



HAL
open science

Le “BSTD” – Une représentation graphique de la brillance et de l’écart type spectral, comme possible représentation de l’évolution du timbre sonore

Mikhail Malt, Emmanuel Jourdan

► To cite this version:

Mikhail Malt, Emmanuel Jourdan. Le “BSTD” – Une représentation graphique de la brillance et de l’écart type spectral, comme possible représentation de l’évolution du timbre sonore. L’analyse musicale aujourd’hui, Delatour, 2015, 9782752102430. hal-01265269

HAL Id: hal-01265269

<https://hal.science/hal-01265269v1>

Submitted on 1 Sep 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LE « BStD » – UNE REPRESENTATION GRAPHIQUE DE LA BRILLANCE ET DE L'ECART TYPE SPECTRAL, COMME POSSIBLE REPRESENTATION DE L'EVOLUTION DU TIMBRE SONORE.

Mikhail Malt
mikhail.malt@ircam.fr
Ircam/Paris IV – MINT

Emmanuel Jourdan
emmanuel.jourdan@ircam.fr
Ircam

Résumé

La BStD (**B**rightness, **S**tandard **D**eviation)¹ est une représentation graphique de l'évolution du timbre et/ou de la texture sonore destinée à être un outil d'aide à l'analyse de musiques électroniques et/ou enregistrées fondés sur le calcul de trois descripteurs audio, la brillance, l'écart type spectral et l'amplitude.

Un descripteur audio est un paramètre, uni ou multidimensionnel, ramenant une dimension particulière de ce signal à un (ou plusieurs) paramètre(s) numérique(s). Son évolution temporelle peut être utilisée pour représenter, caractériser, illustrer, etc. un flux ou une évolution d'événements sonores. Initialement prévus comme des moyens pour identifier des caractéristiques sonores et ensuite pour faciliter l'extraction d'information dans de vastes bases de données de musique enregistrée, ils sont aujourd'hui suffisamment développés, et surtout accessibles, pour que soit considérée leur intégration dans des stratégies analytiques prenant en compte les textes musicaux et le signal audio provenant d'enregistrements.

Lorsque nous parlons d'outil d'aide à l'analyse, nous parlons d'un outil permettant d'obtenir d'autres points de vue que ceux des catégories traditionnelles de la musique (rythme, harmonie, etc.). D'un outil qui puisse nous permettre de visualiser des caractéristiques, qui ne seraient pas explicites par l'analyse ou visualisation, d'autres catégories ou représentations musicales. D'un outil qui puisse nous permettre de trouver des indices pour la segmentation. Soit d'un outil heuristique (du point de vue étymologique), s'insérant dans un processus d'analyse musicale et offrant des propositions de représentations (sujettes à l'interprétation de l'analyste) nous permettant de construire un discours critique sur l'objet musical étudié.

Introduction

Par le biais de la construction d'une représentation, qui réunit trois descripteurs acoustiques, dans un même espace graphique, la BStD (**B**rightness, **S**tandard **D**eviation) est censée mettre en évidence des informations codées séparément dans les descripteurs. Une de nos hypothèses de travail, est le fait que le changement de représentation et la construction de nouvelles représentations, à partir d'une décomposition paramétrique du phénomène étudié, sont des étapes importantes et des aides considérables dans le processus de modélisation et d'analyse. La résolution et la compréhension d'un problème dépendent sensiblement de la représentation choisie [Fink 1996, 1].

Cette représentation a été développée dans un contexte d'étude de l'utilisation des descripteurs acoustiques pour la captation audio, en temps réel, de paramètres sonores pour la musique mixte [Malt&Jourdan 2009] et d'une recherche de méthodes (fondés sur l'analyse audio) pour l'étude de la structure et forme de phénomènes musicaux.

Lorsque nous parlons d'outil d'aide à l'analyse, nous parlons d'un outil qui puisse nous permettre d'avoir (ou de nous donner) d'autres points de vue que ceux des catégories traditionnelles de la musique (rythme, harmonie, etc.). D'un outil qui puisse nous permettre de représenter, de forme graphique, des caractéristiques qui ne seraient pas explicites par l'analyse ou par la visualisation d'autres catégories ou

¹ Brillance, écart type.

représentations musicales. D'un outil qui puisse aussi nous aider dans le travail de segmentation². Soit d'un outil heuristique (du point de vue étymologique), s'insérant dans un processus d'analyse musicale, et offrant des propositions de représentations (sujettes à l'interprétation de l'analyste) nous permettant de construire un discours critique sur l'objet musical étudié.

Les descripteurs audio

Un descripteur audio est un paramètre, uni ou multidimensionnel, caractérisant un aspect particulier du signal sonore, ramenant une dimension particulière, de ce signal, à un (ou plusieurs) paramètre(s) numérique(s). Son évolution temporelle peut être utilisée pour représenter, caractériser, illustrer, etc. un flux ou une évolution d'évènements sonores. L'amplitude, l'énergie, la sonie (l'intensité perçue), la fréquence fondamentale et la hauteur perçue font partie des descripteurs les plus couramment utilisés.

Leurs premières utilisations massives datent des années 1960 [Futrelle&Downie 2002, 215] [Uitdenbogerd&all 2000] lorsque est apparu l'idée d'appliquer les techniques d'extraction, de traitement et de recherche des informations à des données musicales [Kassler 1966]. Initialement prévus comme des moyens pour identifier des caractéristiques sonores et ensuite pour faciliter l'extraction d'information dans de vastes bases de données de musique enregistrée, ils sont aujourd'hui suffisamment développés, et surtout accessibles, par le biais de nombreuses implémentations logicielles³, pour que soit considérée leur intégration dans des stratégies analytiques prenant en compte les textes musicaux et le signal audio provenant d'enregistrements. Des récents travaux investiguent des possibles applications des descripteurs dans des contextes analytiques en musicologie [Crawford & Gibson 2009] et en ethnomusicologie [Tagg 2010], [Moelants & all. 2006].

Parmi les descripteurs communément utilisés [Herrera, Serra & Peeters, 1999] [Peeters 2003], trois d'entre eux ont attiré notre attention, la brillance, l'écart type spectral et l'amplitude.

Le centroïde spectral

La brillance ou centroïde spectral, est la fréquence moyenne pondérée par les amplitudes, issue d'une analyse de Fourier d'un signal sonore numérique (Figure 1). La Brillance μ , instantanée, peut être calculée selon l'expression suivante :

$$\mu = \frac{\sum_{i=0}^{n-1} f[i]a[i]}{\sum_{i=0}^{n-1} a[i]} \quad \text{Équation 1}$$

où

i est l'index de bande de fréquence considérée, $f[i]$ sa fréquence centrale et $a[i]$ son amplitude.

² La segmentation pouvant advenir lorsqu'un changement d'un paramètre précis est significatif et localement stable. Comme cela a déjà été décrit par différents auteurs [Mesnage 1994] et [Mééus 1994], l'activité de segmentation est fondamentale dans le processus d'analyse musicale.

³ Voir, par exemple : [Todorof&all. 1995], [Puckette&all. 1998], [Malcom 1998], [Schnell&Schwarz 2005], Auditory Toolbox, [Bullock 2007], [Cabrera&all. 2007], [Lartillot 2007], [Malt&Jourdan 2008] et [Tristan 2010], entre autres. Nous ne citons que les implémentations actuellement disponibles au public. Il existe d'autres implémentations, pour le moment en utilisation interne par les institutions qui les développent, telles que les diverses librairies utilisées en « Audio Indexation » à l'Ircam (Ircamdescriptor, Ircambeat, Ircamchord, etc., <http://anasynth.ircam.fr/drupal6/?q=english/software>, 09/06/2010).

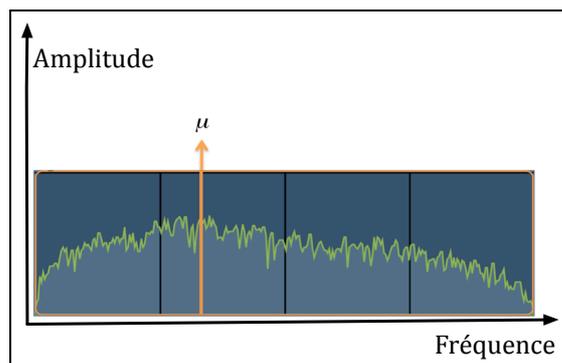


Figure 1 : Représentation de la brillance

Il y a déjà plus de trente ans que John Grey [Grey 1975] avait mis en évidence le fait que la brillance, était un des trois principaux paramètres pour la caractérisation du timbre. Cette idée s'est vue renforcée et développée particulièrement par les travaux de Wessel [Wessel 1979] et McAdams [McAdams&all 1995]. Plus récemment, notamment dans le cadre du MIR (« Music Information Retrieval », soit de l'extraction, traitement et recherche de l'information musicale⁴), plusieurs travaux ont montré la pertinence de l'utilisation de la brillance [Livshin&Rodet 2004] comme étant l'un des 10 descripteurs les plus importants pour l'identification du timbre instrumental et comme un bon estimateur de la brillance spectrale perçue [Shubert&Wolfe 2006, p. 823-824].

L'écart type spectral

L'écart type est la racine carrée de la variance spectrale autour du centroïde spectral. C'est une mesure statistique de la dispersion des données au tour de la moyenne. Dans notre cas précis, c'est la dispersion du spectre au tour du « centroïde spectral ».

$$v = \frac{\sum_{i=0}^{n-1} (f[i] - \mu)^2 a^2[i]}{\sum_{i=0}^{n-1} a^2[i]} \quad \text{Équation 2}$$

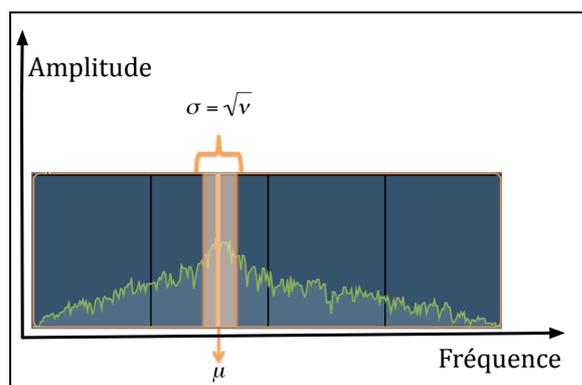


Figure 2 : La variance (v) et l'écart-type (σ) par rapport à la brillance (μ)

Récemment, Frédéric Chiasson [Chiasson 2007] a suggéré que l'écart type spectral pourrait être rapproché du paramètre musical appelé « volume » par Koechlin [Koechlin 1943]. Ce que Koechlin appelle volume étant l'épaisseur du son. Cette hypothèse est utilisée dans l'environnement d'aide à l'orchestration développé à l'Ircam par Grégoire Carpentier [Carpentier 2008] et Damien Tardieu [Tardieu 2007], montrant la pertinence de la proposition de Chiasson.

⁴ Traduction donnée par Michel Fingerhut en communication personnelle aux auteurs le 9 octobre 2007.

« ...l'étendue spectrale correspond à l'écart-type (second moment spectral), appelé couramment spread. D'un point de vue perceptif, il est relié à l'épaisseur spectrale du son... »
[Carpentier 2008, p. 103]

La représentation « BStD »

Ce que nous jugeons important de remarquer est le fait que ces deux descripteurs, la brillance et l'écart-type spectral sont fortement corrélés avec des aspects cognitifs du timbre, exprimant, chacun d'entre eux, une information pertinente sur les caractéristiques du son analysé.

« Même si on constate en pratique une corrélation significative entre le centroïde et l'étendue, ces deux descripteurs conservent toutefois une certaine indépendance. Ainsi, une flûte piccolo dans le registre aigu peut avoir un son très brillant et très pincé, alors qu'un cor dans le grave a un son sourd et large (par comparaison à un orgue par exemple). »
[Carpentier 2008, p. 230]

Notre proposition est de représenter l'évolution de ces deux paramètres, plus l'amplitude, dans une forme graphique où la position verticale (ordonnée) de son barycentre représentera la brillance, la hauteur (autour de la brillance) de cette forme l'écart type spectral et la couleur, l'intensité du signal (Figure 3). Comme nous pouvons l'observer, la BStD est une représentation en trois dimensions. L'abscisse représente le temps d'évolution des paramètres, l'ordonnée l'espace des fréquences et la troisième dimension (soit la couleur ou le niveau de gris) représentant l'amplitude. Il est important de signaler que la forme qui est représenté dans l'axe vertical n'est pas un sonagramme, mais une forme qui par la position de son barycentre indique la brillance et par son étendue verticale le « volume spectral ».

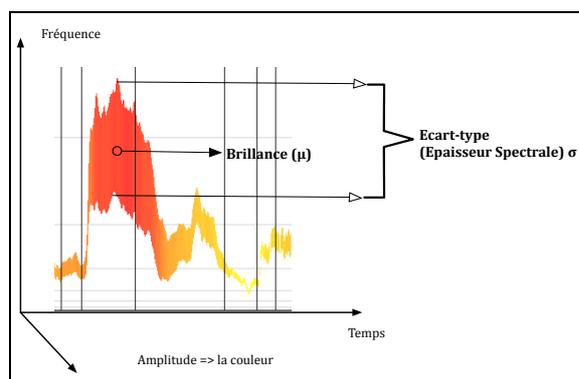


Figure 3 : La représentation graphique « Brillance-écart type »

Cette visualisation et les prochaines se feront dans l'environnement de programmation graphique MaxMSP⁵ en utilisant la librairie de descripteurs audio « Zsa.descriptors »⁶ [Malt&Jourdan 2008]. Il est tout à fait possible et aisé d'implémenter ce type de visualisation dans d'autres environnements tels que MATLAB⁷, PureData⁸, Excel⁹, ou autre.

Dans l'implémentation de la BStD, nous avons été confrontés à un problème majeur. Lorsque l'intensité du signal diminue, le spectre tend au spectre du bruit de fond et tend à s'aplatir, ce qui accroît la valeur de l'écart type. Comme conséquence immédiate, une source avec un écart type spectral bas (une flûte traversière, par exemple) verra la valeur de ce descripteur augmenter au fur et à mesure que son

⁵ © www.cycling74.com.

⁶ Cette librairie d'outils pour l'analyse de descripteurs temps réel est actuellement disponible sur http://www.e--j.com/?page_id=85. Une première version du programme informatique utilisé pour réaliser les BStD de ce texte est incluse dans les exemples de la librairie.

⁷ © The MatWorks, <http://www.mathworks.fr/>.

⁸ Environnement de programmation graphique dédié à la musique, conçu et développé par Miller Puckette, <http://crca.ucsd.edu/~msp/software.html>.

⁹ © Microsoft, <http://www.microsoft.com/france/office/2007/logiciels/excel/overview.msp>.

intensité décroît. Cependant, l'auditeur ne semble pas le percevoir de cette manière. Nous n'avons pas trouvé, à ce jour, une étude de modélisation de l'écart type spectral perçu (ou du volume). Cependant, quelques études s'accordent sur le fait que la perception de la densité spectrale est principalement liée à l'écart type spectral et aux fluctuations de l'énergie du signal [Hanna 2003, p. 76]. [Hartmann & Pumplin 1988] et [Hartmann & Pumplin 1991], montrent que le problème n'est pas simple, puisque la perception de la densité spectrale, que nous avons rattachée à l'écart type spectral est aussi une fonction du contenu spectral du signal [Hanna 2003, p. 108]. En extrapolant les résultats de ces études nous sommes appuyés sur l'hypothèse que la densité spectrale perçue est fonction principalement de l'écart type spectral et de l'énergie du signal¹⁰. De ce fait, pour la représentation, nous calculons un écart type σ' par rapport à l'énergie ($e = a^2[i]$) et par rapport à une fonction $f(e)$ selon l'expression *ad hoc* :

$$\sigma' = f(e) * \sqrt{\frac{\sum_{i=0}^{n-1} (f[i] - \mu)^2 a^2[i]}{\sum_{i=0}^{n-1} a^2[i]}}$$

Équation 3

L'expression $f(e)$, que nous avons construit manuellement, est une courbe qui fonctionne autant comme un seuil qu'une mise à l'échelle pour prendre en compte la non linéarité de la relation écart type/énergie.

Quelques exemples

Voyons ensuite quelques exemples d'application de la BStD. Nous précisons, que les exemples qui suivent ne sont, en aucun cas des analyses exhaustives des pièces présentées, mais seulement des parallèles, à fin pédagogiques, entre des connaissances largement diffusées et l'utilisation de la BStD. Nous présenterons systématiquement la BStD, la représentation en forme d'onde et le spectrogramme¹¹. Même si nous ne discuterons pas en détail ces deux dernières représentations elles sont exposées pour qu'on puisse se rendre compte de l'apport d'information que la BStD nous propose.

Le Boléro de Ravel

Comme premier exemple, nous présentons le Boléro (1928) de Ravel, dans l'interprétation de la Philharmonique de Berlin dirigée par Pierre Boulez, (CD Deutsche Grammophon, 1994).

¹⁰ Nous pourrions supposer que la prise en compte du Kurtosis (quatrième moment statistique spectral) dans cette modélisation puisse être utile, du fait que ce paramètre caractérise la forme spectrale.

¹¹ Les sonagrammes présentés utilisent les paramètres suivants : taille de fenêtre – 4096 échantillons, pas d'avancement – 512 échantillons et fenêtre d'analyse – blackman.

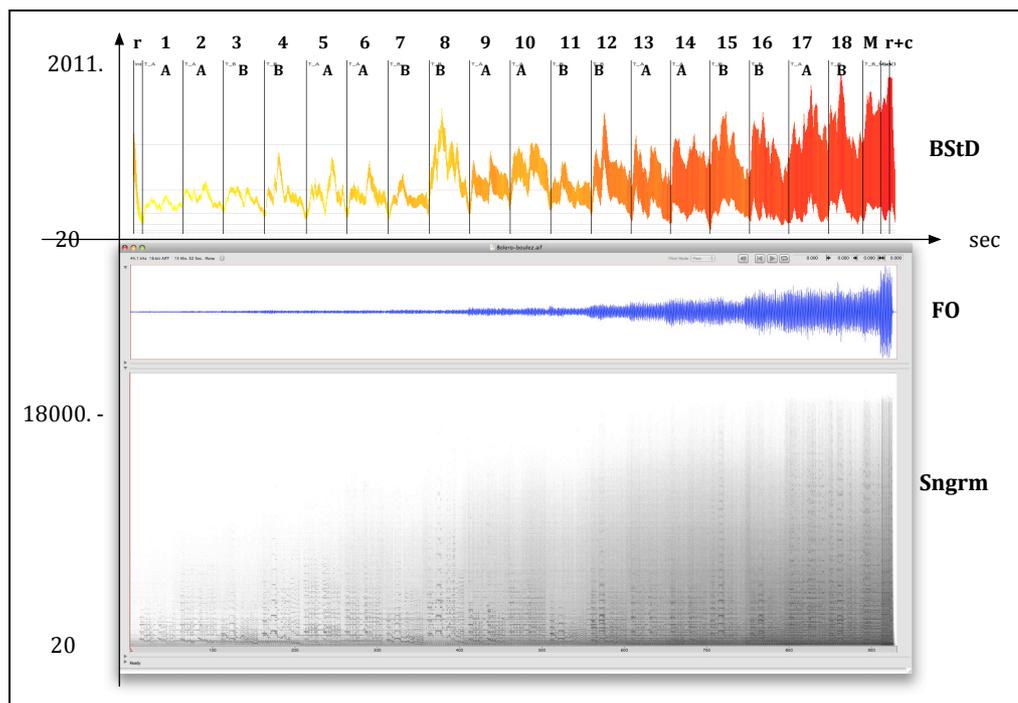


Figure 4 : La BStD du *Bolero* (1928) de Ravel

La Figure 4 présente la BStD accompagnée du sonagramme (Sngm) et de la représentation en forme d'onde (FO) de cette interprétation. Nous avons ajouté des marqueurs signalant les numéros de sections, les thèmes A et B, la ritournelle (r), la modulation finale (M+B) et la coda ritournelle finale (r+c).

Cette œuvre est principalement connue pour être une expérience de Ravel, qui se présente comme une alternance uniforme de deux thèmes (A et B), sur un fond d'ostinato de caisse claire, avec pour unique variation l'instrumentation. Après une courte introduction, Ravel énonce le premier thème (thème A, section 1) par la première flûte. Les deux thèmes passent par un ensemble de combinaisons instrumentales pour arriver, après la 18^{ème} section, à une modulation en mi majeur, où tout l'orchestre est sollicité. L'œuvre se voit comme une lente évolution de la couleur, de l'intensité et de la densité orchestrale (assez facilement repérable dans la représentation en forme d'onde et dans le sonagramme). C'est précisément ce que nous pouvons observer dans la représentation que nous proposons.

La forme proposée par la BStD évolue avec une augmentation de l'écart type spectral et de la brillance. Dans la section 1 (thème A) nous trouvons un écart type spectral au tour de 10 à 30 Hz et une brillance de 300-500 Hz. Dans la dernière section écart type spectral arrive à 500 et 1500 Hz pendant que la brillance arrive à une moyenne autour de 1500 Hz. En plus, du fait d'une basse valeur d'écart type spectral, pendant les premières sections (de 1 à 8), la forme suit de près la forme mélodique des deux thèmes. On ne visualise pas seulement l'évolution de la brillance, mais aussi ses variations locales.

***Atmospheres* de Ligeti**

Un deuxième exemple d'application est la représentation BStd de l'évolution de *Atmospheres* (1961) de Ligeti¹² (Figure 5). Nous utiliserons la version de la création (22 octobre 1961) au festival de Donaueschingen, par Orchestre du Südwestfunk Baden-Baden, direction de Hans Rosbaud.

Nous avons ajouté à l'illustration des marqueurs signalant les sections de A à U, plus quelques indications de temps en secondes, comme repères pour nos commentaires.

¹² Version de la création : 22 octobre 1961, festival de Donaueschingen, par Orchestre du Südwestfunk Baden-Baden, direction : Hans Rosbaud.

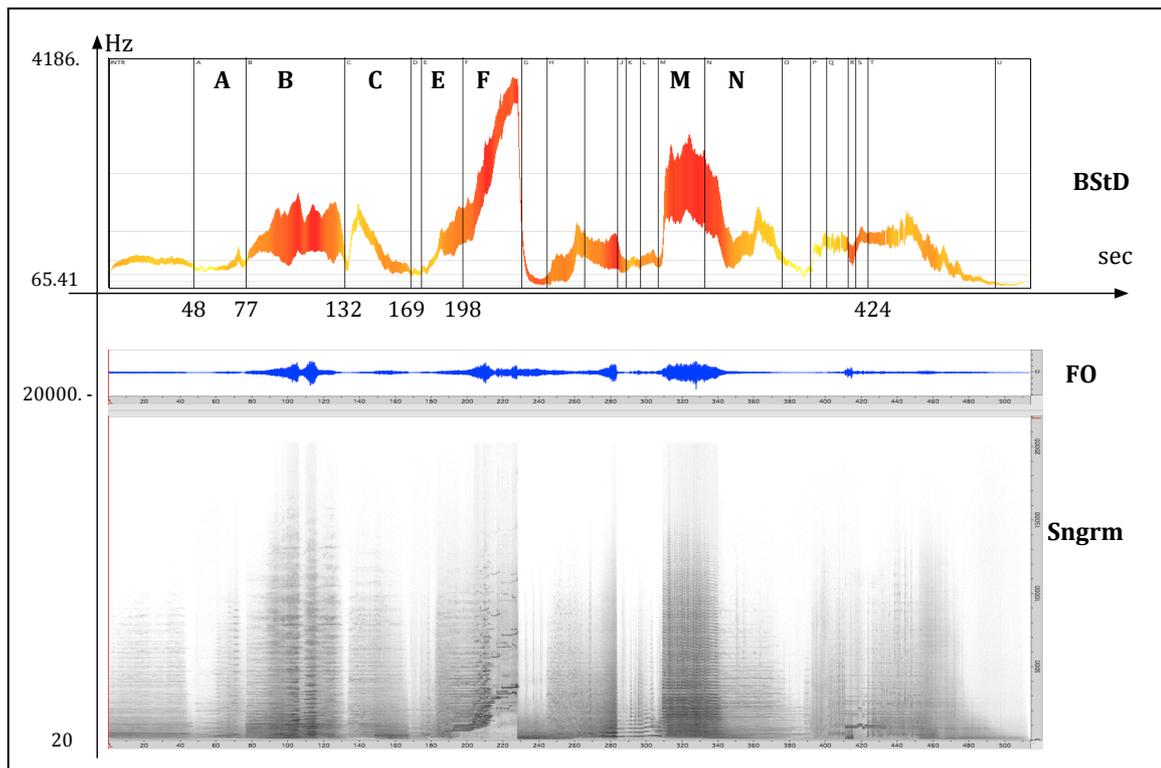


Figure 5 : Représentation de l'évolution de la « BStD » de *Atmospheres* (1961) de Ligeti

L'introduction de cette œuvre (m. 1 à 8) [Ligeti 1961] est orchestrée par des cordes au complet (14 vl1, 14 vl2, 10 alt, 10 vlc et 8 cb), des flûtes (par 4), des clarinettes (par 4), des bassons (par 3), un contrebasson et des cors (6). Cette section devrait avoir, selon la partition (p.2), une attaque *pp* et ensuite un mouvement de dynamique indiquée comme « *dim.....morendo.....* », jusqu'à la mesure 9 (début de la section A).

L'analyse de cet extrait, avec une représentation BStD des premières 88 secondes, est présentée Figure 6. Nous constaterons que, après une attaque *pp*, l'ensemble maintient légèrement la dynamique en effectuant un léger crescendo pendant les premières 15 secondes pour ensuite procéder à un decrescendo très lent, (la troisième dimension, la couleur qui représente l'amplitude évolue du plus foncé vers le plus clair), avec un centroïde qui baisse aussi très lentement et un écart type qui tend, dans les premières 15 secondes, de plutôt augmenter pour revenir à décroître par la suite. A la fin de l'introduction, nous observons la diminution brusque d'intensité et du volume orchestral jusqu'à la mesure 9. A cette mesure (début de la section A) le compositeur opère un filtrage orchestral, l'effectif se réduisant aux altos et aux violoncelles. Les hauteurs jouées au début de cette section (A, m. 9) se concentrent dans un registre entre le mi2 (vlc) et le si3 (vla) avec un centroïde spectral au tour de 250 Hz et un écart-type autour de 5 Hz.

Pour la section A, qui commence « *ppp, senza colore, non vibr.* » (Selon la partition, p. 3, m. 9), nous percevons premièrement une oscillation du centroïde et ensuite une augmentation de densité spectrale et d'intensité. Le premier point se doit principalement à l'instabilité propre à maintenir un même accord avec 20 musiciens (10 alti et 10 violoncelles). Il serait intéressant de comparer cet extrait à une autre interprétation plus actuelle. Le fait que les créations, dans certains contextes, manquent de répétitions suffisantes est un fait connu. Le deuxième point (l'augmentation de la densité spectrale, de l'intensité et du centroïde) se doivent principalement à l'interprétation de la partition (m. 11-12) où nous avons des indications de crescendos non synchronisés et une indication « *poco a poco vibr.* », qui engendrent, au fur et à mesure, un timbre plus brillant (pic de brillance de +/- 990Hz) et un spectre plus dense (pic d'écart type de 1200Hz), symbolisés par le pic dans la BStD autour des 72 secondes (Figure 6) avant l'entrée dans la section B.

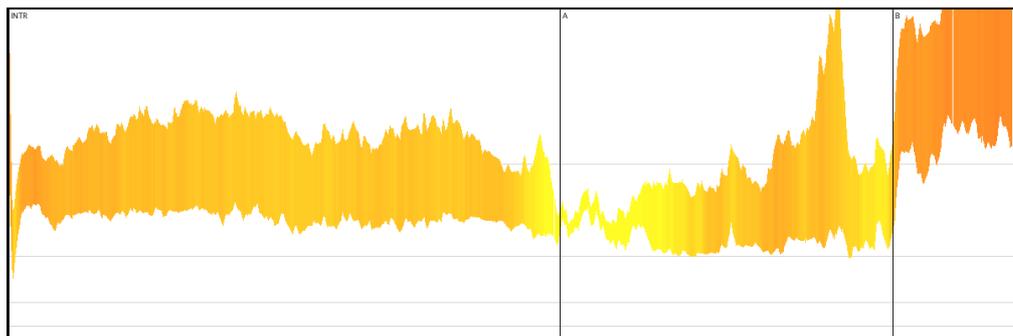


Figure 6 : *Atmospheres*, les premières 88 sec (introduction + section A)

La densification d'instrumentation et le changement de couleur orchestral dans la section B, par l'entrée des bois, des métaux, des violons et des contrebasses en crescendo, sont suggérées par la forme de la section B (Figure 5) entre 77 et 132 secondes.

Si nous connaissons l'œuvre, nous pouvons aussi remarquer la corrélation qui existe entre l'évolution de la forme de la BStD et l'articulation des 22 sections. Certaines sections, comme la B (77'') et C (132''), sont clairement corrélées pendant que pour d'autres, comme sections E et F et sections M et N, il y a une disjonction entre articulation formelle et le processus d'évolution du timbre.

De Natura Sonorum de Parmegiani

Nous présenterons un extrait (de 4'14 à 4'28) de la deuxième partie, *Accidents/Harmoniques*, de cette œuvre composée en 1975. L'enregistrement utilisé est une numérisation du côté A du disque vinyle de la création (3 Juin 1975), réalisée par le « Avant Garde Project »¹³. Le fichier original était en format FLAC. Celui-ci a été converti au format AIFF et traité, pour supprimer le bruit de fond avec l'outil de « noise removal » du logiciel Audiosculpt¹⁴ (Figure 7). Cette opération nous a été nécessaire compte tenu de la présence d'un bruit de fond trop importante, qui engendrait des distorsions considérables dans la forme générée (Figure 8).

Dans cet extrait, à un son tenu, de fondamentale 436.465 Hz avec quatre harmoniques, se superposent trois événements que nous pourrions appeler de « stationnaires », selon la nomenclature des UST¹⁵. Les deux premiers (1^{er} départ à 4'16'', durée 3''544 et le 2^{ème} départ à 4'20 durée 3''267) présentent un caractère percussif avec des caractéristiques temporelles de « freinage », soit une enveloppe d'amplitude ondulatoire et amortie (voir l'évolution de l'amplitude dans la Figure 9). Le troisième événement est une large bande de bruit (départ 4'25''449 durée 0.1''), courte et de caractère percussif¹⁶. Dans la partition de diffusion de *De Natura Sonorum* [Thomas&all, 9], Bernard Parmegiani, ne détaille pas ces trois événements, mais note entre 4'20 et 4'30 deux événements dénommés « vibra ».

¹³ <http://www.avantgardeproject.org/agp140/index.htm>

¹⁴ © Ircam.

¹⁵ Unités sémiotiques temporelles.

¹⁶ Voir dans [Thomas&all, 53-54] l'explication de Bernard Parmegiani, sur la réalisation technique de ces événements.

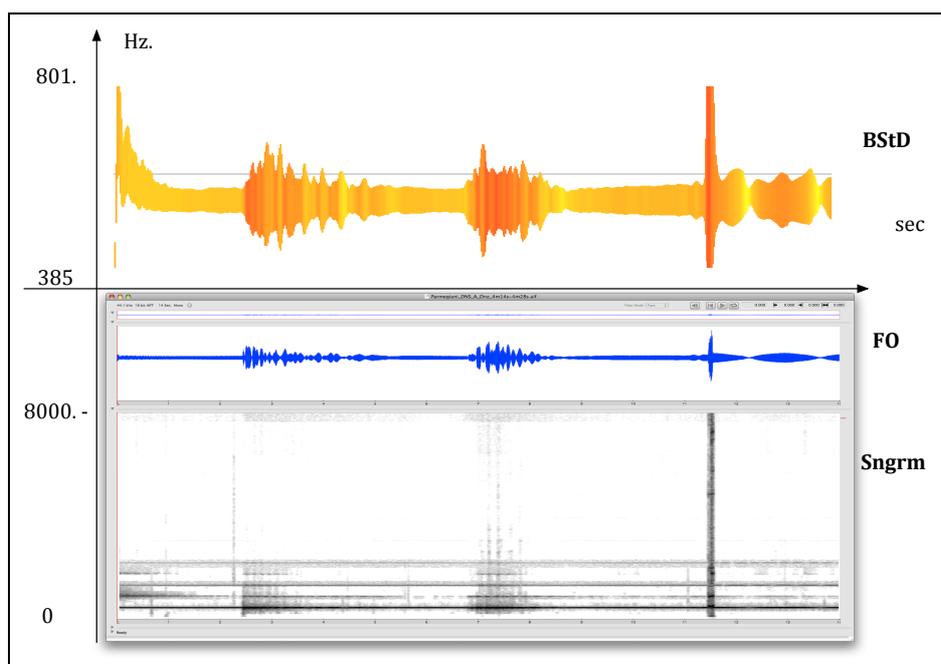


Figure 7 : *De Natura Sonorum*, Parmegiani, (1975), 2. *Accidents/Harmoniques*, (4'14"-4'28")

A la simple écoute, les deux premiers événements se perçoivent aisément comme un phénomène ondulatoire avec un vif changement du timbre sonore.

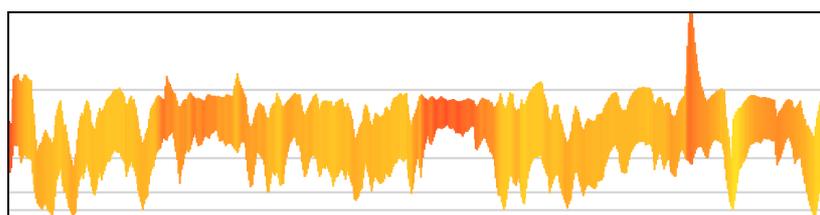


Figure 8 : BStD de l'extrait, 2. *Accidents/Harmoniques*, (4'14"-4'28"), sans suppression du bruit

En suivant l'évolution individuelle des trois descripteurs utilisés (Figure 9) nous constatons que cet extrait est un phénomène physiquement complexe. Même s'il existe une forte corrélation entre les trois descripteurs, chacun décrit un axe paramétrique différent. En les comparant, nous remarquons que la variation d'écart type spectral est proportionnellement plus importante que la variation du centroïde spectral, signifiant le poids de ce paramètre dans l'évolution de ces gestes. Le centroïde varie peu pour les deux premiers gestes. Le pic pour le premier geste de +/- 534 Hz représentant un écart de 343 cents. Pour le deuxième geste le pic est de 456 Hz, soit un écart approximé de 70 cents. Dans les deux cas ces pics sont placés dans les transitoires d'attaque, ce qui signifie qu'ils sont pour ainsi dire cachés. En ce qui concerne la variation et l'ondoiement amorti de l'écart type, celui-ci est beaucoup plus important. L'écart type moyen, entre 4'15'' et le départ du premier geste est de +/- 60 Hz. A 4' 16.769'' l'écart type a un pic de 344.9 Hz, soit une variation de 3028 cents, soit 30 demi-tons. Pour le deuxième geste un pic plus important, de 404.43 Hz, se trouve à 4' 20.09'', soit l'équivalent à 3303 cents. Le troisième événement, Le troisième événement, présente un pic d'écart type de 2451 Hz à 4' 25.529'', caractérisant bien la bande de bruit.

Nous pourrions continuer avec ces descriptions, notamment concernant la BStD, cependant il est aisé de s'apercevoir comment la forme générée épouse visuellement les phénomènes décrits. Ces formes (Figure 7) présentent clairement un aspect d'ondulation amortie, par la variation presque périodique de la dimension de l'écart type spectral et par la variation, également périodique de l'amplitude. Cet extrait est un bon exemple de comment les formes générées par la BStD condensent l'information contenue dans ces trois descripteurs (Figure 9).

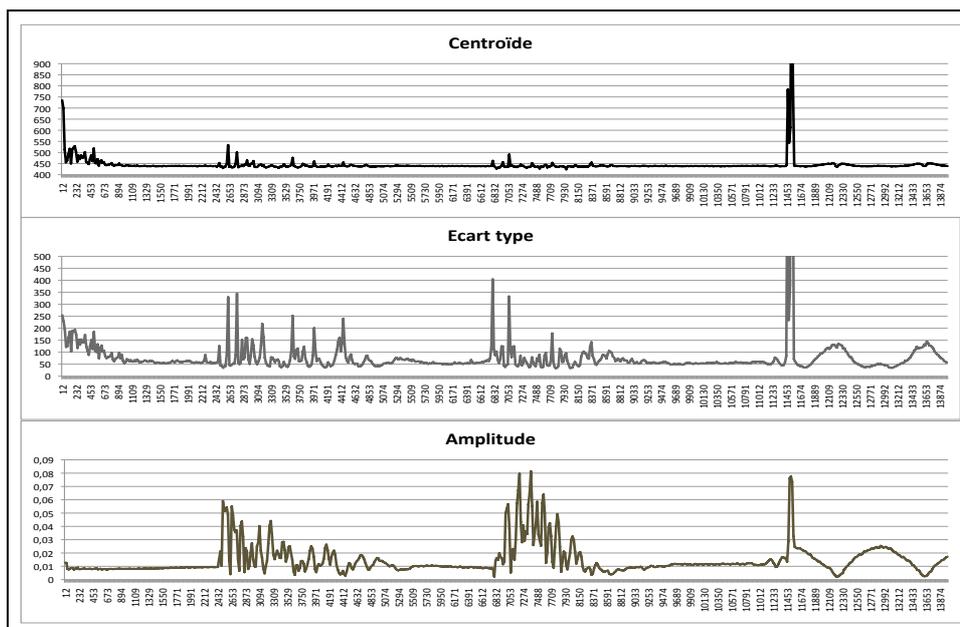


Figure 9 : L'évolution des trois descripteurs pour l'extrait de 2. Accidents/Harmoniques, (4'14"-4'28")

Remarques et conclusions

Le premier problème que nous avons rencontré dans l'utilisation de cette représentation, est en rapport au fait de travailler avec des enregistrements sonores. Cela étant un problème bien connu nous renvoyons le lecteur vers les nombreux ouvrages à ce sujet (notamment le récent ouvrage en ligne de Daniel Leech-Wilkinson, *The Changing Sound of Music: Approaches to Studying Recorded Musical Performance* [Leech-Wilkinson 2009], qui présente une copieuse bibliographie sur le sujet). En plus de la problématique évidente, liée à l'analyse de performances musicales, la représentation que nous proposons est sensible à la prise de son, aux effets de salle et à l'ensemble des traitements de « post-production » (réverbérations, compression, égalisation, etc.) usuels dans la réalisation d'enregistrements du commerce. Par exemple, les effets de salle, notamment la réverbération laissent des « traces » à ne pas confondre avec les formes générées par le phénomène que nous étudions. Ces traces pouvant être différentes entre plusieurs interprétations de la même œuvre. Dans l'implémentation actuelle, nous n'avons pas d'outils pour différencier efficacement entre un signal quelconque et le bruit de fond. Celui-ci devra avoir un niveau assez bas, pour ne pas distordre le résultat analytique, soit la forme générée. Comme corolaire, il est souvent nécessaire de « nettoyer » l'enregistrement avant de l'analyser. La lecture et l'interprétation des formes de la BStD, nécessitent aussi d'un apprentissage.

Cette représentation repose sur des concepts de « couleur » (clair/obscur, brillant/foncé), « d'épaisseur spectrale » (densité) et d'intensité du signal, à partir du calcul de descripteurs statistiques du signal sonore (le centroïde et l'écart type). Comme conséquence immédiate, cette représentation est une approche globale de la surface sonore, soit une description générale de la texture qui ne prend pas en compte les détails de la trame ne géant pas une éventuelle polyphonie.

Un autre point important à prendre en compte est le fait que l'écart type spectral est un descripteur statistique, qui indique la dispersion de l'énergie autour du centroïde spectral, et qui dans certains cas diffère d'une « densité » spectrale perçue. Ce fait corrobore nos remarques sur la nécessité d'une étude sur le rapport entre l'écart type spectral et la perception de densité, dont nous avons signalé le besoin ci-dessus. Dans la suite de cette recherche, nous devrions inclure la valeur du quatrième moment spectral (le Kurtosis) pour intégrer la forme de l'enveloppe spectrale dans le calcul de l'écart type σ' .

Même si cette représentation se fonde sur une simplification paramétrique de la notion de timbre et de texture, le fait de réunir les trois paramètres permet de créer une forme qui synthétise de manière visuelle ces trois descripteurs (la brillance, l'écart type et de l'amplitude), étant un apport significatif d'information par rapport au sonagramme et à la représentation en forme d'onde. La variation, de ces trois paramètres, crée des formes rapidement identifiables, comme nous l'avons montré pour les

variations de couleur et de morphologie dans l'extrait *De Natura Sonorum* (Figure 7) présenté plus haut. Un autre intérêt, commun à l'utilisation des descripteurs acoustiques, et le fait de pouvoir quantifier une partie de la variation de la couleur sonore et le fait que certains phénomènes construisent des formes facilement identifiables.

Finalement, cette représentation a l'avantage de suggérer des points de discontinuités ou similarités, fondés sur les caractéristiques de couleur sonore, devenant ainsi un outil d'aide à la segmentation.

Comme nous l'avions signalé au début de ce texte, la BStD se veut un outil heuristique qui dans le cas de musiques écrites, ou notées, devrait être utilisé en association à d'autres formes d'analyse, à l'évolution d'autres descripteurs (tel que la « pente spectrale », par exemple), à la comparaison à d'autres catégories paramétriques et/ou musicales et n'oubliant pas que la prise en compte de la partition est souvent essentielle, lorsque celle-ci existe.

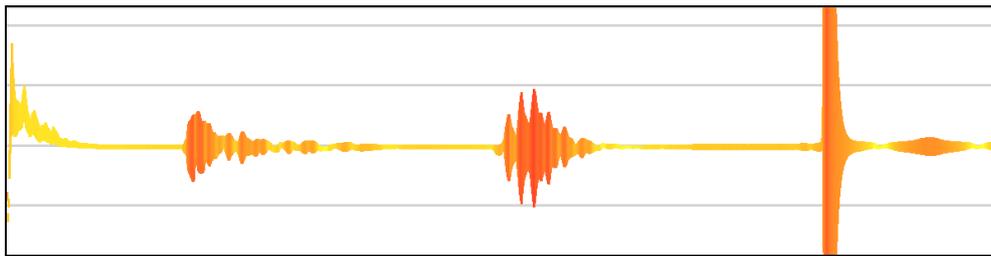


Figure 10 : BStD de l'extrait, 2. *Accidents/Harmoniques*, (4'14"-4'28"). Centroïde et écart type calculés à partir d'une analyse de 24 bandes de Mel.

Cette recherche est un travail en cours. Dans des développements futurs nous aimerions explorer d'autres représentations fondées sur la brillance et l'écart type perceptifs (sur des échelles de Bark ou de Mel), en utilisant la sonie à la place de l'amplitude, et associant la BStD à d'autres descripteurs spectraux (pente spectrale, bandes de Bark et de Mel, flux spectral, rugosité, chromagramme, etc.). La Figure 10 présente un des actuels prototypes en cours, l'analyse du même extrait de Bernard Parmegiani, en utilisant le centroïde et l'écart type calculés à partir d'une analyse de 24 bandes de Mel. La figure est centrée sur la cinquième bande de Mel. Ce travail devrait être réalisé avec une étude en parallèle de la signification « musical » des descripteurs. Hors le contexte physique d'analyse sonore, quel est la corrélation entre les descripteurs et les catégories classiques (harmonie, densité, texture, etc.) de la musique ? Comment pourront-ils être mis en relation avec d'autres théories analytiques, comme par exemple les Unités Sémiotiques Temporelles ? Un autre fait qu'il nous semble pertinent de signaler est la validation perceptive des formes obtenues avec la BStD. Un ensemble d'expériences psychoacoustiques pourraient être développées pour valider nos hypothèses, notamment celle de la perception de l'écart type spectral en fonction de l'intensité du signal.

Références

BULLOCK, JAMIE (2007), « Libxtract: A lightweight library for audio feature extraction », *Proceedings of the ICMC 2007 International Computer Music Conference*, Copenhagen, Danemark.

CABRERA, DENSIL FERGUSON, SAM SCHUBERT, EMERY (2007), « PsySound3: software for acoustical and psychoacoustical analysis of sound recordings », *Proceedings of the 13th International Conference on Auditory Display*, Montreal Canada, p. 356-363.

CARPENTIER, GREGOIRE (2008), *Approche computationnelle de l'orchestration musicale Optimisation multicritère sous contraintes de combinaisons instrumentales dans de grandes banques de sons*, thèse de doctorat, UPMC Paris-6.

CHIASSON, FREDERIC (2007), « L'orchestration selon Koechlin: recherche d'un point de rencontre entre la pratique musicale et les sciences cognitives », in *Actes du Colloque international : « Composer au XXI^{ème} Siècle – Processus et Philosophies »*, OICM, Montréal (Québec) Canada, 28 février – 3 mars.

CRAWFORD, TIM GIBSON, LORNA (2009), *Modern Methods for Musicology: Prospects, Proposals, and Realities*, Ashgate, Great Britain.

DUFOURT, HUGUES (2007), « De la dimension productive de l'intensité et du timbre et leur intégration au système des "éléments porteurs de forme" », *Actes du Colloque international : « Composer au XXI^e Siècle – Processus et Philosophies »*, OICM, Montréal (Québec) Canada, 28 février – 3 mars.

FINK, EUGENE (1996), *Design of Representation-Changing Algorithms*, School of Computer Science, Carnegie Mellon University.

FUTRELLE, JOE DOWNIE, J. STEPHEN (2002), « Interdisciplinary communities and research issues in music information retrieval », *Proceedings of the Third International Conference on Music Information Retrieval*, Paris, France, p. 215–221.

GREY, JOHN (1975), *An exploration of musical timbre*, thèse de doctorat, rapport STAN-M-2, Dept. of Music, Stanford University. <https://ccrma.stanford.edu/STANM/stanms/stanm2/stanm2.pdf> (vérifié le 02/06/2010).

GREY, JOHN (1977), « Multidimensional perceptual scaling of musical timbres », *J. Acoust. Soc. Am.*, 61(5), p. 1270-1277.

HANNA, PIERRE (2003), *Modélisation statistique de sons bruités : étude de la densité spectrale, analyse, transformation musicale et synthèse*, Thèse pour obtenir le grade de docteur spécialité : informatique, soutenue le 12 Décembre 2003, Université Bordeaux I, Ecole doctorale de mathématiques et d'informatique.

HARTMANN, WILLIAM MORRIS PUMPLIN, JON (1988), « Noise power fluctuations and the masking of sine signals », *Journal of Acoustical Society of America*, 83(6), p. 2277–2289.

HARTMANN, WILLIAM MORRIS PUMPLIN, JON (1991), « Periodic signals with minimal power fluctuations », *Journal of Acoustical Society of America*, 90(4), p. 1986–1999.

HERRERA, PERFECTO SERRA, XAVIER PEETERS, GEOFFROY (1999), « Audio Descriptors and Descriptors Schemes in the Context of MPEG-7 », *International Computer Music Conference - ICMC99*, Beijing (China).

JEHAN, TRISTAN (2005), *Creating Music by Listening*, PhD Thesis in Media Arts and Sciences, Massachusetts Institute of Technology.

JEHAN, TRISTAN (2010), *Msp Externals*, <http://web.media.mit.edu/~tristan/maxmsp.html> (09/06/2010).

KASSLER, MICHAEL (1966), Toward Musical Information Retrieval, *Perspectives of New Music*, vol. 4 (2), p. 59-67.

KOECHLIN, CHARLES (1943), *Traité de l'Orchestration*, Max Eschig, Paris.

LARTILLOT, OLIVIER (2007), *MIRtoolbox* tutorial, ISMIR 2007, Viena, Autriche.

LEECH-WILKINSON, DANIEL (2009), *The Changing Sound of Music: Approaches to Studying Recorded Musical Performance*, London, CHARM. <http://charm.rhul.ac.uk/studies/chapters/intro.html> (02/06/2010).

LIGETI, GYORGY (1961), *Atmospheres*, Universal Edition, UE11418, Vienna.

LIVSHIN, ARIE A. RODET, XAVIER (2004), « Musical instrument identification in continuous recordings », *Proc. of the 7th Int. Conference on Digital Audio Effects (DAFX-04)*, Naples, Italy, October 5-8, p. 222-226.

MALT, MIKHAIL JOURDAN, EMMANUEL (2008), « Zsa.Descriptors: a library for real-time descriptors analysis », *SMC08 - 5th Sound and Music Computing Conference*, Berlin, Allemagne, 31 Juillet au 3 Août 2008, p. 134-137.

MALT, MIKHAIL JOURDAN, EMMANUEL (2009), « Real-Time Uses of Low Level Sound Descriptors as Event Detection Functions Using the Max/MSP Zsa.Descriptors Library », *SBCM2009 - 12th Brazilian Symposium on Computer Music*, September 7-9, Recife, Brasil.

MCADAMS, STEPHEN WINSBERG, SUZANNE DONNADIEU, SOPHIE DE SOETE, GEERT KRIMPHOFF, JOCHEN (1995), « Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes », *Psychological Research*, vol. 58, 1995, pp. 177-192.

MEEUS, NICOLAS (1994), « De la forme musicale et de sa segmentation », *Musurgia* vol. 1, pp. 7-23.

MESNAGE, MARCEL (1994), « Techniques de segmentation automatique en analyse Musicale », *Musurgia*, vol. 1 n°1, pp. 39-49.

MOELANTS, DIRK CORNELIS, OLMO LEMAN, MARC GANSEMANS, JOS DE CALUWE, RITA DE TRE, GUY MATTHE, TOM HALLEZ, AXEL (2006), « Problems and Opportunities of Applying Data- & Audio-Mining Techniques to Ethnic Music », *ISMIR 2006 7th International Conference on Music Information Retrieval Proceedings*, Victoria, Canada, p. 1-3.

PEETERS, GEOFFROY (2003), *A Large Set of Audio Features for Sound Description*, rapport technique, Ircam, Paris, France.

PUCKETTE, MILLER S. APEL, THEODORE ZICARELLI, DAVID D. (1998), « Real-time audio analysis tools for Pd and MSP », *Proceedings of the International Computer Music Conference*, p. 109-112.

SCHNELL, NORBERT SCHWARZ, DIEMO (2005), "Gabor, multi-representation real-time analysis/synthesis", *Proc. of the 8th Int. Conference on Digital Audio Effects (DAFx'05)*, Madrid, Spain, September 20-22.

SCHUBERT, EMERY WOLFE, JOE (2006), « Does timbral brightness scale with frequency and spectral centroid? », *Acta Acustica united with Acustica*, vol. 92 (5), p. 820-825.

SLANEY, MALCOLM (1998), *Auditory Toolbox V. 2.0*, <http://rvl4.ecn.purdue.edu/%7Emalcolm/interval/1998-010/> (09/06/2010).

TAGG, PHILIP (2010), *Music's Meanings*, Book in progress, p. 1-190. <http://www.tagg.org/bookxtrax/NonMuso/NonMuso.pdf> (27/05/2010).

TARDIEU, DAMIEN, RODET, XAVIER (2007), « An Instrument Timbre Model For Computer Aided Orchestration », *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, WASPAA*, New Paltz, New York.

THOMAS, JEAN-CHRISTOPHE, NATTIEZ, JEAN-JACQUES, MION, PHILIPPE (1983), *L'Envers d'une œuvre, "De natura sonorum" de Bernard Parmegiani*, Buchet/Chastel, Institut National de la Communication Audiovisuelle, Paris, France

TODOROFF, TODOR, DAUBRESSE, ERIC, FINEBERG, JOSHUA (1995), « Iana~ (a real-time environment for analysis and extraction of frequency components of complex orchestral sounds and its application with a musical realization) », *Proceedings of the International Computer Music Conference*, Banff, Canada.

WESSEL, DAVID L. (1979), « Timbre space as a musical control structure », *Computer Music Journal*, 3(2), p. 45-52.

UITDENBOGERD, ALEXANDRA L., CHATTARAJ, ABHIJIT, ZOBEL, JUSTIN (2000), « Music IR: Past, Present and Future », *International Symposium on Music Information Retrieval Proceedings – ISMIR 2000*.