CFA/VISHNO 2016

Estimation des HRTFs Individuelles sur la Base d'Enregistrements Binauraux en Conditions non Supervisées

M. Maazaoui et O. Warusfel
UMR STMS, IRCAM-CNRS-UPMC Sorbonne Universités, 1 place Igor Stravinsky, 75004 Paris, France
mounira.maazaoui@ircam.fr

Les fonctions de transfert de tête (HRTFs : Head Related Transfer Functions) sont les attributs clés de la spatialisation binaurale sur casque audio. Ces filtres sont spécifiques à chaque individu et requièrent un protocole de mesure rigoureux et complexe en conditions anéchoïques (signaux calibrés, repérage spatial, nombre important de directions, etc.). L'accès à de telles mesures pour le grand public n'est par conséquent pas envisageable, cependant l'utilisation de filtres non individuels peut provoquer des artefacts perceptifs. Différentes recherches ont été consacrées à l'élaboration de méthodes alternatives évitant le recours à des mesures en conditions anéchoïques. Certaines sont basées sur la modélisation numérique, d'autres sur une adaptation paramétrique ou sur la sélection de HRTFs non-individuelles dans une base de données. L'approche proposée dans cette étude relève de cette dernière catégorie en se basant sur un enregistrement binaural de l'auditeur effectué en environnement non contrôlé.

1 Introduction

Les systèmes de spatialisation sonore binaurale sont basés sur l'utilisation de filtres modélisant les transformations que subissent les ondes sonores par effet de diffraction des oreilles, de la tête et du torse de l'auditeur. Ces filtres, dénommés HRTFs (Head Related Transfer Functions : fonctions de transfert de tête) sont propres à la morphologie de chaque individu. Lors d'une synthèse binaurale, l'utilisation de HRTFs non adaptées peut provoquer des artéfacts perceptifs tels qu'un défaut de localisation. Cependant, l'acquisition de HRTFs individuelles nécessite un processus long et complexe reposant en particulier sur des mesures en chambre anéchoïque.

De nombreux travaux de recherche sont menés afin de s'affranchir de la mesure en conditions contrôlées et proposer des solutions alternatives, moins coûteuses en temps et en matériel et adaptées au grand public. Parmi ces méthodes d'individualisation, nous trouvons celles basées sur la modélisation numérique des HRTFs comme FEM (Finite Element Method) et BEM (Boundary Element Method) [6, 5]. Dans cette technique, les filtres sont obtenus par estimation numérique de la transformation appliquée à l'onde sonore se propageant de la source jusqu'à l'entrée des canaux auditifs. Cette méthode nécessite l'acquisition du maillage 3D de l'auditeur ce qui peut être difficile à obtenir. L'individualisation des HRTFs peut aussi être réalisée par l'adaptation de HRTF non-individuelle comme par exemple par morphisme (scaling) fréquentiel [8] ou la rotation de la fonction de directivité [4, 7]. Les auteurs montrent que la différence entre l'orientation du pavillon de deux individus peut se résumer à une combinaison des opérations de décalage par rotation et par un facteur d'échelle sur les fréquences.

Dans l'approche proposée, la méthode d'individualisation de HRTFs repose sur une sélection au sein d'une ou plusieurs bases de données publiques de HRTFs mesurées en conditions anéchoïques [8, 4]. Cette sélection s'effectue à partir d'un enregistrement binaural de l'auditeur acquis dans un environnement non contrôlé (milieu réverbérant, signaux quelconques, points de mesures parcimonieux, sujet et sources en mouvement). L'identification du jeu de HRTFs le plus proche repose sur une métrique dérivée du modèle de localisation auditive par égalisation-annulation (Equalization-Cancellation) [1]. Ce principe consiste à trouver dans le signal enregistré, et à chaque instant le retard interaural de phase ou d'enveloppe et le gain interaural qui minimisent un terme d'erreur résiduelle entre les signaux gauche et droit. L'indice de décision est établi à partir de la recherche du couple d'HRTFs de la base de données dont les paramètres interauraux de retard et de gain permettent d'approcher au mieux cette erreur résiduelle. Les performances de la méthode sont estimées sur un ensemble de situations simulées (milieu réverbérant, source statique ou en mouvement, . . .).

2 Principe de la détection par égalisation-annulation

Cette méthode est dérivée du modèle d'audition de Durlach [3, 1]. Si un signal audio cible est masqué par une source interférente, le système auditif cherche à annuler le signal masquant en multipliant et retardant le signal à une oreille pour que le signal masquant soit identique à gauche et à droite (égalisation), puis en soustrayant les deux signaux (annulation). C'est le principe d'égalisation-annulation (Equalization-Cancellation : EC). Dans le but d'effectuer une tâche automatique de détection, Baskind a dérivé le modèle de Durlach afin d'annuler l'éventuel signal source dans le but de détecter *a posteriori* sa présence ou non [1]. Le signal cible est mélangé à un « bruit » qui correspond principalement au champ réverbéré, décorrélé d'une voie à l'autre sauf dans les basses fréquences.

Supposons une source sonore ponctuelle dans un environnement anéchoïque capturée par deux microphones dans un champ libre. Soient $x(t)$ et $y(t)$ les signaux enregistrés. Le signal $y(t)$ peut être écrit dans le cas sans bruit en fonction du signal $x(t)$ en introduisant un gain A et un retard Δ :

$$y(t) = Ax(t-\Delta) \tag{1}$$

L'erreur d'égalisation-annulation absolue est définie comme l'énergie de la différence des termes gauche et droit après application d'un gain et d'un retard sur le signal $x(t)$ [1] :

$$D_{\alpha,\tau} = \sum_t (y(t) - \alpha x(t-\tau))^2 \tag{2}$$

Cette erreur EC peut être exprimée en utilisant les énergies E_x et E_y des signaux $x(t)$ et $y(t)$ respectivement et leur corrélation croisée C_{xy} :

$$D_{\alpha,\tau} = E_y + \alpha^2 E_x - 2\alpha C_{xy}(\tau) \tag{3}$$

et peut aussi être normalisée comme suit :

$$\varepsilon_{\alpha,\tau} = \frac{D_{\alpha,\tau}}{D_{\alpha,\tau} + S_{\alpha,\tau}} \tag{4}$$

où $S_{\alpha,\tau} = \sum_n (y(t) + \alpha x(t-\tau))^2$. Comme $D_{\alpha,\tau}$ et $S_{\alpha,\tau}$ sont positifs ou nuls $\forall(\alpha,\tau)$, l'erreur EC normalisée est toujours bornée par 0 et 1, et peut s'écrire comme suit :

$$\varepsilon_{\alpha,\tau} = \frac{1}{2} - \frac{\alpha C_{xy}(\tau)}{E_y + \alpha^2 E_x} \tag{5}$$

Le minimum de l'erreur EC normalisée est obtenu pour un retard optimal τ^{opt} et un gain optimal α^{opt} :

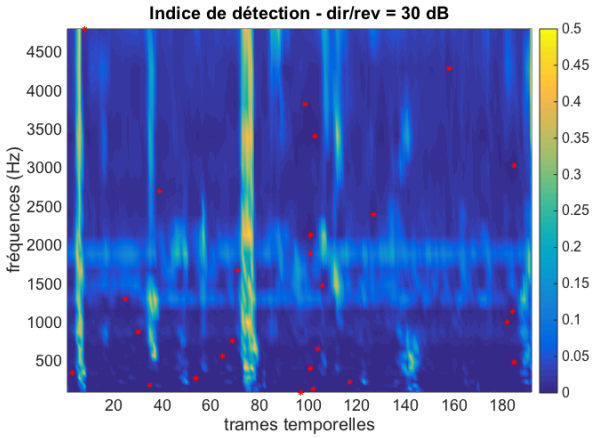
$$\begin{cases} \tau^{opt} = \arg\max_{\tau} \left\{ C_{xy}(\tau) \right\} \\ \alpha^{opt} = sgn\left(C_{xy}(\tau^{opt}) \right) \sqrt{\frac{E_y}{E_x}} \end{cases} \quad (6)$$

où $sgn(.)$ est la fonction signe. Par conséquent, l'erreur EC normalisée optimale est :

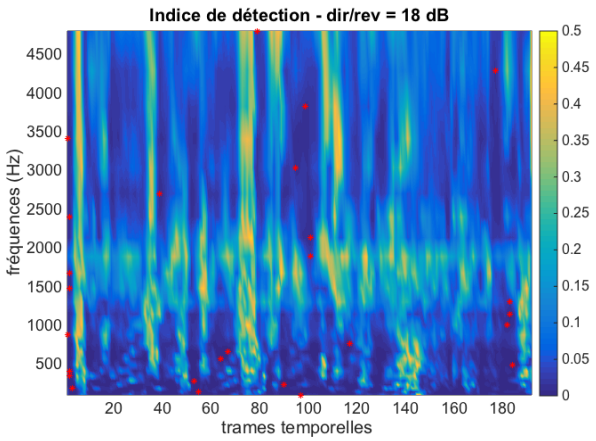
$$\varepsilon_{\alpha^{opt},\tau^{opt}} = \varepsilon_{min} = \frac{1}{2}\left(1 - \left|\rho_{xy}\left(\tau^{opt}\right)\right|\right) \quad (7)$$

La méthode de la minimisation de l'erreur EC normalisée convient aux problèmes de détection parce qu'elle donne un indice, ε_{min}, qui est borné et indépendant des puissances des signaux à l'entrée. Cet indice est appelé *indice de détection* et est bornée par 0 et $\frac{1}{2}$ (*cf.* Figure 1) :

- si $\varepsilon_{min} = 0$ (les tons bleus de la Figure 1) : les signaux ne contiennent pas de bruit additionnel, le modèle gain-retard est une description précise de la réalité parce que $y(t) = \alpha^{opt} x(t - \tau^{opt}), \forall t$.
- si $\varepsilon_{min} = \frac{1}{2}$ (les tons jaunes de la Figure 1) : les signaux $x(t)$ et $y(t)$ sont complètement décorrélés et le modèle gain-retard ne convient pas aux signaux observés.



(a)



(b)

FIGURE 1 – L'indice de détection ε_{min} estimé sur un signal de parole pour des fréquences $f \in [100\text{ Hz}, 5000\text{ Hz}]$ et un temps de réverbération de 250 ms (a) champ direct sur réverbéré = 30 dB (b) champ direct sur réverbéré = 12 dB

3 Individualisation des HRTF par égalisation-annulation

Supposons que nous disposons d'une base de données de réponses impulsionnelles de tête HRIRs (Head Related Impulse Responses) d'un sujet enregistrées pour D directions (θ, ϕ) $\left\{\mathbf{h}^n_{L,R}\left(\theta_i, \phi_j\right)\right\}_{1\le i\le D_{az}, 1\le j\le D_{el}}$, les indices L, R font référence aux oreilles gauche (left) et droite (right). Les directions peuvent être divisées en D_{az} angles d'azimut et D_{el} angles d'élévation tels que le nombre total de directions est $D = D_{az} \times D_{el}$. Rappelons que les HRTF sont la représentation fréquentielle des HRIRs.

Soient $\mathbf{x}_L$ et $\mathbf{x}_R$ un signal audio binaural enregistré avec deux microphones placés dans les oreilles d'un sujet, où $\mathbf{x}_L = [x_L(t_1), \ldots, x_L(t_T)]$, $\mathbf{x}_R = [x_R(t_1), \ldots, x_R(t_T)]$ sont les signaux échantillonnés reçus respectivement par les oreilles gauche et droite et T est le nombre d'échantillons de l'enregistrement. $\mathbf{x}_L$ et $\mathbf{x}_R$ peuvent être modélisés comme la convolution d'un signal source et d'une HRIR dont la direction correspond à la direction d'arrivée (DOA : direction-of-arrival) du signal source audio avec un bruit additionnel correspondant au bruit de fond électronique et à la réverbération tardive. Nous supposons que la réponse impulsionnelle peut être modélisée à chaque fréquence par un gain et un retard pur.

Le principe de la méthode d'égalisation-annulation adaptée à la détection de la DOA d'une source cible consiste à chercher dans une base de données de HRTFs une différence interaurale de phase (IPD) et une différence interaurale de gain (IGD) qui minimisent la différence résiduelle entre les signaux gauche et droit après égalisation-annulation. Cette différence résiduelle est appelée *indice de décision*. Le couple d'HRTFs qui vérifie cette condition est le couple de filtres optimal, sa direction correspond à la DOA de la source cible [1].

Nous étendons cette idée à l'individualisation des HRTFs. Supposons que la base de données des HRIRs a été mesurée pour N sujets : $\left\{\mathbf{h}^n_{L,R}\left(\theta_i, \phi_j\right)\right\}^{1\le n\le N}_{1\le i\le D_{az}, 1\le j\le D_{el}}$. La différence résiduelle entre les signaux gauche et droit après égalisation-annulation est minimale non seulement pour les HRTFs dont les IPD et IGD correspondent à la DOA du signal source, mais aussi pour ceux correspondant au sujet le plus proche de l'auditeur. La sélection de la HRTF optimale est réalisée en 4 étapes : pré-traitement de la base de données des HRTFs et du signal binaural enregistré, estimation des paramètres d'égalisation-annulation de la base de données et du signal binaural enregistré, estimation de l'indice de décision et estimation de la HRTF optimale (direction et sujet).

3.1 Pré-traitement

Chaque HRIR $\mathbf{h}^n_{L,R}\left(\theta_i, \phi_j\right)$ de la base de données est filtrée en utilisant un banc de filtre gammatone pour une plage de fréquence $[f_{min}, f_{max}]$. Nous obtenons $\mathbf{h}^n_{L,R}\left(f, \theta_i, \phi_j\right) = \mathbf{h}^n_{L,R}\left(\theta_i, \phi_j\right) \star \mathbf{g}(f)$ où $\mathbf{g}(f) = [g(t_1, f), \ldots, g(t_{T'}, f)]$, $g(t, f) = t^{n-1} e^{-2\pi bt} \cos(2\pi f t)$ est un filtre gammatone avec des fréquences centrales f, $f_{min} \le f \le f_{max}$, un ordre n, une longueur T' et une largeur de bande b, t est le temps en seconde. Le signal binaural enregistré fait l'objet du même pré-traitement.

3.2 Estimation des paramètres égalisation-annulation

Les paramètres EC (*cf.* section 2) estimés sur les signaux et leurs enveloppes sont :

1. pour la base de données des HRTFs : le retard optimal $\tau^{opt}\left(n,\theta_i,\phi_j,f\right)$ et le gain optimal $\alpha^{opt}\left(n,\theta_i,\phi_j,f\right)$ liés à la HRTF $\mathbf{h}^n_{L,R}\left(f,\theta_i,\phi_j\right)$ pour tous les sujets $n=1,\dots N$, toutes les fréquences $f \in [f_{min}, f_{max}]$ et toutes les directions de la base de données (θ_i,ϕ_j), $1 \leq i \leq D_{az}, 1 \leq j \leq D_{el}$.
2. pour le signal binaural enregistré : l'erreur d'égalisation-annulation normalisée de la source cible $\varepsilon^t_{\tau^t,\alpha^t}(f)$, la corrélation croisée des signaux gauche et droit $C^t_{LR}(f)$ et les énergies des signaux gauche et droit $E^t_L(f)$ et $E^t_R(f)$, l'exposant ^{t} fait référence à la source cible (target).

Les paramètres EC de la base de données des HRTFs sont estimés hors-ligne et stockés pour être utilisés dans l'estimation de l'indice de décision à l'étape suivante.

3.3 Estimation de l'indice de décision

Le processus d'individualisation consiste à trouver dans la base de données des HRTFs le retard optimal $\tau^{opt}(n^*,\theta^*,\phi^*,f)$ et le gain optimal $\alpha^{opt}(n^*,\theta^*,\phi^*,f)$, liés à la direction optimale (θ^*,ϕ^*) et au sujet n^*, qui minimisent le résidu intéraural ou l'indice de décision suivant [2, 1] :

$$d_{EC}(f,\theta,\phi,n) = \varepsilon_{\alpha(n,\theta,\phi,f),\tau(n,\theta,\phi,f)}(f) - \varepsilon^t_{\tau^t,\alpha^t}(f) \qquad (8)$$

où

$$\varepsilon_{\alpha(n,\theta,\phi,f),\tau(n,\theta,\phi,f)} = \frac{1}{2} - \frac{\alpha^{opt}(n,\theta,\phi,f)\,C^t_{LR}\left(\tau^{opt}(n,\theta,\phi,f)\right)}{E^t_R(f) + \alpha^{opt}(n,\theta,\phi,f)^2\,E^t_L(f)} \qquad (9)$$

Ceci peut être interprété comme une décision au sens du maximum de vraisemblance [1]. En effet, estimer l'indice $d_{EC}(f,\theta,\phi,n)$ pour une direction donnée (θ,ϕ) et un sujet donné n revient à estimer l'erreur EC en supposant que la source vient de la direction (θ,ϕ) et que l'auditeur est le sujet n. Dans cette hypothèse, l'erreur EC est liée à la puissance du bruit résiduel gauche et droit, qui représente en réalité les portions du signal qui ne pouvaient pas être modélisées par égalisation-annulation avec le retard et le gain associés à la direction (θ,ϕ) et au sujet n. Par conséquent, la décision consiste à choisir la direction et le sujet les plus vraisemblables qui maximisent le rapport signal sur bruit.

Afin de tenir compte de la différence du mécanisme auditif entre les basses et hautes fréquences, l'erreur EC normalisée d'enveloppe est introduite. Ses paramètres EC sont estimés sur les enveloppes des signaux gauche et droit. L'indice de décision correspondant est noté $d^{env}_{EC}(f,\theta,\phi)$. Ces deux indices de décision sont combinés comme suit [1] :

$$D_{EC}(f,\theta,\phi,n) = \frac{w_f.d_{EC}(f,\theta,\phi,n) + w^{env}_f.d^{env}_{EC}(f,\theta,\phi,n)}{w_f + w^{env}_f} \qquad (10)$$

où w_f et w^{env}_f sont des facteurs de pondération. Un indice de décision global est estimé en intégrant l'équation 10 sur la plage de fréquences [1] :

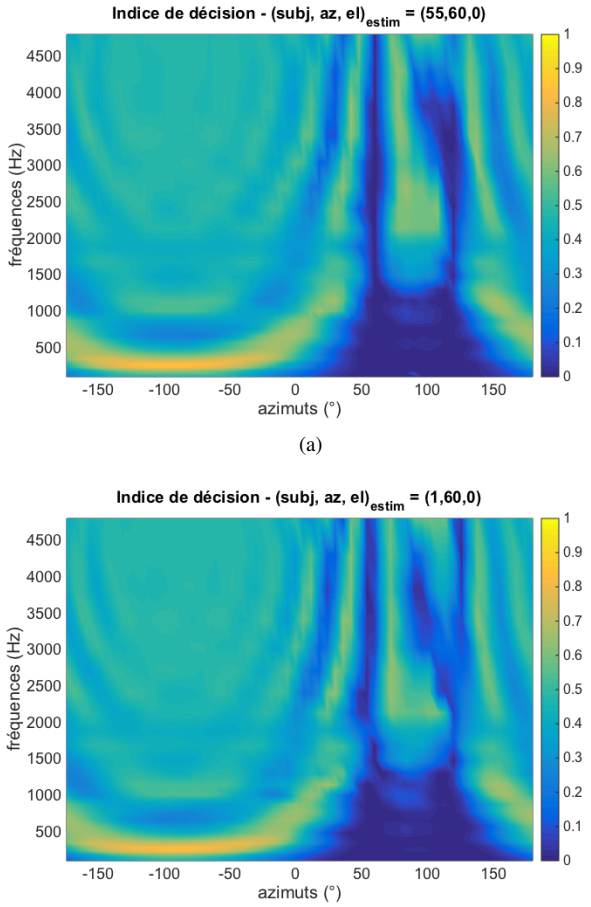


FIGURE 2 – L'indice de décision $D_{EC}(f,\theta,\phi^*,n)$ par bandes de fréquences $f \in [100\text{ Hz},\ 5000\text{ Hz}]$ et à l'élévation optimale (a) pour le sujet optimal $n = n*$, (b) pour un sujet différent du sujet optimal $n \neq n*$

$$D_{EC}(\theta,\phi,n) = \frac{\sum_f \left[w_f.d_{EC}(f,\theta,\phi,n) + w^{env}_f.d^{env}_{EC}(f,\theta,\phi,n)\right]}{\sum_f \left(w_f + w^{env}_f\right)} \qquad (11)$$

La Figure 2 présente l'indice de décision $D_{EC}(f,\theta,\phi,n)$ par bande de fréquence dans le cas où le sujet cible a bien été estimé et dans le cas où l'algorithme confond le sujet cible avec un autre sujet.

3.4 Estimation de la HRTF individualisée

La HRTF optimale est associée à la direction optimale (θ^*,ϕ^*) et au sujet optimal n^* qui minimisent l'indice de décision global $D_{EC}(\theta,\phi,n)$:

$$(\theta^*,\phi^*,n^*) = \underset{\theta,\phi,n}{\arg\min}\{D_{EC}(\theta,\phi,n)\} \qquad (12)$$

4 Expériences et résultats

4.1 Base de données des HRTFs

L'algorithme est évalué pour des situations de synthèse avec des signaux sources cibles statiques et mobiles. Dans les deux cas, le signal audio anéchoïque source est convolué avec une réponse impulsionnelle de salle binaurale synthétique composée d'un trajet direct et d'un champ

réverbéré diffus et générée par la librairie Spat˜ [9]. Nous avons utilisé les HRTFs de la base de données BILI enregistrée à l'IRCAM [10] qui comprend les filtres de 56 sujets mesurés dans 1680 directions.

Pour le cas des signaux sources statiques, un signal de parole de 3s a été utilisé comme signal source. Les HRTFs sont choisies parmi l'ensemble des sujets et selon soixante directions. Ces directions couvrent l'ensemble des soixante azimuts disponibles dans la base de données. Pour chaque azimut de 0° à 360°, un angle d'élévation est sélectionné selon une distribution normale centrée à l'élévation 0° et bornée entre -40° et 40°. Le signal binaural est simulé par convolution du signal de parole avec une réponse binaurale de salle (BRIR) synthétisée sous forme d'un trajet direct (utilisant la HRTF sélectionnée) et d'un champ diffus. Le temps de réverbération considéré est de 250ms et le rapport champ direct sur champ réverbéré varie de 30dB à 0dB. L'espace de recherche est constitué de la base des HRTFs BiLi-IRCAM, c'est à dire celle qui a servi à générer les stimuli tests.

Pour le cas des signaux sources cibles mobiles, nous avons simulé une source mobile dont l'azimut varie de 0° à 360° à une élévation constante de 0° et une vitesse angulaire constante de 10°/s. Le signal source est un bruit blanc, le temps de réverbération simulé est de $1,25$s et le rapport champ direct sur champ réverbéré (dir/rev) de 12dB.

Différents espaces de recherche sont successivement considérés cette fois. D'une part, les 56 sujets de la base des HRTFs BiLi-IRCAM ayant servi à générer les signaux binauraux. D'autre part, une base de données de 14 sujets mesurés à Orange-Labs (BiLi-ORANGE) selon une grille spatiale de 1560 directions et enfin les 74 sujets de la base Crossmod-Listen dont les HRTFs sont disponibles selon 710 directions. Pour chacune des ces bases, les HRTFs sont égalisées en champ diffus. Certains sujets ayant été mesurés dans chacune de ces bases, il sera intéressant d'observer si le modèle permet de retrouver ces sujets dans ces bases bien qu'elles n'aient pas directement servi à construire les signaux binauraux.

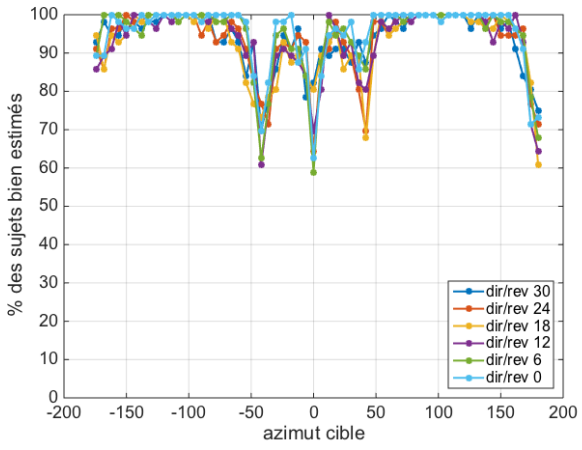


FIGURE 3 – Estimation en conditions statiques pour un signal de parole. Taux des sujets bien estimés en fonction de l'azimut cible et pour un temps de réverbération de 250ms et un rapport champ direct sur champ réverbéré = 30, 24, 18, 12, 6 and 0 dB

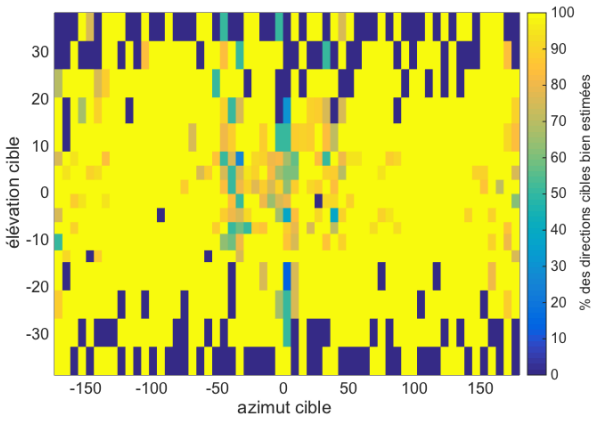


FIGURE 4 – Estimation en conditions statiques pour un signal de parole. Matrice des distributions des directions cibles et pourcentage des directions bien estimés pour un temps de réverbération 250ms et un rapport champ direct sur champ réverbéré de 6 dB - source statique

4.2 Résultats et analyses

4.2.1 Source statique

Nous commençons l'évaluation de l'algorithme d'individualisation avec le cas des sources cibles fixes. Les résultats sont obtenus avec une analyse à court-terme du signal binaural dont la BRIR est supposé stationnaire au court du temps : les paramètres EC de la source cible sont estimés sur des fenêtres d'analyse courtes, la fenêtre temporelle donnant le meilleur indice de détection est sélectionnée ainsi que ses paramètres EC (*cf.* Figure 1, trames indiquées par des astérisques rouges).

La figure 3 montre le taux des sujets bien estimés pour tous les azimuts et les différents rapports champ direct sur champ réverbéré évalués. D'après cette figure et pour les configurations testées, les sujets sont généralement bien estimés avec l'apparition de plages de directions d'estimation privilégiées. On peut notamment observer que les taux chutent considérablement pour les directions frontales (azimut = 0°) et arrière (azimut = 180°) pour lesquelles le modèle ne permet pas de distinguer les sujets.

La figure 4 présente la distribution des directions d'arrivées de la source cible ainsi que les directions bien estimées pour un rapport champ direct sur champ réverbéré de 6 dB. Elle montre généralement une bonne estimation de la direction d'arrivée de la source cible, cette performance baisse pour certaines directions comme présenté dans la Figure 3.

4.2.2 Source en mouvement

Dans le cas où la source cible est en mouvement, le signal binaural est découpé en fenêtres d'analyse à moyen terme (500ms), pendant lesquelles le canal acoustique (BRIR) est supposé stationnaire. Au sein de cette fenêtre, l'analyse est menée de manière similaire au cas statique précédemment exposé.

Les Figures 5 et 6 représentent l'azimut, l'élévation et le sujet estimés pour les deux sujets communs aux trois bases de HRTFs considérées. Le signal binaural a été synthétisé en utilisant les HRTFs des sujets 54 et 55 de la base de données BiLi-IRCAM. Ces sujets correspondent respectivement aux

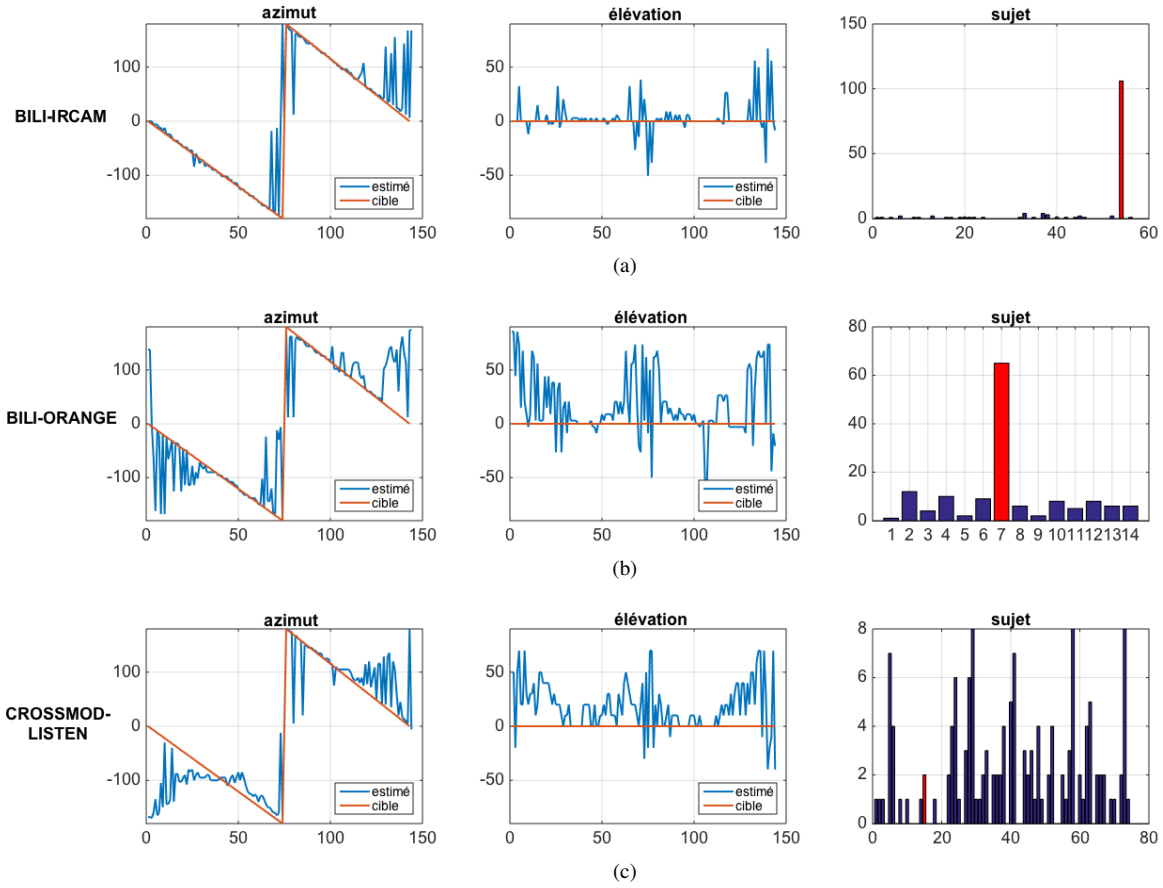


FIGURE 5 – Estimation de la direction instantanée (azimut et direction) de la source et histogramme des sujets détectés dans le cas d'une source en mouvement et en milieu réverbéré (TR = 1,25s, dir/rev = 12dB). Le signal binaural (bruit) est généré à partir des HRTFs du sujet 54 de la base BiLi-IRCAM. Les espaces de recherche du sujet sont successivement la base BiLi-IRCAM (5a), la base BiLi-ORANGE (5b) où le même sujet a été mesuré (sujet 7) et la base Crossmod-Listen (5c) dans laquelle le même sujet a été mesuré (sujet 15)

sujets 7 et 8 de la base de données BiLi-ORANGE (14 sujets) et 15 et 16 de la base de données Crossmod-Listen (74 sujets).

Dans le cas de la base de données BILI-IRCAM (Figures 5a et 6a), l'azimut est bien estimé comparé aux autres bases de données. Ceci peut s'expliquer par le fait que le signal binaural a été synthétisé à partir des HRTFs de cette même base de données. L'estimation de la direction se passe moins bien pour la base de données BILI-ORANGE (Figures 5b et 6b) mais le sujet est toutefois bien estimé. L'algorithme échoue à trouver les sujets cibles dans la base de données Crossmod-Listen (Figures 5c et 6c).

4.2.3 Discussion

Le modèle exploité dans cette étude vise à estimer conjointement la direction d'arrivée de la source et le sujet. Néanmoins, dans le cadre spécifique de cette étude, la motivation principale est l'identification du sujet le plus vraisemblablement associé à la séquence binaurale enregistrée plutôt que l'exactitude de la direction instantanée de la source. A ce titre, il est intéressant de constater en figures 5b et 6b, que malgré le bruit important d'estimation de la direction instantanée (notamment en élévation), le sujet correspondant au sujet cible se dégage statistiquement de manière nette sur l'ensemble de la séquence. En revanche, dans le cas de la base de données Crossmod-Listen, le modèle échoue à retrouver le sujet malgré une estimation moins bruitée de la direction instantanée. L'hypothèse peut être celle d'une résolution spatiale insuffisante de cette base de données. Une étude plus systématique doit être menée pour évaluer l'influence de la grille spatiale de la base de données. Par ailleurs, dans le cas où le sujet cible n'appartient pas à la base, une étude objective et perceptive devrait être réalisée pour évaluer si les sujets sélectionnés sont effectivement proches du sujet cible.

5 Conclusion

Dans cet article, nous avons présenté une méthode de sélection des HRTFs basée sur un enregistrement binaural effectué en conditions non anéchoïques. Cette méthode s'appuie sur un modèle de localisation auditive par égalisation-annulation et sur une recherche exhaustive dans des bases de données de HRTFs pour estimer conjointement la direction instantanée de la source cible et le sujet le plus vraisemblable. Cette méthode donne des résultats encourageants dans le cas de situations simulées avec une source fixe ou en mouvement. Dans la suite, nous nous intéresserons à l'évaluation de cet algorithme en utilisant des signaux issus d'enregistrements binauraux en environnement réel non simulé et non contrôlé.

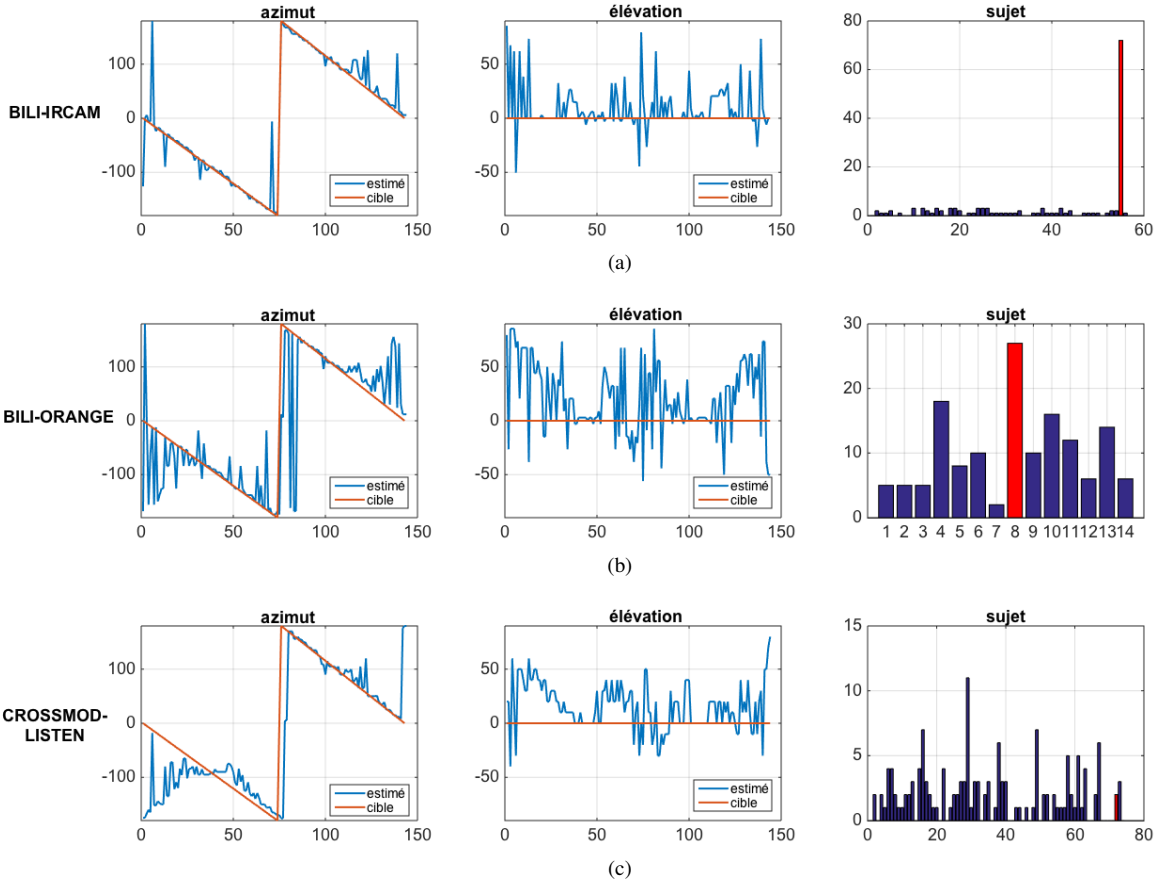


FIGURE 6 – Estimation de la direction instantanée (azimut et direction) de la source et histogramme des sujets détectés dans le cas d'une source en mouvement et en milieu réverbéré (TR = 1,25s, dir/rev = 12dB). Le signal binaural (bruit) est généré à partir des HRTFs du sujet 55 de la base BiLi-IRCAM. Les espaces de recherche du sujet sont successivement la base BiLi-IRCAM (6a), la base BiLi-ORANGE (6b) où le même sujet a été mesuré (sujet 8) et la base Crossmod-Listen (6c) dans laquelle le même sujet a été mesuré (sujet 72)

Remerciements

Cette étude a été menée dans le cadre du projet de Collaboratif FUI-AAP14 BiLi (Binaural Listening, www.bili-project.org) soutenu par le Pôle Cap Digital et cofinancé par la BPI, la Région Ile de France et la Ville de Paris.

Références

[1] Alexis Baskind. *Modèles et méthodes de description spatiale des scènes sonores - Application aux enregistrements binauraux*. PhD thesis, Université Paris 6, 2003.

[2] S. Bertet, J. Daniel, E. Parizet, and O. Warusfel. Investigation on Localisation Accuracy for First and Higher Order Ambisonics Reproduced Sound Sources. *Acta Acustica united with Acustica*, 99 :642–657, 2013.

[3] N. I. Durlach. Equalization and cancellation theory of binaural masking level differences. *Journal of The Acoustical Society of America*, 35 :1206–1218, 1963.

[4] 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, 2009.

[5] Yuvi Kahana. *Numerical modeling of the head-related transfer function*. PhD thesis, University of Southampton, Faculty of engineering and applied science, Institute of sound and vibration research, 2000.

[6] B. F. G. Katz. *Measurement and calculation of individual head-related transfer functions using a boundary element model including the measurement and effect of skin and hair impedance*. PhD thesis, The Pennsylvania state university, 1998.

[7] K. Maki and S. Furukawa. Reducing individual differences in the external-ear transfer functions of the mongolian gerbil. *The Journal of the Acoustical Society of America*, 118 :2392–2404, 2005.

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

[9] M. Noisternig T. Carpentier and O. Warusfel. Twenty years of ircam spat : Looking back, looking forward. *41st International Computer Muic Conference (ICMC)*, 2015.

[10] M. Noisternig T. Carpentier, H. Bahu and O. Warusfel. Measurement of a head-related transfer function database with high spatial resolution. *Forum Acousticium*, 2014.