

Évaluation Objective et Subjective de Différentes Méthodes de Lissage des HRTF

F. Rugeles^a, M. Emerit^a et B. Katz^b

^aOrange Labs, 4 Rue du clos Courtel, 35510 Cesson Sevigné, France

^bLIMSI-CNRS, BP 133, 91403 Orsay, France

felipe.rugelesospina@orange.com

Les Head Related Transfer Functions (HRTF) sont des filtres linéaires décrivant les effets acoustiques produits par la tête et le torse dans les signaux qui sont reçus au niveau des tympans. Ces filtres peuvent être obtenus numériquement par la mesure de la réponse acoustique d'un individu en chambre anéchoïque. Plusieurs méthodes telles que l'analyse en composantes principales ou indépendantes, ou la décomposition en harmoniques sphériques ont été utilisées dans la littérature pour analyser et traiter ces données numériques afin de comprendre leur rôle dans la perception des sons. Généralement, un pré-traitement visant à réduire la variance des données est réalisé en amont afin de réduire la complexité du problème. Pour assurer que ces pré-traitements ne dégradent pas pour autant la qualité perceptive des HRTF plusieurs méthodes de lissage ont été évaluées. Ces traitements ont été comparés entre eux avec des critères objectifs comme les différences spectrales et avec une série de tests perceptifs. Dans cet article, les différentes techniques de lissage testées sont décrites dans un premier temps. Puis, la méthodologie d'évaluation objective et subjective retenue est présentée. Enfin, les résultats des tests sont discutés.

1 Introduction

L'audition est un des moyens par lequel les êtres vivants recueillent des informations sur l'environnement qui les entoure. Grâce aux indices binauraux (différences de temps interaurale et de niveau interaurale) et monauraux (variation spectrale de signaux perçus en fonction de l'élévation) [15] et selon certaines conditions, l'écoute permet de localiser une source dans l'espace.

La synthèse binaurale a pour but de reproduire les phénomènes de retard, réflexion, diffraction et atténuation subits par le signal provenant d'une source dans l'espace jusqu'à son arrivée à l'entrée des canaux auditifs. Pour chaque direction de l'espace et pour chaque oreille, ceci peut être représenté comme un système linéaire et invariant dans le temps, dans le cas où la personne est statique. Ce système est complètement caractérisé par sa fonction de transfert appelée en anglais Head Related Transfer Function (HRTF) [2, 14] ou par son équivalent dans le domaine temporel, le Head Related Impulse Response (HRIR).

Ces fonctions de transfert sont spécifiques à chaque personne. Elles dépendent de la forme du torse, de la tête et du pavillon d'un individu. Chaque HRTF est définie par le triplet : morphologie, direction et oreille. Pour étudier les caractéristiques de ces fonctions qui permettent la localisation des sons, il est nécessaire de mesurer les réponses impulsionnelles d'un grand nombre de personnes à différents endroits de l'espace.

Au cours des dernières années, plusieurs campagnes de mesure de HRIR ont été effectuées [3, 6, 9, 14] afin de constituer des bases de données de réponses impulsionnelles de sujets. Une fois mesurée, la HRIR pour une oreille est souvent modélisée par un retard pur et un filtre à phase minimale [11]. Le filtre à phase minimale est par construction défini uniquement par son spectre.

Souvent, chacun des spectres est égalisé par la moyenne des spectres (spectre du champ diffus). L'ensemble des spectres des HRTF égalisés champs diffus a été défini [13] comme Directional Transfer Function (DTF). Les DTF décrivent uniquement la dépendance directionnelle des spectres. Elles sont appliquées aux signaux en calculant leurs filtre à phase minimale puis en ajoutant le retard pur des HRTF d'origine.

L'intérêt de ces études est de connaître quelles sont les parties du spectre des DTF qui sont nécessaires au système auditif humain pour localiser "correctement" un son dans l'espace. Définir le terme "correctement" demande à lui seul une étude approfondie, mais ce n'est pas l'objet de cet article.

Il est connu que les performances de localisation du

système auditif humain dépendent fortement du type de signaux émis [2]. Dans cet article, l'influence sur la perception de différentes méthodes de lissage est évaluée. Les signaux utilisés sont des sons large bande filtrés par différentes DTF dont les spectres ont été dégradés. L'idée est d'essayer de simplifier au maximum les DTF par lissage sans dégrader pour autant la perception de la synthèse binaurale résultante. Le but est de réduire la complexité des études qui visent à caractériser les DTF. Il convient aussi de noter que le lissage permet de simplifier et d'améliorer l'implémentation numérique de la synthèse binaurale.

Dans cet article, différentes techniques de lissage des DTF ont été comparées, en prenant en compte les principes de fonctionnement du système auditif humain. Pour comparer les méthodes entre elles, des tests subjectifs et un critère objectif a été pris en compte. Les résultats globaux montrent qu'il est possible de lisser les filtres obtenus dans la campagne de mesure sans pour autant dégrader la qualité de la synthèse binaurale des signaux large bande. De plus, ils montrent une forte corrélation entre le critère objectif choisi et les résultats des tests subjectifs.

L'article est organisé de la manière suivante : dans la section 2, les techniques utilisées pour le lissage spectrale des DTF sont expliquées en détail. Ensuite, le protocole expérimental pour les tests subjectifs est décrit dans la section 3. Puis, les résultats des tests objectifs et subjectifs sont présentés et commentés dans la section 4. Enfin, des conclusions sont tirées des résultats obtenus et des ouvertures pour d'autres recherches sont proposées.

2 Lissage spectrale des Directional Transfer Functions

Le système d'audition humain est composé de trois parties : l'oreille externe, l'oreille moyenne et l'oreille interne. La première partie est composée du pavillon et du conduit auditif externe. L'oreille moyenne est reliée à l'oreille externe par le tympan. Elle est elle-même composée par les trois osselets (le marteau, l'enclume et l'étrier), la caisse du tympan et par deux fenêtres, la fenêtre ovale et la fenêtre ronde, qui lui permettent de communiquer avec l'oreille interne. Cette dernière partie du système auditif contient la cochlée, qui est l'organe de l'audition, et le nerf auditif qui sert de canal pour transmettre les informations perçues au cerveau.

La perception des signaux se fait dans une partie de la cochlée appelée l'organe de Corti, où se trouvent des structures sensorielles de l'audition. Ces structures ont

le nom de cellules ciliées et sont réparties le long des deux spires et demi qui constituent la cochlée. Les hautes fréquences sont perçues à l'entrée de l'organe d'audition et les basses fréquences sont perçues à l'apex de la cochlée [7]. La répartition des cellules sensorielles est beaucoup plus importante dans les basses fréquences que dans les hautes fréquences et peut se modéliser comme une fonction logarithmique [17].

2.1 Méthodes de lissage

Les algorithmes de lissage proposés dans cet article visent à inclure la répartition logarithmique des cellules sensorielles de l'audition dans la résolution fréquentielle des filtres binauraux. En effet, le but de notre traitement est d'éliminer les informations présentes dans les filtres mesurés en chambre anéchoïque qui sont peu ou pas perceptibles par les individus.

On a retenu à cet effet, deux méthodes de lissage. La première consiste à faire un lissage spectral par bande de fréquence fixe et la deuxième réalise un lissage spectral par fenêtres glissantes de tailles variables.

2.1.1 Lissage par bande de fréquence

La première étape pour réaliser ce type de lissage est de diviser l'espace de fréquences en sous bandes. Le premier type de partition de l'espace de fréquence utilise l'échelle de bandes critiques proposée par Zwicker [18, 19], les bandes de Bark. Celles-ci peuvent s'approcher avec la formule (1).

$$z_c = 13 \arctan(7.6 * 10^{-4} f) + 3.5 \arctan\left(\frac{f}{7500}\right)^2 \quad (1)$$

Avec f en Hertz et z_c en Bark. Les limites des bandes de Bark sont définies comme les valeurs de fréquences f pour lesquelles z_c a une valeur entière positive.

Le deuxième type de partition considéré est la division des fréquences en octaves. La fréquence de référence qui est standardisée pour cette division est $f_0 = 1000\text{Hz}$ [1]. Pour calculer les fréquences limites de la partition il faut utiliser la formule (2).

$$B_i(k) = 2^{\frac{k}{i}} f_0 \quad (2)$$

Les échelles d'octave, tiers d'octave, sixième d'octave et douzième d'octave correspondent à $i = 1, 3, 6, 12$ respectivement. Pour notre étude, k a une valeur entière (positive ou négative) de sorte que $\frac{k}{i}$ appartienne à l'intervalle $[-5; 5]$.

Une fois la partition en sous-bandes de l'espace de fréquences est choisie, le lissage des DTF est fait de la manière suivante. Pour une fréquence f appartenant à une sous bande β ayant un nombre d'éléments W , et pour une fonction de transfert H définie par rapport aux fréquences, on a :

$$H_s(f) = \mathbb{E}[|H(\beta)|] = \frac{1}{W} \sum_{f_i \in \beta} |H(f_i)| \quad (3)$$

Avec H_s étant la fonction de transfert lissée et \mathbb{E} étant l'espérance de premier ordre. Ce lissage est ainsi appliqué à chaque fréquence où la DTF est définie.

2.1.2 Lissage par fenêtre glissante de taille variable

Les principes de ce lissage ont été présentés dans les travaux de Smith [16] et ont pour but de faire un lissage par rapport aux bandes critiques de l'oreille. Le principe est d'appliquer un filtre glissant dont la longueur varie en fonction de la fréquence centrale de la fenêtre. Pour cette étude, le filtre choisi est une fenêtre de Hanning dont la taille est déterminée par une valeur α qui représente un pourcentage de la fréquence centrale de la fenêtre de lissage. Ainsi, le lissage prend en compte une bande de fréquence qui est proportionnelle à la fréquence du point à calculer et fait un lissage de type logarithmique. Ceci peut s'exprimer comme :

$$H_{vb}(f) = \frac{1}{\sum_{i=0}^{2\alpha f} w(i)} \sum_{f_i=f(1-\alpha)}^{f(1+\alpha)} |H(f_i)| * w(f_i - f(1-\alpha)) \quad (4)$$

$$w(n) = \frac{1}{2} \left(1 - \cos\left(2\pi \frac{n}{2\alpha f}\right) \right), 0 < n < 2\alpha f \quad (5)$$

Avec H_{vb} étant la fonction de transfert lissée et w la fenêtre de Hanning qui est appliquée pour le lissage. Ce traitement est ainsi calculé à chaque fréquence où la DTF est définie.

2.2 Critère objectif de mesure

Pour quantifier les dégradations qui sont créées par les méthodes de lissages préalablement définies, il est nécessaire de définir un critère objectif de mesure qui prenne en compte les différences entre le spectre original et le spectre lissé.

2.2.1 Inter Subject Spectral Difference (ISSD)

Le critère de l'ISSD comme mesure objective pour comparer des DTF a été pris en compte dans un premier temps par Middlebrooks [12], puis a été utilisé par Guillon [8]. Ce critère est utilisé comme un premier indice de la distance entre deux spectres de DTF.

Le calcul de l'ISSD pour une direction (θ, ϕ) donnée est fait comme suit. Soit les fonctions de transfert $DTF_A(\theta, \phi, f)$ et $DTF_B(\theta, \phi, f)$ de deux sujet A et B exprimées en dB pour chaque fréquence f . La différence spectrale entre les sujet pour cette direction est égale à :

$$ISSD_{\theta, \phi} = \text{var}(DTF_A(\theta, \phi, f) - DTF_B(\theta, \phi, f)) \quad (6)$$

Où var est la variance calculée par rapport au fréquences. L'ISSD globale est calculée comme la moyenne de toutes les ISSD directionnelles. Celui-ci s'exprime en dB^2 .

$$ISSD = \sum_{\theta} \sum_{\phi} ISSD_{\theta, \phi} \quad (7)$$

2.2.2 Inter Subject logSpectral Difference (logISSD)

Une variante du critère de l'ISSD proposée dans ces travaux est le critère du logISSD. Ce critère repose, une fois de plus, sur les bases de la perception des sons.

Le critère de l'ISSD calcule la variance d'une différence spectrale entre deux DTF. Or, ces DTF sont calculées par une transformée de Fourier discrète des HRIR mesurées en chambre anéchoïque et échantillonnées à une fréquence F_e . Cette transformation donne les valeurs complexes de la

HRTF pour un ensemble de valeurs de fréquences réparties de manière linéaire entre 0 et F_e . Le nombre de valeurs calculées dépend de la taille de la transformée choisie. Donc, le critère de l'ISSD donne le même poids aux différences spectrales qui sont présentes dans les basses fréquences et dans les hautes fréquences.

Il est plus pertinent de donner plus d'importance aux erreurs présentes dans les basses fréquences étant donné que le système auditif a une meilleure résolution perceptive dans cette zone du spectre.

Donc, un ré-échantillonnage des DTF suivant une échelle logarithmique est réalisé avant de calculer le critère du ISSD décrit par Guillon [8]. Ainsi, plus de poids est donné aux erreurs commises dans les basses fréquences et moins de poids est attribué aux erreurs commises dans les hautes fréquences.

Un critère très similaire à été décrit par Huopaniemi [10] appelé le Spectral Distance Measure (SDM). Celui-ci calcule une valeur RMS au lieu d'une variance de la différence des spectres égalisés et ré-échantillonnés.

Le SDM représente donc un écart type de la différence entre deux spectres et le logISSD représente la variance de cette différence qui est calculée indépendamment fréquence par fréquence.

3 Test subjectifs

3.1 Méthodes testées

Cinq algorithmes de lissage ont été choisis pour faire l'analyse subjective et objective du lissage de DTF. Les algorithmes choisis sont :

- Lissage par bande de Bark
- Lissage par bande de sixième d'octave (1/6 oct)
- Lissage par bande de douzième d'octave (1/12 oct)
- Lissage par fenêtre glissante de taille variable à 10% ($\alpha = 10\%$)
- Lissage par fenêtre glissante de taille variable à 20% ($\alpha = 20\%$)

Après avoir fait des traitements dans le spectre des filtres, leur réponse impulsionnelle est calculée en utilisant des filtres à phase minimale et en ajoutant le retard pur des HRTF originales.

Les méthodes de lissage par bande de fréquence donnent des spectres qui ont des changements brusques (pentes raides dans le spectre). Ceci crée des réponses impulsionnelles qui sont très longues lors du calcul des filtres à phase minimale. Pour éviter ce phénomène, la méthode de lissage par fenêtre glissante avec un paramètre de $\alpha = 4\%$ a été appliquée après toutes les méthodes de lissage par bandes de fréquences pour créer des spectres sans des changements brusques de valeurs. Cette valeur a été choisie expérimentalement après quelques tests.

3.2 Protocole de test

Les différentes méthodes de lissages ont été évaluées subjectivement avec un test MUSHRA [4]. Dans celui-ci, il est nécessaire de classer les signaux présentés par rapport à un signal de référence. Pour notre analyse, 8 signaux doivent être classés par rapport au signal de référence. Il y a les cinq

signaux correspondant aux méthodes de lissage testées, le signal de référence caché, et deux ancres basses composées du signal de référence filtré par des filtres passe-bas à 3,5 kHz et 7kHz.

Le test a été réalisé dans une cabine d'écoute avec des casques Beyerdynamic 990 PRO 250 Ohm, avec 12 experts habitués à faire de tests d'audio spatialisée. Les stimuli présentés dans le test sont des bruits blancs gaussiens de 500ms, ayant un « fade in » et un « fade out » de 10ms, convolués par une DTF non individualisée du sujet 1 de la base de données d'Orange Labs.

L'utilisation de DTF non individualisés pour le test subjectif est justifié par le fait que les signaux qui sont présentés aux auditeurs viennent d'une seule source à la fois et chaque auditeur ne connaît pas la position à laquelle la source est positionnée à l'origine. Donc, chaque personne doit donner une note à chaque condition par rapport aux différences de spatialisation qu'il perçoit entre cette condition et le signal de référence.

Le test comporte une phase d'entraînement composée de deux échantillons suivie du test composé de neuf échantillons. Un échantillon correspond à une direction de DTF. Les différentes positions des DTF prises en compte pour le test subjectif sont affichées dans le tableau 1. Elles ont été prises arbitrairement en priorisant l'hémisphère frontal de l'auditeur où il perçoit plus facilement les différences de spatialisation des sources [5].

TABEAU 1 – Élévation et azimuth des DTF choisis pour le test subjectif (Les azimuths 90, -90, 0 et 180 correspondent à la gauche, la droite, en face et à l'arrière de l'individu)

Élévation	-30	-30	-30	-30	0	0	30	30	30	45	90
Azimuth	-45	-15	0	135	0	90	-135	45	105	-90	0

La consigne globale donnée pour le test est de noter la qualité de chaque signal par rapport à la référence en prenant en compte les différences de spatialisation qui se présentent entre cette condition et la référence. Il est demandé au sujet de ne pas mettre deux fois la même note. Cette note varie entre 0 et 100, 100 représentant la note pour laquelle il n'y a pas de dégradation par rapport à la référence. L'interface de test est représentée en figure 1.

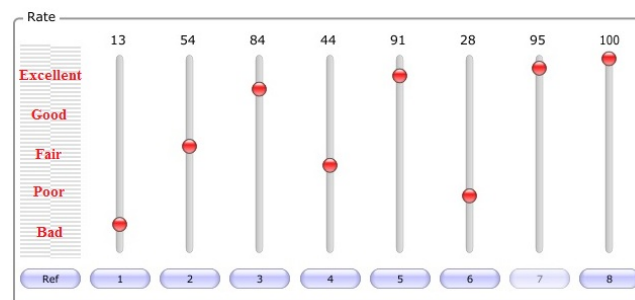


FIGURE 1 – Fenêtre de classification pour un échantillon du test MUSHRA.

La plage de valeurs est divisée en 5 intervalles de qualité : "Bad" (0 à 20), "Poor" (20 à 40), "Fair" (40 à 60), "Good" (60 à 80) et "Excellent" (80 à 100).

4 Résultats

4.1 Résultats objectifs

Pour illustrer le résultat issu des différents traitements, les figures 2 et 3 montrent les spectres obtenus avec les diverses méthodes de lissage et le plan horizontal d'un même jeu de DTF respectivement. Les dissimilitudes entre les différents lissages sont plus facilement remarquables dans les zones avec de nombreux pics et creux importants (partie droite de chaque sous-ensemble de la figure 3).

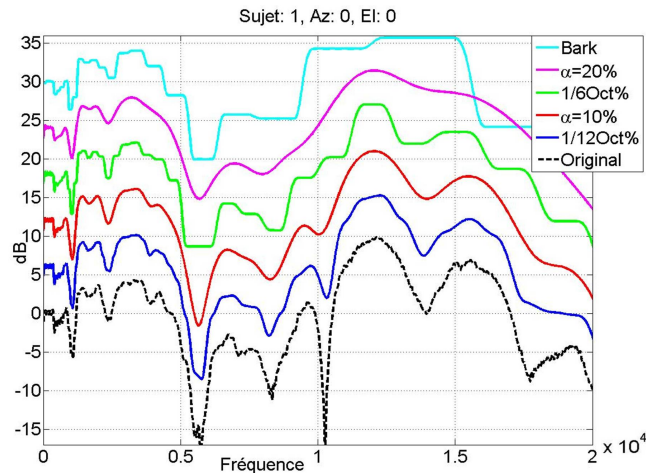


FIGURE 2 – Spectres (en dB) des différentes HRTF utilisées dans le test subjectif pour une direction.

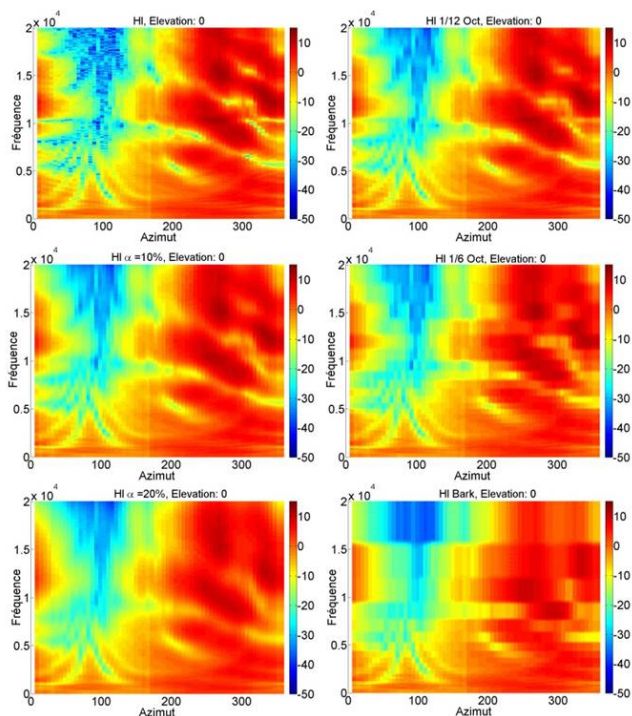


FIGURE 3 – Gain en dB dans le plan horizontal des HRTF (oreille gauche) utilisées dans le test subjectif. De gauche à droite et du haut vers le bas sont représentés les HRTF sans lissage, 1/12 oct, $\alpha = 10\%$, 1/6 oct, $\alpha = 20\%$ et Bark respectivement.

Pour quantifier la dégradation qui est faite par chaque méthode de lissage, le critère du logISSD a été calculé pour

la DTF qui a été utilisée pour faire les tests subjectifs. Elle correspond aux mesures du sujet 1 de la base de données d'Orange Labs. Cette base de données est une base de données non publiée qui est constituée de 8 individus et qui comprend 695 mesures dans l'espace. Les résultats calculés pour cet individu se montrent dans le tableau 2. De même, la moyenne et la variance de toute la base d'Orange Labs est donnée pour information.

TABEAU 2 – Valeurs du logISSD (en dB^2) calculé suivant plusieurs types de lissages pour une personne et pour toute la base de données d'Orange Labs (OL) (valeur moyenne et variance)

Lissage/BD	Sujet 1 (OL)	Moyenne (OL)	Variance (OL)
1/12 Octave	0.647	0.710	$4.818 * 10^{-3}$
$\alpha = 10\%$	0.827	0.880	$7.083 * 10^{-3}$
1/6 Octave	1.277	1.307	$1.237 * 10^{-2}$
$\alpha = 20\%$	1.627	1.677	$1.308 * 10^{-2}$
Bark	2.760	2.834	$5.441 * 10^{-3}$

Les résultats objectifs permettent de faire une première classification des différentes méthodes de lissage testées. En effet, le lissage par bandes de douzième octave a la valeur de logISSD plus petite de tous les traitements testés. Elle est suivie par le lissage par fenêtre glissante de taille variable ($\alpha = 10\%$), le lissage par bandes de sixième octave, le lissage par fenêtre glissante de taille variable ($\alpha = 20\%$) et le lissage par bandes de Bark respectivement. Cette classification est cohérente avec les distorsions qui se voient à l'oeil nu dans la figure 3 en comparant les différents types lissages avec les HRTF mesurées (sans lissage).

4.2 Résultats subjectifs

Les résultats des tests subjectifs sont présentés sur la figure 4. Les valeurs présentées sont issues des résultats obtenus pour les 12 sujets de tests et pour les 9 directions de l'espace testées.

Les tests subjectifs montrent que deux des lissages (lissage par bandes de douzième octave et le lissage par fenêtre glissante de taille variable ($\alpha = 10\%$)) ont des valeurs médianes et moyennes qui sont très proches des valeurs obtenues par la référence. La première méthode de lissage semble être meilleure que la deuxième parce que les valeurs médiane, moyenne et du Q3 (le troisième quartile) sont légèrement plus élevée pour la première que pour la deuxième. Ces deux méthodes de lissage créent moins de dégradations de spatialisation que toutes les autres méthodes de lissage testées.

Les autres méthodes de lissage (sixième d'octave, $\alpha = 20\%$ et Bark) créent de plus en plus de dégradations de spatialisation respectivement. En effet, les résultats obtenus par ces trois méthodes varient beaucoup par rapport à la direction choisie et par rapport aux critères d'évaluation de chaque sujet. Ainsi, le lissage par bandes de sixième octave a des performances "Excellentes" ou "Bonnes", la méthode de lissage par fenêtre glissante de taille variable $\alpha = 20\%$

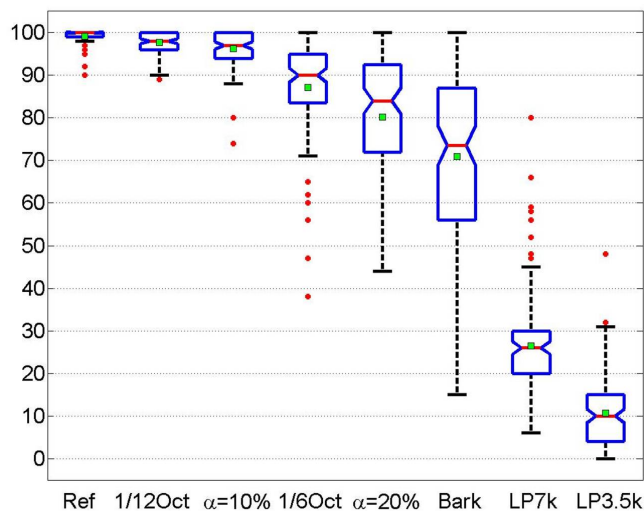


FIGURE 4 – Résultats du test MUSHRA pour chacune des conditions du test subjectifs. Ils sont exprimés par la médiane (ligne rouge), son intervalle de confiance à 95% (lignes bleues en diagonale), la moyenne (carré vert) et les quartiles Q1 et Q3 (lignes bleues horizontales).

a des performances "Excellentes" ou "Assez Bonnes" et le lissage par bandes de Bark a des performances "Excellentes" ou "Mauvaises".

4.3 Analyse des résultats

Les résultats objectifs et subjectifs sont cohérents par rapport à la classification des différentes méthodes de lissage testées. En effet, les deux résultats donnent l'ordre suivant pour les méthodes de lissage, en partant de celle qui crée le moins de dégradation à celle qui crée le plus de dégradation : 1/12 Octave, $\alpha = 10\%$, 1/6 Octave, $\alpha = 20\%$ et Bark.

Le calcul de la corrélation entre les moyennes des résultats objectifs et subjectifs donne une corrélation de 0.9667. Cette valeur de corrélation, qui est proche de 1, montre que le critère objectif du logISSD est fortement corrélé aux dégradations perçues entre une HRTF et sa version dégradée.

Les valeurs moyennes des résultats subjectifs ont été représentées en fonction des valeurs moyennes des résultats objectifs en figure 5.

Les valeurs des filtres passe-bas n'ont pas été incluses dans la représentation parce que leur valeur objective dépend du gain des filtres au delà de la fréquence de coupure. A titre indicatif, les filtres passe bas appliqués ont un gain de $-96dB$ dans cette zone là. Leur valeur de logISSD, pour le sujet 1 de la base d'Orange Labs, est de 607.13 pour le filtre passe-bas à 7kHz et de 953.10 pour le filtre passe-bas à 3.5kHz.

La figure 5 montre qu'avec les résultats des différents algorithmes de lissage testés il est possible de modéliser la dégradation spatiale obtenue à partir de la valeur du logISSD. En effet, une ligne de tendance ayant l'équation $y = -11.51x + 102.3$ permet de modéliser les données obtenues dans cette étude avec une corrélation de 93.45%. Cette fonction permettrait d'avoir une idée de la dégradation spatiale causée par une autre méthode lissage qu'à partir de sa valeur de logISSD. Pour avoir des résultats plus cohérents avec le test MUSHRA il faut reformuler l'équation pour ne pas dépasser la limite haute : $y = \min(-11.51x + 102.3, 100)$.

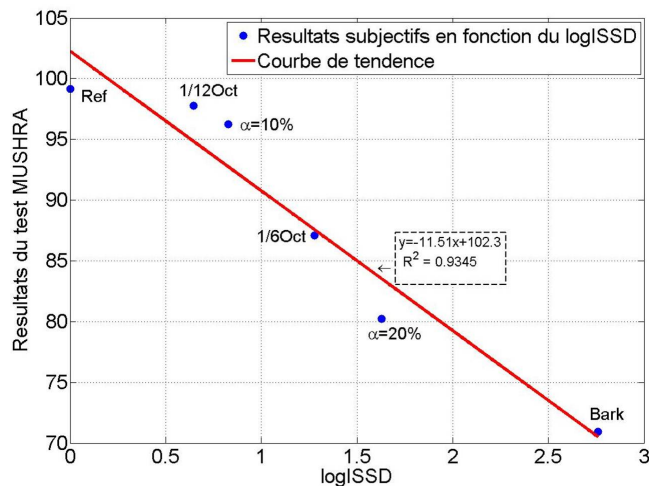


FIGURE 5 – Résultats moyens subjectifs en fonction des résultats moyens objectifs. Une ligne de tendance est montrée en rouge avec son équation et le coefficient de corrélation avec les données.

5 Conclusion

Les résultats obtenus dans cette étude montrent qu'il est possible de lisser des HRTF mesurées en chambre sourde par des bandes de douzième d'octave ou par une méthode de fenêtre glissante à taille variable ($\alpha = 10\%$) en ayant une excellente qualité dans la spatialisation de la synthèse binaurale. Ces méthodes donnent des bonnes performances pour toutes les directions testées dans cette étude.

Un de ces deux algorithmes de lissages peut donc être inclus dans l'étape de pré-traitement des HRTF issues d'une campagne de mesure. Ceci permet de réduire la variance des données avant de faire des analyses statistiques telles que la décomposition en composantes principales ou en harmoniques sphériques.

D'autre part, les résultats obtenus montrent que le critère objectif du logISSD est fortement corrélé avec les résultats obtenus dans les tests subjectifs. En effet, une fonction pour relier les valeurs du critère objectif avec a été trouvée. Il faut préciser que le critère objectif a été utilisé pour comparer une HRTF et sa version dégradée seulement avec des méthodes de lissage par bandes de fréquence ou par fenêtre glissante de taille variable.

D'autres études restent à faire avec d'autres méthodes de lissage ou autres algorithmes d'individualisation de HRTF pour valider la pertinence du critère objectif du logISSD dans la comparaison de HRTF mesurés et de ses version dégradées.

6 Remerciements

Ces travaux ont été menés dans le cadre du projet "BiLi" (www.bili-project.org), financé par le FIU, et avec le support de "CAP DIGITAL - PARIS REGION".

Références

- [1] ISO 266 :1997 Acoustics–Preferred frequencies. pages 1–3.

- [2] Jens Blauert. *Spatial Hearing : The Psychophysics of Human Sound Localization*. MIT Press, 1997.
- [3] Jens Blauert, Marc Brueggen, Adelbert W. Bronkhorst, Rob Drullman, Gerard Reynaud, Lionel Pellioux, Winfried Kriebber, and Roland Sottek. The AUDIS catalog of human HRTFs. *The Journal of the Acoustical Society of America*, 103(5) :3082–3082, May 1998.
- [4] International Telecommunications Union Radiocommunication Sector BS.1534. Method for the subjective assessment of intermediate quality level coding systems– general requirements. Technical report, Geneva 2001.
- [5] S Carlile, C Jin, and J Leung. Performance measures of the spatial fidelity of virtual auditory space : effects of filter compression and spatial sampling. pages 375–379, 2002.
- [6] Bill Gardner and Keith Martin. HRTF measurements of a KEMAR dummy-head microphone. Technical report, MIT Media Lab Perceptual Computing, 1994.
- [7] C. Daniel Geisler. *From Sound to Synapse : Physiology of the Mammalian Ear*. Oxford University Press, January 1998.
- [8] P. Guillon. *Individualisation des indices spectraux pour la synthèse binaurale : recherche et exploitation des similarités inter-individuelles pour l'adaptation ou la reconstruction de HRTF*. PhD thesis, Université du Maine, November 2009.
- [9] Navarun Gupta, A. Barreto, M. Joshi, and J.C. Agudelo. HRTF database at FIU DSP lab. In *2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pages 169–172, March 2010.
- [10] Jyri Huopaniemi and Matti Karjalainen. Review of digital filter design and implementation methods for 3-d sound. In *Audio Engineering Society Convention 102*. Audio Engineering Society, 1997.
- [11] Doris J. Kistler and Frederic L. Wightman. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *The Journal of the Acoustical Society of America*, 91(3) :1637–1647, 1992.
- [12] John C. Middlebrooks. Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. *The Journal of the Acoustical Society of America*, 106 :1493, 1999.
- [13] John C. Middlebrooks, James C. Makous, and David M. Green. Directional sensitivity of sound-pressure levels in the human ear canal. *The Journal of the Acoustical Society of America*, 86(1) :89–108, 1989.
- [14] Henrik Møller, Michael Friis Sørensen, Dorte Hammershøi, and Clemen Boje Jensen. Head-related transfer functions of human subjects. *Journal of the Audio Engineering Society*, 43(5) :300–321, May 1995.
- [15] William H. Slattery III and John C. Middlebrooks. Monaural sound localization : Acute versus chronic unilateral impairment. *Hearing Research*, 75 :38–46, 1994.
- [16] Smith. *Techniques for Digital Filter Design and System Identification with Application to the Violin*. PhD thesis, Stanford University, 1983.
- [17] Bosun Xie. *Head-Related Transfer Function and Virtual Auditory Display*. J ROSS PUB Incorporated, May 2013.
- [18] E. Zwicker. Subdivision of the audible frequency range into critical bands (frequenzgruppen). *The Journal of the Acoustical Society of America*, 33(2) :248–248, 1961.
- [19] Eberhard Zwicker and Ernst Terhardt. Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. *The Journal of the Acoustical Society of America*, 68 :1523, 1980.