



Des HRTF aux Object-RTF : Système de prise de son 3D pour dispositifs nomades

J. Palacino et R. Nicol
Orange Labs, 2 Avenue Pierre Marzin, 22300 Lannion, France
julianpalacino@gmail.com

Les systèmes de prise de son 3D disponibles aujourd'hui reposent sur 2 principales stratégies : soit la captation de tout ou partie des indices naturels de localisation auditive (stéréophonie ou technologie binaurale), soit la décomposition spatiale de l'onde acoustique sur une base de fonctions propres à partir d'un réseau microphonique (Higher Order Ambisonics par exemple). Une approche plus récente propose d'extraire directement l'information de direction des sources sonores à partir d'une prise de son par un microphone ambisonique à l'ordre 1 (Pulkki [14] et Berge [1]) permettant en même temps, la compression de l'information et la restitution spatiale. L'extraction de la direction est basée sur les propriétés du format B et les descripteurs de localisation qui lui sont associés [3, 4]. Dans un autre but, Keyrouz [5, 6] et MacDonald [8] ont présenté une méthode qui exploite l'information spatiale contenue dans les signaux binauraux en la comparant avec l'ensemble d'une base de HRTF (Head Related Transfer Function) pour identifier la direction d'une source sonore. Nous proposons ici une nouvelle approche de prise de son 3D inspirée du procédé de Pulkki et Berge, mais utilisant une procédure de localisation proche de celle de Keyrouz et MacDonald. Plus précisément, nous généralisons cette dernière en utilisant un dispositif de captation constitué de deux ou trois microphones intégrés à un objet diffractant dont on définit les fonctions de transfert directionnelles ou "Object-RTF" (Objet Related Transfert Function) à l'image des HRTF. Après une description du procédé, ses performances de localisation sont évaluées sur une grille de critères objectifs qui ont été introduits dans [11].

1 Introduction

Il existe aujourd'hui plusieurs méthodes pour capter l'information spatiale d'une scène sonore. Néanmoins, à cause de leur complexité de mise en œuvre et de la taille des dispositifs, ces technologies n'ont pas atteint le grand public. Afin de favoriser leur dissémination, l'étude présentée ici s'inscrit dans des travaux visant à doter les terminaux mobiles de solutions adaptées de prise de son 3D. Cet objectif impose des contraintes d'encombrement et de mobilité qui sont liées à ce type de terminaux.

Les systèmes de prise de son 3D disponibles aujourd'hui reposent sur trois principales stratégies. La première s'intéresse à la captation et la reproduction des indices de localisation naturelle. Un exemple est la technologie binaurale qui consiste à capter les signaux acoustiques à l'entrée du conduit auditif. La prise de son est couramment réalisée par une tête artificielle. La seconde méthode vise la décomposition physique des propriétés spatiales du champ acoustique, par exemple sur une base de fonctions propres. Il s'agit notamment de la technologie ambisonique qui repose sur des réseaux sphériques de microphones. Une troisième approche propose d'extraire pour chaque source simplement deux informations : la direction et un signal représentatif de la source sonore, à partir d'une prise de son par un dispositif microphonique donné [1] [14].

Cette dernière solution, qui s'apparente à une représentation « objet » de la scène sonore, offre deux avantages : d'une part il s'agit d'une représentation efficace en termes de compression de l'information, d'autre part elle est compatible avec différents formats de reproduction dans la mesure où, grâce à l'information de localisation des sources, il suffit de synthétiser des sources virtuelles selon le format correspondant au système de restitution disponible.

Dans notre contexte d'application aux terminaux mobiles, les deux premières stratégies de prise de son 3D sont inadaptées au vu des contraintes d'encombrement et de mobilité. Nous avons donc opté par la troisième méthode, en imposant comme contrainte supplémentaire l'utilisation d'un dispositif de captation à faible nombre de capteurs. Nous avons étudié comment un dispositif

microphonique composé de trois capteurs cardioïdes est capable de capter l'intégralité de l'information spatiale de la scène sonore. En aval de la prise de son, un algorithme de post-traitement spatial permet d'extraire, en tirant parti à la fois de la directivité et des différences de temps, d'une part les paramètres de localisation de sources sonores et d'autre part, un signal de référence représentatif du signal acoustique émis par les sources.

Dans une première étape, les microphones ont été considérés comme de capteurs idéaux et l'influence acoustique du support microphonique n'a pas été prise en compte. Dans cet article, nous décrivons comment le processus de localisation de sources peut être adapté pour intégrer les caractéristiques réelles des microphones, incluant potentiellement l'influence du support microphonique et du corps du terminal. Ces caractéristiques sont décrites sous la forme de fonctions de transfert dites « Object-RTF » constituant une généralisation du concept de HRTF (Head Related Transfer Function).

Le dispositif de prise de son est d'abord brièvement rappelé. Ensuite, le concept d'Object-RTF est introduit. Puis, son utilisation pour la localisation des sources est décrite. Pour l'identification de la direction, trois critères pour la mesure de distance sont proposés. Leur intérêt est comparé dans la dernière partie et les performances de l'algorithme de localisation sont évaluées en termes d'erreur angulaire et de complexité algorithmique.

2 Travaux antérieurs

Afin de proposer une méthode permettant l'enregistrement d'une scène spatiale deux méthodes analogues aux techniques de Berge [1] et Pulkki [14] ont été proposées dans [10, 11]. Ces méthodes cherchent principalement à représenter une scène sonore complexe, grâce à l'extraction des positions de sources la composant et d'un signal décrivant l'ensemble de ces sources.

Les méthodes proposées sont associées à un dispositif microphonique qui par ces caractéristiques remplit le cahier de charges pour une intégration en mobilité, à

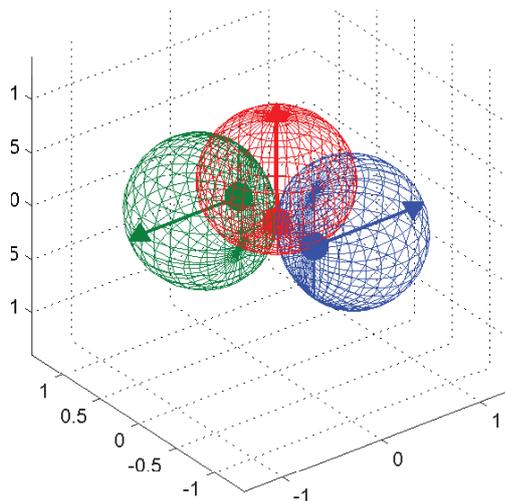


FIGURE 1 – Configuration microphonique.

savoir sa faible taille et son faible nombre de capteurs.

Le dispositif microphonique est composé de trois capteurs cardioïdes (cf. figure 1). Le premier microphone pointe vers les x positifs (droite), le deuxième vers les x négatifs (gauche) et le troisième vers l'axe z (haut). Les deux microphones sur le plan horizontal sont écartés de 1 cm du centre selon l'axe y . Ce dispositif permet l'obtention aisée du signal représentatif de la scène sonore et la localisation de la direction des sources sur toute la sphère 3D avec uniquement 3 capteurs. Dans le cas d'une prise de son pour une diffusion sur un système multicanal classique (5.1, 7.1) ou pour une application de téléconférence, où toutes les sources se trouvent a priori sur le même plan horizontal, l'élévation n'est plus nécessaire et le système peut se réduire aux seuls deux microphones placés sur le plan horizontal.

2.1 Utilisation de la directivité pour l'extraction du signal de référence

Nous considérons le signal de pression acoustique S_o au centre du dispositif microphonique et en l'absence de celui-ci. S_o est suffisant pour caractériser complètement la scène sonore. A condition d'utiliser un dispositif comme celui décrit précédemment, une estimation de S_o peut être obtenue grâce à la directivité de microphones utilisés.

La directivité d'un microphone cardioïde est exprimée par la relation

$$M = \frac{1}{2}(1 + \vec{d}_s \cdot \vec{d}_p) \quad (1)$$

où \vec{d}_s est le vecteur définissant la direction de la source et \vec{d}_p est le vecteur déterminant la direction de pointage du microphone. Dans le cas présent, les directions de pointage \vec{d}_{pn} des trois microphones peuvent être exprimées dans la base de coordonnées cartésiennes \mathfrak{B}_c

$$\vec{d}_{p1} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}_{\mathfrak{B}_c} \quad \vec{d}_{p2} = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}_{\mathfrak{B}_c} \quad \vec{d}_{p3} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}_{\mathfrak{B}_c} \quad (2)$$

et considérant que la direction de la source est exprimée dans la base \mathfrak{B}_s en coordonnées sphériques ou dans la base de coordonnées cartésiennes \mathfrak{B}_c

$$\vec{d}_s = \begin{pmatrix} \theta \\ \phi \\ r \end{pmatrix}_{\mathfrak{B}_s} = \begin{pmatrix} r \cos \theta \cos \phi \\ r \sin \theta \cos \phi \\ r \sin \phi \end{pmatrix}_{\mathfrak{B}_c} \quad (3)$$

les relations de directivité sont définies par :

$$M_1 = \frac{1}{2}(1 + r \cos \phi \cos \theta), \quad (4a)$$

$$M_2 = \frac{1}{2}(1 - r \cos \phi \cos \theta), \quad (4b)$$

$$M_3 = \frac{1}{2}(1 + r \sin \phi). \quad (4c)$$

Pour alléger les notations, il est possible de définir $r = 1$ car ceci ne modifie pas la direction de pointage. La source dont la direction est pointée par le vecteur \vec{d}_s génère un signal S_0 à l'origine du repère. Considérant que les trois microphones y sont placés, les signaux S_n issus de chacune des capsules correspondent au signal S_0 pondéré par la fonction de directivité associée

$$S_n = M_n S_0. \quad (5)$$

Compte tenu des relations (4), le signal de référence S_0 cherché s'obtient :

$$S_0 = S_1 + S_2. \quad (6)$$

Ce signal pourra être utilisé comme porteur de toute l'information sonore de la scène analysée.

2.2 Localisation des sources basée sur la directivité de microphones

Dans [12] nous avons proposé d'utiliser la directivité des capteurs afin de localiser la source. Dans l'équation (5), le terme M qui définit la directivité est une pondération directionnelle du signal S_o . M dépend de la position de la source en coordonnées sphériques θ et ϕ et est propre au microphone. Connaissant donc S_o et les fonctions directionnelles $M(\theta, \phi)$, il ne reste qu'à résoudre une équation à deux inconnues pour connaître les coordonnées de la direction de la source. La position obtenue avec cette méthode présente une ambiguïté avant-arrière liée à l'inversion d'une fonction cosinus.

Dans [11], la méthode est étendue en considérant aussi le cas des signaux ambisoniques à l'ordre 1. Les signaux du format B de l'ambisonique permettent de reconstruire virtuellement un dispositif présentant les caractéristiques du capteur décrit précédemment. Dans ce cas, étant tributaire du format d'entrée, le microphone obtenu virtuellement est coïncidant et la même ambiguïté avant-arrière est rencontrée.

2.3 Localisation des sources basée sur le vecteur intensité acoustique

Le vecteur intensité acoustique est colinéaire avec la direction de propagation de l'onde acoustique. Par extension, ce vecteur pointe la position de la source qui la génère et peut être dérivé du format

B de l'ambisonique. Ces informations permettent la localisation de la source selon la méthode [14]. Les résultats obtenus par cette technique présentent une ambiguïté droite-gauche liée à l'inversion d'une tangente.

2.4 Résolution de l'ambiguïté

Afin de localiser la direction de la source sur toute la sphère 3D, il est nécessaire de résoudre les ambiguïtés liées aux techniques décrites. Dans le premier cas, présenté dans le paragraphe 2.2, l'ambiguïté de localisation est résolue par la dissymétrie sur l'axe y du dispositif en exploitant la différence de phase entre les signaux captés par les deux microphones placés sur le plan horizontal.

Dans le cas où les signaux du format B de l'ambisonique à l'ordre 1 sont utilisés, il est nécessaire de conjuguer les deux méthodes décrites précédemment. Les deux résultats présentent des ambiguïtés qui s'avèrent complémentaires. Il est possible, en les comparant, de trouver une solution unique et sans ambiguïté.

3 Les Object-RTF

Les HRTF ou fonctions de transfert liées à la tête sont les réponses en fréquence directionnelles décrivant le trajet acoustique entre une source sonore et l'entrée du conduit auditif en fonction de la direction de la provenance de l'onde. Le système auditif exploite les indices spectraux engendrés par les diffractions sur la tête, le torse et les pavillons des oreilles ainsi que les différences interaurales pour déterminer la position d'une source. Tout microphone présente une directivité ainsi qu'une réponse en fréquence intrinsèque à ses dimensions et à sa conception. Le champ sonore capté par celui-ci est aussi modifié par les diffractions et réflexions engendrées par les objets qui l'entourent, tels que les préamplificateurs ou leur support. Lorsque le microphone est inséré ou intégré à un objet, sa réponse en fréquence et sa directivité sont modifiées et peuvent être mesurées. Pour généraliser le concept d'HRTF, nous définissons les Object-RTF (*Objet Related Transfert Function*) comme les réponses en fréquence d'un microphone et dépendant de la fréquence et de la direction de la source en azimut θ et en élévation ϕ . Les Object-RTF peuvent être mesurées de manière similaire aux HRTF en utilisant des méthodes existantes [7, 9, 13]. Il s'agit principalement de mesurer la réponse du dispositif pour un ensemble de directions situées sur la sphère 3D. Dans le cas d'un microphone directionnel parfait, sa réponse est identique autour de son axe de révolution. C'est-à-dire que le signal émis par une source sonore est identique quelle que soit sa position sur le cercle défini autour de l'axe de révolution du microphone. Connaissant la directivité du microphone, ainsi que le signal capté par un microphone omnidirectionnel au même point (signal de référence), il est possible de déterminer la position du cercle contenant la source. Si un deuxième microphone est utilisé, la solution se limite à deux points à l'intersection de deux cercles.

Une fois que le dispositif est inséré dans un objet, sa directivité et sa réponse en fréquence sont modifiées, on parle alors d'Object-RTF. En fonction de la géométrie de l'objet, la symétrie de la directivité des microphones est conservée ou non. Dans le cas où la géométrie est cassée, connaissant les Object-RTF ainsi que le signal de référence, l'analyse permet de déterminer une solution unique. Ainsi, l'algorithme de localisation proposé par MacDonald [8] et Keyrou [5, 6] montre la possibilité d'utiliser un jeu de HRTF afin d'obtenir une bonne localisation sans aucune connaissance a priori de la source à localiser utilisant un set de HRTF.

4 Analyse de scène sonore basée sur les Object-RTF

L'algorithme de localisation de sources basée sur les Object-RTF est la généralisation de l'utilisation des HRTF pour la localisation de sources tel qu'elle a été définie par MacDonald [8] et Keyrou [5, 6].

4.1 Les HRTF dans la localisation des sources

Les algorithmes proposés par MacDonald [8] et Keyrou [5, 6] sont utilisés pour la localisation des sources sonores à l'aide des HRTF.

Dans le domaine fréquentiel le signal issu du microphone placé à l'oreille gauche est caractérisé par :

$$S_L(f, \hat{\theta}, \hat{\phi}) = S_o(f) HRTF_L(f, \hat{\theta}, \hat{\phi}) \quad (7)$$

où $S_o(t)$ est la réponse en fréquence du signal de pression acoustique pouvant être mesuré entre les deux oreilles et en absence de la tête, et $HRTF_L$ la réponse en fréquence de l'oreille gauche pour une source placée à $(\hat{\theta}, \hat{\phi})$. Par la suite, les indices L et R désignent les grandeurs correspondant aux oreilles gauche et droite respectivement.

Nous constatons qu'en multipliant le signal d'une oreille par l'HRTF de l'oreille opposée, mesurée à la direction de la source, on obtient :

$$\lambda(f, \hat{\theta}, \hat{\phi}) = S_L(f, \hat{\theta}, \hat{\phi}) HRTF_R(f, \hat{\theta}, \hat{\phi}) = S_o(f) HRTF_L(f, \hat{\theta}, \hat{\phi}) HRTF_R(f, \hat{\theta}, \hat{\phi}) \quad (8)$$

et que

$$\chi(f, \hat{\theta}, \hat{\phi}) = S_R(f, \hat{\theta}, \hat{\phi}) HRTF_L(f, \hat{\theta}, \hat{\phi}) = S_o(f) HRTF_R(f, \hat{\theta}, \hat{\phi}) HRTF_L(f, \hat{\theta}, \hat{\phi}) \quad (9)$$

alors

$$\lambda(f, \hat{\theta}, \hat{\phi}) = \chi(f, \hat{\theta}, \hat{\phi}) \quad (10)$$

Pour identifier la direction de la source, il est nécessaire d'effectuer le produit du signal avec l'ensemble des directions (θ, ϕ) pour lesquelles les HRTF sont connues et d'identifier la direction pour laquelle l'égalité (10) est vérifiée parmi cet ensemble.

Cette méthode permet d'obtenir une localisation avec des erreurs inférieures à 5° avec l'utilisation de 2 microphones et inférieures à 2° avec 4 microphones pour un rapport signal à bruit proche de 40 dB.

4.2 Les Object-RTF dans la localisation des sources

Cette méthode part de l'hypothèse qu'il n'existe qu'une seule source à chaque instant par bande fréquentielle. Le traitement s'effectue alors sur des fenêtres temporelles dont la taille doit être déterminée en fonction de l'écart des capteurs et en fonction du nombre d'échantillons fréquentiels souhaité. Il est également possible de rajouter des zéros « zeroppadding » en fonction de la discrétisation spectrale souhaitée.

La relation (10) peut être généralisée à toute paire microphonique m, n en remplaçant les HRTF par les Object-RTF de la façon suivante :

$$\lambda(f, \hat{\theta}, \hat{\phi}) = S_m(f, \hat{\theta}, \hat{\phi}) ObRTF_n(f, \theta', \phi') \quad (11a)$$

$$\chi(f, \hat{\theta}, \hat{\phi}) = S_n(f, \hat{\theta}, \hat{\phi}) ObRTF_m(f, \theta', \phi') \quad (11b)$$

ou S_x sont les signaux issus d'un capteur x et $ObRTF_x(f, \theta', \phi')$ est la réponse en fréquence du capteur x pour la direction (f, θ', ϕ') . Comme précédemment la relation (10) est vérifiée

$$\text{ssi } (\hat{\theta}, \hat{\phi}) = (\theta', \phi')$$

Ce cas théorique n'est pas toujours atteint. On cherche alors à trouver la meilleure estimation de la direction (θ', ϕ') parmi l'ensemble des directions (θ, ϕ) minimisant la distance δ entre ces directions.

Considérant $\beta_{\hat{n}, m}$ comme le produit du signal S_n généré par une source placée à $(\hat{\theta}, \hat{\phi})$ avec l'ensemble des $ObRTF_m$ pour toutes les directions disponibles. C'est-à-dire :

$$\beta_{\hat{n}, m} = S_n(f, \hat{\theta}, \hat{\phi}) ObRTF_m(f, \theta, \phi), \quad (12)$$

où $\beta_{\hat{n}, m}$ est représenté par un vecteur de longueur L correspondant au nombre des directions des Object-RTF.

On définit alors une distance $\delta(n, m)$ entre chaque paire d'éléments $\beta_{\hat{n}, m}(l)$ et $\beta_{\hat{m}, n}(l)$

$$\delta(n, m) = \delta(\beta_{\hat{n}, m}, \beta_{\hat{m}, n}) \quad (13)$$

On considère alors que la position de la source $(\hat{\theta}, \hat{\phi})$ est la direction où $\delta(n, m)$ est minimale.

Lorsque plus de deux capteurs sont utilisés simultanément, l'analyse s'effectue sur l'ensemble des paires microphoniques. La direction retenue est celle pour laquelle la valeur $\delta(m, n)$ est la plus faible.

4.3 Définition du critère de distance δ

Les données manipulées étant des valeurs complexes, la norme des vecteurs n'est pas le meilleur indicateur de ressemblance de ces derniers, trois indicateurs de distance ont été alors comparés.

Distance angulaire : Dans un premier temps, nous avons proposé l'utilisation du cosinus de l'angle entre les deux valeurs complexes comme indicateur de distance δ entre ces deux valeurs. Celui-ci est calculé grâce au produit scalaire.

$$\delta(A, B) = \cos(\widehat{AB}) = \frac{A \cdot B}{|A||B|} \quad (14)$$

avec

$$A = S_m(f, \hat{\theta}, \hat{\phi}) ObRTF_n(f, \theta, \phi)$$

$$B = S_n(f, \hat{\theta}, \hat{\phi}) ObRTF_m(f, \theta, \phi)$$

Cette relation présente l'avantage de fournir un indice normalisé de ressemblance qui vaut 0 lorsque les deux vecteurs sont colinéaires et qui vaut 1 lorsqu'ils sont perpendiculaires. Néanmoins, cette méthode ne rend pas compte de la norme des vecteurs.

Tanimoto : Afin de rassembler la magnitude et l'angle des vecteurs dans un seul indicateur. Il est proposé d'utiliser un deuxième indicateur de distance $D_T(A, B)$ [2] issu de l'indice de Tanimoto $T(A, B)$ qui est défini par :

$$D_T(A, B) = 1 - T(A, B) \quad (15)$$

$$= 1 - \frac{A \cdot B}{|A|^2 + |B|^2 - A \cdot B} \quad (16)$$

où \cdot définit le produit scalaire et $|\cdot|$ la norme du vecteur.

Quotient : Une dernière mesure de distance, appelée ici la méthode du quotient, est proposée. Partant de l'hypothèse définie dans l'équation (10) avec les paramétrés de (11) :

$$\frac{S_n(f, \hat{\theta}, \hat{\phi}) ObRTF_m(f, \hat{\theta}, \hat{\phi})}{S_m(f, \hat{\theta}, \hat{\phi}) ObRTF_n(f, \hat{\theta}, \hat{\phi})} = 1 \quad (17)$$

la distance $\delta(\beta_{\hat{n}, m}, \beta_{\hat{m}, n})$ est alors définie par :

$$\delta = \frac{S_n(f, \hat{\theta}, \hat{\phi}) ObRTF_m(f, \hat{\theta}, \hat{\phi})}{S_m(f, \hat{\theta}, \hat{\phi}) ObRTF_n(f, \hat{\theta}, \hat{\phi})} - 1 \quad (18)$$

permettant de ne calculer qu'une seule fois les rapports $\frac{ObRTF_n(f, \theta, \phi)}{ObRTF_m(f, \theta, \phi)}$ car ils sont constants tout au long de l'analyse et sont indépendants des signaux S_n . Ce critère permet de réduire considérablement le nombre d'opérations en allégeant ainsi la complexité.

5 Evaluation des performances des Object-RTF pour la localisation des sources

Des simulations numériques ont été mises en place pour effectuer les évaluations. Le signal test est une source large bande tournant autour du microphone à des élévations variant par palier de 20° partant de l'hémisphère sud vers le zénith.

Nous étudions notamment ici, le meilleur critère permettant d'évaluer la distance δ (cf. équation (13))

5.1 Critères d'évaluation

Les critères d'évaluations introduits dans [11] sont utilisés. E_t est la distance angulaire totale sur la sphère par rapport à la position réelle de la source et E_{75} est l'erreur E_t obtenue pour au moins 75 % du spectre calculé en tiers d'octaves.

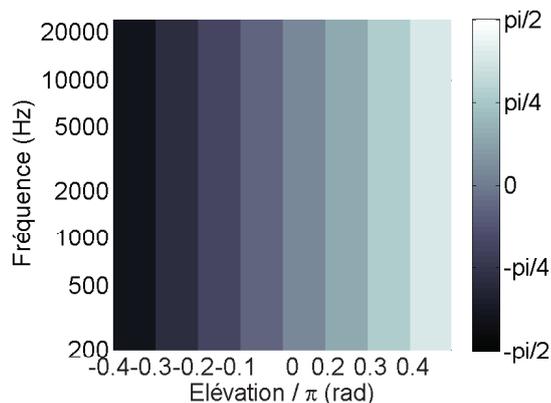


FIGURE 2 – Élévation calculée, δ évalué avec le coefficient de Tanimoto

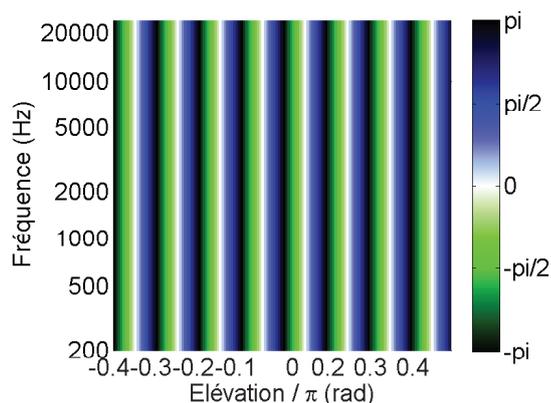


FIGURE 3 – Azimuth calculé, δ évalué avec le coefficient de Tanimoto

5.2 Performances

Les critères de distance définis dans le paragraphe 4.3 ont été évalués. L'utilisation des critères de Tanimoto et du quotient permettent une localisation correcte et sans ambiguïté dans une configuration idéale (cf. figures 2 et 3). L'utilisation de la distance angulaire laisse apparaître des nombreuses erreurs de localisation (cf. figure 4) où l' E_{75} moyen est de 31° . Ces erreurs apparaissent notamment lorsque la source s'éloigne de l'azimut de 0° et elles augmentent avec l'élévation pour atteindre des erreurs maximums de localisation E_{75} de 130° .

Ces résultats nous permettent d'écarter l'utilisation du critère défini par la distance angulaire, nous permettant de limiter les études ultérieures aux coefficients de Tanimoto et au critère du quotient.

Rapport signal à bruit : Les résultats de localisation obtenus avec les coefficients retenus (ie. Tanimoto et Quotient) sont identiques quelque soit le rapport signal à bruit. L'erreur E_{75} moyennée sur l'ensemble des directions est de $64,5^\circ$ lorsque le rapport signal à bruit (RSB) est de 20 dB et atteint $10,0^\circ$ avec un RSB de 30 dB. Cette erreur est nulle quand la source sonore se trouve au même azimut de la source perturbatrice et maximale lorsqu'elles sont diamétralement opposées.

Si uniquement les deux microphones placés sur le

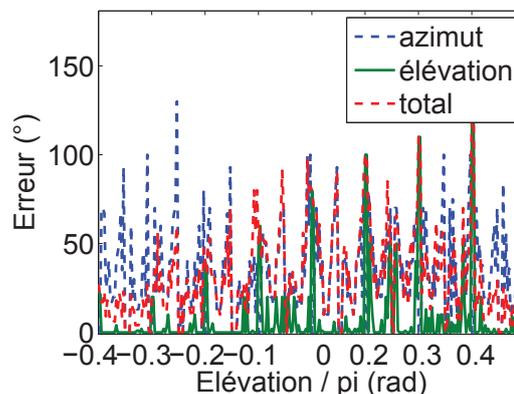


FIGURE 4 – Erreur de localisation E_{75} , δ calculé avec le critère du quotient

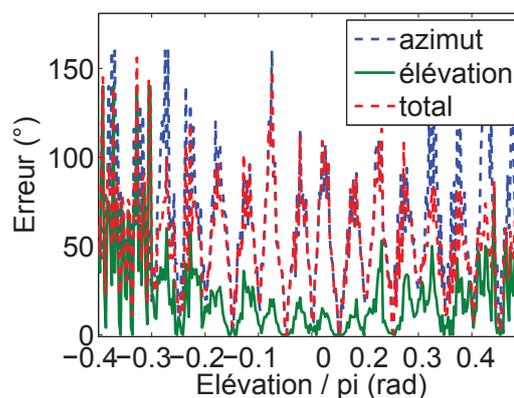


FIGURE 5 – Erreur de localisation E_{75} , résultats obtenus avec un RSB de 30 dB pour une source perturbatrice placée à $\theta = 0$ et $\phi = 0$

plan horizontal sont utilisés, l'indicateur E_{75} moyen atteint $64,5^\circ$ pour un RSB de 20 dB et de 15° pour un RSB de 30 dB. Lorsque le RSB est de 40 dB le E_{75} moyen est de $0,5^\circ$ et l'erreur maximum est de 26° confortant ainsi les résultats de Keyruz et McDonald [6, 8].

Complexité : L'utilisation de la méthode du quotient permet le calcul de la matrice $\frac{ObRTF_n(f, \theta, \phi)}{ObRTF_m(f, \theta, \phi)}$ (cf. paragraphe 4.3) ce qui permet de gagner considérablement en complexité algorithmique. En effet, les calculs réalisés avec cette méthode sont effectués dans environ un tiers du temps utilisé par la méthode de Tanimoto.

6 Conclusion

Afin de proposer une nouvelle méthode permettant la captation du son spatialisé avec des terminaux mobiles, nous avons développé une méthode permettant l'encodage de la scène sonore avec un signal audio représentatif de la scène et des paramètres de localisation des sources qui la composent.

Un dispositif microphonique composé de 3 capsules a été proposé auparavant ainsi que les algorithmes de localisation associés. Ces derniers étant très

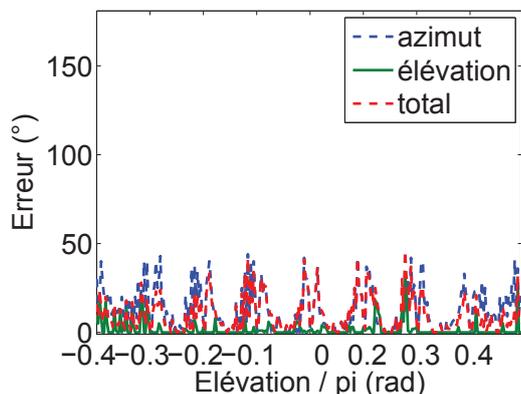


FIGURE 6 – Erreur de localisation E_{75} , résultats obtenus avec un RSB de 30 dB pour une source perturbatrice placée à $\theta = 0$ et $\phi = 0$

contraignants, une nouvelle méthode appelée méthode des Objets-RTF a été proposée ici. Elle permettant de prendre en compte les caractéristiques réelles des capteurs. Celle-ci est une généralisation de la méthode de HRTF de Keyrouz et McDonald.

La direction de la source est obtenue en cherchant la similitude parmi l'ensemble de object-RTF par un critère de distance. Nous avons proposé et évalué ici trois critères permettant la mesure de cette distance (ie. Distance Angulaire, Tanimoto et Quotient). Dans des condition idéales les critères de Tanimoto et du quotient donnent des résultats parfaits. Dans les mêmes conditions, le troisième critère donne des résultats comportant une erreur moyenne de 31° ce qui permet de l'écarter. La robustesse de la méthode été évalué en présence d'une source perturbatrice avec les deux critères de distance retenus. les résultats sont identiques pour les deux critères quel que soit le rapport signal à bruit. Les erreurs moyennes de localisations rencontrées sont inférieures à 16° pour un RSB de 30 dB avec l'utilisation de 2 capteurs et de 10° lorsque 3 capteurs sont utilisés.

En termes de complexité, la méthode du quotient est à privilégier car elle permet une réduction considérable du temps de calcul par rapport à la méthode de Tanimoto.

Des études complémentaires doivent être effectuées afin d'évaluer les performances de cette méthode avec des dispositifs réels ce qui implique la mesure des Object-RTF du dispositif microphonique.

Références

- [1] S. Berge and N. Barrett. A new method for b-format to binaural transcoding. In *AES 40th international conference*, Tokyo, Japan, Oct. 2010.
- [2] R. Camacho and al. Assessing the effect of 2D fingerprint filtering on ILP-Based structure-activity relationships toxicity studies in drug design. In *5th International Conference on Practical Applications of Computational Biology & Bioinformatics (PACBB 2011)*, volume 93 of *Advances in Intelligent and Soft Computing*, pages 355–363. Springer Berlin Heidelberg, Jan. 2011.
- [3] M. A. Gerzon. The design of precisely coincident microphone arrays for stereo and surround sound. *Audio Engineering Society Convention 50*, 50 :50, 1975.
- [4] M. A. Gerzon and P. G. Craven. Coincident microphone simulation covering three dimensional space and yielding various directional outputs, Aug. 1977.
- [5] F. Keyrouz and K. Diepold. An enhanced binaural 3D sound localization algorithm. In *Signal Processing and Information Technology, 2006 IEEE International Symposium on*, page 662–665, 2006.
- [6] F. Keyrouz, K. Diepold, and P. Dewilde. Robust 3d robotic sound localization using state-space hrtf inversion. In *Robotics and Biomimetics, 2006. ROBIO'06. IEEE International Conference on*, page 245–250, 2006.
- [7] E. H. Langendijk and A. W. Bronkhorst. Fidelity of three-dimensional-sound reproduction using a virtual auditory display. *The Journal of the Acoustical Society of America*, 107 :528, 2000.
- [8] J. A. MacDonald. A localization algorithm based on head-related transfer functions. *The Journal of the Acoustical Society of America*, 123(6) :4290, 2008.
- [9] P. Majdak, P. Balazs, and B. Laback. Multiple exponential sweep method for fast measurement of head related transfer functions. *Journal Audio Engineering Society*, 2007.
- [10] J. Palacino and R. Nicol. Full 3D sound pick-up with a small microphone array : Prototype outline and preliminary assessment. In *Proceedings of the International Conference on Acoustics*, Merano, Mar. 2013.
- [11] J. Palacino and R. Nicol. Spatial sound pick-up with a low number of microphones. In *Proceedings of Meetings on Acoustics*, volume 19, page 055078, Montreal, June 2013.
- [12] J. Palacino, R. Nicol, M. Emerit, and L. Gros. Perceptual assessment of binaural decoding of first-order ambisonics. In *Acoustics 2012*, Nantes, France, Apr. 2012.
- [13] J.-M. PERNAUX. *Spatialisation du son par les techniques binaurales : application aux services de télécommunications*. PhD, I.N.P.G, Grenoble, May 2003.
- [14] V. Pulkki. Directional audio coding in spatial sound reproduction and stereo upmixing. In *Proc. of the AES 28th Int. Conf*, Pitea, Sweden, 2006.