel 2003, 2010].

celles-ci en niveaux d'abstraction croissants, les niveaux *physique* , *signal*, *contrôle*, *symbolique* et *sémantique* (cf. Figure 1) [Vinet 2003]. Le niveau symbolique, entendu au sens informatique, comprend les structures relatives à des échelles de valeurs discrètes des paramètres du sonore (hauteurs, intensités, durées et occurrences temporelles) et le niveau sémantique s'attache à toute description textuelle du contenu musical, leur combinaison constituant l'ensemble des informations contenues dans la partition. Quant au niveau *contrôle*, il rend compte de l'action corporelle de l'interprète, instanciant les paramètres musicaux sous forme continue, tant dans leur temporalité que dans leurs valeurs. L'ordre croissant d'abstraction de ces niveaux est en rapport direct avec la décroissance de la quantité d'information par unité de temps et de la bande passante nécessaire au codage des représentations correspondantes. Les niveaux d'abstraction élevés font référence à des connaissances implicites, relevant de théories musicales ou de conventions culturelles, alors que les représentations de plus bas niveau, plus volumineuses, véhiculent en elles-mêmes l'intégralité des informations nécessaires. La conversion entre niveaux différents est ainsi assimilable, dans le sens croissant d'abstraction, à un processus d'analyse ou extraction d'information et dans le sens décroissant à un processus de synthèse ou génération d'information, la descente du contrôle au signal correspondant notamment à une notion élargie d'instrument, comme nous le verrons dans la suite.

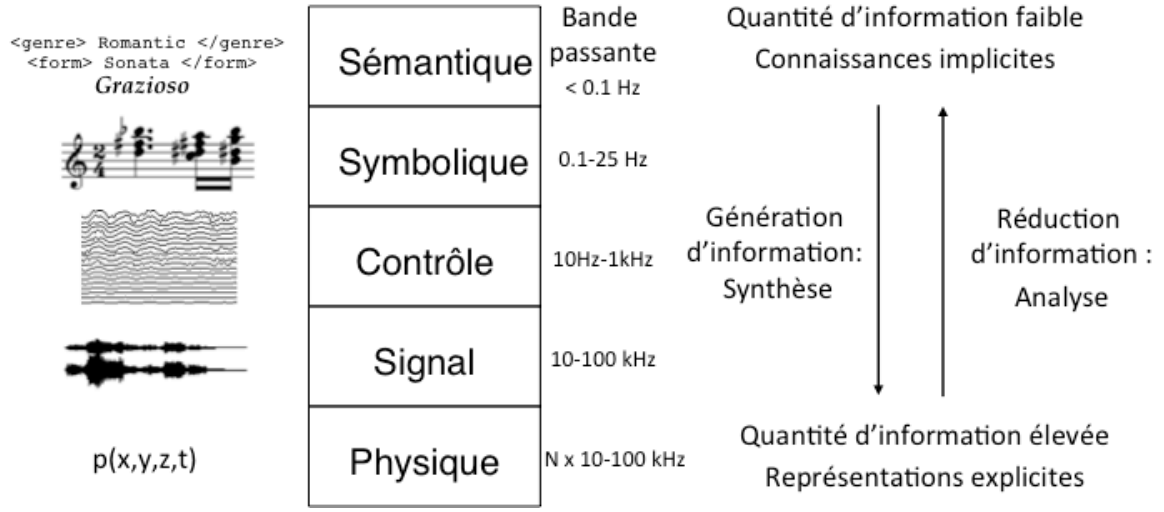

Figure 1 : les niveaux de représentation des informations musicales

Sous l'éclairage de ce modèle, notre problématique se traduit le plus souvent par la nécessité d'accès des paramètres de transformation de signaux sonores au niveau symbolique, impliquant d'abord une phase d'analyse (signal/symbolique), un traitement dans le domaine symbolique puis une regénération du signal par synthèse (symbolique/signal). Si la transformation concerne les paramètres continus de l'interprétation (vibrato et tremolo, variation de tempo, crescendo/decrescendo), la même structure d'analyse/synthèse s'applique entre les niveaux signal et contrôle. Il est à noter que la phase d'analyse tend à reconstituer les représentations intermédiaires issues des étapes de production technique mentionnées plus haut.

Les verrous scientifiques et technologiques actuels en traitement de signal audio liés à la réalisation des traitements par le contenu concernent en premier lieu les

problématiques d'analyse, dont la section suivante propose une synthèse sous forme d'état de l'art résumé.

Problématiques, stratégies et outils d'analyse

L'analyse des signaux sonores fait intervenir les notions suivantes :

- la *segmentation* d'un flux sonore consiste à le découper en segments temporels dont les bornes sont déterminées par des événements particuliers – tels que début et fin de note - ou définissent un intervalle temporel de stabilité d'un ou plusieurs paramètres sonores (hauteur, intensité, instrumentation, etc.) ; un cas particulier concerne les segmentations régulières relatives aux musiques pulsées. La segmentation peut être simple ou se décomposer de manière hiérarchique pour refléter les structures temporelles imbriquées inhérentes à la musique (parties et mouvements, mesures, rythmes).
- les descripteurs, notion introduite dans la norme MPEG7 [MPEG7 2002], rendent compte des différentes caractéristiques du son : hauteur, intensité, contenu spectral, tempo, instrumentation, etc. Leur contenu peut selon les cas prendre plusieurs formes, numérique, textuelle - nom d'instrument par exemple - voire celles de structures plus complexes. Il peut être défini comme une fonction du temps ou comme une constante sur un segment donné. En lien avec la section précédente, on distingue les descripteurs de bas niveau, sous forme numérique et facilement extractibles automatiquement par analyse de signal [Peeters 2004, 2011b] des descripteurs de haut niveau, en lien avec des grandeurs et catégories pertinentes du point de vue de la cognition humaine, mais dont l'extraction automatisée à partir du signal peut présenter une difficulté variable, selon notamment la nature et le caractère plus ou moins explicite de la relation psychophysique existant entre grandeur mesurée et percept [Susini 2012].
- les possibilités d'analyse de flux sonores *en temps réel*, notamment issus de signaux produits en direct, sont beaucoup plus contraintes, en termes de latence admissible, de puissance de calcul et de nécessité de prise en compte des informations de manière causale, c'est-à-dire au fur et à mesure qu'elles arrivent, que celles d'analyses *en temps différé* de sons pré-enregistrés.

Deux cas principaux peuvent être distingués selon qu'ils concernent l'analyse :

1. de sons *monophoniques*, c'est-à-dire issus d'une seule source acoustique ou électronique. Lorsqu'en particulier celle-ci est un instrument de musique, l'extraction de ses principaux descripteurs musicaux - hauteur, intensité, attributs du timbre- comme fonctions continues du temps, de même que sa segmentation en notes sont réalisables avec de bonnes performances en termes de coût de calcul et de taux d'erreur, dans certaines conditions de prise de son.
2. de sons *polyphoniques*, c'est-à-dire comportant une superposition de sons différents. Il serait idéalement souhaitable de décomposer la scène sonore en autant de flux monophoniques pour les traiter indépendamment et leur appliquer les analyses correspondantes, mais l'état de l'art de la recherche en séparation de sources ne le permet pas dans le cas général. Les performances d'analyse s'améliorent en fonction des informations connues a priori, telles que le nombre et la nature des sources en présence [Vincent 2010]. Lorsqu'une représentation de la partition est connue, des techniques d'alignement automatique identifient avec une bonne précision le positionnement temporel de chaque note jouée dans le signal audio [Kaprykowsky 2006]. L'aboutissement

récent de travaux de recherche sur l'analyse de fréquences fondamentales multiples permet depuis peu une transcription polyphonique de qualité satisfaisante dans certaines conditions [Yeh 2010]. D'autres stratégies d'analyse consistent à calculer des descripteurs s'appliquant à l'intégralité du flux polyphonique et font l'objet de nombreux résultats de recherche portant notamment sur la structure temporelle globale du morceau [Peeters 2007], l'analyse de tempo [Peeters 2011a], le contenu harmonique [Gomez 2004], etc.

Cet état de l'art résumé et concentré sur des notions simples ne doit cependant pas occulter les multiples problématiques inhérentes à l'analyse musicale en général. L'extraction d'informations numériques relatives aux principales catégories de la musique occidentales (hauteurs, intensités, structures temporelles) devient insuffisante pour l'analyse d'œuvres récentes fondées sur un vocabulaire étendu à d'autres caractéristiques du sonore (timbre, modes de jeux, échelles microtonales, sons électroniques). Ceci est d'autant plus vrai des « musiques de bruit » acousmatiques dans lesquelles la référence instrumentale a disparu et l'organisation des hauteurs ne joue plus un rôle structurant [Schaeffer 1966]. De plus, si l'on se restreint au champ de la tradition instrumentale occidentale, la référence à la partition ne constitue, dans le champ de la sémiologie musicale, qu'un niveau neutre, consignant une prescription d'interprétation, d'une tripartition qui le distingue des niveaux relatifs à sa composition –poïésis - et à sa réception - esthésis [Nattiez 1975].

Ainsi, il existe de multiples points de vue et stratégies d'analyse possibles d'un même contenu musical et l'approche technique pour la réalisation d'outils d'analyse automatisée, tant destinés aux musicologues qu'aux créateurs, doit être de proposer une architecture modulaire, reposant sur un ensemble aussi exhaustif que possible de descripteurs élémentaires caractérisant les différents aspects du sonore, et qui puissent être combinés dans une perspective d'analyse particulière. C'est en particulier le cas développements récents menés à l'Ircam autour du module *ircamDescriptor* qui fournit l'extraction d'un grand nombre de descripteurs audio [Peeters 2004], de l'application *Audiosculpt* [3]pour l'édition des sons et de l'éditeur *MuBu* (multibuffer) intégré à l'application Max/MSP.

[3] http://forumnet.ircam.fr/product/audiosculpt/

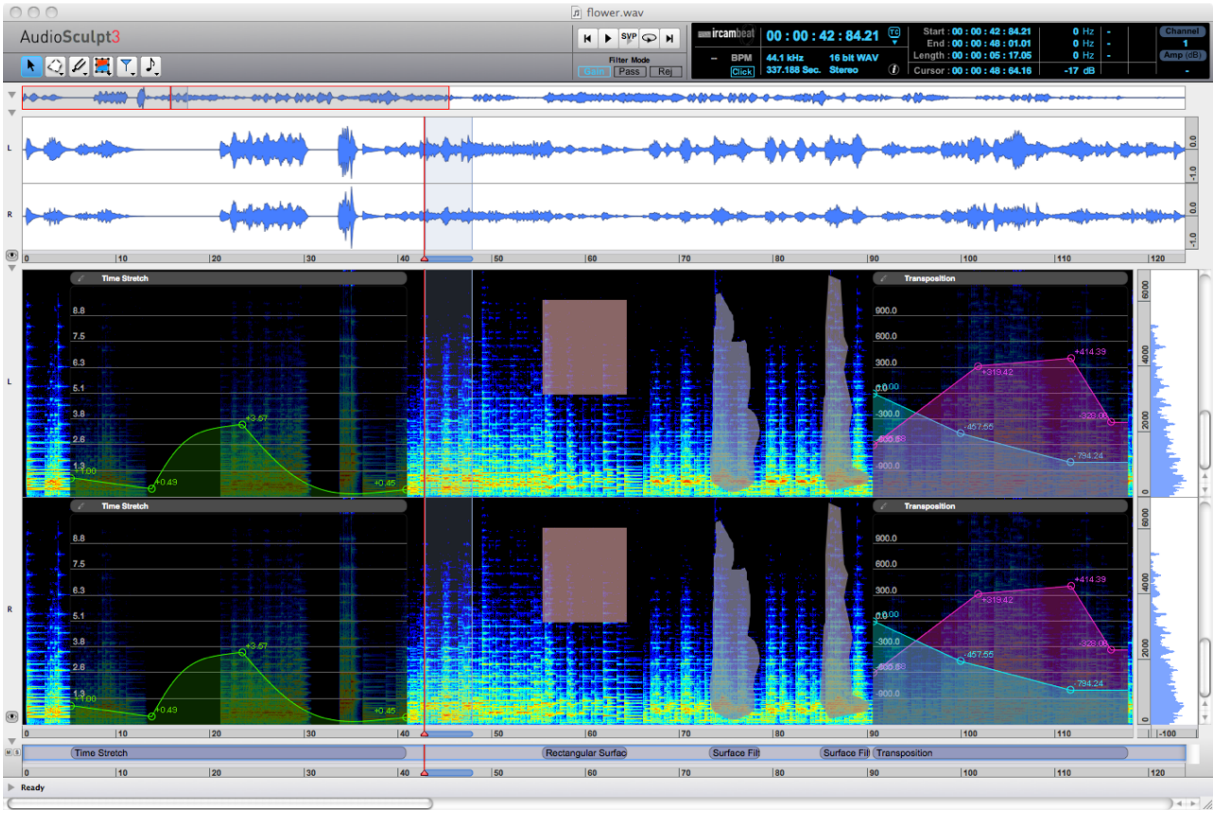

Figure 2 : interface du logiciel Audiosculpt, ©Ircam

Audiosculpt (Figure 2) comprend un éditeur graphique permettant la superposition sur une même base temporelle de multiples analyses combiné au moteur de traitement SuperVP, fondé sur le modèle de vocodeur de phase mentionné plus haut et proposant de nombreux traitements de haute qualité sur le son manipulé. La possibilité d'édition de l'ensemble des paramètres produits répond à la nécessité d'ajustement manuel de résultats parfois incorrects de modules d'analyse automatisée. MuBu est un objet graphique intégré à l'environnement modulaire Max, destinée à la réalisation d'algorithmes de traitement des informations musicale en temps réel [Schnell 2009].

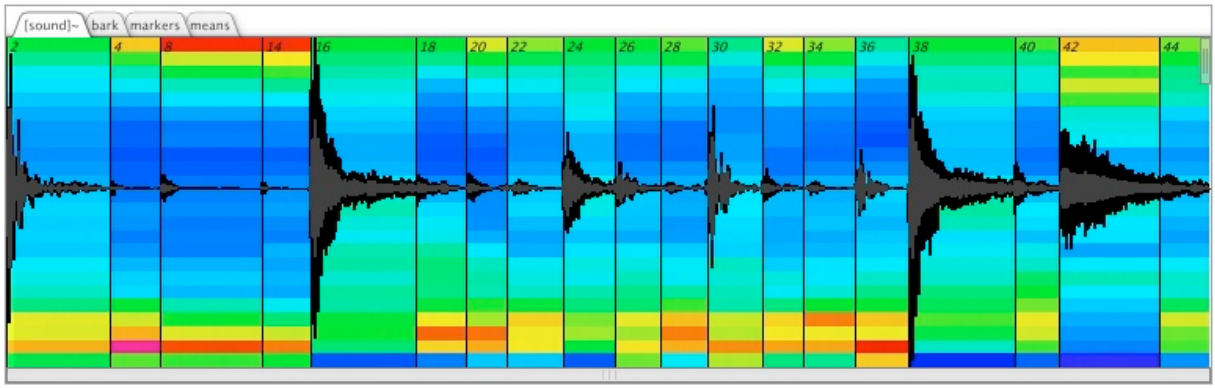

Figure 3 : Editeur du logiciel MuBu visualisant de multiples analyses d'un son sur une même référence temporelle, ©Ircam

Proposant des fonctions d'édition similaires à celles d'Audiosculpt, il intervient notamment dans la réalisation de dispositifs d'interaction sonore faisant appel à des contenus musicaux pré-enregistrés en fournissant un accès efficace pour la synthèse temps réel aux structures temporelles issues de leurs différentes analyses. Le traitement

modulaire des informations résultant de l'analyse du signal sous forme symbolique ou continue est quant à lui possible à partir d'environnements informatiques spécialisés pour l'aide à la composition et à l'analyse musicologique, tels que le logiciel OpenMusic de l'Ircam [Agon 1998].

Applications et cas d'usage

Cette partie présente des exemples représentatifs d'applications et cas d'usage issus de travaux de recherche récents, menés notamment à l'Ircam, et illustrant différentes possibilités nouvelles de manipulation par le contenu des signaux musicaux.

Navigation dans la structure temporelle des morceaux

Les recherches sur la segmentation d'enregistrements de musiques polyphoniques permettent une analyse et un appariement des principales parties d'un morceau (introduction, refrain, couplets, etc.)[Peeters 2007]. Une application, représentant l'une des fonctions du démonstrateur MuMa[4] réalisé par la société Exalead-Dassault Systèmes dans le cadre du projet Quaero[5], est la réalisation d'interfaces d'écoute proposant une navigation dans le morceau à partir de la visualisation de cette structure temporelle (Figure 4).

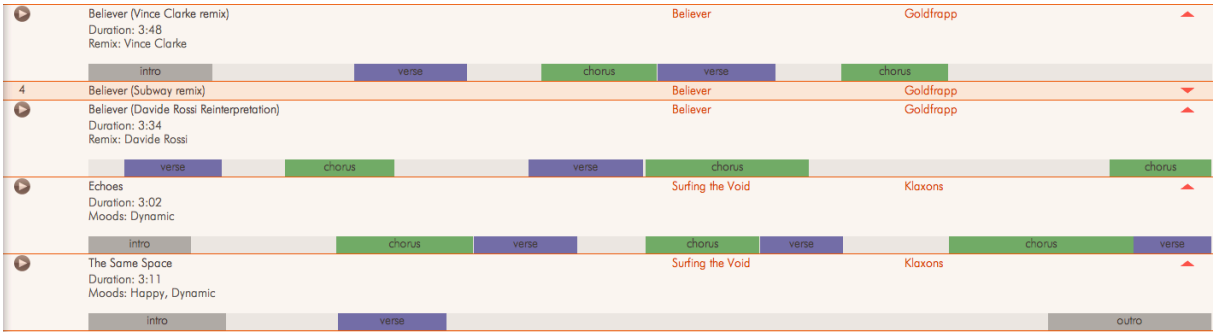

Figure 4 : Extrait d'interface de l'application MuMa, © Exalead-Dassault Systèmes

Edition des sons

Plusieurs applications avancées, dont Audiosculpt et Melodyne[6] de Celemony proposent des fonctions d'édition graphique aux niveaux symbolique et contrôle à partir de l'analyse/resynthèse de fichiers sons. L'interface de Melodyne (cf. Figure 5) visualise simultanément l'information symbolique discrète (valeur de note, position temporelle) ainsi que les variations continues d'amplitude et de fréquence fondamentale au cours du temps, et permet de les modifier par édition graphique et copier-coller.

[4] http://muma.labs.exalead.com/
[5] http://www.quaero.org/
[6] http://www.celemony.com

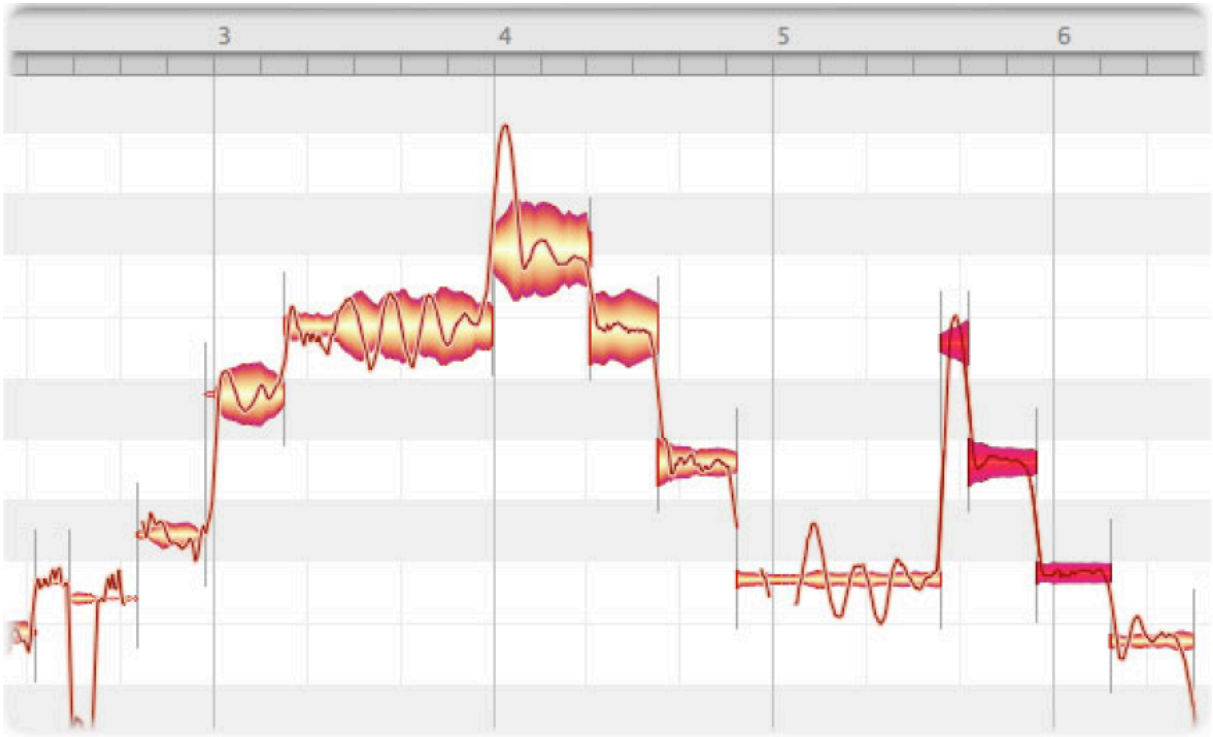

Figure 5 : Éditeur du logiciel Melodyne, ©Celemony

Les moteurs d'analyse/synthèse utilisés par ces applications sont également potentiellement automatisables par des langages de programmation pour modifier ou étendre à l'infini le contenu musical monophonique à partir de son analyse symbolique : mélodies, échelles de hauteurs, création de voix supplémentaires transposées, changement des structures rythmiques, etc.

L'intégration dans l'application de séquenceur Live 9 d'Ableton[7] de récents résultats de recherches sur la transcription d'enregistrements polyphoniques se traduit par la possibilité de production d'une séquence polyphonique au format MIDI à partir d'un son de départ. Il s'agit ici non plus de transformer celui-ci, mais de le compléter par des sons de synthèse synchrones et définis à partir des valeurs de hauteur des notes jouées.

Suivi de tempo et synchronisation temporelle

Les fonctions précédemment décrites reposent sur une segmentation en événements successifs. L'analyse du tempo associée à la détection des temps forts du morceau (débuts de mesure) rend compte d'un autre aspect de sa temporalité liée à son caractère périodique et permet de nombreuses applications consistant soit à l'étendre à des séquences rythmiques additionnelles, soit à l'instar de l'application Traktor[8] de Native Instruments pour les DJ de pouvoir mixer de manière synchronisée des morceaux différents, en passant progressivement du tempo du premier à celui du second.

Suivi de partition pour les œuvres mixtes et l'accompagnement automatique

De nombreuses œuvres musicales contemporaines, dites mixtes, reposent sur la combinaison de parties instrumentales et de sons électroniques issus de procédés de synthèse ou du traitement en temps réel des sons des instruments sur scène. Différentes

[7] http://www.ableton.com
[8] http://www.native-instruments.com/#/en/products/dj/traktor/

stratégies ont été expérimentées pour synchroniser les parties acoustiques et électroniques, les plus anciennes consistant à diffuser une bande son pour les parties électroniques et à lui asservir l'interprétation instrumentale, avec la perte d'expressivité qui en résulte pour cette dernière. Les travaux menés à l'Ircam sur le suivi de partition [Orio 2003] et qui ont récemment abouti au logiciel antescofo [Cont 2012] consistent à l'inverse à laisser toute liberté à l'interprétation en synchronisant automatiquement le déclenchement de sons électroniques produits par l'ordinateur au jeu instrumental, au gré de ses variations. L'algorithme dispose d'une représentation de la partition jouée par l'instrumentiste et analyse son jeu en temps réel en le comparant à sa référence pour en inférer la position temporelle à chaque instant et déclencher les effets sonores à la note près. Les perfectionnements du logiciel comportent une analyse continue du tempo en cours et permettent ainsi d'adapter le déroulement temporel des sons électroniques à celui de l'interprétation.

Une autre application de cette même technologie, qui fait l'objet de plusieurs produits en cours d'élaboration, est l'accompagnement automatique, ou *music minus one,* offrant la possibilité à un instrumentiste soliste de s'entraîner avec un accompagnement pré-enregistré, celui-ci s'adaptant aux variations de son interprétation.

Extensions de la notion d'instrument à l'interaction performance/son

L'extension de la notion d'instrument de musique aux sons électroniques peut être conçue comme la combinaison d'une fonction de captation du geste avec celle d'une synthèse sonore en temps réel. L'établissement d'une correspondance ou *mapping* entre les signaux issus de la captation gestuelle et les paramètres de synthèse est constitutive des caractéristiques de cet instrument étendu [Wanderley 1999] et peut prendre plusieurs formes, la plus simple étant, à l'instar d'un jeu au clavier, le déclenchement d'un son pré-enregistré à un instant particulier. Des recherches actuelles sur l'interaction sonore tendent à généraliser cette problématique en explicitant et mettant en œuvre les relations entre les structures temporelles inhérentes au geste d'une part, au son d'autre part [Schnell 2011]. Ces travaux se fondent notamment sur un algorithme de suivi de geste et de formes temporelles, analogue au suivi de partition évoqué plus haut pour des signaux continus, qui reconnaît un geste effectué à partir d'une référence précédemment apprise et fournit à chaque instant sa position temporelle par rapport à la référence [Bevilacqua 2010]. A partir d'une synchronisation préétablie entre le geste de référence et un son pré-enregistré, il est ainsi possible de resynchroniser en temps réel ce son avec le nouveau geste. Le système est de plus compatible avec une notion de la performance étendue à d'autres paramètres que le geste comme les descripteurs continus du son d'un instrument ou d'une voix. Ces travaux, qui s'appuient notamment sur l'éditeur MuBu présenté plus haut, font l'objet de nombreuses expérimentations et applications pour la réalisation de dispositifs interactifs : musique mixte à partir d'instruments dotés de capteurs, déclenchement d'effets sonores au théâtre par suivi de la voix des acteurs, création de situations d'interaction sonore à partir de la manipulation d'objets de la vie courante[9,10,11]

[9] http://www.brunozamborlin.com/mogees/
[10] http://www.urbanmusicalgame.net
[11] http://interlude.ircam.fr/

Synthèse sonore par corpus

Issu des techniques de synthèse de la parole, un nouveau mode de synthèse sonore temps réel fondé sur l'utilisation de bases de données de descripteurs musicaux est la synthèse concaténative par corpus, moteur du logiciel CataRT de l'Ircam [Schwarz 2007]. Un ensemble de sons de départ est segmenté automatiquement en unités de courte durée (typiquement 0,5 s), chaque unité du corpus étant analysée selon un ensemble de descripteurs. La synthèse peut alors être contrôlée selon plusieurs modes. Le premier prend un son cible en entrée, calcule ses descripteurs à chaque instant et recherche l'unité du corpus la plus proche selon un critère de similarité défini selon la combinaison choisie de descripteurs (similarité en hauteur, et/ou en timbre, etc.). Il en résulte un son issu du matériau du corpus et dont la dynamique suit celle du son cible. Le second mode de contrôle utilise l'interface graphique de CataRT (cf. Figure 6), le déplacement de la souris dans une zone bidimensionnelle définie par deux axes de descripteurs sélectionnés produisant une synthèse sonore par concaténation des unités les plus proches. Il s'agit bien d'une application d'analyse/synthèse par le contenu, le corpus initial étant déconstruit par segmentation puis reconstruit par synthèse selon la séquence cible dans l'espace des descripteurs.

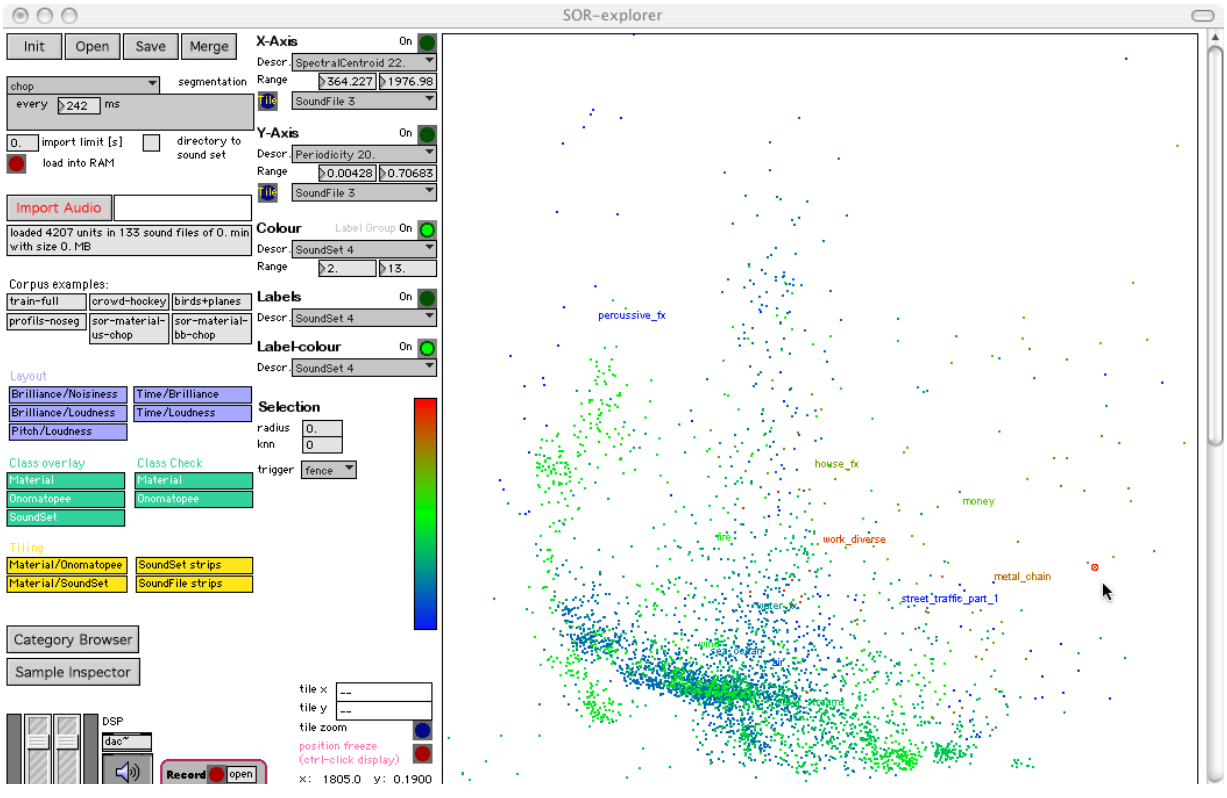

Figure 6 : Interface du logiciel CataRT, ©Ircam

Interaction symbolique pour l'improvisation et les nouveaux instruments

Une autre application relative au traitement dans le domaine symbolique des informations musicales est l'environnement OMax, conçu pour l'interaction instrument-ordinateur dans le contexte de la musique improvisée [Assayag 2006]. A l'inverse de la situation de musique écrite, aucune information préalable n'est disponible sur le contenu de l'interprétation. Le système effectue une analyse symbolique en temps réel du jeu instrumental et code celui-ci selon un algorithme, dit *oracle des facteurs*, dédié à la représentation de la séquence des symboles sous forme de relations multiples entre sous-séquences. Cette représentation s'étoffe au fil du jeu et peut être utilisée en lecture pour produire de nouvelles séquences de notes, constituant des variantes dans le même « style » musical que la séquence de référence. Le système a été utilisé dans de nombreuses situations expérimentales d'improvisation.

Une autre application novatrice du système, en cours d'expérimentation dans le cadre du projet de recherche Sample Orchestrator 2[12], consiste à appliquer ce modèle non plus à l'improvisation, mais à la création de nouveaux types d'instruments étendus reposant sur l'analyse de tout un corpus de morceaux pré-analysés. Le système compare en temps réel le jeu de l'instrumentiste aux structures du corpus pour lui adjoindre un accompagnement dans le style des séquences analysées.

Perspectives

Les exemples présentés illustrent des résultats récents de travaux de recherches en cours, sous la forme, pour la plupart d'entre eux, de prototypes expérimentaux dont les fonctions sont vouées à être progressivement intégrées dans des applications de plus large diffusion. Le traitement par le contenu des sons musicaux n'en est qu'à ses prémisses et ses développements laissent entrevoir la perspective d'un renouvellement profond des possibilités expressives des technologies musicales, tant auprès des créateurs et artistes professionnels que du grand public. Une tendance importante qui en résulte est la généralisation du matériau de base musical, jusqu'à présent principalement fondé sur des sons isolés -notes, échantillons - à des phrases instrumentales et morceaux polyphoniques constitués, exécutés en direct ou issus d'enregistrements, selon des modes d'extraction, de recombinaison et de synchronisation combinables à l'infini. Dans le champ de la création contemporaine, on pourra y voir, selon l'orientation esthétique, une réintégration radicale de la modernité dans la postmodernité… ou exactement l'inverse.

[12] http://sor2.ircam.fr

Glossaire des logiciels et projets de l'Ircam

Antescofo

Antescofo[13] est un système de suivi de partition modulaire et un langage de programmation synchrone pour la composition musicale. Il effectue une reconnaissance automatique en temps réel de l'interprétation - position dans la partition et tempo - permettant ainsi de synchroniser une performance instrumentale avec celle d'une partition virtuelle informatique. Antescofo réunit la description des parties instrumentales et électroniques dans la même partition, grâce à un langage synchrone conçu pour la pratique de musique mixte, visant à augmenter l'expressivité de l'écriture des processus temps réel, sous une forme adaptée au langage musical. Après le chargement de la partition, Antescofo, qui prend comme entrée un flux audio polyphonique, est capable de suivre la position et le tempo des musiciens en temps réel et de synchroniser les actions programmées pour la partie d'informatique musicale (déclenchement et contrôle de la partie électronique).

Audiosculpt

AudioSculpt[14] est une application pour Macintosh permettant de « sculpter » littéralement un son de manière visuelle. Après une phase d'analyse, le son s'affiche sous la forme d'un sonagramme, et l'utilisateur peut dessiner les modifications qu'il veut lui appliquer. Les traitements principaux sont le filtrage, la synthèse croisée, la transposition, la dilatation et compression temporelles, le débruitage. Plusieurs types d'analyses montrent le contenu spectral d'un son et l'utilisateur peut ensuite modifier celui-ci par plusieurs méthodes : dessiner des filtres, déplacer des régions du sonagramme en temps et fréquence, ou appliquer l'une des nombreuses transformations de haute qualité.

CataRT

Reposant sur la synthèse sonore concaténative par corpus, CataRT[15] propose une exploration interactive et en temps réel d'une base de données sonore et une composition granulaire ciblée par des caractéristiques sonores précises. Il permet aux compositeurs et musiciens d'atteindre de nouvelles sonorités, et aux designers sonores de rapidement explorer un corpus sonore constitué d'un grand nombre d´enregistrements. CataRT existe en application standalone ou en système modulaire dans Max. L'interaction repose sur une interface simple consistant en l'affichage d'une projection 2D de l'espace de descripteurs, offrant une navigation avec la souris et dans laquelle les grains sont sélectionnés et joués par proximité géométrique.

IrcamDescriptor

Ircamdescriptor est un logiciel dédié à l'extraction automatique d'un grand nombre de descripteurs sonores et musicaux à partir de l'analyse d'un fichier son. Il est disponible sous forme de bibliothèque C++ pouvant être intégrée dans un environnement logiciel, ainsi que sous la forme de l'objet ircamdescriptor~pour Max fournissant l'extraction en temps réel d'une quarantaine de descripteurs à partir d'un signal audio[16].

---

[13] http://forumnet.ircam.fr/product/antescofo/
[14] http://forumnet.ircam.fr/product/audiosculpt/
[15] http://forumnet.ircam.fr/product/catart-standalone/
[16] http://forumnet.ircam.fr/product/max-sound-box/

Max

Max[17] (anciennement Max/MSP) est un environnement visuel pour la programmation d'applications interactives temps réel. C'est actuellement la référence mondiale pour la création d'installations sonores interactives. Max est la combinaison du logiciel Max (Ircam/Cycling'74) pour le contrôle temps réel d'applications musicales et multimédia interactives par MIDI, de MSP, une bibliothèque d'objets pour l'analyse, la synthèse et le traitement du signal audio en temps réel et de Jitter qui est un ensemble d'objets vidéo, matriciels et graphiques 3D. Max est conçu pour les musiciens, les designers sonores, les enseignants et les chercheurs qui souhaitent développer des programmes interactifs temps réel. Max est développé et édité par la société californienne Cycling'74 sous licence exclusive de l'Ircam.

MuBu

MuBu[18] pour *multi-buffer* est une collection d'objets Max destinée à l'édition de données temporelles de toutes sortes et leur utilisation pour différentes applications de traitement temps réel dont la synthèse sonore. MuBu contient des pistes multiples de données alignées à des structures de données complexes telles que :

- Données audio segmentées avec descripteurs et annotation
- Données de mouvement de captation annotées
- Données audio et de signaux de captation gestuelle synchronisées.

Chaque piste d'un buffer MuBu peut représenter un flux de données échantillonné ou une séquence d'événements temporels étiquetés, comme par exemple des échantillons audio, des descripteurs audio, des données de mouvement de captation, des marqueurs, des segments et des événements musicaux.

MuMa

Muma[19] est un prototype d'application web, développé par la société Exalead-Dassault Systèmes dans le cadre du projet Quaero, destiné à illustrer les nouvelles possibilités de navigation par le contenu dans des bases de morceaux de musique enregistrés, sur la base d'un ensemble de descripteurs musicaux automatiquement extraits par analyse de signal.

[17] http://cycling74.com/products/max/
[18] http://forumnet.ircam.fr/product/mubu/
[19] http://muma.labs.exalead.com/

OMax

OMax[20] est un environnement pour l'improvisation avec ordinateur qui analyse, modélise et réimprovise en temps réel le jeu d'un ou de plusieurs instrumentistes, en audio ou en MIDI. OMax est basé sur une représentation informatique nommée "Oracle des facteurs", un graphe qui interconnecte tous les motifs des plus petits aux plus grands et fournit donc une carte de navigation dans la logique motivique apprise de l'instrumentiste, engendrant ainsi un grand nombre de variations cohérentes stylistiquement. OMax base sa reconnaissance soit sur des notes (suivi de hauteurs), soit sur des timbres (suivi spectral).

OpenMusic

OpenMusic[21] est un environnement de programmation visuelle pour la création d'applications de composition et d'analyse musicale assistées par ordinateur. OpenMusic offre à l'utilisateur de nombreux modules qui peuvent être associés à des fonctions mathématiques ou musicales, représentées par des icônes. L'utilisateur peut relier ces modules entre eux et créer un programme appelé «patch» qui va générer ou transformer des structures musicales. Les patchs peuvent s'emboîter les uns dans les autres pour constituer des programmes et créer des structures de plus en plus élaborées. OpenMusic est aujourd'hui utilisé par un grand nombre de compositeurs et de musicologues. Il est enseigné dans les principaux centres d'informatique musicale ainsi que plusieurs universités en Europe et aux États-Unis.

Sample Orchestrator 2

Sample Orchestrator 2[22] est un projet de recherche et développement soutenu par l'Agence nationale de la recherche. Mené par un consortium coordonné par l'Ircam et associant la société Univers sons et le Conservatoire national de musique et de danse de Paris, il vise la réalisation d'une nouvelle génération d'application d'échantillonneur musical étendant les possibilités actuelles de la synthèse, du traitement sonore et des instruments de musique selon trois directions de recherche parallèles :

- l'élaboration de modèles de signaux paramétriques pour la synthèse d'instruments de musique ;
- l'élaboration de nouvelles méthodes de spatialisation hybrides combinant échantillonnage spatial à haute résolution et modèles paramétriques ;
- l'élaboration de nouvelles formes d'instruments fournissant un accompagnement interactif à partir de l'analyse de corpus musicaux pré-enregistrés.

SuperVP

SuperVP est une bibliothèque de traitement de signal reposant sur un vocodeur de phase perfectionné. Elle permet un grand nombre de transformations du signal avec une très grande qualité sonore (étirement temporel, transposition de la fréquence fondamentale et de l'enveloppe spectrale, débruitage, re-mixage des composantes sinusoïdales, bruitées et transitoires, dilatation de l'enveloppe spectrale, synthèse croisée généralisée, synthèse croisée en mode source et filtre…). Elle donne accès à un vaste jeu de paramètres qui fournissent un contrôle complet, et à grain fin, du résultat

[20] http://forumnet.ircam.fr/product/omax/
[21] http://forumnet.ircam.fr/product/openmusic/
[22] http://sor2.ircam.fr

d'algorithmes différents. En sus des algorithmes de transformation sonores, la bibliothèque comprend une collection importante d'algorithmes d'analyse du signal (fréquence fondamentale, détection des débuts de notes, spectrogramme, spectrogramme réassigné, enveloppe, spectrale…). SuperVP est disponible sous forme de bibliothèque C++ pour intégration dans des environnements logiciels, ou sous forme d'objets de traitement temps réel pour l'environnement Max[23]. C'est aussi le moteur principal du logiciel Audiosculpt. De nombreux produits commerciaux l'utilisent pour le traitement audio de haute qualité.

Références

Agon Carlos, *OpenMusic: Un langage visuel pour la composition musicale assistée par ordinateur*. Thèse de doctorat, Université Pierre et Marie Curie, Paris (1998)

Allen Jont B. and Rabiner Lawrence R. "A unified approach to short-time Fourier analysis and synthesis" *Proceedings of the IEEE*, Vol. 65, N° 11 (1977)

Assayag Gérard, Bloch Georges, Chemillier Marc, Dubnov Shlomo, "Omax Brothers : a Dynamic Topology of Agents for Improvization Learning", Workshop on Audio and Music Computing for Multimedia, ACM Multimedia, (2006)

Bevilacqua Frédéric, Zamborlin Bruno, Sypniewski Anthony, Schnell Norbert, Guédy Fabrice, Rasamimanana Nicolas, "Continuous realtime gesture following and recognition" in *Embodied Communication and Human-Computer Interaction*, Lecture Notes in Computer Science (LNCS) Vol. 5934, Springer Verlag (2010)

Cont Arshia, "Synchronisme musical et musiques mixtes: Du temps écrit au temps produit", in *Circuit Musiques Contemporaines*, Vol. 22 N°1 (2012)

Gomez Emilia, "Estimating the tonality of polyphonic audio files: cognitive versus machine learning modelling strategies", in *Proc. International Conference on Music Information Retrieval* (2004)

Kaprykowsky Hagen, Rodet Xavier, "Globally Optimal Short-Time Dynamic Time Warping Application to Score to Audio Alignment", in *Proc. Int. Conf. on Audio, Signal and Speech Processing* (2006)

Moorer James Andy "The use of the phase vocoder in computer music applications", in *Journal of the Audio Engineering Society*, vol. 26, n° 1/2, p. 42-45 (1978)

Nattiez Jean-Jacques, *Fondements d'une sémiologie de la musique*, 10-18, Union générale d'éditions, Paris (1975)

Orio Nicola, Lemouton Serge, Schwarz Diemo, Schnell Norbert, "Score Following: State of the Art and New Developments", in *Proc. New Interfaces for Musical Expression* (2003)

MPEG-7, Information Technology - Multimedia Content Description Interface - Part 4: Audio, ISO/IEC JTC 1/SC 29, ISO/IEC FDIS 15938-4 (2002)

Peeters Geoffroy, "A large set of audio features for sound description (similarity and classification) in the Cuidado project", *Technical Report version 1.0*, IRCAM – Centre Pompidou, Paris, France (2004)

Peeters Geoffroy, "Sequence representation of music structure using Higher-Order Similarity Matrix and Maximum-Likelihood approach", in *Proc. International Conference on Music Information Retrieval* (2007)

Peeters Geoffroy, Papadopoulos Hélène, "Simultaneous beat and downbeat-tracking using a probabilistic framework: theory and large-scale evaluation", in *IEEE. Trans. on Audio, Speech and Language Processing*, Vol. 19, n° 6, (2011a)

Peeters Geoffroy, Giordano Bruno L., Susini Patrick, Misdariis Nicolas, McAdams Stephen, "The Timbre Toolbox: Audio descriptors of musical signals", in *Journal of the Acoustical Society of America*, Vol. 5. N°130 (2011b)

[23] http://forumnet.ircam.fr/product/supervp-pour-max/

Roebel Axel, “A new approach to transient processing in the phase vocoder”, in *Proc. International Conference on Digital Audio Effects (DAFx’03)*, p. 344-349 (2003)

Roebel Axel, “A Shape-Invariant Phase Vocoder for Speech Transformation”, in *Proc. Int. Conf. on Digital Audio Effects* (2010)

Schaeffer Pierre, Traité des objets musicaux, Editions du Seuil, Paris, France (1966)

Schnell Norbert, Roebel Axel, Schwarz Diemo, Peeters Geoffroy, Borghesi Riccardo, “MuBu & Friends – Assembling Tools for Content Based Real-Time Interactive Audio Processing in Max/MSP”, in *Proc. of the Int. Conf. on Computer Music* (2009)

Schnell Norbert, Bevilacqua Frédéric, Guédy Fabrice, Rasamimanana Nicolas, “Playing and Replaying – Sound, Gesture and Music Analysis and Re-Synthesis for the Interactive Control and Re-Embodiment of Recorded Music”, in *Klang und Begriff, Gemessene Interpretation - Computergestützte Aufführungsanalyse im Kreuzverhör der Disziplinen*, Mainz, Schott Verlag (2011)

Schwarz Diemo, “Corpus-based concatenative synthesis », in *IEEE Signal Processing Magazine*, 24, 2, p. 92-104, Special Section: Signal Processing for Sound Synthesis (2007)

Susini Patrick, Lemaitre Guillaume, McAdams Stephen, “Psychological measurement for sound description and evaluation”, in *Measurements with persons*, Ed. Berglund B., Rossi G.B., Townsend J.T., Pendrill L.R., Scientific Psychology Series, Psychology Press, Taylor and Francis (2012)

Vincent Emmanuel, Ono Nobutaka, “Music Source Separation and its Application to MIR”*, Proc. International Conference on Music Information Retrieval (*2010)

Vinet Hugues, “The Representation Levels of Musical Information”, in *Lecture Notes in Computer Science*, N° 2771, Springer Verlag (2003)

Wanderley Marcelo, Depalle Philippe, “Contrôle gestuel de la synthèse sonore”, in Interfaces homme-machine et création musicale, Ed. Vinet H. et Delalande F., Hermes Science Publications (1999)

Yeh Chungsinh, Roebel Axel, Rodet Xavier, “Multiple Fundamental Frequency Estimation and Polyphony Inference of Polyphonic Music Signals”, in *IEEE Transactions on Audio, Speech and Language Processing*, Vol. 18, n° 6, p.1116-1126 (2010)