UTILISATION D'ESPACES PERCEPTIFS POUR LA SYNTHÈSE ET LA TRANSFORMATION SONORE

JOURNÉES D'ÉTUDES DE L'ACI "ESPACES SONORES", AVRIL 2002, PARIS VIII

Vincent Verfaille
CNRS - LMA, 31, chemin Joseph Aiguier, 13402 Marseille Cedex 20
verfaille@lma.cnrs-mrs.fr

Résumé

On s'intéresse dans cette étude aux espaces perceptifs. Ils correspondent à des représentations du son dans des espaces dont les dimensions font sens à notre perception. Ils sont utilisés depuis de nombreuses années pour la synthèse sonore, avec un regain d'intérêt depuis quelques années que la synthèse en temps-réel est accessible à tous. Les applications multimédias nécessitant codage, compression, classification, segmentation y font aussi appel. De plus, ils permettent de transformer le son selon ces dimensions perceptives et de contrôler des effets par les paramètres de ces dimensions.

1 Introduction

On appelle espaces perceptifs des espaces de représentation du son, constitués à partir d'indices perceptifs (psychoacoustiques ou autres, faisant sens à la perception auditive). Ces espaces perceptifs sont tout d'abord employés pour le contrôle de sons de synthèse en temps-réel par des paramètres perceptifs. En effet, on cherche manipuler simplement le matériau sonore ; des représentations du son sont donc indispensables. Pour des sons provenant d'instruments réels, la représentation du son se fait par apprentissage (*feedback* ou retour tactile, sonore, visuel) et par l'utilisation de représentations (notes dans une partition, de plus en plus complexe, afin de décrire au mieux l'interprétation désirée). Dans le cas d'instruments virtuels, il faut que le geste et ce qu'il induit soit directement compréhensible par l'utilisateur. Ceci implique de faire apparaître dans les fonctions de correspondance entre le transducteur gestuel et le modèle de synthèse à la fois des espaces perceptifs et des retours (tactiles, visuels, sonores, cf. [12]).

Ensuite, les analyses réalisées en vue du codage, de la compression de données audionumériques, de la classification de sons et de la segmentation de flux audio font aussi appel plus ou moins directement à des espaces perceptifs. En effet, la recherche de définitions et de descriptions du son (par des paramètres qui en sont caractéristiques) conduit à étudier la perception que l'on a du son et à utiliser les résultats connus concernant la perception auditive.

Une voie d'investigation plus récente consiste en l'analyse-synthèse et les transformations basées sur le contenu. Il s'agit de rendre les schémas d'analyse-synthèse plus performants en utilisant une étape d'analyse du signal qui utilise des connaissances psychoacoustiques. Dans le cadre de ma recherche doctorale, ceci apparaît sous la forme d'effets adaptatifs. J'utilise des descriptions perceptives du son, soit pour modifier ces sons par des effets audionumériques ou transformations, soit simplement pour piloter des effets dont l'évolution des paramètres reste cohérente avec le son et les gestes musicaux qui le composent.

Afin de ne pas s'avancer dans des domaines que nous ne connaissons pas, précisons dans quel cadre cette étude se place (cf. Fig.1, [8]). Rappelons qu'en partant de l'intention musicale de l'interprète ou du compositeur et par le biais d'instruments, on produit des sons, qui sont alors perçus à un premier niveau (psychoacoustique), avant d'être intégrés en données plus complexes (flux, structures mélodiques et harmoniques, formes, etc). Dans ce qui suit, nous nous intéressons à la partie encadrée de la figure, en lien avec la physique (traitement du signal, théorie de l'information) et ne nous aventurons pas vers la compréhension et l'analyse des intentions musicales, domaine des psychologues et musicologues.

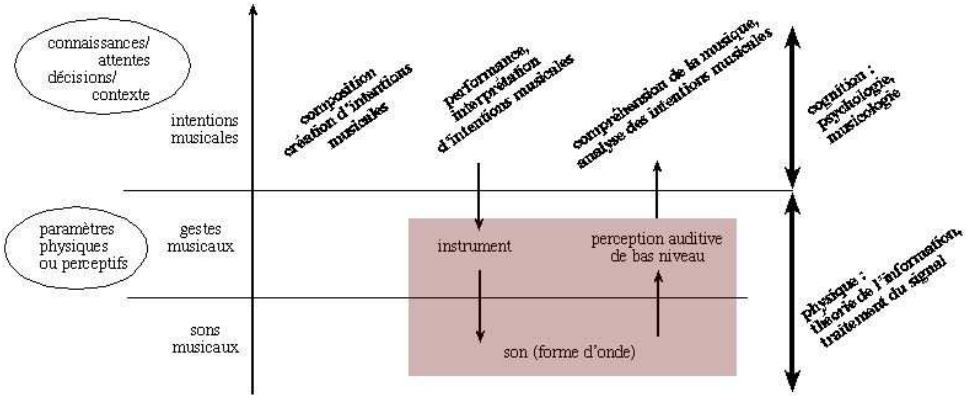

Figure 1: *De l'intention au son musical, en passant par le geste (Métois, 1996)*

2 Cartographie psychoacoustique

2.1 Définitions

On utilise une représentation des sons communément admise, provenant de la psychoacoustique. Cette représentation, ou cartographie psychoacoustique, est le pendant de la partition complète. Le son y est représenté par la sensation physique que l'on en a, selon plusieurs critères ou attributs psychoacoustiques. Ce sont la hauteur (note), la sonie (niveau sonore perçu, nuance de jeu), le timbre (instrument, modes de jeu), la localisation (spatiale), la durée (localisation et trajectoire temporelle) et la qualité. Dans les modes de jeu instrumentaux, des attributs perceptifs tels que vibrato, tremolo ou rugosité peuvent apparaître.

2.2 Multi-dimensionnalité

Certains de ces attributs sont eux-mêmes multi-dimensionnels. Ainsi, la localisation spatiale fait appel à un espace à trois dimensions : l'espace physique, représenté par ses coordonnées cartésiennes (x,y,z) ou sphériques (distance, élévation, azimut). Le timbre quant à lui, selon qu'il concerne des sons harmoniques ou percussifs, s'est vu attribuer plusieurs dimensions. Pour les sons harmoniques par exemple, on distingue deux sous-espaces et cinq paramètres (ou dimensions) : le temps (logarithme du temps d'attaque), et le spectre fréquentiel (quatre paramètres : le centroïde spectral, l'écart spectral harmonique, la déviation spectrale harmonique et la variation spectrale harmonique) [22, 18]. Pour les sons percussifs [20], on distingue deux sous-espaces et trois paramètres : le temps (log du temps d'attaque et centroïde temporel) et le spectre fréquentiel (centroïde spectral).

2.3 Espace de timbre

Comme expliqué brièvement dans la partie précédente, le timbre comporte plusieurs sous-dimensions. Un ensemble d'études converge pour donner à l'espace de timbre des sons instrumentaux les trois dimensions suivantes :

1. la brillance, caractéristique de la distribution de l'énergie spectrale, et corrélée au centroïde (ou centre de gravité spectrale) [15, 21, 17] ;

2. les propriétés spectrales fines, à savoir la cohérence des micro-variations des composantes spectrales, mesurées par le flux spectral et la synchronicité des attaques et décroissances des harmoniques [21] (la synchronicité en question est statique pour les instruments à vents et dynamique pour les cuivres) ;

3. le transitoire d'attaque qui confère le caractère d'explosivité ou non de l'attaque d'un son ; il est décrit par la synchronicité et l'harmonicité des partiels dans différents parties du spectre.

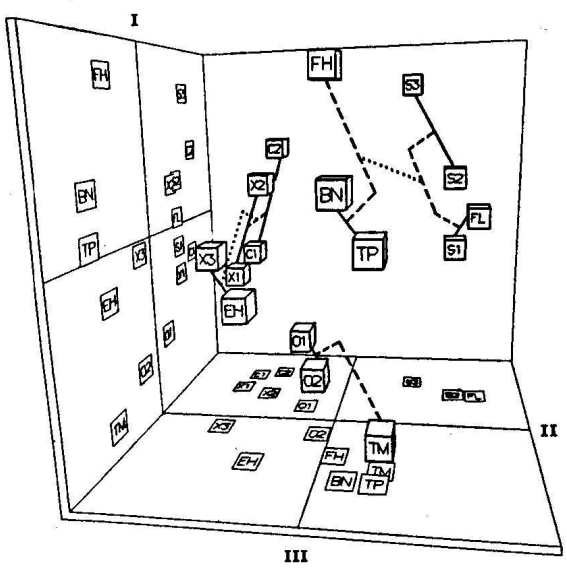

Figure 2: *Espace à 3 dimensions extrait de l'analyse de proximité des jugements de similarité (d'après Grey 1977 [15]).*

2.4 Non-orthogonalité de l'espace perceptif du son

Le timbre se définit aussi à partir d'attributs perceptifs plus ou moins généraux, comme la présence ou non d'un vibrato, d'un tremolo, de rugosité, d'harmonicité, de consonance ou dissonance. Qu'en est-il du lien entre les différents axes ? Par de simples expériences, on peut montrer que certaines dimensions ne sont absolument pas indépendantes :

- **exp 1** Prenons le son d'une note percussive ralentie temporellement. Elle n'est plus reconnaissable en tant que telle : le problème vient-il uniquement du traitement audionumérique ? N'y a-t-il pas aussi un problème de définition de ce qu'est un son percussif ralenti par rapport au son non ralenti ?
- **exp 2** Soit un vibrato ralenti temporellement : il devient méconnaissable comme vibrato, et révèle sa nature intrinsèque, à savoir une lente modulation de fréquence.
- **exp 3** Soit un son riche en harmonique, pas parfaitement harmonique. Si l'on filtre en partie son spectre, de façon à ne sélectionner qu'une partie des harmoniques tout en conservant une assez bonne reconnaissance du timbre, la hauteur perçue varie légèrement.

2.5 Types de cartographies

On envisage deux types de cartographies : statique (ie. à court terme) et dynamique (ie. à moyen ou long terme). La première correspond à une représentation pour des sons courts, isolés, presque hors contexte musical si l'on prend par exemple des sons instrumentaux classiques. Un son est alors dans cette représentation un objet, placé quelque part dans une cartographie. La seconde correspond à l'évolution d'une représentation dans un espace, à une trajectoire.

Prenons tout de suite un exemple : la cartographie psychoacoustique est très adaptée à la représentation de zones de timbres d'instruments. Par contre, une phrase musicale complexe jouée par un instrument peut évoluer dans des zones différentes dans cette cartographie. Ainsi, un violon peut produire des sons entretenus (reconnaissance des cordes entretenues), des sons secs percussifs (cordes pincées), des harmoniques (cordes en général), des crissements (on perd alors la sensation d'entendre un violon si on n'entend que cet extrait). Aussi, certaines cartographies permettent de suivre des trajectoires, et seront plus adaptées à décrire l'évolution de sons : c'est ce que propose Métois avec sa méthode **psymbesis** [8]. On retrouve cette idée avec la synthèse imitative [2].

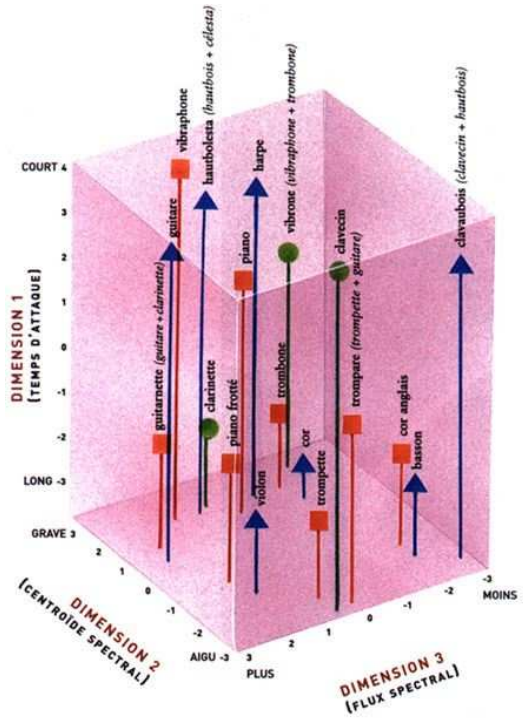

Figure 3: *Espace de timbre issu de l'analyse d'échelle (AMDS) tri-dimensionnelle de jugements de dissimilarité de 21 sons synthétiques [21] (d'après Krumhansl, 1989).*

3 Espaces perceptifs et synthèse

Dans le cadre de la synthèse sonore numérique, nous présentons la chaîne de traitement de l'information allant du geste physique au son numérique. Ensuite, nous présentons des espaces de contrôle de la hauteur et du timbre, décrivons leurs limites et proposons des solutions pour y pallier.

3.1 Contrôle gestuel de la synthèse sonore

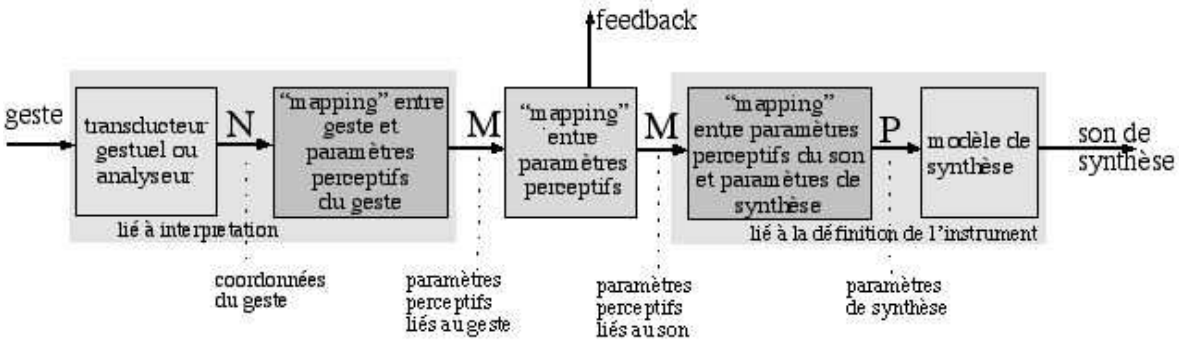

Figure 4: *Chaîne de traitement de l'information du geste au son, à l'aide d'espaces perceptifs (Arfib et al. 2002)*

Depuis l'avènement du temps-réel, on cherche à piloter au mieux des modèles de synthèse sonore numériques à partir de gestes. Dans ce but, on construit une chaîne de traitement de l'information (cf. Fig.4). Dans cette chaîne, on fait apparaître d'un côté ce qui a trait à l'interprétation et à l'intention (à gauche, le geste), et de l'autre côté ce qui a trait à l'instrument de synthèse. Le premier se perçoit comme un geste, une intention, le second comme un son. On comprend le son en l'intègrant depuis la perception : à un certain niveau d'abstraction

(deuxième niveau), on a une représentation (psychoacoustique) bien plus parlante que le signal lui-même (premier niveau), avant ce qui concerne la cognition (mémoire à plus long terme et cognition, troisième niveau).

Lorsqu'on joue d'un instrument que l'on découvre, on cherche à comprendre le son par l'intention, par le jeu que l'on peut en avoir. Deux intentions coexistent, l'une dans le geste, l'autre dans l'espace perceptif. Les indices perceptifs ont ceci d'intéressant qu'ils peuvent être mis en lien directement avec le geste. Ceci permet de jouer directement sur le son par le geste, et de fusionner le *feedback* auditif (la perception que l'on a du son) avec l'espace du geste.

3.2 Contrôle de la hauteur (chroma)

Pour représenter la hauteur tonale, on utilise l'échelle des chromas. Pour affiner la représentation des hauteurs (linéaire) par le biais des notes, Drobisch postule en 1852 une représentation circulaire dite de chromas (cf. Fig.5). Celle-ci sera affinée par Shepard [19] en 1982 avec l'hélice des chromas, puis avec le tore des chromas, prenant en compte les relations intra-chromatiques.

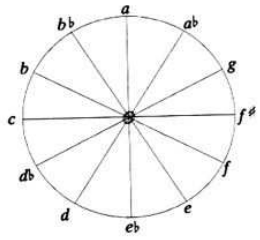

Figure 5: *Chroma selon Dobisch (1852)*

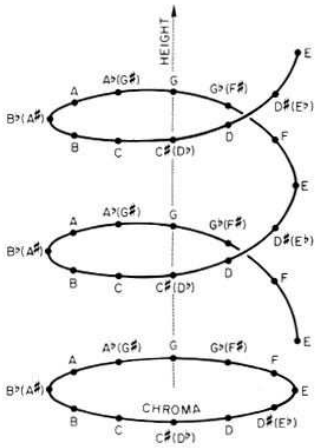

Figure 6: *Chroma selon Shepard (1982)*

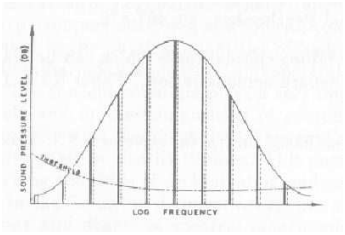

Figure 7: *Utilisation de l'échelle des chroma de Shepard par Risset*

Dans ses illusions sonores, Risset [11] utilise le cercle des chromas et une structure de sons harmoniques dont l'enveloppe spectrale est fixe. Lorsque le son monte, toutes les harmoniques montent, le chroma change (ou tourne sur le cercle). Cependant, l'enveloppe spectrale étant fixe, la hauteur spectrale est fixe. Le son semble monter ou descendre indéfiniment, tout en restant à une hauteur moyenne fixe.

L'hélice des chromas est implémentée dans le Voicer, instrument de synthèse vocale présenté par Loic Kessous [3, 14]. Elle sert de contrôle pour la fréquence fondamentale.

3.3 Espaces d'état

Au cours de ses recherches doctorales, Métois [8] a utilisé la dynamique pour représenter les états d'un son (micro et méso-temporel). D'une représentation figée dans le temps et délimitée à des unités temporelles de quelques dixièmes de secondes (notes), on passe à une représentation d'espace d'états (ms), où les vecteurs de base sont constitués de paramètres perceptifs (notamment la hauteur, la sonie, la brillance). Les données sont ensuite séparés en agrégats (*clusters*) selon ce qu'elles représentent : les parties harmoniques ou les attaques de notes (cf. Fig.8).

Cette méthode s'inspire des méthodes d'analyse des systèmes dynamiques ; une telle application est des plus logique lorsqu'elle s'applique à des instruments de musique qui sont par essence des systèmes dynamiques. Les applications rendues accessibles par cette représentation sont le contrôle d'instruments de synthèse aussi bien que des modèles d'instruments existants.

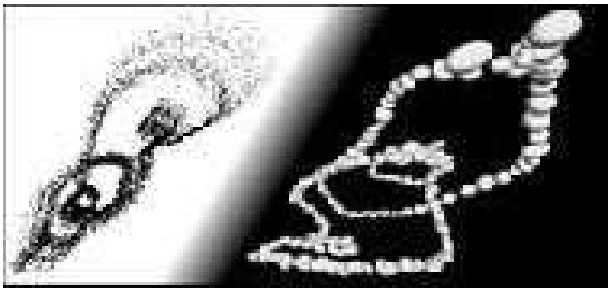

Figure 8: *Espace de phase (d'état) d'une phrase instrumentale (figure de gauche), et sa représentation par des agrégats de Gaussiennes (figure de droite), pour la synthèse*

3.4 Espace de contrôle du timbre (de sons instrumentaux)

Voici quelques exemples d'espace de contrôle du timbre de sons par le geste (contrôle effectué en manipulant des paramètres perceptifs) :

- Wessel (1979) : hauteur, sonie et timbre (brillance) [15, 7] ;
- Beauchamp (1982) : sonie et timbre (synthèse additive par une base réduite d'harmoniques, obtenue par algorithmes génétiques), hauteur constante [29] ;
- McAdams, Cunibile (1992) : timbre (logarithme du temps d'attaque, flux spectral, brillance), hauteur constante et sonies égalisées [21] ;
- Métois (1996) : hauteur, sonie, timbre (brillance) dans un espace d'état [8] ;
- Drame, Wessel (1998) : hauteur, sonie et timbre (brillance), avec une synthèse additive des harmoniques par réseaux de neurones ou bases de données [2] ;
- Jehan, Schoner (1999) : hauteur, sonie et timbre (brillance), les données formant des agrégats pour un meilleur contrôle des différentes parties d'un son [25, 26] ;
- Kessous (2001) : hauteur et timbre (formant, voyelle) avec le *Voicer*, un instrument de voix chantée [3, 14] ;
- Morier (stage de DEA, 2002) : hauteur, sonie, timbre (brillance), dérivées temporelles (prise en compte de l'évolution, de l'état du son ; réseaux de neurones et bases de données).

Jensen [13] propose un modèle de timbre instrumental complet, prenant notamment en compte les variations des harmoniques ou partiels en fréquence et en module autour d'une valeur ou courbe moyenne. Cependant, son contrôle gestuel n'a pas encore été clairement exploité.

Remarquons qu'aucun de ces modèles ne comportent de spatialisation ni de gestion particulière d'organisation de flux temporels, de rythmes.

3.5 Limites de l'espace de timbre

L'espace de timbre a plusieurs limites, comme l'indiquent très justement Keller et Berger [6] dans leur étude sur les espaces de représentation des sons. Tout d'abord, il est uniquement adapté aux sons musicaux, instrumentaux, et exclu tout son hors de la classe instrumentale. Les dimensions du timbre utilisées ne sont alors représentatives que de sons continus et impulsifs. D'autre part, l'hypothèse de base de l'indépendance des dimensions ne prend pas en compte d'autres paramètres perceptifs, parmi lesquels vibrato, tremolo, rugosité. Enfin, ces espaces ne prennent pas en compte le contexte : la durée des sons utilisés pour l'étude, les ensembles de sons, la tâche donnée à l'auditeur, l'expérience des sujets à l'expérimentation.

3.6 Espace de contrôle de sons écologiques et électroacoustiques

Etant données les limites de l'espace de timbre, des recherches ont lieu dans d'autres directions, pour des sons électroacoustiques et écologiques. Nous pensons notamment à deux études, l'une effectuée par Keller & Berger [6] et portant esur des espaces tenant compte de caractéristiques micro et méso-temporelles (événements temporels) et spatiales ; l'autre effectuée par McAdams [1] portant sur ce qu'il appelle la psychomécanique, ou reconnaissance d'actions et d'objets par la perception du matériau et du mode d'excitation de celui-ci lors de la production de son.

D'autres études portent sur la représentation des sons granulaires, sur la perception des sons dans l'espace, sur la perception de l'organisation des sons dans le temps. Néanmoins, nous ne développerons pas ces questions pour l'instant.

4 Compression, classification, recherche de sons, segmentation de flux

En ce qui concerne la compression de sons, le standard MPEG-7 [18] utilise des descripteurs présentés en 2.2. Ce standard permettra la compression son et image avec une haute définition, jusqu'ici non encore atteinte. Afin de vérifier la représentativité de ces indices dans le cadre de recherche d'extraits sonores, deux expériences ont été menées, l'une à l'**IRCAM** avec le **Studio-On-Line** [10](recherche un extrait sonore à partir de ses critères perceptifs), l'autre à Barcelone au **Music Technology Group** sous forme d'une extension au programme **SMS** [31].

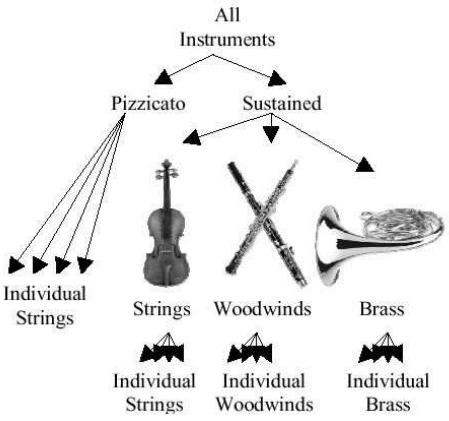

Figure 9: *1ère taxinomie hiérarchique (Martin, Kim, 1998), intuitive*

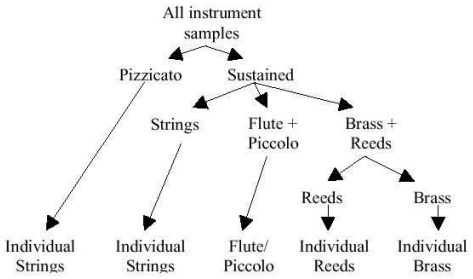

Figure 10: *2ème taxinomie hiérarchique (Martin, Kim, 1998), issue de l'analyse*

Les méthodes de classification automatique d'extraits sonores s'appuient aussi sur des représentations perceptives. Scheirer et Slaney [24] proposent ainsi un discriminateur parole-musique pour des applications d'archivage radiophoniques. Martin et Kim [16] effectuent l'identification d'instruments de musique par taxinomie hiérarchique, basée sur des indices acoustiques et un modèle d'audition. La figure Fig.9 présente la taxinomie à priori (d'après les connaissance des familles d'instruments) et la figure Fig.10 la taxinomie à posteriori. Les regroupements sont légèrement différents dans la hiérarchie (la flûte est séparée des autres bois, qui sont eux regroupés avec les cuivres, avant d'en être séparés au niveau inférieur) : on peut se demander cependant quelle est la part de ce résultat dû à la méthode d'analyse statistique et celle due à la perception. Les indices utilisés ne sont pas tous perceptifs, mais amènent une classification à l'aide de modèles statistiques et selon des critères perceptifs (reconnaissance d'une voix, d'un instrument).

Enfin, la segmentation de flux audionumériques en atomes ou éléments unitaires (notes, grains, etc.) fait elle aussi appel à des paramètres perceptifs [23, 9].

5 Espaces perceptifs et effets audionumériques

5.1 Définition des effets adaptatifs

Un effet adaptatif est un effet dont les paramètres de contrôle sont pilotés par des paramètres extraits du son entrant (cf. Fig.11) [28, 27, 30]. Ainsi, nous automatisons le contrôle en fonction de propriétés intrinsèques du son.

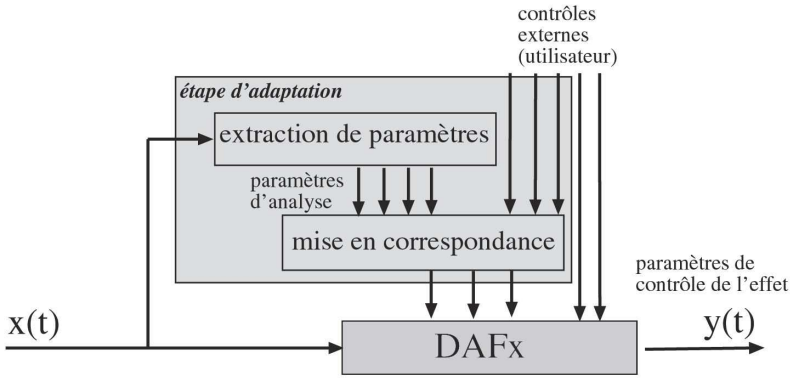

Figure 11: *Principe d'un effet audionumérique adaptatif [28]*

5.2 Contrôle des effets adaptatifs

Le contrôle habituellement confié à l'utilisateur lui étant retiré, nous proposons de le réintroduire dans la chaîne de traitement, mais cette fois-ci à un niveau plus élevé. Ceci est illustré par le diagramme complet (cf. Fig.12), qui prend en compte la gestualisation du contrôle et de l'automation du contrôle [4]. Ceci nous permet d'envisager de passer d'une configuration d'effet adaptatif à une autre.

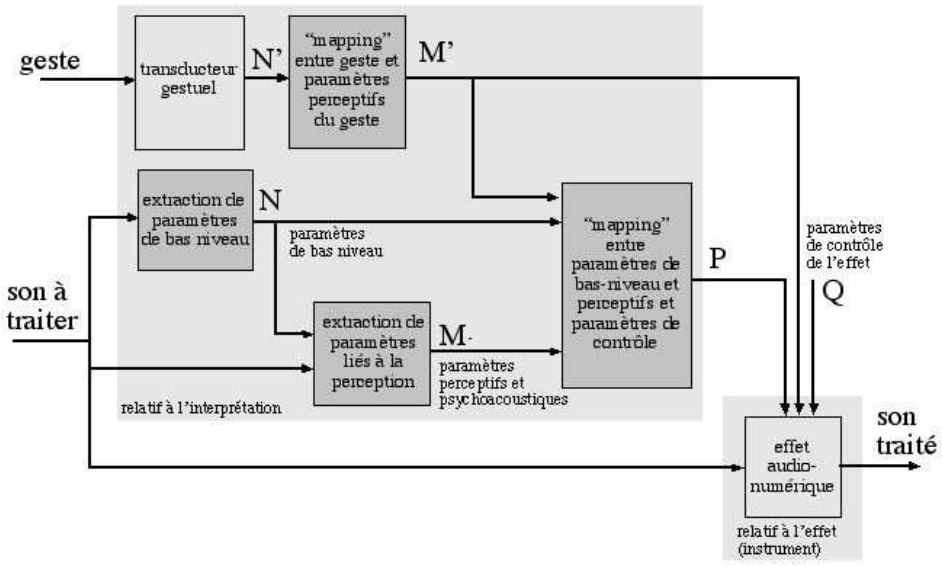

Figure 12: *Diagramme de l'effet audionumérique adaptatif contrôlé gestuellement [4]*

5.3 Paramètres de contrôle

Les paramètres de contrôle sont soit des indices de bas niveau (telle la fréquence fondamentale F_0, l'énergie calculée par RMS, le centroïde ou centre de gravité spectrale, la balance de voisement, cf. Fig.13), soit des indices de haut niveau. Ces derniers correspondent à des paramètres nécessitant une analyse approfondie du signal utilisant des connaissances sur le signal, en partie psychaocoustiques. Ce peuvent être les trajets des harmoniques, l'harmonicité, l'ouverture spectrale, le synchronisme des harmoniques... Ils sont aussi utilisés pour la segmentation de signaux audionumériques [23]. Les indices perceptifs, utilisés pour la définition d'espace perceptif, sont quant à eux la sonie, la hauteur, des paramètres descriptifs du timbre (vibrato, tremolo, rugosité, etc.). Ces indices de contrôle utilisés sont formalisés dans les standards **SDIF**et **MPEG-7** [18], et dans le programme **PsySound** [5].

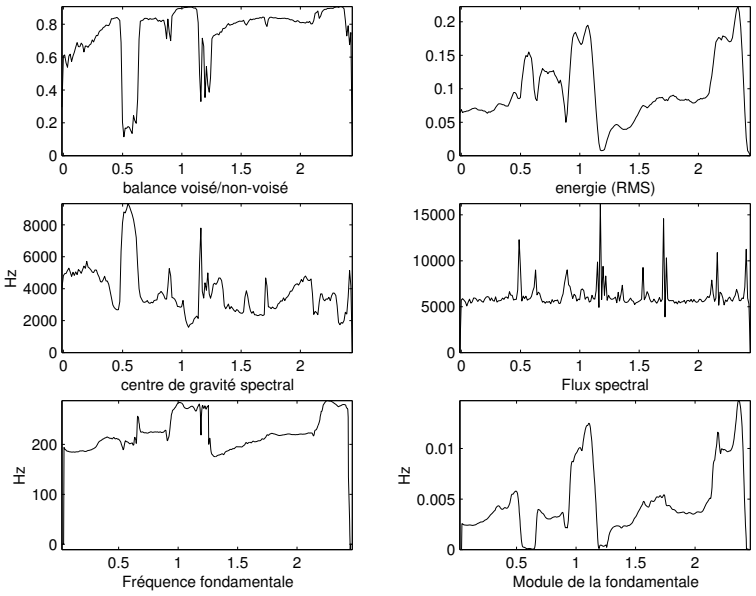

Figure 13: *Exemple d'indices globaux (voix chantée) : F0 (fréquence fondamentale), énergie par RMS, centroïde ou centre de gravité spectrale (CGS), balance de voisement*

5.4 Exemples d'utilisation

Les exemples qui vont être donnés sont accompagnés de sons. Ils portent sur le changement de durée d'un son avec conservation du timbre, sur la robotisation (ou le changement de qualité d'une voix), sur le changement de hauteur avec préservation du timbre et sur le délai granulaire adaptatif.

5.4.1 Changement de durée et conservation d'attributs perceptifs

Le changement de la durée d'un son se fait par un algorithme de contraction-étirement temporel (*pitch-shifting*) sélectif, implémenté à l'aide d'un vocodeur de phase. L'aspect sélectif de l'algorithme que nous utilisons vient du fait que le facteur de contraction ou d'étirement du son peut varier au cours du temps, et dépend d'une courbe donnée par l'utilisateur.

Prenons un premier exemple avec un son instrumental, une phrase musicale de flûte (**Piste 01**). Sa forme d'onde est donnée Fig.14 et sa fréquence fondamentale Fig.15. On effectue tout d'abord une contraction de la note la plus aigüe et un étirement des autres notes (**Piste 02**, forme d'onde Fig.16 et courbe de contrôle Fig.17), puis la transformation inverse, à savoir un étirement de la note la plus aigüe et une contraction des autres notes (**Piste 03**, forme d'onde Fig.18 et courbe de contrôle Fig.19). L'étirement se fait jusqu'à un facteur 4 et la contraction un facteur $1/4$.

Dès lors que l'on applique un changement de durée (adaptatif ou non) à un son, l'intelligibilité du contenu peut être altérée. Prenons pour exemple l'extrait sonore **Piste 04**, constitué d'une phrase dans un langage imaginaire. Sa transformation effectuée en utilisant une balance voisé-non voisé comme contrôle donne un texte

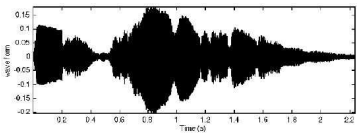

Figure 14: *Forme d'onde du son de flûte original (***Piste 01***)*

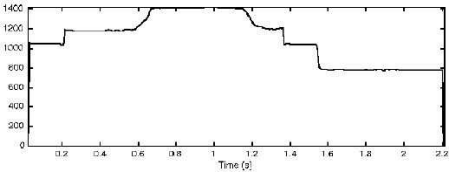

Figure 15: *Fréquence fondamentale du son de flûte (***Piste 01***)*

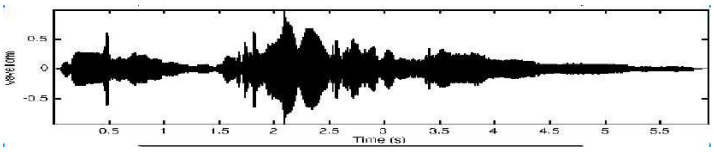

Figure 16: *Forme d'onde du son après étirement des notes basses, contraction de la note la plus aigüe (***Piste 02***)*

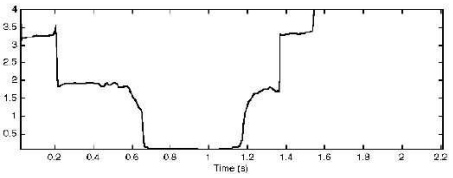

Figure 17: *Courbe de contrôle de l'étirement-contraction (***Piste 02***)*

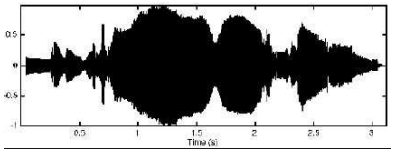

Figure 18: *Forme d'onde du son après étirement de la note la plus aigüe, contraction des notes basses (***Piste 03***)*

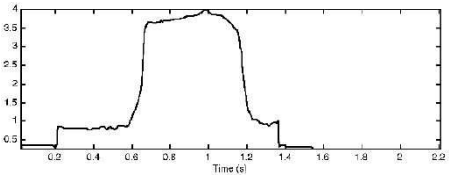

Figure 19: *Courbe de contrôle de l'étirement-contraction (***Piste 03***)*

prononcé différemment mais toujours compréhensible (**Piste 05**, et avec une loi non-linéaire de transformation de la courbe de contrôle **Piste 06**).

Un second exemple est donné avec l'extrait de voix de Pierre Schaeffer parlant de la musique électroacoustique (**Piste 07**). Nous lui avons appliqué différentes courbes de contrôles, et les sons obtenus donnent tous une expressivité différente :

- **Piste 08** : précipité, les mots en deviennent presque hachés ;
- **Piste 09** : pressé, sans pause de respiration entre les mots ;
- **Piste 10** : ralenti par moment, sans respiration ; semble réfléchir à ce qu'il dit en le disant ;
- **Piste 11** : précipité, mais néanmoins avec de longues pauses entre les morceaux de phrase ;
- **Piste 12** : ralenti aux pauses respiratoires et au milieu de certains mots ; semble réfléchir entre chaque mot.

Un troisième exemple porte sur la perception du vibrato. Prenons la **Piste 13** dont le vibrato exagéré a été appliqué par synthèse. Lorsqu'on ralenti le son (d'un facteur 4 pour l'exemple **Piste 14**), le vibrato n'est plus perçu comme qualité psychoacoustique du son, mais comme une modulation de hauteur, ce qu'il est aussi effectivement. Il n'est plus reconnaissable en tant que tel. Il devient alors nécessaire de le retirer du son original (par un changement de hauteur d'amplitude égale et en opposition de phase à celle du vibrato original), de procéder à l'étirement temporel, puis d'appliquer enfin un vibrato avec les propriétés d'amplitude et de fréquence originales (ce qui donne le résultat **Piste 15**).

Ces exemples montrent que pour garder l'intégrité des caractères perceptifs d'un son lors de sa contraction ou son étirement temporel, il faut utiliser des algorithmes plus complexes, permettant de reconnaître et d'extraire un vibrato, mais aussi un trémolo, une rugosité, et de reconnaître une voyelle d'une consonne pour les modifier de manières différentes.

5.4.2 Robotisation adaptative d'une voix

L'effet de robotisation est obtenu en mettant à zéro la phase d'un grain de son et en effectuant un bourrage de zéros (*zero-padding*) entre deux grains traités [27, 32]. On obtient un train d'onde, ce qui donne l'effet robotique, tout en conservant l'aspect formantique du son original à l'aide d'une fenêtre de traitement entre 256 et 1024 échantillons.

Prenons l'exemple de la voix de Schaeffer précédemment utilisée (**Piste 07** et **Piste 16**). Sa forme d'onde est donnée Fig.20 et le RMS Fig.21. Les extraits **Piste 17** et **Piste 18** correspondent à deux robotisations classiques à des hauteurs constantes mais différentes. La hauteur est donnée par le pas entre deux grains transformés. La robotisation adaptative consiste à faire varier ce pas, en fonction du RMS par exemple. Trois exemples différents sont données (**Piste 19**, **Piste 20**, **Piste 21**), tous basés sur le RMS (modifié par une fonction non linéaire), dont les hauteurs limites aussi sont différentes. Les voix de robot obtenues sont significativement différentes, par la hauteur mais surtout par l'expressivité.

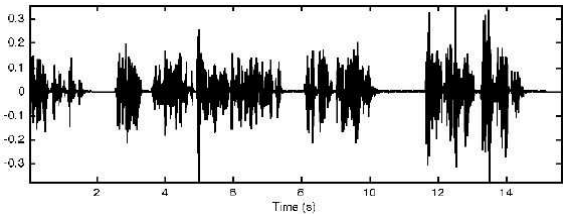

Figure 20: *Forme d'onde de l'extrait de voix de Pierre Schaeffer (***Piste 07***,* **Piste 16***)*

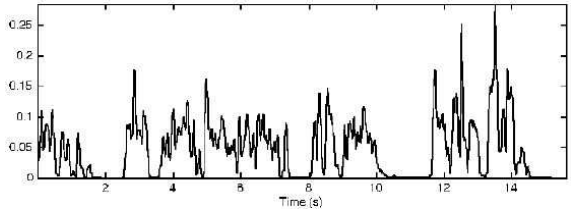

Figure 21: *Energie calculée par RMS de l'extrait de voix de Pierre Schaeffer (***Piste 07***,* **Piste 16***)*

Prenons à nouveau le langage imaginaire (**Piste 04**, **Piste 22**). La robotisation classique (**Piste 23**) ne surprend en rien. Les robotisations adaptatives présentent elles des expressivités bien distinctes. Les courbes de contrôle utilisées sont : pour la **Piste 24**, la fréquence fondamentale (la hauteur est donc la même, seul le timbre subit une modification) ; pour la **Piste 25**, l'opposée du centroïde ; pour la **Piste 26** l'opposée de la fréquence fondamentale ; pour la **Piste 27** l'opposé du RMS.

5.4.3 Changement de hauteur adaptatif avec préservation du timbre

Nous utilisons un algorithme de changement de hauteur lui aussi présenté dans [32], dans une version adaptative. Un changement de hauteur constant d'un ton (**Piste 28**) avec préservation du timbre ne sonne pas tout à fait pareil qu'un changement de hauteur adaptatif, contrôlé par la fréquence fondamentale, variant entre -1 et $+1$ demi-ton (**Piste 29**) ou contrôlé par l'opposée de la fréquence fondamentale (**Piste 30**). L'expressivité de la voix originale s'en trouve amoindrie ou renforcée ; la préservation de la forme spectrale induit effectivement une préservation du timbre.

5.4.4 Délai granulaire adaptatif

Un délai granulaire adaptatif est un retard appliqué à des grains en fonction d'un paramètre. Cela nécessite la gestion de plusieurs lignes à retard, de longueur et niveau de réinjection différents. En temps différé, par lecture/écriture dans des fichiers, le nombre de lignes à retard est limité par la fréquence d'échantillonnage en ce qui concerne la longueur de la ligne à retard, et par la quantification des échantillons (généralement, on utilise des sons échantillonnés à $44,1\ kHz$, codés sur 16 bits) en ce qui concerne le gain de réinjection. Les paramètres de contrôle sont la longueur de la ligne, le gain de réinjection, et taille du grain.

Les exemples que nous donnons sont obtenus à taille de grain constante (2048 échantillons). Tout d'abord, on applique au son de guitare **Piste 32** un écho traditionnel (**Piste 33**), avec un gain de réinjection de 0.4 et un délai de $0.3\ s$. On applique ensuite un écho dont seul el gain de réinjection est contrôlé par une courbe dérivée du RMS, afin d'appliquer l'écho uniquement aux attaques (**Piste 35**) ou uniquement aux parties harmoniques (**Piste 34**). La forme d'onde du son original est donnée Fig.22, le RMS Fig.23 et la courbe de contrôle correspondant à un lissage du RMS après troncature Fig.24. La longueur des lignes à retard est fixe.

Le deuxième exemple est cette fois le délai granulaire appliqué à la langue imaginaire (**Piste 04**, **Piste 22** et **Piste 36**). Si l'on fixe le gain à 0.4 et que le temps de délai varie entre 0.2 et 1 seconde, en utilisant comme courbe de contrôle la fréquence fondamentale, on obtient un son de synthèse granulaire dont la hauteur monte

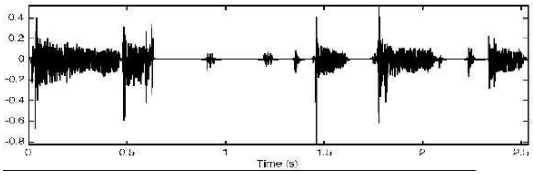

Figure 22: *Forme d'onde du son de guitare (***Piste 32***)*

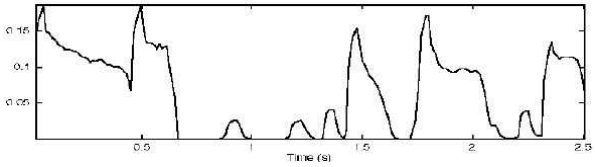

Figure 23: *Energie par RMS du son de guitare (***Piste 32***)*

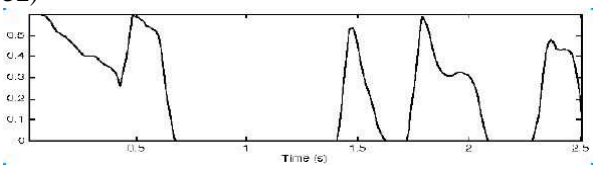

Figure 24: *Courbe de contrôle après troncature et lissage du RMS du son de guitare*

(**Piste 35**), étiré dans le temps au fur et à mesure de ses répétitions. Si au contraire on fixe le temps de délai à 0.2 seconde et fait varier le gain de réinjection en fonction de la balance de voisement, on obtient une voix dont seules les consonnes sont répétées (**Piste 36**).

Ces différents exemples de délai granulaire adaptatif montrent l'intérêt musical qu'apporte ces nouveaux effets audionumériques pour la composition et le traitement du jeu en direct.

5.4.5 Modifications des attributs perceptifs

Pour conclure sur les effets adaptatifs, on remarquera, suite aux exemples sonores présentés, la cohérence entre le son et l'effet qui lui est appliqué. Celle-ci découle du fait que les contrôles de l'effet dépendent du son par le biais de paramètres qui en sont caractéristiques. Bien utilisés, ces effets permettent notamment de modifier des paramètres perceptifs (la hauteur, le timbre, la durée) tout en conservant une intelligibité, une clarté dans le propos musical. On comprend ainsi mieux pourquoi et comment les effets et les espaces perceptifs (à priori sans rapport) peuvent être mis en relation étroite.

6 Conclusions, Perspectives

On notera l'intérêt de proposer des espaces perceptifs plus complets. En effet, l'espace de timbre, n'étant pour l'instant principalement développés que pour des sons musicaux instrumentaux est mal adapté aux autres sons. De plus, l'hypothèse sous-jacente d'orthogonalité de l'espace (hauteur, sonie, timbre), n'est pas tout à fait vérifée (ces dimensions ne sont pas tout à fait indépendantes). L'utilisation d'autres critères perceptifs (ex : vibrato, rugosité) permettent d'envisager de modifier la dimension temporelle de sons tout en conservant ses propriétés perceptives. Une voie de recherche est la prise en compte de propriétés micro et méso-temporelles, car elle semble nécessaire notamment pour mieux aborder la finesse de description d'un timbre ainsi que la représentation de sons écologiques. Une voie de recherche complémentaire consiste à prendre en compte l'évolution à plus long terme (macro-temporel) du son : la dynamique du système "onde sonore" étudié. Enfin, la prise en compte de la localisation spatiale devrait permettre une description plus complète de l'espace perceptif des sons. En ce qui concerne les effets adaptatifs, de nouveaux effets devraient voir le jour, portant notamment sur des modifications du timbre (changement de voyelle pour la voix, changement du type de transition pour des sons instrumentaux), sur la spatialisation, sur l'extraction de critères perceptifs tel le vibrato avant l'application de changement de durée d'un son. Ceux-ci respecteront au mieux les attributs perceptifs, ou les modifieront de manière plus contrôlée. Enfin, la gestualisation du contrôle de ces effets est déjà en cours, en vue de *morphing* entre effets adaptatifs et entre configurations d'un effet adaptatif.

Références

[1] McAdams S. and. *Audition : Physiologie, Perception et Cognition*, pages 283–344. Richelle M., Requin J. and Robert M., Eds., Presses Universitaires de France, Paris, 1994.

[2] Drame C., Wessel D., and Wright M. Removing the time axis from spectral model analysis-based additive synthesis: Neural networks vs. memory-based machine learning. In *Proc. ICMC, Ann Arbor, Michigan*, 1998.

[3] Arfib D. L'espace perceptif dans la relation du son au geste. In *Proc. Journées d'Informatique Musicale*, Bourges, 2001.

[4] Arfib D., Couturier J. M., Kessous L., and Verfaille V. Strategies of mapping between gesture parameters and synthesis model parameters using perceptual spaces. *Organised Sound*, Mapping Strategies Issue in 2002.

[5] Cabrera D. "psysound": a computer program for the psychoacoustical analysis of music. In *Proc. Australasian Computer Music Conference, MikroPolyphonie*, volume 5, Wellington, New Zealand, 1999.

[6] Keller D. and Berger J. Everyday sounds: synthesis parameters and perceptual correlates. In *Proc. of the VIII Brazilian Symposium of Computer Music*, Fortaleza, Brazil, 2001.

[7] Wessel D. Timbre space as a musical control structure. *Computer Music Journal*, 3(2):45–52, 1979.

[8] Metois E. *Musical Sound Information: Musical gesture and Embedding synthesis*. PhD thesis, Massachusetts Institute of Technology, 1996.

[9] Tzanetakis G. and Cook P. Multifeature audio segmentation for browsing and annotation. In *Proc.IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, WASPAA99*, New Paltz, NY, USA, 1999.

[10] IRCAM. Studio on line, http://sol.ircam.fr, 2000.

[11] Risset J.-C. and Wessel D. L. *Exploration of timbre by analysis and synthesis*, pages 113–169. D. Deutsch, Academic Press, New York, 1999.

[12] Couturier J.-M. Espaces interactifs visuels et sonores pour le contrôle des sons musicaux. In *Actes des Journées d'étude Espaces Sonores*, Avril 2002.

[13] Jensen K. *Timbre Models of Musical Sounds*. PhD thesis, University of Copenhagen, 1999. DIKU Report 99/7.

[14] Kessous L. Instruments bi-manuels et espaces sonores. In *Actes des Journées d'étude Espaces Sonores*, Avril 2002.

[15] Grey J. M. *An exploration of musical timbre*. PhD thesis, Stanford University, 1975.

[16] K. Martin. Musical instrument identification: A pattern-recognition approach, 1998.

[17] Iverson P. Auditory stream segregation by musical timbre: Effects of static and dynamic acoustic attributes. *Journal of Experimental Psychology: Human Perception and Performance*, 21:751–763, 1995.

[18] G. Peeters, S. McAdams, and P. Herrera. Instrument sound description in the context of mpeg, 2000.

[19] Shepard R. *Pitch perception and measurement*, chapter 13. Cook P. R. Ed., MIT Press, 1999.

[20] Lakatos S. A common perceptual space for harmonic and percussive timbres. *Perception & Psychophysics*, 62:1426–1439, 2000.

[21] McAdams S. and J.C. Cunibile. Perception of timbral analogies. In *Philosophical Transactions of the Royal Society*, volume series B 336, pages 383–389, London, 1992.

[22] McAdams S., Winsberg S., de Soete G., and Krimphoff J. Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes. *Psychological Research*, 58:177–192, 1995.

[23] Rossignol S., Rodet X., Soumagne J., Collette J.-L., and Depalle P. Feature extraction and temporal segmentation of acoustic signals, 1998.

[24] E. Scheirer and M. Slaney. Construction and evaluation of a robust multifeature speech/music discriminator. In *Proc. ICASSP '97*, pages 1331–1334, Munich, Germany, 1997.

[25] Jehan T. and Schoner B. An audio-driven perceptually meaningful timbre synthesizer. In *Proc. International Computer Music Conference*, Havana, Cuba, 2001.

[26] Jehan T. and Schoner B. An audio-driven, spectral analysis-based, perceptually meaningful timbre synthesizer. In *AES 11Oth convention*, Amsterdam, Netherland, 2001.

[27] Verfaille V. Réalisation d'effets audionumériques adaptatifs en temps réel et hors temps réel. In *Journées d'Informatique Musicale, 9e édition*, 2002.

[28] Verfaille V. and Arfib D. Adafx: Adaptive digital audio effects. In *Proceedings of the DAFx'01 Workshop in Limerick, Ireland*, December 2001.

[29] Beauchamp J. W. Synthesis by spectral amplitude and "brightness" matching of analyzed musical instrument tones. *J. Audio Eng. Soc.*, 30(6):396–406, 1982.

[30] Amatriain X., Bonada J., Loscos A., Arcos J. L., and Verfaille V. Addressing the content level in audio and music transformations. *Journal of New Music Research*, 2002.

[31] Serra X. *Musical Sound Modeling with Sinusoids plus Noise*. Poli G. D., Picialli A., Pope S. T. and Roads C. Eds., Swets & Zeitlinger, 1996.

[32] Udo Zolzer, editor. *DAFX - Digital Audio Effects*. John Wiley & Sons, 2002.