



**HAL**  
open science

# Sélection de HRTF dans une Base de Données en Utilisant des Paramètres Morphologiques pour la Synthèse Binaurale

David Schönstein, Brian F,g Katz

► **To cite this version:**

David Schönstein, Brian F,g Katz. Sélection de HRTF dans une Base de Données en Utilisant des Paramètres Morphologiques pour la Synthèse Binaurale. 10ème Congrès Français d'Acoustique, Apr 2010, Lyon, France. hal-00551145

**HAL Id: hal-00551145**

**<https://hal.science/hal-00551145v1>**

Submitted on 2 Jan 2011

**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.

## Sélection de HRTF dans une Base de Données en Utilisant des Paramètres Morphologiques pour la Synthèse Binaurale

David Schönstein<sup>1</sup>, Brian FG Katz<sup>2</sup>

<sup>1</sup>ARKAMYS, 31 rue Pouchet, 75017, Paris, dschonstein@arkamys.com

<sup>2</sup>LIMSI-CNRS, BP 133, Université Paris XI, F-91403, Orsay Cedex, brian.katz@limsi.fr

Les environnements virtuels auditifs deviennent de plus en plus pertinents pour des applications comme la téléconférence, les prothèses auditives, les jeux vidéo, et l'écoute immersive en général. Afin de permettre une interprétation de haute fidélité d'une scène sonore dans de tels environnements, le contenu audio doit être traité avec les filtres acoustiques de l'auditeur, nommés *Head-Related Transfer Function* (HRTF). Le défi actuel, étant donné la difficulté de mesurer les HRTF d'un auditeur, est d'être capable d'effectuer une sélection de HRTF, à partir d'une base de données préexistante de HRTF, et de donner à l'auditeur les HRTF qui ressemblent le plus possible à ses propres HRTF en utilisant des données simples, comme uniquement une photo de l'oreille. Un processus est décrit dans lequel une base de données de HRTF mesurant 46 auditeurs différents est réduite au moyen de différentes techniques telles que l'analyse en composantes principales ou un décalage des HRTF dans le domaine fréquentiel. De multiples régressions linéaires sont effectuées sur les paramètres morphologiques les plus importants afin de prédire les HRTF préférées d'un auditeur. Les HRTF prédites sont ensuite comparées à un classement subjectif des sujets dans la base de données selon leurs préférences. Les résultats montrent que ce processus est en mesure de prédire les HRTF pour un auditeur de façon significativement meilleure que si les HRTF avaient été choisies au hasard.

## 1 Introduction

Les humains ont la capacité d'encoder l'information directionnelle d'un son incident avec une fonction de transfert acoustique. La tête, les oreilles externes, le corps d'un auditeur transforment l'information spectrale d'un son dans l'espace par ce qu'on appelle *Head-Related Transfer Function* (HRTF), et ceci nous permet de percevoir notre environnement acoustique en fonction de la position, distance, etc. des sources sonores.

Pour les applications de la synthèse binaurale comme la téléconférence, les prothèses auditives, les jeux vidéo, et l'écoute immersive en général, un rendu haute fidélité de la scène sonore est préférable. L'utilisation des HRTF qui ressemblent le plus possible aux filtres acoustiques de l'auditeur permet ce rendu haute fidélité. Plusieurs études dans la littérature démontrent l'intérêt de ce qu'on appelle des HRTF individualisées [1-4], surtout en termes de précision dans les tests de localisation. En revanche, il existe des études qui montrent que les propres HRTF d'un auditeur ne sont pas toujours préférées par rapport à d'autres HRTF [5].

Les HRTF individualisées peuvent être obtenues via des mesures avec microphones dans les oreilles de l'auditeur [6-8], ou même via une simulation numérique [9, 10]. Malgré la qualité de leur rendu, ces méthodes restent encore aujourd'hui très laborieuses. Il existe d'autres possibilités orientées vers une utilisation grand public comme la mesure rapide des HRTF via le principe de réciprocité [11] ou interpolation [12], l'adaptation de HRTF non individualisées [12, 13], ou la sélection des meilleures HRTF dans une base de données [14-18].

Cette étude vise à poursuivre la troisième méthode, notamment inspirée par des techniques qui utilisent les paramètres morphologiques [17-20]. La plupart de ces

techniques se focalisent sur une validation avec test de localisation. L'étude actuelle cherche à valider deux différentes méthodes pour la sélection de HRTF dans une base de données, non pas avec des tests de localisation, ni avec des HRTF interpolées, mais avec une évaluation qui est plus adaptée aux applications citées : en utilisant un test d'écoute et les résultats des préférences des sujets.

## 2 Méthode

### 2.1 Test d'écoute

L'objet de cette étude était de trouver un système de sélection d'une HRTF parmi les éléments d'une base de données. Cette base de données est composée de HRTF et des paramètres morphologiques des sujets associés. Pour mieux évaluer les méthodes de sélection, un classement subjectif des HRTF dans la base de données a été effectué.

Les HRTF utilisées venaient de la base de données publique du projet LISTEN [21]. 45 sujets ont participé à cette évaluation. Les mesures « brutes » (sans égalisation du champ diffus) de 46 sujets (un sujet n'a pas participé dans le test) ont été choisies pour ce test. Les HRTF ont été décomposées pour extraire les composants « phase-minimum » et « phase-excès ». La partie « phase-excès » a été remplacée par un retard pur, qui représente l'*Interaural Time Difference* (ITD) de chaque sujet. L'ITD a été déterminée selon le calcul du maximum du coefficient de corrélation croisée *interaural* (maxIACC) des enveloppes énergétiques.

De cette façon, le délai est apparié pour le sujet et l'information spectrale de filtrage sur la HRTF est variée. Cela revient, en simulation, à changer les oreilles de chaque sujet tout en gardant une même géométrie de la tête.

Le signal utilisé pour le test était une synthèse binaurale d'un bruit blanc à large bande de 0,23 secondes modelé en utilisant une fenêtre de Hanning. Le signal de test a été présenté à des positions fixes le long de deux trajectoires présentées en séquence :

1. Un cercle dans le plan horizontal (élévation = 0°) en incréments de 30°. La trajectoire commence à 0° azimut et 0° élévation et on effectue deux rotations (durée 6 secondes).
2. Un arc dans le plan médian (azimut = 0°) à partir de l'élévation -45° à l'avant jusqu'à -45° à l'arrière en incréments de 15°. La trajectoire commence à l'avant à élévation -45°, on fait varier l'élévation vers l'arrière et on revient par le même chemin à la position initiale (durée 9 secondes).

Pour chaque sujet, 46 signaux, correspondant aux 46 différentes HRTF de la base de données (incluant la propre HRTF du sujet), ont été présentés en utilisant une interface graphique. Pour chaque HRTF, les sujets ont dû faire un choix : *excellent*, *moyen* ou *mauvais*. Les sujets ont été invités à sélectionner de 5 à 10 HRTF comme *excellent*, afin d'obtenir une statistique utile pour l'analyse des résultats. Les sujets ne savaient pas quelle HRTF parmi les 46 correspondait à leur propre mesure. Les sujets ont été autorisés à écouter les signaux à plusieurs reprises, et dans n'importe quel ordre. La durée du test était d'environ 35 minutes.

La Figure 1 présente les résultats des tests d'écoute pour tous les sujets. Les sujets sont représentés sur l'axe horizontal et les HRTF classées sur l'axe vertical. La couleur des cercles indique le classement des HRTF. La diagonale montre clairement que tous les sujets, sauf 3, ont classé leurs propres HRTF comme *excellent*.

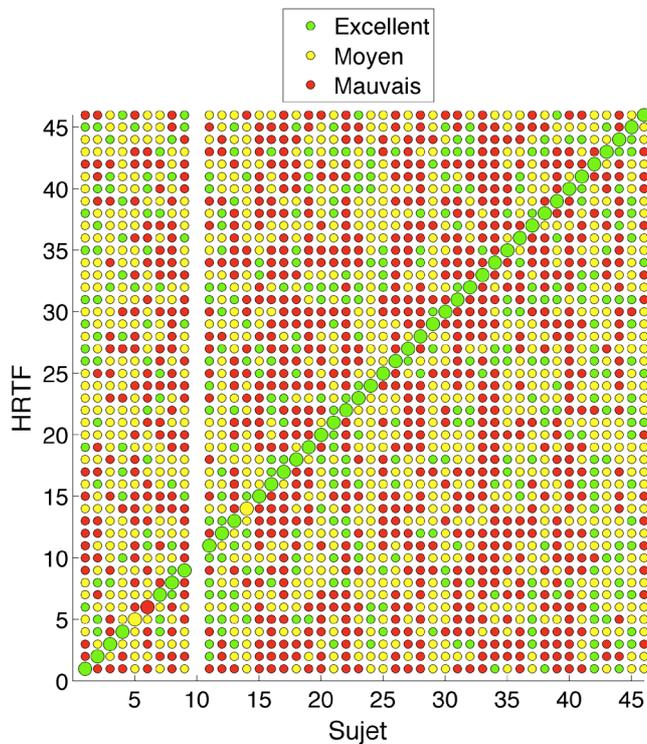


Figure 1 : Classement des HRTF par les sujets. Le sujet numéro 10 n'a pas classé les HRTF.

## 2.2 Analyse en composante principale

Les mesures de HRTF LISTEN [21] ont été prises à des positions dans l'espace correspondant à des angles d'élévation compris entre -45° et 90° par incréments de 15° et angles d'azimut commençant à 0° par incréments de 15°. Les incréments en azimut ont été progressivement augmentés pour les angles d'élévation au-dessus de 45°, afin d'échantillonner l'espace de façon égale, pour un total de 187 positions. Aucune interpolation n'a été utilisée dans cette étude puisque il était considéré comme important de ne pas avoir la qualité d'interpolation comme variable dans les données, ce qui veut dire que les HRTF mesurées n'échantillonnaient pas l'espace d'une manière parfaitement égale. Les mesures brutes de HRTF ont été normalisées par la racine carrée moyenne entre les oreilles gauche et droite pour toutes les positions, et ensuite normalisées par la racine carrée moyenne pour toutes les positions pour les oreilles gauche et droite sur tous les sujets.

Les HRTF ont été transformées ensuite en ce que l'on appelle la *Directional Transfer Function* (DTF) [22], qui contient uniquement la part des HRTF qui présentent une dépendance directionnelle : les DTF sont les spectres d'amplitude des HRTF divisés par la moyenne des spectres pour toutes les positions dans l'espace pour chaque oreille. De plus, la résolution fréquentielle limitée du système auditif est prise en compte par un lissage des DTF en bande critique. Dans cette étude, la distribution spatiale des mesures de HRTF dans l'espace n'a pas été prise en compte pour le calcul de la DTF.

Les DTF traitées de 512 points, ont été ensuite concaténées pour toutes les positions pour chaque oreille. Les vecteurs des spectres concaténés de chaque oreille ont été ensuite eux même concaténés, ce qui a donné un vecteur par sujet, formant ainsi une rangée d'une matrice de dimension 46x191488 contenant tous les sujets. Les colonnes de cette matrice représentent les magnitudes en fréquence pour toutes les DTF. Une analyse en composante principale (ACP) a été réalisée sur cette matrice. L'objectif de l'ACP était de réduire la dimensionnalité de la base de données de HRTF qui présente une certaine redondance, tout en conservant la part la plus significative de la variabilité observée [23, 24]. L'analyse a donné une nouvelle matrice de données (les scores) qui représentaient les données originales projetées sur de nouveaux axes (les composantes principales) dans les dimensions de la matrice originale qui redisent la variance totale pour chaque colonne de magnitude en fréquence. Chaque rangée de cette nouvelle matrice de scores représentait toujours un sujet, et chaque colonne représentait les données pour chaque axe dans les dimensions de l'espace classé en fonction de la quantité de variance des données originales décrites. Chaque colonne dans la matrice des scores a été utilisée ensuite comme des coordonnées dans une dimension de l'espace. En prenant par exemple trois colonnes, les sujets ont pu être représentés dans un espace tridimensionnel. La distance euclidienne entre les sujets représente une mesure de leurs similarités en fonction des HRTF mesurées.

## 2.3 Choix des paramètres morphologiques les plus importants

La base de données LISTEN inclut, pour chaque sujet, 22 paramètres morphologiques définis en [25] (Figure 2). Cela représente tous les paramètres morphologiques sauf :  $x_{13}$ ,  $x_{14}$ ,  $x_{15}$ ,  $d_8$  et  $\theta_2$ . En utilisant une méthode de *machine*

learning nommée *feature selection*, on souhaitait trouver un sous-ensemble de caractéristiques pertinentes, les paramètres morphologiques les plus importants. Après des évaluations préliminaires, une meilleure performance de *feature selection* a été trouvée en groupant les résultats en deux catégories : les HRTF classées comme *excellent* et *moyen* versus les HRTF classées comme *mauvais*.

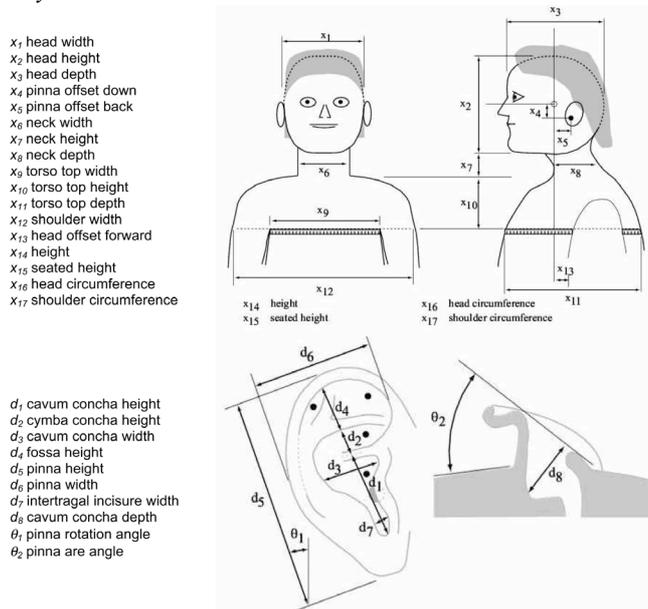


Figure 2 : Paramètres morphologiques de la base de données CIPIC dont 22 ont été utilisés dans cette étude. Figure prise de [25].

Une normalisation des données morphologiques a ensuite été réalisée pour convenir à la méthode de *feature selection*. Les valeurs morphologiques de l'ensemble des HRTF ont été divisées par les données de chaque sujet. Avec cette normalisation, les valeurs représentaient le pourcentage d'un paramètre morphologique d'un sujet par rapport à un autre. Ces valeurs étaient liées avec les classements de HRTF pour chaque pair de sujet pour l'analyse.

La méthode de *feature selection* choisie pour trouver les paramètres morphologiques les plus importants a été celle des machines à vecteur de supports (*support vector machines* - SVM) qui construit un ensemble d'hyperplans dans un espace de dimension élevée pour classer les données normalisées. Cette technique a été utilisée par exemple pour trouver des gènes de cancer [26]. Avec cette méthode, les paramètres ont donc été classés du plus au moins important. Une variable dans la classification avec SVM est la valeur de complexité  $C$ , qui contrôle la tolérance d'erreurs de classifications dans l'analyse, et introduit une fonction de pénalité [27]. Une valeur de  $C$  nulle indique que la fonction de pénalité n'est pas prise en compte, et une valeur de  $C$  élevée ( $C \rightarrow \infty$ ) indique que la fonction de pénalité est dominante. Selon les différentes valeurs de  $C$ , le classement des paramètres morphologiques changeait. Une valeur de  $C = 1$  a été utilisée pour la classification avec SVM, choisie pour avoir une tolérance élevée des erreurs de classifications et pour donner les meilleurs résultats (section 3.1). En utilisant cette méthode, les dix premiers paramètres morphologiques les plus importants, du plus important au moins important, correspondaient à :  $x_{11}$ ,  $x_2$ ,  $x_8$ ,  $d_5$ ,  $x_3$ ,  $d_4$ ,  $x_{12}$ ,  $d_2$ ,  $d_1$  et  $x_6$ .

### 3.1 Analyse du *Scaling* Fréquentiel

[13] a proposé une technique d'adaptation de HRTF qui applique un décalage global en fréquence sur la HRTF entière pour que les creux et pics correspondent entre les HRTF du sujet et les HRTF non-individualisées. Cette méthode repose sur l'idée que les variations entre les HRTF d'un sujet à l'autre sont liées à la taille des cavités du pavillon qui jouent sur les résonances dans l'oreille. Une augmentation ou diminution de la taille d'une oreille par un facteur modifie ces résonances et décale les creux et pics vers les basses ou hautes fréquences respectivement.

La méthode de [15] a été utilisée pour calculer le degré de décalage (*scaling factor* global) entre deux sujets de la base de données. Afin de ne pas sur-échantillonner dans les hautes fréquences, les DTF ont été traitées avec une banque de filtres passe-bande (70 par octave) pour échantillonner les composantes en fréquence sur des intervalles égaux en échelle logarithmique base 2, soit en échelle d'octave. Pour chaque direction, la DTF d'un sujet est soustraite, fréquence par fréquence, de la DTF de l'autre sujet, à la même direction pour calculer ce qu'on appelle l'*Inter-Subject Spectral Difference* (ISSD). La variance de cette différence est calculée sur la bande de fréquence de 3.7 kHz à 12.9 kHz et l'ISSD globale est finalement définie comme la moyenne, pour toutes les directions, des ISSD directionnelles. Le *scaling factor* est donc la valeur qui donne le minimum ISSD global.

Une fois le *scaling factor* calculé pour chaque pair de sujet, une matrice de dissemblance a été créée et utilisée pour une analyse par *Multi-Dimensional Scaling* (MDS). Comme il n'y avait pas de notion de direction de transformation avec les *scaling factor*, seulement la magnitude a été prise en compte. Le *scaling factor*, nommé alpha ( $\alpha$ ), peut être en forme de  $\alpha$  ou  $1/\alpha$  correspondant à un décalage de HRTF A vers B ou HRTF B vers A. Vu que les deux formes de *scaling factor* représentent la disparité fréquentielle entre HRTF A et B, et que le choix sera arbitraire, le facteur  $\alpha > 1$  a été utilisé pour l'analyse.

Le MDS est une technique statistique qui a donné en effet le même type de matrice de coordonnées euclidiennes pour décrire les distances entre les HRTF dans un espace multidimensionnel, comme pour l'ACP. Dans l'analyse, cette méthode a seulement été utilisée pour l'oreille droite, puisque dans notre étude et celle de [13] il existait une forte corrélation (coefficient de corrélation de 0,83 et 0,95 respectivement) entre les *scaling factor* des deux oreilles.

## 3 Résultats

### 3.1 Validation

La validation des espaces multidimensionnels a été faite pour chaque sujet en classant les HRTF des sujets dans la base de données en fonction des distances euclidiennes pour toutes les dimensions possibles. Le classement des DTF pour chaque sujet a été comparé avec les résultats du test d'écoute. Pour chaque sujet, le rang a été calculé (une valeur entre 1 et 45, avec la valeur 1 représentant la meilleure HRTF) pour les HRTF classées comme *excellent*, *moyen* ou *mauvais* en utilisant l'espace multidimensionnel. Ensuite, les valeurs de rang de toutes les HRTF classées soit comme *excellent*, *moyen* ou *mauvais* pour tous les sujets ont été groupées ensemble et la distribution a été analysée pour chaque dimension. Comme les distributions des valeurs de

rang ne suivaient pas une distribution normale, une analyse de variance Kruskal-Wallis a été utilisée pour tester l'égalité des médianes.

La Figure 3 montre les distributions des rangs les plus statistiquement différents pour les deux méthodes de réduction. Les 45 premières dimensions ont été utilisées dans l'ACP, ce qui représentait 79 % de la variance des données des DTF. Pour l'analyse par MDS, les 26 premières dimensions ont été prises, ce qui représentait une reconstruction des données de la matrice de dissemblance. L'erreur maximale de la reconstruction était de 0,13 avec une valeur maximale de 0,23 dans la matrice de dissemblance. La distribution dans la figure pour *excellent* par exemple représente les valeurs de rang pour les HRTF classées comme *excellent* pour tous les sujets. Les distributions devenaient légèrement plus statistiquement significatives en augmentant le nombre de dimensions puisque les composants principaux (pour la méthode de l'ACP) décrivaient plus de variance des données. Globalement, ces deux méthodes ont des résultats comparables mais traitent les données des DTF très différemment. Il est intéressant que l'information contenue dans les *scaling factor*, qui représente les différences entre les sujets sur seulement un aspect de la DTF, puisse être utilisée dans les analyses des données pour créer un espace multidimensionnel performant. Cet espace est aussi cohérent que l'espace utilisant toutes les informations du spectre (ACP des DTF).

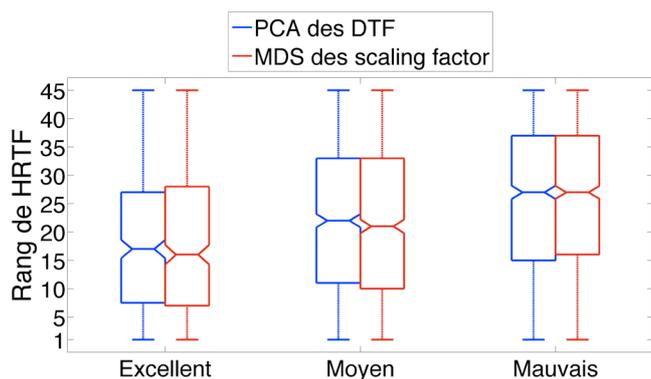


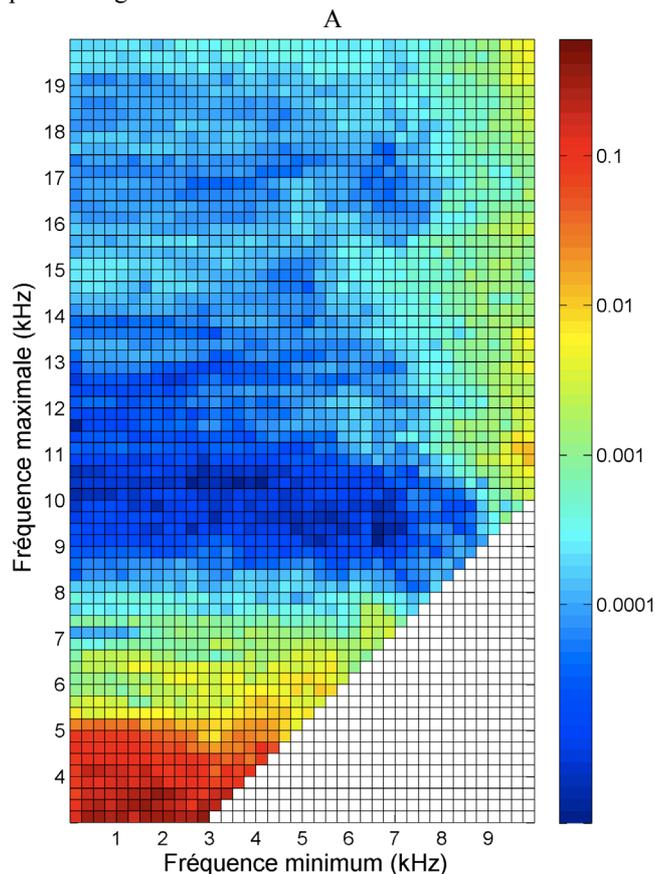
Figure 3 : distributions des rangs des HRTF pour les deux méthodes : l'ACP des DTF et l'analyse par MDS des *scaling factor*.

Étant plus intéressés par l'application de cette méthode pour la prédiction des meilleure HRTF pour un sujet donné, nous avons utilisé une métrique de validation plus subtile et mieux adaptée pour des applications de la synthèse binaurale. Cette métrique est calculée en vérifiant le pourcentage de HRTF classées dans le test d'écoute comme *excellent*, parmi les dix meilleures HRTF du classement de l'espace multidimensionnel de l'ACP des DTF. Ce pourcentage peut être comparé avec le pourcentage global de HRTF classées dans le test d'écoute comme *excellent* pour chaque sujet; correspondant au pourcentage de HRTF classées comme *excellent* qui existera dans une sélection de dix HRTF choisies au hasard. Cette comparaison a été faite pour toutes les dimensions et pour chaque sujet. Un test Student a été utilisé pour vérifier si les moyennes des distributions de valeurs de pourcentage étaient statistiquement différentes pour chaque nombre de dimension différent. La *p-value* des tests Student, nommé *p-value excellent*, représente la probabilité d'obtenir une

statistique de test au moins aussi extrême que celle qui a été effectivement observée, en supposant que l'hypothèse nulle (que les valeurs de pourcentage viennent de la même population) est vraie.

La *p-value excellent* a été utilisée pour trouver les bornes de fréquence des DTF pour créer la matrice de l'ACP donnant la valeur la plus petite, soit la plus statistiquement significative. La Figure 4(A) montre les minimums des valeurs *p-value excellent* pour toutes les dimensions en fonction des différentes bornes de fréquence utilisées. La fréquence minimum est représentée sur l'axe horizontal et la fréquence maximale sur l'axe vertical. Les bornes de fréquence optimales sont : 0 Hz et 11500 Hz. Ces bornes correspondent à une valeur de *p-value excellent* de  $3,5 \times 10^{-6}$ . Ces bornes de fréquence peuvent être interprétées comme la partie du DTF perceptiblement le plus important en termes de préférence des sujets dans le test d'écoute.

La Figure 4(B) montre le nombre de dimensions utilisées pour chaque borne de fréquence avec une valeur maximale de 45. Le nombre de dimensions utilisé pour les bornes de fréquence optimales était 35. Une corrélation existe entre la valeur *p-value excellent* représentée sur la Figure 4(A) et la valeur du nombre de dimensions utilisé dans l'analyse sur la Figure 4(B). Plus la valeur *p-value excellent* est petite, correspondant à un espace multidimensionnel plus performant, plus de dimensions sont utilisées dans l'analyse représentant un plus grand pourcentage de la variance des données.





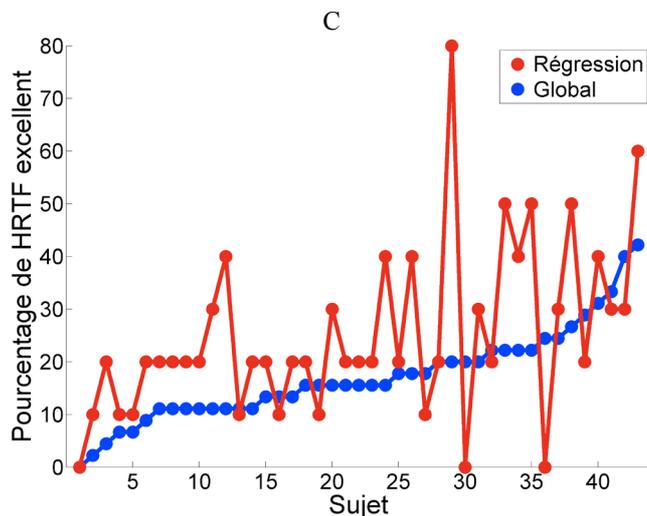


Figure 5 : validation de la régression avec paramètres morphologiques. (A) Minimum  $p$ -value excellent en fonction du nombre de paramètres utilisés pour l'ACP. (B)  $p$ -value excellent en fonction du nombre de dimensions utilisées pour une régression avec les 4 premiers paramètres morphologiques. (C) Pourcentage de HRTF classée comme excellent sur un choix de dix HRTF en utilisant la méthode proposée (en rouge) par rapport à une sélection aléatoire (en bleu).

## 4 Conclusion

Cette étude a validé avec succès deux méthodes pour décrire les différences de HRTF dans une base de données. Cette validation a été effectuée avec les résultats d'un test d'écoute qui s'appuie sur les préférences subjectives sans utiliser l'interpolation de HRTF. La technique proposée a été employée pour sélectionner les paramètres morphologiques les plus importants, ainsi que des bornes de fréquence pertinentes. En comparaison au classement aléatoire, la technique décrite dans cette étude a permis de prédire de façon statistiquement plus significative les classements subjectifs des HRTF. Cette méthode, toujours un travail en cours, pourrait donc être utilisée pour faciliter la sélection d'une HRTF dans une base de données. Elle permettra des applications de la synthèse binaurale au service du grand public en utilisant simplement une photo de l'oreille de l'auditeur.

## Références

[1] Moller H. *et al.*, "Binaural technique: Do we need individual recordings?", *Journal of the Audio Engineering Society*, 44, 451-469 (1996).

[2] Wenzel E. M. *et al.*, "Localization Using Nonindividualized Head-Related Transfer-Functions", *Journal of the Acoustical Society of America*, 94, 111-123 (1993).

[3] Middlebrooks J. C., "Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency", *Journal of the Acoustical Society of America*, 106, 1493-1510 (1999).

[4] Gardner M. B., Gardner R. S., "Problem of Localization in Median Plane - Effect of Pinnae Cavity Occlusion", *Journal of the Acoustical Society of America*, 53, 400-408 (1973).

[5] Usher J., Martens W., "Perceived naturalness of speech sounds presented using personalized versus

non-personalized HRTFs", *13th International Conference on Auditory Display*, Montreal, Canada, (2007).

[6] Pralong D., Carlile S., "Measuring the human head-related transfer functions – a novel method for the construction and calibration of a miniature in-ear recording-system", *Journal of the Acoustical Society of America*, 95, 3435-3444 (1994).

[7] Wightman F., Kistler D., "Measurement and validation of human HRTFs for use in hearing research", *Acta Acustica United with Acustica*, 91, 429-439 (2005).

[8] Wightman F. L., Kistler D. J., "Headphone Simulation of Free-Field Listening. I: Stimulus Synthesis", *Journal of the Acoustical Society of America*, 85, 858-867 (1989).

[9] Kahana Y., Nelson P. A., "Boundary element simulations of the transfer function of human heads and baffled pinnae using accurate geometric models", *Journal of Sound and Vibration*, 300, 552-579 (2007).

[10] Katz B. F. G., "Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation", *Journal of the Acoustical Society of America*, 110, 2440-2448 (2001).

[11] Zotkin D. N. *et al.*, "Fast head-related transfer function measurement via reciprocity", *Journal of the Acoustical Society of America*, 120, 2202-2215 (2006).

[12] Guillon P., "Individualisation des indices spectraux pour la synthèse binaurale : recherche et exploitation des similarités inter-individuelles pour l'adaptation ou la reconstruction de HRTF", *thèse de doctorat*, Université du Maine (2009).

[13] Middlebrooks J. C., "Individual differences in external-ear transfer functions reduced by scaling in frequency", *Journal of the Acoustical Society of America*, 106, 1480-1492 (1999).

[14] Seeber B., Fastl H., "Subjective selection of non-individualhead-related transfer functions", *9th International Conference on Auditory Display*, Boston, (2003).

[15] Iwaya Y., "Individualization of head-related transfer functions with tournament-style listening test: Listening with other's ears", *Acoustical Science & Technology*, 27, 340-343 (2006).

[16] Wightman F., Kistler D., "Multidimensional scaling analysis of head-related transfer functions", *IEEE Digita Audio Workshop*, 98-101 (1993).

[17] Zotkin D. N. *et al.*, "Rendering localized spatial audio in a virtual auditory space", *IEEE Transactions on Multimedia*, 6, 553-564 (2004).

[18] Jin C. *et al.*, "Enabling individualized virtual auditory space using morphological measurements", *IEEE International Conference on Multimedia Information Processing*, December 13-15, (2000).

[19] Nishino T. *et al.*, "Estimation of HRTFs on the horizontal plane using physical features", *Applied Acoustics*, 68, 897-908 (2007).

[20] Xu S. *et al.*, "Improved method to individualize head-related transfer function using anthropometric measurements", *Acoustical Science and Technology*, 29, 388-390 (2008).

- [21] "LISTEN HRTF database", <http://recherche.ircam.fr/equipes/salles/listen/>
- [22] Middlebrooks J. C., Green D. M., "Directional dependence of the interaural envelope delays", *Journal of the Acoustical Society of America*, 87, 2149-2162 (1990).
- [23] Kistler D. J., Wightman F. L., "A model of head-related transfer-functions based on principal components-analysis and minimum-phase reconstruction", *Journal of the Acoustical Society of America*, 91, 1637-1647 (1992).
- [24] Martens W., "Principal components analysis and resynthesis of spectral cues to perceived direction", *Proceedings of the International Computer Music Conference*, 274-281, (1987).
- [25] Algazi V. R. *et al.*, "The CIPIC HRTF Database", *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, New York, October 21-24, 99-102, (2001).
- [26] Guyon I. *et al.*, "Gene selection for cancer classification using support vector machines", *Machine Learning*, 46, 389-422 (2002).
- [27] Gunn R. S., "Support Vector Machines for Classification and Regression" *Technical Report*, Image Speech and Intelligent Systems Research Group, University of Southampton, (1997).