

10ème Congrès Français d'Acoustique

Lyon, 12-16 Avril 2010

Rôle du contrôle et du timbre dans la perception du naturel de sons de clarinette

Philippe Guillemain¹ et Adrien Merer^{1,2}

¹CNRS - Laboratoire de mécanique et d'acoustique, 31 chemin Joseph-Aiguier 13402 Marseille Cedex 20

² Aix-Marseille Universités

{guillemain,merer}@lma.cnrs-mrs.fr

La conception d'un modèle de synthèse de sons d'instruments de musique passe par l'évaluation des sons obtenus. Cette évaluation fait appel à des critères subjectifs comme le "réalisme" ou le "naturel" et implique de comparer ces sons de synthèse avec des sons produits par un instrumentiste, sur instrument "réel". Pour mieux comprendre ces notions subjectives, en vue de l'amélioration perceptive du modèle, il est donc nécessaire de prendre en compte tous les aspects qui différencient ces sons depuis leur production jusqu'à leur restitution. Dans l'étude que nous présentons ici, il s'agit de déterminer l'importance respective des propriétés "instrument" et "contrôle" dans la perception du naturel des sons issus d'un modèle physique d'instruments à vent. Quatre modalités sont étudiées, combinant sons acoustiques produits par une bouche artificielle et sons produits par synthèse, joués soit par un musicien soit par une commande robotisée. Les conclusions de cette étude sont basées sur un protocole de comparaison par paires où il est demandé aux auditeurs de juger sur une échelle continue, lequel des deux sons qui lui sont présentés, est le plus naturel. La première observation est que le contrôle, corrélé acoustiquement aux micro-variations de pression d'alimentation, influe d'avantage la perception du naturel des sons, que l'instrument. Ce résultat est à mettre en parallèle avec des expériences conduites dans le domaine de la parole dont les résultats démontrent que l'accentuation portée sur les mots par le locuteur est un facteur considérable dans la perception du naturel. On montre aussi que, dans une moindre mesure, l'instrument influe sur le naturel par le biais de la fréquence de coupure du spectre, confirmant ainsi les résultats d'autres études montrant que la perception du naturel des sons semble liée à la fréquence de coupure des enceintes avec lesquelles ils sont écoutés.

1 Introduction

L'évaluation de la qualité sonore d'un instrument de synthèse passe, lorsque celui-ci est dédié à la reproduction d'instruments acoustiques existants, par une évaluation non seulement de qualités purement timbrales mais aussi de qualités reliées aux possibilités d'expressivité musicale offertes par l'instrument de synthèse. Il s'agit alors de qualités reliées non plus au timbre mais d'avantage à ses variations dynamiques par rapport au contrôle exercé sur l'instrument. Ces qualités tant timbrales qu'expressives peuvent se regrouper dans la notion subjective de "naturel" qui vise à quantifier si un dispositif de synthèse peut tromper le cerveau humain. Cette notion a été abondamment développée dans le contexte de la synthèse de la parole mais très peu d'études s'intéressent à la musique ([1, 2]). Tout comme la parole, de part la très grande palette de nuances et l'expressivité qu'ils offrent, les instruments à son entretenu constituent un terrain d'investigation particulièrement intéressant. Dans cette étude, nous nous intéressons aux rôles joués par les modalités timbre et contrôle dans l'évaluation subjective du naturel de sons de type clarinette produits par un instrument acoustique et un instrument de synthèse reposant sur un modèle physique simplifié du fonctionnement de l'instrument.

Après avoir brièvement présenté le modèle utilisé et justifié son utilisation dans le contexte de l'étude, nous présentons les hypothèses et la principale question posée : d'un son synthétique contrôlé par un musicien ou d'un son acoustique contrôlé par une commande robotisée, lequel sera perçu comme plus naturel ? En découle la présentation du choix des stimuli et de la méthode de jugement utilisée. Finalement les résultats d'une analyse multidimensionnelle et leurs corrélats acoustiques sont présentés.

2 Comparaison entre sons acoustiques et sons de synthèse

Sont présentés ici brièvement le modèle physique utilisé et la cohérence de l'évolution de descripteurs de timbre en fonction de la pression d'alimentation entre les sons naturels et les sons de synthèse, étape préliminaire permettant d'évaluer objectivement le comportement du modèle.

2.1 Modèle physique

Le modèle physique utilisé pour générer les sons de synthèse repose sur trois éléments couplés. Le pre-

mier est caractéristique de la perce de l'instrument, le deuxième l'ouverture du canal d'anche et le troisième le couplage non linéaire existant entre les deux premiers. Selon [3], des variables adimensionnées sont utilisées.

Les variables internes au modèle, $p_e(t)$, $u_e(t)$ et $x(t)$ représentent respectivement la pression acoustique, le débit acoustique et le déplacement de l'anche. Leurs versions en lettres majuscules représentent leurs transformées de Fourier. Les paramètres de contrôle $\gamma(t)$ et $\zeta(t)$ sont reliés respectivement à la pression d'alimentation et à la pression exercée par les lèvres sur l'anche.

2.1.1 Impédance du résonateur

On considère un tuyau cylindrique de longueur L . Son impédance d'entrée, reliant côté fréquentiel pression et débit acoustiques s'exprime :

$$Z_e(\omega) = \frac{P_e(\omega)}{U_e(\omega)} = i \tan(k(\omega)L) \quad (1)$$

où le nombre d'onde $k(\omega)$ tient compte des pertes viscothermiques. Cette impédance présente des pics à des fréquences quasiment multiples impaires de la fondamentale.

2.1.2 Déplacement de l'anche

Le modèle classique exprimant le déplacement $x(t)$ de l'anche par rapport à sa position d'équilibre en fonction d'une sollicitation de pression $p_e(t)$ utilisé ici est :

$$\frac{1}{\omega_r^2} \frac{d^2 x(t)}{dt^2} + \frac{q_r}{\omega_r} \frac{dx(t)}{dt} + x(t) = p_e(t) \quad (2)$$

où $\omega_r = 2\pi f_r$ et $1/q_r$ sont respectivement la pulsation et le facteur de qualité de la première fréquence de résonance de l'anche.

2.1.3 Caractéristique non linéaire

Le modèle classique, unidimensionnel, repose sur l'équation de Bernoulli stationnaire et relie dans le domaine temporel le débit acoustique (produit entre la vitesse du jet et la surface d'ouverture) à la différence de pression entre la bouche du musicien et le bec de l'instrument. La surface d'ouverture s'exprime :

$$S(t) = \zeta(t)\Theta(1 - \gamma(t) + x(t))(1 - \gamma(t) + x(t))$$

où Θ est la fonction de Heaviside. Le paramètre sans dimension $\zeta(t)$ est proportionnel à la racine carrée de la hauteur du canal d'anche au repos. Le paramètre sans dimension $\gamma(t)$ est proportionnel à la pression d'alimentation.

À partir de la surface d'ouverture $S(t)$, en supposant un écoulement de Bernoulli, le débit acoustique s'exprime sous la forme :

$$u_e(t) = S(t)\text{sign}(\gamma(t) - p_e(t))\sqrt{|\gamma(t) - p_e(t)|} \quad (3)$$

2.1.4 Schéma numérique pour le modèle complet

Le système d'équations constitué de l'impédance d'entrée du résonateur, du déplacement de l'anche et

de la caractéristique non linéaire est résolu dans le domaine des signaux échantillonnés à partir du schéma numérique détaillé dans [4]. En supposant un rayonnement monopolaire, la pression externe est calculée par l'approximation : $p_{ext}(t) = \frac{d}{dt}(p_e(t) + u_e(t))$

2.2 Comportement du modèle

Nous étudions à présent la cohérence de la variation de deux descripteurs de timbre entre les sons issus du modèle et ceux issus de l'instrument acoustique. Les descripteurs considérés s'intéressent au timbre et sont le barycentre du spectre, classiquement utilisé dans la caractérisation du lien entre sonie et brillance et le rapport énergétique entre harmoniques impairs et pairs, descripteur plus spécifique à la clarinette [5].

Afin d'actionner les deux dispositifs à l'aide de la même commande, une bouche artificielle permettant simultanément de piloter la pression d'alimentation et d'assurer un appui constant sur l'anche est utilisée [6]. La mesure fait appel à une excitation composée d'une lente rampe linéaire de pression fenêtrée à l'aide d'une porte rectangulaire permettant d'une part de balayer l'ensemble des nuances de jeu de l'instrument en régime quasi-permanent et d'autre part de simuler des transitoires d'attaque et d'extinction.

La figure 1(a) présente, pour l'instrument acoustique, respectivement de haut en bas, la pression externe, le barycentre du spectre et le rapport impair/pair en fonction du temps. La figure 1(b) présente les mêmes grandeurs pour l'instrument de synthèse.

La justification physique des variations observées sort du cadre de cet article, mais il est important de noter des comportements similaires entre son acoustique et son de synthèse, avec toutefois une plus faible amplitude des variations et des valeurs moyennes pour le son de synthèse.

Cet exemple montre la cohérence dans l'évolution de descripteurs de timbre en fonction du contrôle, tant en régime transitoire qu'en régime permanent.

2.3 Hypothèses

Nous posons comme hypothèse que le son produit par une clarinette jouée par un musicien sera jugé plus naturel que celui produit par une commande robotisée sur le modèle de synthèse. En revanche nous n'avons pas d'hypothèse sur le naturel d'un son produit par une commande robotisée sur la clarinette par rapport à un son obtenu par un musicien sur le modèle de synthèse. En plus de la comparaison entre l'influence du "contrôle" par rapport à celle de "l'instrument", on cherche à évaluer la stabilité des résultats obtenus sous l'influence de deux autres paramètres pour deux états différents :

- Le profil de pression
 - + Note courte (transitoire d'attaque)
 - + Crescendo

Suivant le type de note, on met à l'épreuve différents paramètres du modèle. Les crescendi donnent à priori plus d'importance aux aspects dynamiques du timbre et permettent de juger de la cohérence du lien entre sonie et brillance. Les

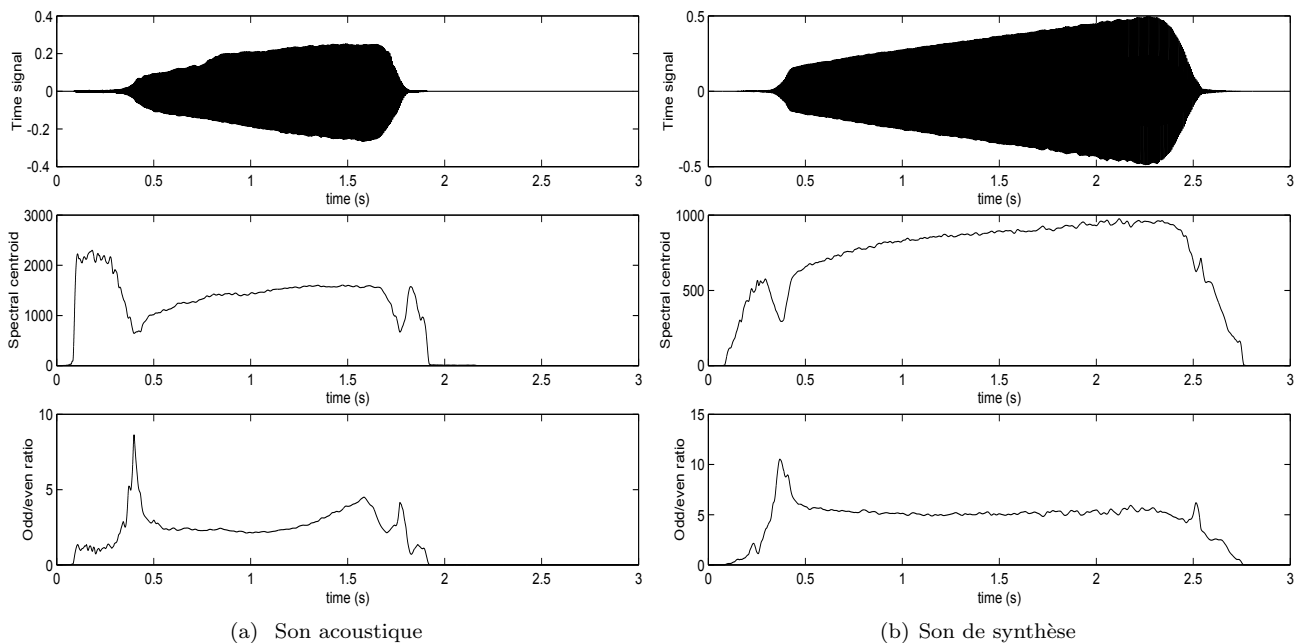


FIGURE 1 – Pression externe et descripteurs de timbre (Barycentre du spectre et rapport impair/pair) dans les cas d’un son de clarinette acoustique (a) et synthétique (b).

notes courtes permettent d’évaluer l’influence perceptive du transitoire d’attaque.

- Le registre
- + Grave
- + Aigu

Le modèle physique est basé sur des hypothèses qui simplifient le comportement de la clarinette dans le second registre : on teste donc à la fois la pertinence de ces simplifications et le rôle du nombre de composantes spectrales (une note grave contient plus de composantes qu’une note aigu).

Le corpus qui est testé est donc composé de 8 sons pour chaque registre avec toutes les combinaisons ”Instrument” / ”Contrôle” / ”Profil de pression” possibles.

3 Protocole

3.1 Stimuli

Afin de s’affranchir des problèmes de directivité de l’instrument acoustique et pour avoir une réverbération naturelle, les stimuli ont été enregistrés dans la même pièce (cf. figure 2) à un mètre à l’aide d’un micro cardioïde Neumann KM184 via une interface audio Motu UltraLite-mk3 (échantillonnage 44100 Hz sur 16 bits). Les profils de pression joués par le musicien ont été enregistrés à l’aide d’un contrôleur Yamaha wx5. Les versions ”robotisées” ont été obtenues en supprimant les irrégularités locales de ces profils ”musiciens” tout en conservant l’évolution globale. Le modèle de synthèse a été paramétré par l’expérimentateur de façon à produire des sons les plus similaires possibles de ceux de la clarinette (clarinette Sib résine Yamaha ycl250). Les sons de synthèse sont restitués via la carte son, sur une enceinte Tannoy Precision 8 (ampli Alesis RA300), le niveau de diffusion correspondant à celui de la clarinette au niveau du microphone (égalisation du niveau effectuée par

l’expérimentateur).

Les stimuli ont été présentés au casque (Stax 3R202 et préampli SRM310) en diotique, en utilisant la carte son interne de l’ordinateur (Imac G4) dans une cabine audiométrique. Le niveau d’écoute était libre mais fixe durant tout le test (réglage effectué durant la phase d’entraînement).

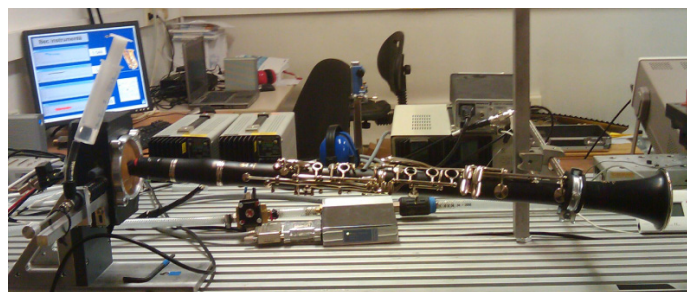


FIGURE 2 – Dispositif utilisé pour générer les stimuli.

Utilisation d’une bouche artificielle et d’un haut-parleur pour restituer les sons issus du modèle de synthèse.

3.2 Sujets

19 sujets (3 femmes, 16 hommes) ont participé au test, moyenne d’âge 26 ans (min 20 ans max 48). 15 des sujets jouent d’un instrument régulièrement (trois sujets jouent de la clarinette non-professionnellement).

3.3 Méthode

Le jugement du naturel est obtenu par un test de comparaison par paire (chaque stimuli est évalué en regard de chacun des autres) sur une échelle continue non graduée. Les auditeurs doivent juger lequel des 2 sons est

le plus naturel et à quel point il est plus naturel (l'échelle va de "bien plus naturel" à "bien moins naturel") sachant qu'ils n'étaient pas informés de la provenance des sons. Les consignes ne donnent pas de définition du naturel, mais en revanche, il est précisé une manière de considérer le naturel par son utilisation en dans le contexte d'une phrase tirée de [7] "J'oubliais que c'était un personnage en images de synthèse, je le trouvais NATUREL"

L'ordre de présentation des paires et l'ordre des sons au sein d'une paire est aléatoire (chaque paire n'est présentée qu'une fois). Les sujets peuvent réécouter chaque paire autant de fois qu'ils le souhaitent (le nombre de réécoutes étant enregistré).

En raison de la durée du test et d'une éventuelle baisse d'attention, le test est séparé en deux parties dont une avec les notes jouées dans le premier registre (Fa4) et l'autre avec les notes du second registre (Do6), la moitié des sujets ayant commencé par le registre aigu.

4 Résultats

4.1 Classement des différentes modalités

Dans ce qui suit, on notera *C* (clarinette) et *S* (synthèse) les instruments, *M* (musicien) et *R* (robot) le contrôle ainsi que *Att* (attaque) et *Cre* (crescendo) pour le profil dynamique de pression d'alimentation. Pour chaque sujet, on obtient une "note" de naturel de chaque son sur une échelle de -7 à 7. Un son ayant obtenu une note de 7 indique qu'il a été jugé *beaucoup* plus naturel que tous les autres sons. *beaucoup* signifie que l'auditeur à utilisé à chaque fois la position extrême du curseur.

4.1.1 Registre aigu

La figure 3 regroupe les résultats bruts obtenus pour chaque sujet et chaque son.

L'analyse de variance à 3 facteurs indique que les différences de jugement sont significatives pour le "contrôle" ($p < 0,01$) ainsi que pour l'interaction entre "contrôle" et "instrument". Les sons *S_M_* ont donc été jugés plus naturels que les sons *C_R_*, l'interaction entre les facteurs "instruments" et "contrôle" illustre quand à elle, la nette préférence des sons *C_M_* par rapports aux sons *S_R_*. A noter que pour le facteur "instrument", on obtient $p = 0,018$ ce qui n'est pas significatif au seuil de 1% mais incite à approfondir l'expérience.

4.1.2 Registre grave

Les résultats obtenus pour le registre grave (cf figure 4) sont beaucoup moins différenciés dans le sens où les moyennes sont beaucoup plus proches que pour le registre aigu. Seul le "facteur" instrument est significatif ($p < 0.01$) ce qui indique que globalement, les sons issus de la clarinette ont été jugés plus naturel que ceux issus du modèle de synthèse.

Il est important de noter que la variabilité inter-individus est relativement importante avec pour certains

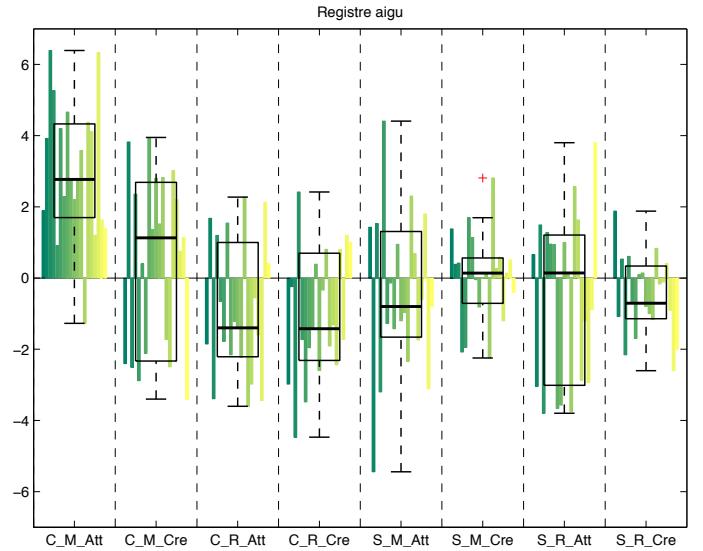


FIGURE 3 – Résultats obtenus pour le registre aigu.

Les barres verticales correspondent aux réponses de chaque sujet (une valeur positive indique que le son a été jugé plus naturel que la majorité des autres sons).

En noir : diagrammes "boîtes à moustaches" (le trait épais correspond à la médiane)

sons, des distributions de réponses qui s'éloignent de la normalité de façon non-négligeable.

4.2 Analyse multidimensionnelle

En considérant les données collectées comme étant le reflet de dissemblances entre les stimuli, on peut procéder à une analyse multidimensionnelle.

Pour le registre aigu, quatre dimensions sont nécessaires pour représenter les distances inter-sons avec moins de 1% d'erreur.

La figure 5 est la projection des sons sur les deux premières dimensions obtenues pour analyse multidimensionnel non-métrique avec comme critère de minimisation le stress de Kruskal et 200 répliques.

La dimension 1 regroupe d'une part les sons de clarinette (à gauche) et d'autres part les sons de synthèse (à droite). La dimension 2 fait la distinction du "contrôle" en regroupant les sons issus de commandes robotisées (en haut) et ceux joués par le musicien (en bas). Il est à noter que des résultats proches sont obtenus pour le registre grave même si les facteurs "instrument" et "contrôle" se distinguent moins clairement dans l'espace obtenu compte tenu de la trop grande dispersion.

Nous avons procédé au calcul de plusieurs descripteurs de signal (sur la base des définitions de [8]) pour chacun des huit sons. La première dimension est corrélée avec le "kurtosis" ($r = -0.84$) ainsi qu'avec l'étalement spectral ($r = 0.81$) qui sont tous les deux des mesures de la largeur de bande du signal. La seconde dimension est principalement corrélée au "roll-off" qui correspond à la fréquence de coupure supérieure du spectre. Ces observations peuvent être rapprochées des résultats obtenus dans [1] où des filtrages sont appliqués à différents signaux. Les auteurs montrent une nette diminution du naturel quand la bande passante est réduite.

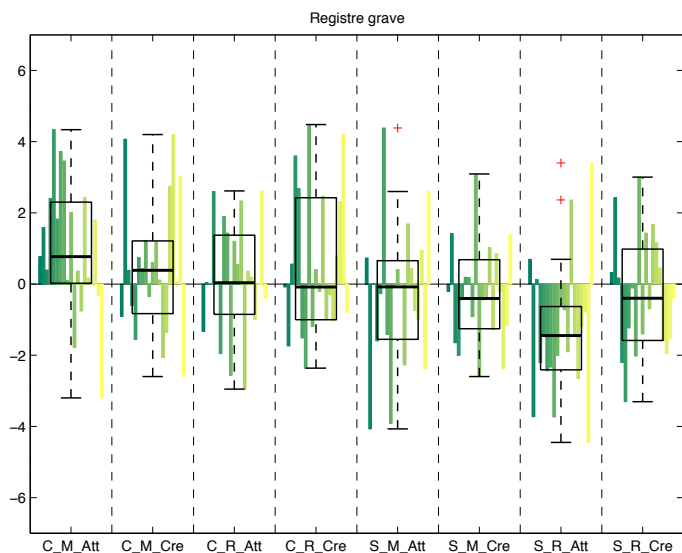


FIGURE 4 – Résultats obtenus pour le registre grave.
Les différences de jugement entre les sons sont beaucoup moins importantes que pour le registre aigu (cf. figure 3)

5 Conclusion - Perspectives

Nous avons présenté une première approche visant à évaluer les sons produits par un modèle de synthèse par le biais de la notion de naturel. Bien qu'imparfait, le protocole mis en place a permis de mettre en évidence l'importance relative du contrôle et de l'instrument dans la perception du naturel. Parmi les aspects qu'il sera nécessaire d'aborder dans tout travail à venir, on se doit de mentionner en premier le problème de l'appui des lèvres sur l'anche. A l'heure actuelle, la bouche artificielle utilisée pour générer les sons de clarinette ne permet pas d'effectuer un contrôle dynamique de la force d'appui (et de la position) de la lèvre sur l'anche alors que ce contrôle joue un rôle important dans le timbre notamment en terme de brillance. Dans la présente expérience, nous avons dû laisser ce paramètre fixe au cours de la note ce qui nous éloigne des conditions "réelles" de jeu. Un aspect brièvement abordé ici et qu'il sera de toute évidence nécessaire d'approfondir concerne le "calage" des paramètres du modèle de synthèse par rapport à la clarinette. Comme discuté dans [5] de très petites modifications des paramètres du modèle peuvent entraîner de fortes modifications dans son comportement et donc dans le son produit surtout si on considère différents régimes avec un même jeu de paramètres. Il serait donc intéressant de tester différentes stratégies de "calage" et envisager les résultats obtenus notamment en terme de variabilité, sous l'angle de ces considérations.

Enfin, il est évident qu'il sera nécessaire d'étendre cette étude au cas des phrases musicales ne serait-ce que pour les transitions entre notes.

Remerciements

Cette étude a en grande partie été réalisée par Sébastien Bel dans le cadre de son stage de Master 2

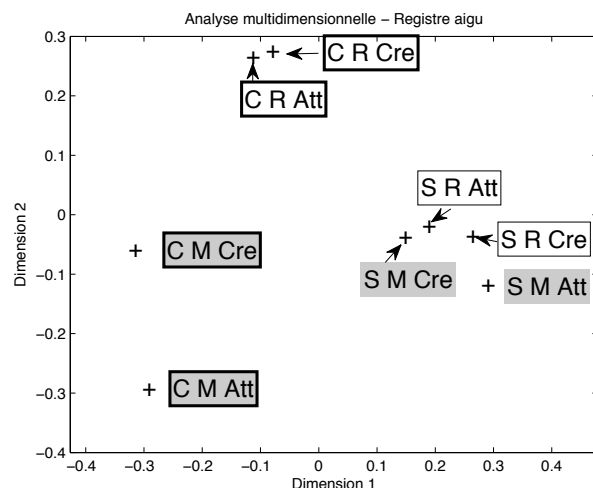


FIGURE 5 – Résultats de l'analyse multidimensionnelle.

La position exacte des sons dans l'espace est matérialisée par les croix. Le fond gris repère les sons joués par le musicien tandis que les cadre noir repère les sons de clarinette.

avec l'aide de D. Ferrand pour l'utilisation de la bouche artificielle.

Références

- [1] B. C. J. Moore and C.-T. Tan, "Perceived naturalness of spectrally distorted speech and music," *The Journal of the Acoustical Society of America*, vol. 114, no. 1, pp. 408–419, 2003.
- [2] J. Bensa, D. Dubois, R. Kronland-Martinet, and S. Ystad, "Perceptive and cognitive evaluation of a piano synthesis model," in *CCMMR* (Springer, ed.), pp. 232–245, 2004.
- [3] J. Kergomard, "Elementary considerations on reed-instruments oscillations", in *Mechanics of Musical Instruments*, edited by Hirschberg *et al.*, Lectures notes CISM, (Springer, New York 1995).
- [4] Ph. Guillemain, J. Kergomard, Th. Voinier, "Real-time synthesis of clarinet-like instruments using digital impedance models", *J. Acoust. Soc. Am*, Vol. 118(1), july 2005, pp.483-494.
- [5] M. Barthet, P. Guillemain, R. Kronland-Martinet, and S. Ystad, "From clarinet control to timbre perception," *Acta Acustica (under revision)*, 2009.
- [6] D. Ferrand, C. Vergez, B. Fabre, F. Blanc "High-precision regulation of a pressure controlled artificial mouth : the case of recorder-like musical instruments", *Acta Acustica united with Acustica*, accepté pour publication, numéro spécial, 2010.
- [7] P. Susini, O. Houix, N. Misdariis, B. Smith, and S. Langlois, "Instruction's effect on semantic scale ratings of interior car sounds," *Applied Acoustics*, vol. In Press, Corrected Proof, pp. –, 2008.
- [8] G. Peeters, "A large set audio features for sound description (similarity and classification) in the cuidado project," tech. rep., IRCAM, 2004