

Caractérisation du timbre des sons complexes. II. Analyses acoustiques et quantification psychophysique

J. KRIMPHOFF^{*,**}, S. McADAMS^{*,***} et S. WINSBERG^{*}

^{*} IRCAM, 1 place Stravinsky, 75004 Paris, France

^{**} Laboratoire d'Acoustique de l'Université du Maine, LAUM/CNRS, U.F.R. Sciences, Av. Olivier Messiaen, BP. 535, 72017 Le Mans cedex, France

^{***} Laboratoire de Psychologie Expérimentale, CNRS, Université René Descartes Paris V, 28 rue Serpente, 75006 Paris, France

Abstract: A timbre space represents the mental organization of sound events at equal pitch, loudness, and duration. The geometric distance between two timbres corresponds to their degree of perceived dissimilarity. The dimensions of such a three-dimensional space, established by Krumhansl [1] for a set of 21 synthesized sounds, were investigated with respect to their acoustic characteristics. Several acoustical parameters based on the temporal and frequency properties of the sounds were calculated. The high degree of correlation of several parameters with the perceptual axes lend support to previous interpretations of the qualitative character of two perceptual dimensions and their semantic attributes. The perceptual dimensions “*brightness*” and “*rapidity of attack*” turn out to be quantitatively explainable by the center of gravity of the sound spectrum (CGS) and the rise time on a logarithmic scale (LTM), respectively. The third dimension, initially called “*spectral flux*” corresponds partially to the standard deviation of the time-averaged harmonic amplitudes from a spectral envelope (IRR). A new verbal descriptor, “*spectral fine structure*” seems to fit better with the results of acoustic analyses.

1. INTRODUCTION

Il est généralement admis que le timbre est un attribut multidimensionnel. Plusieurs auteurs (par exemple [1]-[3]) ont étudié les différentes dimensions du timbre à l'aide des techniques multidimensionnelles qui, à partir de jugements de dissemblance entre paires de timbres, établissent une représentation spatiale des relations perceptives entre les timbres. En faisant correspondre les dissemblances subjectives aux distances métriques, on représente les timbres par des points dans un “espace des timbres”. Les dimensions d'un tel espace peuvent être interprétées comme les dimensions perceptives du timbre. Des interprétations qualitatives de ces dimensions en termes d'attributs verbaux comme la “*brillance*” et la “*qualité d'attaque*” sont communes à la plupart des études.

En revanche, une interprétation quantitative des dimensions perceptives en termes de grandeurs acoustiques objectives des timbres utilisés n'a pu être trouvée que pour une seule dimension perceptive, la “*brillance*”, qui est fortement corrélée avec le centre de gravité du spectre [3].

2. ANALYSES MULTIDIMENSIONNELLES ET DIMENSIONS PERCEPTIVES DU TIMBRE

Un espace de timbres à trois dimensions a été établi pour un ensemble de 21 timbres synthétiques. Il s'agit de timbres issus d'une synthèse par modulation de fréquence imitant des instruments de musique ou leurs hybrides (par exemple: le timbre “piano frotté” réuni les caractéristiques perceptives de deux timbres: “piano” et “cordes frottées”). Les dissemblances subjectives entre les différents timbres, présentés par

paires à une hauteur commune (mi^3) correspondant à une fréquence fondamentale d'environ 311 Hz, ont été reliées à des distances métriques au sein d'un espace quasi-euclidien [4]. Les dimensions de cet espace étaient interprétées de manière qualitative selon leurs caractéristiques acoustiques. Une première dimension a été interprétée comme le degré de "brillance" des timbres et une deuxième, comme la "rapidité d'attaque". Il était plus difficile d'associer un seul attribut verbal à la troisième et dernière dimension, puisqu'elle semblait représenter les variations du contenu spectral au cours du temps. Cette dimension, étant liée à un critère spectro-temporels, a été nommée "flux spectral" par Krumhansl [1].

3. ANALYSES ACOUSTIQUES ET PARAMETRES ACOUSTIQUES DU TIMBRE

Les timbres ont été analysés par vocodeur de phase. Puisque les sons présentent une structure harmonique régulière et une faible variation de la fréquence fondamentale (<3% = 1/4 de ton), une technique synchronisée à la hauteur (Pitch-Synchronous Phase-Vocoder Method, [5]) qui néglige la partie non-harmonique du signal, semblait appropriée. Cette technique extrait, par transformée de Fourier glissante à court-terme, les amplitudes $A_k(t)$, les fréquences $f_k(t)=kf_1+\Delta f_k(t)$ et les phases θ_k des n harmoniques (k=1 correspond à la fondamentale du son). A l'intérieur d'une fenêtre d'analyse de longueur constante, le signal s(t) est analysé par Transformée de Fourier Rapide. Pour obtenir une bonne restitution des variations rapides du signal s(t) lors de la resynthèse du signal $s_n(t)$ (équation 1), une longueur de fenêtre de deux périodes ($2 / f_1 = 2 / 311 \text{ Hz} = 6.4 \text{ ms}$) a été choisie et le pas de déplacement de la fenêtre était d'une demi-période.

$$s_n(t) = \sum_{k=1}^n A_k(t) \cos(2\pi \int (kf_a + \Delta f_k(t)) dt + \theta_k) \tag{1}$$

Bien qu'une telle synthèse ne prenne en compte que la partie harmonique du signal, aucune différence audible entre le signal original et sa synthèse avec n = 30 harmoniques, correspondant à une fréquence maximale effective de 9330 Hz, ne pouvait être constatée pour les sons analysés.

3.1 La "brillance" - Centre de gravité spectrale (CGS)

Afin de pouvoir expliquer la dimension perceptive "brillance" de manière quantitative, le centre de gravité spectrale (CGS) a été calculé de différentes manières à partir des résultats des analyses par vocodeur de phase. Trois modèles extraient un spectre d'amplitudes globales A_k , en prenant la valeur maximale, la moyenne ou la moyenne quadratique des amplitudes $A_k(t)$, pour ensuite calculer le centre de gravité spectrale (équation 2).

$$CGS = \frac{\sum_{k=1}^N kA_k}{\sum_{k=1}^N A_k} \tag{2}$$

(Centre de gravité du spectre)

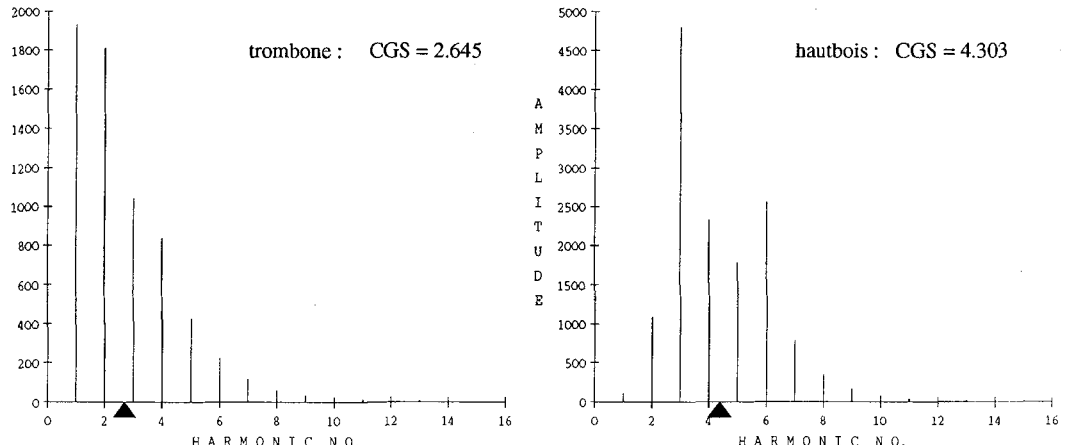


FIGURE 1. Le spectre moyen pour les timbres trombone (gauche) et hautbois (droite) et leurs CGS respectifs (triangles)

Par exemple, le CGS prend la valeur 1 pour un son pur et $CGS=n/2$ pour un son de n harmoniques à amplitudes égales (voir Fig. 1). Les trois modèles présentent des corrélations significatives avec les coordonnées des timbres sur la dimension perceptive "brillance", mais un autre modèle s'est avéré plus performant. Au lieu de calculer le CGS à partir des amplitudes globales A_k , on détermine d'abord le CGS(t) à l'intérieur de chaque fenêtre d'analyse pour ensuite prendre sa moyenne temporelle sur toute la durée du son. Les corrélations avec la dimension "brillance" permettent d'expliquer 88% de la variance ($r=0.9359$, $dl=20$, $p<0.01$) et le paramètre reste parfaitement orthogonal par rapport aux deux autres dimensions perceptives ($r<0.0738$).

3.2 La "rapidité d'attaque" - Logarithme du temps de montée (LTM)

La seconde dimension perceptive, appelée "rapidité d'attaque" semble être liée aux aspects temporels des sons. Afin de les regrouper, quatre enveloppes temporelles ont été calculées selon différentes manières: en prenant la valeur absolue maximale, la moyenne quadratique sur une période, ou en se servant des résultats de l'analyse par vocodeur et en faisant la somme linéaire ou la racine de la somme quadratique de toutes les n amplitudes des harmoniques $A_k(t)$. Pour ces enveloppes, les temps de montée ont été évalués. Le temps de montée est défini comme l'intervalle de temps entre le début de la perception du son t_{seuil} et l'instant t_{max} quand l'enveloppe atteint sa valeur maximale (Fig. 2). Le début de perception d'un son est défini comme l'instant où l'enveloppe passe le seuil de perception. Ce seuil n'est pas le seuil d'audition absolu comme on aurait tendance à l'admettre, mais il dépend linéairement de l'amplitude maximale [6]. Plusieurs seuils (de -2dB jusqu'à -23dB de l'amplitude maximale) ont été testés pour les 8 enveloppes différentes. Une telle définition d'un seuil relatif à l'amplitude maximale a en même temps l'avantage de rendre la définition du temps de montée indépendant du niveau global. On calcule le logarithme du temps de montée (LTM, équation 3).

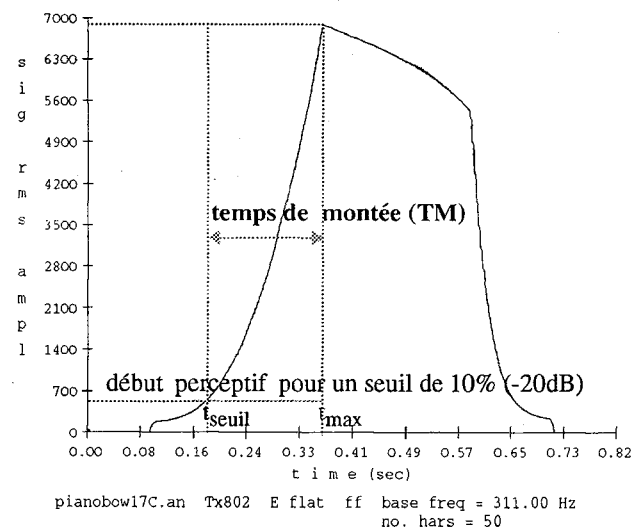


FIGURE 2. La définition du temps de montée (TM) illustrée pour le timbre hybride "piano frotté".

$$LTM = \log_{10} (t_{max} - t_{seuil}) \quad (3)$$

Le LTM est fortement corrélé avec la deuxième dimension perceptive, interprétée comme "enveloppe temporelle" et explique 89% de la variance ($r=-0.9443$, $dl=20$, $p<0.01$), alors que les corrélations avec les autres dimensions perceptives restent très faibles ($r<0.0321$).

3.3 La "structure fine du spectre" - Irrégularité par rapport à une enveloppe globale

Un dernier paramètre acoustique semble expliquer la troisième dimension perceptive. En désaccord avec l'interprétation initiale comme "flux spectral", il s'agit d'un paramètre purement spectral puisque les paramètres testés qui mesuraient les fluctuations temporelles dans le spectre ne permettaient d'expliquer que 34% de la variance.

Par contre, un paramètre qui a été conçu afin de décrire le degré d'irrégularité du spectre (IRR, équation 4) présente des corrélations fortement significatives ($r=-0.8539$, $dl=20$, $p<0.01$) avec la troisième dimension perceptive et permet d'expliquer 73% de la variance. Ce paramètre mesure les écarts des amplitudes des harmoniques par rapport à une enveloppe spectrale globale (voir Fig. 3). L'enveloppe spectrale globale se calcule en prenant la moyenne de trois raies d'amplitude A_k adjacentes en décibel (déterminées de la même façon comme dans §3.1). Ce paramètre reste également indépendant des deux autres axes perceptifs ($r<0.1708$).

$$IRR = \log_{10} \left(\sum_{k=2}^{n-1} \left| 20 \log_{10}(A_k) - \frac{20 \log_{10}(A_{k+1}) + 20 \log_{10}(A_k) + 20 \log_{10}(A_{k-1})}{3} \right| \right) \quad (\text{Irrégularité du spectre}) \quad (4)$$

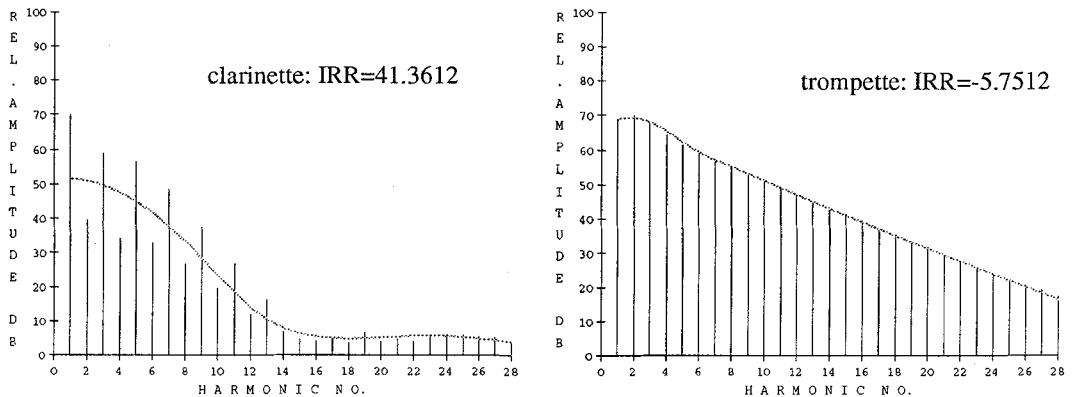


FIGURE 3. Spectre de raies des sons clarinette et trompette. La ligne continue en gris représente une enveloppe, calculée en prenant la moyenne de trois raies adjacentes. Le paramètre IRR représente le logarithme de la somme des différences entre le spectre de raies et l'enveloppe.

4. CONCLUSION

Il s'agissait d'établir les paramètres acoustiques qui sous-tendent les dimensions perceptives d'un espace de 21 timbres synthétiques. Pour chacune des trois dimensions, des paramètres acoustiques ont pu être dégagés qui ne sont corrélés qu'à cette dimension perceptive et à aucune des deux autres. Ces paramètres acoustiques - choisis physiquement afin d'être orthogonaux dès le départ - le restent donc perceptivement et permettent d'associer à chaque dimension perceptive un paramètre physique.

Il reste à étudier si on peut généraliser les paramètres dégagés à des classes de timbres plus larges comme des timbres d'instruments de musique réels, des sons percussifs ou des bruits. Finalement, ceci peut être profitable pour de futures recherches sur la modélisation de la perception des sons complexes. De plus, une application de nouveaux algorithmes d'analyse multidimensionnelle est maintenant envisageable en se servant des axes physiques comme contraintes afin d'ajuster les dimensions perceptives et d'améliorer la validité des représentations multidimensionnelles.

5. BIBLIOGRAPHIE

- [1] Krumhansl, C. L., "Why is musical timbre so hard to understand ?", in *Structure and perception of electroacoustic sound and music*, (eds.S.Nielzen & O.Olsson, Elsevier, Amsterdam, excerpta Medica 846, 1989), pp. 43-53.
- [2] Grey, J.M., "Multidimensional perceptual scaling of musical timbres", *J.Acoust. Soc. Am.* **61** (5) (1977) pp. 1270-1277.
- [3] Grey, J.M. & Gordon, J.W., "Perceptual effects of spectral modifications on musical timbres", *J.Acoust. Soc. Am.* **63** (5) (1978) pp. 1493-1500.
- [4] Winsberg, S. & Carroll J.D., "A Quasi-Nonmetric Method for multidimensional scaling via an extended Euclidean model.", *Psychometrika*, **54** (1989) pp. 217-229.
- [5] Beauchamp, J.W., "Unix Workstation Software for Analysis, Graphics, Modifications, and Synthesis of Musical Sounds," preprint 3479 of 94th Convention AES in Berlin, Germany (1993).
- [6] Vos, J. & Rasch, R. . "The perceptual onset of musical tones", *Perception & Psychophysics*, **29** (4) (1981) pp. 323-335.