

Influence du “ head-tracking ” sur l’externalisation en écoute binaurale non-individualisée

Etienne Hendrickx, Peter Stitt, Jean-Christophe Messonnier, Jean-Marc Lyzwa, Brian F. G. Katz, Catherine De Boishéraud

► To cite this version:

Etienne Hendrickx, Peter Stitt, Jean-Christophe Messonnier, Jean-Marc Lyzwa, Brian F. G. Katz, et al.. Influence du “ head-tracking ” sur l’externalisation en écoute binaurale non-individualisée. 3èmes Journées Perception Sonore, Jun 2017, Brest, France. 2017, <<http://www.univ-brest.fr/JPS2017>>. <hal-01543693>

HAL Id: hal-01543693

<http://hal.univ-brest.fr/hal-01543693>

Submitted on 21 Jun 2017

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.

Etienne Hendrickx (UBO Lab-STICC), Peter Stitt (LIMSIS), Jean-Christophe Messonnier (CNSMDP), Jean-Marc Lyzwa (CNSMDP), Brian F.G. Katz (UPMC, Institut d'Alembert) et Catherine de Boishéraud (CNSMDP)

INTRODUCTION

Binaural : Recréer au niveau des oreilles de l'auditeur une scène sonore réaliste, à l'aide d'un casque

Le binaural peut être :

Naturel ou Synthétique

Individualisé ou Non-Individualisé

La présente étude se focalise sur la **synthèse binaurale non-individualisée** (nombreux avantages)

Apparitions fréquentes d'artefacts...

- Notamment l'*internalisation* des sources sonores (\neq externalisation)
- Surtout dans le cas de sources frontales et arrière...

En binaural, le head tracking permet :

- Une meilleure localisation en azimut et en élévation
- Une réduction du nombre d'inversions avant/arrière...

- ... et une meilleure externalisation ???

→ Résultats d'études précédentes pas très concluants :

Author	Result	Notes
Loomis et al. (1990)	Oui !	Tests informels...
Kawaura et al. (1991)	Oui !	3 sujets seulement...
Begault et al. (2001)	Non !	Stimuli courts (<3s) Mouvements de tête non contrôlés et probablement limités
Wenzel (1995)	Un peu ...	Résultats moyennés sur toutes les positions testées

EXPÉRIENCE

• Étude limitée au plan horizontal

Différences par rapport aux études précédentes :

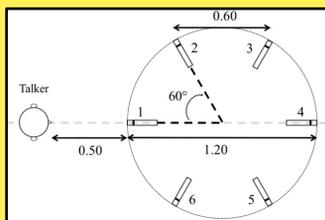
- Analyse azimut par azimut
- Stimulus plus long (9s)
- Mouvements de tête plus larges ($\pm 90^\circ$) et contrôlés
- Étude de l'impact sur l'externalisation *après* le mouvement de tête, et non *pendant*

3 Hypothèses :

- Micro-mouvements involontaires de la tête → améliore l'externalisation s'il y a un head-tracker
- Larges mouvements de tête avec head tracking → améliore considérablement l'externalisation
- Larges mouvements de tête sans head tracking → diminue considérablement l'externalisation

Stimulus

- Extrait (9s) d'un poème lu par un homme
- Enregistré dans un studio du CNSMDP
- Arbre multicanal (6 cardioïdes)
- « Binauralisé » sur 6 enceintes virtuelles



Distances en mètres

Test perceptif

- Salle d'écoute du CNSMDP, lumières éteintes
- Casque Sennheiser HD600
- Head Tracker Hedrot et moteur binaural Bipan
- 10 sujets experts
- Pour chaque présentation d'un stimulus, noter l'externalisation perçue :
0 = voix perçue à l'intérieur de la tête
5 = voix externalisée et lointaine

Variables Indépendantes :

- Conditions (x4) :
S0 : Tête immobile – pas de HT (head tracking)
ST : Tête immobile – avec HT
M0 : Larges mvts de tête ($\pm 90^\circ$) – pas de HT
MT : Larges mvts de tête ($\pm 90^\circ$) – avec HT
- Orientations (x12): $0^\circ, 30^\circ, 60^\circ$, etc.
- HRTFs non individualisées (x3)

Test répété 5 fois de suite

(4h de test par sujet sur plusieurs jours)

CONCLUSIONS

Larges mouvements de tête = meilleure externalisation, aussi bien pour des sujets experts que naïfs

Larges mouvements de sources = meilleure externalisation également, mais dans une moindre mesure

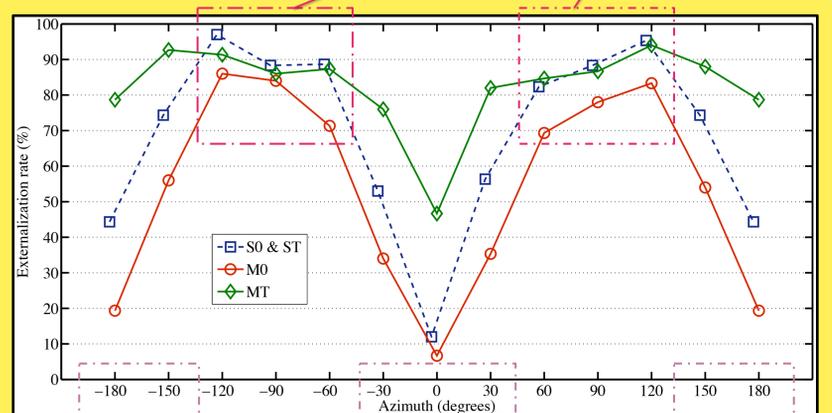
Études envisagées : → Influence de la modalité visuelle (dans le noir / salle avec lumières allumées / salle avec lumières allumées et indices visuels cohérents)
→ Amélioration de l'externalisation à 0° et 180° en entraînant le sujet, avec mouvements de tête plus longs/plus amples

RÉSULTATS

Tests de Wilcoxon avec corrections de Bonferroni

Pas de différence sig. entre S0 et ST
→ Résultats groupés

Azimuths latéraux :
- Bonne externalisation
- Différence non sig. entre MT, S0 et ST
- $M0 \leq$ autres conditions



Azimuths frontaux et arrière :
- Différences sig. marquées : $MT > S0$ & $ST > M0$
- Externalisation particulièrement faible à 0° et 180°

● Larges mvts de tête avec head tracking = meilleure externalisation pour sources frontales et arrière
Externalisation robuste, immédiate et observée avec les 3 HRTFs
Amélioration nettement plus marquée que dans les études précédentes → Pourquoi ?

● Pas d'influence du head tracker lorsque la tête est fixe

● Larges mouvements de tête sans head tracker = chute importante de l'externalisation
→ Situation fréquente !

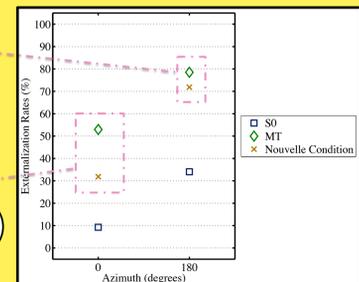
EXPÉRIENCE SUPPLÉMENTAIRE

Expérience quasi similaire mais :

- 9 Sujets naïfs (et non experts)
- 0° et 180° uniquement
- Conditions ST et M0 supprimées
- Ajout d'une condition supplémentaire :
Tête immobile – pas de HT – rotation automatique de la scène autour du sujet ($\pm 90^\circ$)

Amélioration de l'ext. à 180° :
mvts de tête légèrement > mvts de sources ($p = 0.048$)

Amélioration de l'ext. à 0° :
mvts de tête > mvts de sources



$MT > S0$ à 0° et 180°

● Larges mvts de tête avec head tracking = meilleure externalisation

● Larges mvts de sources = meilleure externalisation :
• à 180° : comparable aux mvts de tête
• à 0° : pas aussi efficace que les mvts de tête

- E. Hendrickx, P. Stitt, J.-C. Messonnier, J.-M. Lyzwa, B. F. G. Katz et C. de Boishéraud (2017). Improvement of externalization by listener and source movement. *J. Audio Eng. Soc.* (accepté le 5 mai 2017).
- E. Hendrickx, P. Stitt, J.-C. Messonnier, J.-M. Lyzwa, B. F. G. Katz et C. de Boishéraud (2017). Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis. *J. Acoust. Soc. Am.*, 141(3), 2011-2023.
- P. Stitt, E. Hendrickx, J.-C. Messonnier et B.F.G. Katz (2016). The role of head tracking in binaural rendering. *Proceedings of the 29th Tonmeisterstagung, VDT International Convention*, Cologne, Allemagne, 350-355.
- Loomis, J. M., Hebert, C., & Cicinelli, J. G. (1990). Active localization of virtual sounds. *J. Acoust. Soc. Am.*, 88(4), 1757-1764.
- Kawaura, J. I., Suzuki, Y., Asano, F., & Sone, T. (1991). Sound localization in headphone reproduction by simulating transfer functions from the sound source to the external ear. *J. Acoust. Soc. Am. Jpn.* (E), 12(5), 203-216.
- Begault, D. R., Wenzel, E. M., & Anderson, M. R. (2001). Direct comparison of the impact of head tracking, reverberation, and individualized HRTFs on the spatial perception of a virtual speech source. *J. Audio Eng. Soc.*, 49(10), 904-916.
- Wenzel, E. M. (1995). The relative contribution of interaural time and magnitude cues to dynamic sound localization. In *Applications of Signal Processing to Audio and Acoustics, 1995.*, IEEE ASSP Workshop on (pp. 80-83). IEEE.
- Baskind, A., Carpentier, T., Noisternig, M., Warusfel, O., & Lyzwa, J. M. (2012). Binaural and transaural spatialization techniques in multichannel 5.1 production. In *27th Tonmeisterstagung, VDT International Convention*.
- Brimijoin, W. O., Boyd, A. W., & Akeroyd, M. A. (2013). The contribution of head movement to the externalization and internalization of sounds. *PLoS one*, 8(12), e83068.

CFA '18 LE HAVRE ■ 23-27 avril 2018
14^{ème} Congrès Français d'Acoustique



Influence du 'head-tracking' sur l'externalisation en écoute binaurale non- individualisée

E. Hendrickx^a, P. Stitt^b, J.-C. Messonnier^c, J.-M. Lyzwa^c, B.F.G. Katz^d et C. De Boishéraud^c

^aLab-STICC, Université de Brest, 6 avenue Le Gorgeu, 29200 Brest, France

^bAudio Acoustics Group, LIMSI, CNRS, Université Paris-Saclay, 91400 Orsay, France

^cConservatoire National Supérieur de Musique et de Danse de Paris, 209, avenue Jean-Jaurès, 75019 Paris, France

^dSorbonne Universités, UPMC, Univ Paris 06, CNRS, Institut d'Alembert, 75005 Paris, France
etienne.hendrickx@univ-brest.fr

La reproduction binaurale a pour but de recréer une scène sonore réaliste aux oreilles de l'auditeur, à l'aide d'un casque audio. Malheureusement, plusieurs études montrent une baisse de l'externalisation lorsque des sujets écoutent des contenus binauraux avec HRTFs non-individualisées (les sources sonores sont perçues à l'intérieur de la tête, alors qu'elles devraient être localisées à l'extérieur de la tête).

La présente étude examine si de larges mouvements de tête couplés à un système de « head-tracking » peuvent compenser cette baisse d'externalisation. Pour chaque présentation, un stimulus (voix d'homme) était présenté au casque à divers azimuts, avec différentes HRTFs non-individualisées pour le rendu binaural. Le head-tracker pouvait être soit actif, soit inactif, et les sujets devaient soit bouger la tête, soit rester immobiles. Après chaque présentation, les sujets devaient évaluer le degré d'externalisation perçue. Les résultats ont montré que de larges mouvements de tête peuvent considérablement améliorer l'externalisation, et que cette externalisation persiste quand bien même le sujet a cessé de bouger la tête. Ces résultats ont été obtenus avec des sujets experts et naïfs.

Une autre expérience a été menée, dans laquelle les rôles étaient inversés : les sujets gardaient la tête fixe tandis que les sources sonores tournaient automatiquement autour d'eux. Les résultats ont montré que de tels mouvements de sources peuvent aussi améliorer l'externalisation, mais dans une moindre mesure que les mouvements de tête.

1 Introduction

La technique binaurale permet de recréer une scène sonore réaliste à l'aide d'un casque audio, en reproduisant au niveau des oreilles de l'auditeur les signaux de pression sonore correspondant à une scène sonore donnée.

Un enregistrement binaural peut être effectué de plusieurs manières. Les sources sonores réelles peuvent être captées par une tête artificielle, ou alors par des microphones placés dans les oreilles d'un auditeur. Une autre technique consiste à filtrer des sons anéchoïques avec les HRTFs (*Head-Related Transfer Functions*, ou fonctions de transfert de la tête) d'un individu : on parle alors de *synthèse binaurale*. Cette technique présente l'avantage d'être compatible avec le « head-tracking », permettant ainsi la prise en compte des mouvements de tête de l'utilisateur. Idéalement, les HRTFs utilisées pour la synthèse binaurale devraient être individualisées. Malheureusement, mesurer ses propres HRTFs est une opération onéreuse et délicate.

Une solution plus simple consiste à écouter du binaural « à travers les oreilles » d'un autre individu, dont les HRTFs ont déjà été mesurées. Plusieurs études ont cependant montré que l'utilisation d'HRTFs non-individualisées pouvait provoquer un certain nombre de problèmes, notamment l'augmentation du nombre d'inversions avant/arrière [1] ou encore l'internalisation des sources sonores [2, 3], c'est-à-dire que les sources sonores ont tendance à être perçues à l'intérieur de la tête (*internalisées*), alors qu'elles devraient être perçues à l'extérieur de la tête (*externalisées*), comme le sont les sources qui nous entourent au quotidien.

Plusieurs études ont montré que le problème d'inversion avant-arrière pouvait être atténué efficacement en ajoutant au système de restitution binaurale un head-tracker et en demandant au sujet de bouger la tête [4, 5].

Le problème d'internalisation semble en revanche plus difficile à résoudre. Mendonça *et al.* ont montré qu'un entraînement d'un mois avec une HRTF donnée améliorait l'externalisation dans le plan vertical. Par contre, dans le plan horizontal, l'amélioration était très faible [6]. Ajouter de la réverbération peut également améliorer l'externalisation d'une source sonore [4, 7]. D'autres études ont envisagé l'utilisation d'un head-tracker, mais les résultats ont été jusqu'à maintenant plutôt contradictoires [8].

Les rares études suggérant que les mouvements de tête avec head-tracking améliorent l'externalisation posent un certain nombre de soucis : étude informelle [9], nombre insuffisant de sujets [10], utilisation d'HRTFs individualisées

uniquement [11], etc.

D'autres études affirment que l'effet des mouvements de tête avec head-tracker est faible [12], voire inexistant [4]. Cependant, ces études utilisaient des stimuli de 2-3 s, et Begault *et al.* [4] reconnaissent que des durées aussi courtes ont peut-être empêché les sujets de pleinement profiter des indices apportés par les mouvements de tête. De plus, les résultats de [12] et [4] étaient moyennés sur l'ensemble des positions testées avant d'être analysés. Comme les sources latérales sont en général très bien externalisées, même sans head-tracking [2, 3], c'est plutôt pour les sources frontales et arrières qu'on peut s'attendre à observer une amélioration de l'externalisation grâce au head-tracking. Il est donc possible que des améliorations faibles voire inexistantes obtenues pour des sources latérales aient masqué des améliorations plus importantes obtenues pour des sources frontales ou arrières.

Deux expériences ont donc été mises en place, avec des stimuli plus longs (8 s) et une analyse des résultats position par position. L'Expérience I a été conduite par des sujets experts, tandis que l'Expérience II a été conduite par des sujets naïfs.

2 Expérience I

2.1 Stimuli

Le stimulus était un extrait de 8 s du poème *L'Albatros* de Charles Baudelaire, lu par un homme.

Le stimulus a été enregistré dans un studio du Conservatoire de Paris ($\approx 30 \text{ m}^2$, $\text{RT}_{60} \approx 0.24 \text{ s}$) avec un arbre à 6 microphones présenté dans la Figure 1 (voir [13] pour plus de précisions). Une comparaison informelle de plusieurs arbres microphoniques, menée avec une dizaine de sujets, avait suggéré que cette configuration offrait la reproduction la plus naturelle une fois binauralisée et restituée au casque.

La binauralisation a été effectuée de sorte à placer le sujet « au centre » de l'arbre microphonique. Par exemple, le signal du microphone 1 était convolué avec une HRTF mesurée à 0° , le signal du microphone 2 avec une HRTF à 60° , etc. Les six signaux binauraux obtenus (un par microphone) étaient ensuite sommés pour créer les signaux gauche et droite alimentant le casque audio.

Le rendu binaural utilisait le moteur *Bipan* [14] tandis que le head-tracking utilisait la solution open-source *Headrot* [15]. Le head-tracker était attaché en permanence

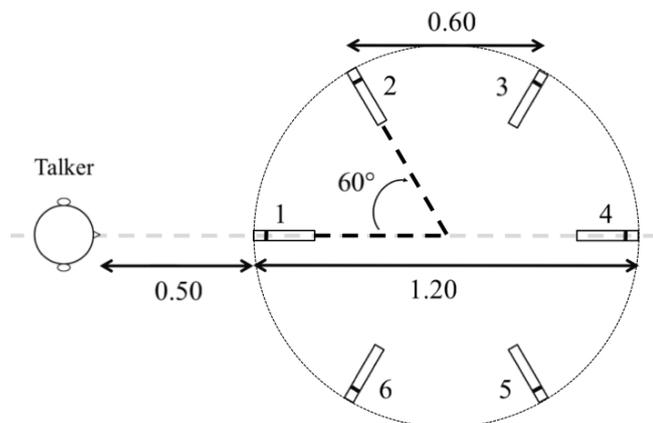


FIGURE 1 – Arbre microphonique utilisé pour l’enregistrement du stimulus, avec un microphone principal (microphone 1) capturant principalement le son direct, et cinq autres microphones (microphones 2-6) permettant d’enregistrer la réverbération. Les microphones étaient des DPA 4021 (cardioïdes), placés à une hauteur de 1.65 m (hauteur de la bouche de l’acteur). Dimensions en mètres.

à la tête du sujet et était relié à l’ordinateur via une carte USB Teensy 3. La latence du système était en moyenne de 48.1 ms (SD = 5.3 ms), ce qui est inférieure aux seuils de détectabilité obtenus par Brungart *et al.* [16].

Trois HRTFs non-individualisées différentes, choisies dans la base de données publique LISTEN [17], ont été utilisées pour le test, afin d’étudier si l’impact du head-tracking dépendait de l’HRTF utilisée.

Différentes orientations de la scène sonore ont également été testées. L’orientation d’un stimulus est ici définie par la position à laquelle le signal obtenu avec le microphone principal (microphone 1) est rendu. Par exemple, une orientation à $+30^\circ$ signifie que le signal du microphone 1 est rendu à $+30^\circ$ (le microphone 2 est alors rendu à $+90^\circ$, le microphone 3 à $+150^\circ$, etc.). Les orientations testées couvraient la totalité du plan horizontal, par pas de 30° et avec une élévation nulle.

Le test perceptif s’est déroulé dans une cabine insonorisée du Conservatoire de Paris (niveau de bruit de fond ≈ 25 dB A). Les lumières étaient éteintes et le sujet était assis au centre de la pièce. Les stimuli étaient présentés avec un casque Sennheiser HD 600. Le niveau a été réglé à ≈ 65 dB A en plaçant le casque sur une tête artificielle Neumann KU 100. La lecture des stimuli, l’interface graphique et l’enregistrement des données étaient contrôlés par un logiciel implémenté sous Max/MSP sur un ordinateur MacBook Pro, relié à une carte son RME Fireface 800.

2.2 Sujets et Protocole

Dix sujets ont participé à l’expérience (4 femmes et 5 hommes, de 22 à 57 ans). Tous les sujets étaient des ingénieurs du son professionnels, habitués à écouter en binaural. Aucun d’eux n’avait d’expérience particulière des tests auditifs en laboratoire, ni de problème d’audition connu.

Les sujets devaient soit garder la tête fixe, droit devant, soit tourner la tête de gauche à droite entre $\pm 90^\circ$. Le head-tracker était soit actif, soit inactif. Le sujet devait donc évaluer 4 conditions différentes :

SØ : statique (tête fixe), sans head-tracking.

ST : statique (tête fixe), avec head-tracking¹.

MØ : avec mouvements de tête, sans head-tracking.

MT : avec mouvements de tête, avec head-tracking.

Pour les conditions avec mouvements (**MØ** and **MT**), la présentation d’un stimulus était divisée en trois phases :

1. 5.5 s de stimulus, durant lesquelles les sujets devaient tourner la tête vers la gauche (-90°) puis vers la droite ($+90^\circ$) avant de revenir à leur position initiale (0°).
2. 1 s de silence. A la fin de ce silence, tous les mouvements de tête devaient être terminés et les sujets devaient tous avoir retrouvé leur position initiale (0°), tête fixe.
3. 2.5 s de stimulus durant lesquelles la tête des sujets devait rester fixe.

Une fois ces trois phases achevées, le sujet devait rapporter à quel point la source sonore était externalisée, en utilisant une échelle à 6 points affichée sur un écran d’ordinateur (voir Tableau 1). Une fois que le sujet avait répondu, le stimulus suivant était diffusé.

TABLEAU 1 – Échelle à six points utilisée pour évaluer l’externalisation.

Note	Externalisation
0	La source est perçue au centre de la tête.
1	La source n’est pas perçue au centre de la tête, mais néanmoins dans ma tête.
2	La source est perçue au niveau de mes oreilles, ou sur mon crâne.
3	La source est externalisée mais proche de la tête.
4	La source est externalisée et à portée de main.
5	La source est externalisée et lointaine.

Il est à noter que les sujets devaient rapporter à quel point la source avait été externalisée *durant les 2.5 dernières secondes* du stimulus, c’est-à-dire à partir du moment où les sujets retrouvaient la position de départ à 0° , tête fixe. En d’autres termes, les sujets devaient juger de l’externalisation *après* avoir bougé la tête, contrairement aux études précédentes qui demandaient d’évaluer l’externalisation *pendant* les mouvements de tête. Ce choix de procédure permettait d’étudier si l’externalisation apportée par les mouvements de tête persistaient quand bien même le sujet s’arrêta de bouger.

Pour les conditions sans mouvement de tête (**SØ** et **ST**), la procédure était la même sauf que les sujets devaient garder la tête fixe et regarder droit devant eux durant la totalité de la présentation du stimulus.

Pour chacune des quatre conditions, il y avait 3 HRTFs \times 12 orientations, soit 36 essais, répétés 5 fois de suite. Les sujets mettaient environ 1 h pour passer une condition, et les sujets ne passaient jamais plus d’une condition dans la même

1. Cette condition peut paraître surprenante, cependant des études ont suggéré que les micro-mouvements de tête qu’un individu effectue inconsciemment (même lorsqu’on lui demande de garder la tête fixe) peuvent potentiellement améliorer l’externalisation [18].

journée. L'ordre des conditions était aléatoire et différent pour chaque sujet. Au sein d'une condition, l'enchaînement des orientations était aléatoire et différent pour chaque sujet et pour chaque répétition du test. L'HRTF utilisée pour le rendu binaural variait toujours d'un essai à l'autre pour minimiser d'éventuels effets d'apprentissage.

2.3 Résultats

L'analyse des mouvements de tête montre que les sujets ont relativement bien respecté les consignes pour chacune des quatre conditions.

Un test de Friedman n'a pas montré de différence significative entre les HRTFs ($p = 0.735$). Les résultats seront donc présentés moyennés sur l'ensemble des trois HRTFs.

La Figure 2 montre les notes d'externalisation pour chaque condition et chaque orientation, moyennées sur l'ensemble des sujets et des HRTFs².

Il peut être observé que :

- Les résultats obtenus pour les conditions sans mouvement de tête sont très semblables, que le head tracking soit actif (**ST**) ou inactif (**SØ**).
- L'externalisation la plus faible a été obtenue pour la condition **MØ** (mouvements de tête sans head-tracking).
- Pour les orientations latérales ($\pm 60^\circ$, $\pm 90^\circ$, $\pm 120^\circ$), l'externalisation a été globalement élevée et les mouvements de tête avec head tracking (**MT**) n'ont pas amélioré l'externalisation par rapport aux conditions sans mouvement de tête (**SØ** and **ST**).
- Pour les orientations frontales et arrières (0° , $\pm 30^\circ$, $\pm 150^\circ$, 180°) l'externalisation a été plus modérée pour les conditions sans mouvement de tête (**SØ** and **ST**). Aux orientations $\pm 30^\circ$, $\pm 150^\circ$ et 180° , les mouvements de tête avec head tracking (**MT**) ont permis de maintenir une externalisation importante, comparable à celle obtenue pour les orientations latérales. A 0° , l'externalisation pour la condition **MT** n'est pas très élevée (moyenne = 2.2), mais reste nettement supérieure à celle observée pour les autres conditions (moyennes entre 0.6 et 1.2).

3 Expérience II

Dans l'Expérience I, les sujets étaient des ingénieurs de son professionnels, habitués à l'écoute de contenus binauraux. Une seconde expérience a été mise en place avec des sujets naïfs.

Un système de head-tracking nécessite des logiciels et du matériel supplémentaire. Une alternative plus simple consiste à imiter des mouvements de tête en déplaçant les sources sonores virtuelles autour du sujet. Dans [19], de

2. Les moyennes des sujets ont été au préalable normalisées, de sorte que la moyenne de chaque sujet sur tous les essais soit égale à la moyenne globale (c'est-à-dire la moyenne sur tous les essais et sur tous les sujets). Une telle normalisation permet de se focaliser sur les différences d'externalisation d'une condition à l'autre, et de minimiser les biais dus au fait que certains sujets, indépendamment de la condition testée, externalisent plus que d'autres.

légères rotations de tête ($\pm 4^\circ$) étaient simulées en générant aléatoirement des fluctuations statistiquement similaires aux micro-mouvements de tête qu'un individu effectue inconsciemment, même lorsqu'on lui demande de garder la tête fixe. Ces fluctuations étaient ensuite appliquées au rendu binaural. Une légère augmentation de la sensation d'espace des sujets a pu être observée. Des simulations de mouvements de tête encore plus petits ($\pm 2^\circ$) ont également permis d'accroître l'externalisation dans [18], mais pour seulement 21% des sujets.

Les améliorations observées dans [19] et [18] paraissent plutôt faibles. Des effets plus importants auraient peut-être été observés si les rotations de sources autour du sujet avaient été plus larges, par exemple $\pm 90^\circ$.

Les objectifs de cette seconde expérience est 1) de déterminer si les mouvements de tête peuvent également améliorer l'externalisation chez des sujets naïfs n'ayant aucune expérience du binaural, 2) de déterminer si une amélioration de l'externalisation peut également être observée lorsque les mouvements ne sont non plus effectués par le sujet, mais par les sources virtuelles.

3.1 Protocole

Le protocole de l'Expérience II était très similaire à celui de l'Expérience I. Les sujets écoutaient le même stimulus, mais avec deux orientations seulement : 0° et 180° (où l'externalisation est la plus faible d'après l'Expérience I).

Neuf sujets naïfs ont passé le test (4 femmes et 5 hommes, entre 18 et 50 ans). Aucun d'eux n'avait d'expérience particulière des tests auditifs, ni même de l'écoute binaurale. Aucun sujet n'a rapporté de problème d'audition particulier.

Les sujets passaient les conditions **SØ** et **MT**, ainsi qu'une nouvelle condition notée **Add**. Il y avait donc au total trois conditions pour l'Expérience II :

- **SØ** : tête fixe, sans head-tracking, pas de mouvement de sources.
- **MT** : avec larges mouvements de tête, avec head-tracking, pas de mouvement de sources.
- **Add** : tête fixe, sans head-tracking, mais avec larges mouvements de sources.

Pour la condition **Add**, les sujets devaient garder la tête fixe et regarder droit devant eux durant la totalité du stimulus. Durant les 5.5 s premières secondes du stimulus :

- Orientation 0° : le signal obtenu à partir du microphone principal (microphone 1) tournait automatiquement autour du sujet, d'abord vers la droite (de 0° à $+90^\circ$) puis vers la gauche (de $+90^\circ$ à -90°), avant de revenir à sa position initiale (0°). Les signaux des autres microphones (microphones 2–6 dans la Figure 1) tournaient également de sorte que la scène sonore globale reste cohérente. La rotation de la scène avait été programmée avec une vitesse angulaire constante et une durée de 5.5 s. Ainsi, le mouvement *relatif* de la scène sonore *par rapport* à la tête du sujet était très semblable à celui observé pour la condition **MT**, à l'orientation 0° .
- Orientation 180° : le signal obtenu à partir du microphone principal (microphone 1) tournait

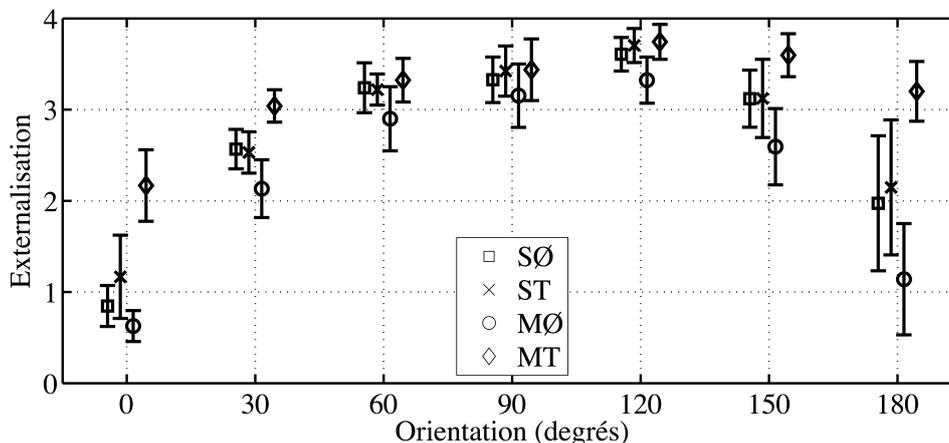


FIGURE 2 – Notes d'externalisation obtenues lors de l'Expérience I, pour chaque condition et chaque orientation, moyennées sur l'ensemble des sujets et des HRTFs. Moyennes normalisées et intervalles de confiance à 95%. Les résultats ont été repliés de part et d'autre du plan médian en raison de la symétrie gauche-droite des résultats.

automatiquement autour du sujet, d'abord vers la gauche (de 180° à -90°) puis vers la droite (de -90° à $+90^\circ$), avant de revenir à sa position initiale (180°). Les signaux des autres microphones tournaient également de sorte que la scène sonore globale reste cohérente. La rotation de la scène avait été programmée avec une vitesse angulaire constante et une durée de 5.5 s. Ainsi, le mouvement *relatif* de la scène sonore *par rapport* à la tête du sujet était très semblable à celui observé pour la condition **MT**, à l'orientation 180° .

Le but de la condition **Add** n'est évidemment pas de reproduire des objets se déplaçant autour du sujet de manière réaliste, mais plutôt d'étudier si des changements continus d'indices binauraux et spectraux semblables aux changements provoqués par une rotation gauche-droite de la tête suffisent à améliorer l'externalisation.

Pour chaque condition, il y avait 3 HRTFs \times 2 orientations, soit 6 essais, répétés 10 fois de suite. Chaque condition nécessitait environ 20 minutes de test, et tous les sujets ont passé les trois conditions sur trois jours différents. L'ordre des conditions était aléatoire et différent pour chaque sujet. Pour une condition donnée, les différentes orientations étaient présentées dans un ordre aléatoire, différent pour chaque sujet et pour chaque répétition du test.

3.2 Résultats

Comme pour l'Expérience I, l'analyse des mouvements de tête montre que les sujets ont relativement bien respecté les consignes pour chacune des conditions, et un test de Friedman n'a pas montré de différence significative entre les trois HRTFs ($p = 0.535$).

La Figure 3 montre les notes d'externalisation pour chaque condition et chacune des deux orientations, moyennées sur l'ensemble des sujets et des HRTFs.

Les résultats montrent que :

- À 0° , l'externalisation la plus importante est obtenue pour la condition **MT** (mouvements de tête avec head-tracking), l'externalisation la moins importante est obtenue pour la condition **SØ** (tête fixe, sans head-tracking), tandis que l'externalisation pour la

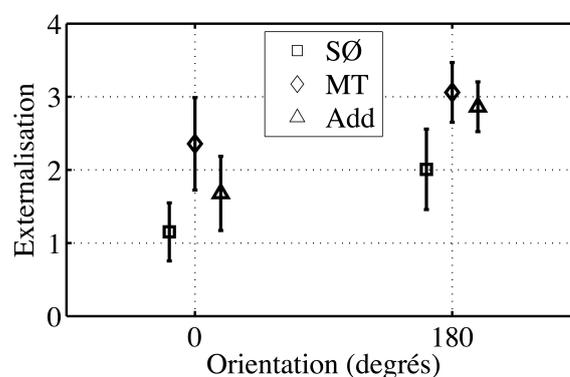


FIGURE 3 – Notes d'externalisation obtenues lors de l'Expérience II, pour chaque condition et chaque orientation, moyennées sur l'ensemble des sujets et des HRTFs. Moyennes normalisées et intervalles de confiance à 95%.

condition **Add** (mouvements de sources) se situe « à mi-chemin » entre les deux autres conditions.

- À 180° , comme à 0° , l'externalisation la plus importante est obtenue pour la condition **MT** et l'externalisation la moins importante est obtenue pour la condition **SØ**. Par contre, l'avantage des mouvements de tête (**MT**) sur les mouvements de sources (**Add**) est bien moins évident.

4 Discussion

Aussi bien dans l'Expérience I que dans l'Expérience II, de larges mouvements de tête avec head tracking ont permis d'améliorer l'externalisation substantiellement pour la plupart des sujets. Dans ces deux expériences, les sujets devaient juger de l'externalisation *après* avoir bougé la tête, contrairement aux études précédentes qui demandaient d'évaluer l'externalisation *pendant* que le sujet bougeait la tête. Le fait d'avoir obtenu une meilleure externalisation pour les conditions **MT** montre ainsi que l'externalisation apportée par des mouvements de tête avec head-tracking *persiste* malgré l'arrêt des mouvements de tête. Il n'est donc pas nécessaire de bouger sans cesse la tête pour mieux

externaliser et un simple aller-retour à $\pm 90^\circ$ permet déjà d'observer des améliorations importantes et robustes.

L'Expérience I montre également que l'amélioration apportée par de larges mouvements avec head tracking (**MT**) est encore plus importante par rapport à une situation où le sujet bouge la tête sans head tracking (**MØ**), ce qui arrive souvent au quotidien (lorsqu'on écoute au casque dans la rue par exemple).

Par contre, aucune amélioration n'a pu être observée pour la condition **ST** par rapport à la condition **SØ**, ce qui suggère que les micro-mouvements involontaires de la tête n'améliorent pas l'externalisation lorsqu'ils sont couplés à un système de head tracking. Ce résultat n'est pas surprenant : une analyse plus en détails des mouvements des sujets durant la condition **ST** montrent en effet que les rotations de tête étaient la plupart du temps inférieures aux MAMA (*Minimum Audible Movement Angles*) rapportés dans la littérature [20].

L'amélioration de l'externalisation due au head tracking a été observée aussi bien avec des sujets experts (Expérience I) que naïfs (Expérience II). Le head-tracking apparaît donc comme un moyen efficace d'améliorer l'externalisation, indépendamment de l'expérience du sujet. De plus, aucune des deux expériences n'a montré de différence significative entre les trois HRTFs non-individualisées utilisées, ce qui suggère que l'efficacité du head tracking est indépendante de l'HRTF choisie.

Ces résultats sont en contradiction avec les études de Begault et al. [4] et Wenzel [12] : dans [4], aucun effet significatif du head tracking sur l'externalisation n'avait été obtenu, et dans [12], l'amélioration du taux d'externalisation grâce au head tracking était significative mais néanmoins faible : de 74.5% à 83.5%.

Plusieurs hypothèses ont déjà été évoquées dans l'introduction pour expliquer ces résultats contradictoires : la présente étude a peut-être obtenu des améliorations plus importantes à cause du stimulus utilisé, qui était plus long (8 s au lieu de 2–3 s), ce qui permettait non seulement aux sujets d'effectuer des mouvements de tête plus larges, mais leur donnait également plus de temps pour pleinement profiter des indices apportés par ces mouvements. Une autre différence importante est que les précédentes études présentaient des résultats moyennés sur des positions frontales, arrières et latérales. Or, comme le montrent les résultats de la présente étude, le head tracking n'a d'intérêt vis-à-vis de l'externalisation que pour les source frontales et arrières, puisque sur les côtés les sources sont déjà très bien externalisées, même sans head-tracking. Il est donc possible que les améliorations faibles voire inexistantes obtenues avec les sources latérales aient masqué dans [4] et [12] des améliorations bien plus marquées obtenues avec les sources frontales et arrières.

Les améliorations importantes observées dans la présente étude peuvent également être expliquées par les systèmes d'enregistrement et de diffusion choisis : un stimulus enregistré avec un arbre microphonique dans une salle à faible réverbération, convolué avec des HRTFs mesurées en chambre anéchoïque (cette « binauralisation » correspond donc à une simulation de haut-parleurs virtuels dans une salle d'écoute anéchoïque). En utilisant une reproduction binaurale plus « idéale » (par exemple, en convoluant une source anéchoïque avec des *Binaural Room Impulse Responses*, comme dans [4]), en enregistrant la voix

d'homme dans une salle plus réverbérante, ou alors en simulant lors de la binauralisation une salle d'écoute avec réverbération, l'externalisation aurait probablement été plus importante pour les conditions sans head-tracking, minimisant ainsi le bénéfice du head-tracking.

Les résultats obtenus lors de la condition **Add** de l'Expérience II montrent que de larges mouvements de sources peuvent également améliorer l'externalisation. Pour l'orientation 180° , l'amélioration apportée par les mouvements de sources était semblable à celle apportée par les mouvements de tête. Pour l'orientation 0° , l'amélioration apportée par les mouvements de sources (**Add**) était bien moins prononcée que celle apportée par les mouvements de tête (**MT**), mais néanmoins significative.

Le consensus actuel est qu'un stimulus sera d'autant mieux externalisé qu'il sera réaliste [11]. Ainsi, les mouvements de tête amélioreraient l'externalisation lorsque le tête est « trackée » tout simplement parce que le tracking permet aux sources virtuelles de bouger de manière cohérente et donc réaliste par rapport aux mouvements de tête du sujet. Pourtant, durant la condition **Add**, les mouvements de sources provoquaient des changements d'indices binauraux et spectraux semblables à ceux de la condition **MT**, sauf que cette fois-ci les mouvements étaient irréalistes et non initiés par le sujet lui-même. Ces mouvements ont quand même amélioré l'externalisation, parfois substantiellement. Cela suggère que les améliorations dues aux mouvements de tête ne sont pas simplement une histoire de cohérence entre les mouvements du sujet et des sources : une part de l'amélioration semble intrinsèquement liée aux changements continus des indices binauraux et spectraux.

À nouveau, les sujets devaient lors de la condition **Add** juger de l'externalisation *après* les mouvements de sources et non *pendant*. L'amélioration de l'externalisation obtenue montre donc qu'une scène sonore, externalisée grâce aux mouvements des sources qui la composent, peut rester externalisée même lorsque les mouvements cessent. En pratique, cela suggère que les mouvements de sources peuvent être une alternative intéressante pour les individus souhaitant écouter du binaural avec une meilleure externalisation, mais ne disposant pas de head-tracker.

5 Conclusion

Dans la présente étude, un stimulus de voix était diffusé au casque à diverses orientations dans le plan horizontal, avec différentes HRTFs non-individualisées pour le rendu binaural. Le « head-tracker » pouvait être soit actif, soit inactif, et les sujets devaient soit bouger la tête, soit rester immobiles. Après chaque présentation d'un stimulus, les sujets devaient évaluer le degré d'externalisation perçue. Les résultats montrent que :

- Les mouvements de tête avec head tracking peuvent améliorer l'externalisation substantiellement pour des sources frontales et arrières, par rapport à une situation où l'auditeur ne bouge pas la tête, et ce aussi bien pour des sujets experts que naïfs.
- Les mouvements de tête avec head tracking peuvent améliorer l'externalisation dans une plus large mesure et à tous les azimuts, par rapport à une

situation où l'auditeur bouge la tête sans head-tracker (une situation que l'on rencontre fréquemment au quotidien).

Les résultats suggèrent également que ces améliorations sont robustes dans le temps, puisqu'elles persistent quand bien même le sujet a cessé tout mouvement.

Une autre expérience a été menée, dans laquelle les rôles étaient inversés : cette fois-ci, les sujets gardaient la tête fixe tandis que les sources sonores tournaient automatiquement autour d'eux. À 180°, l'efficacité des mouvements de sources était très semblable à celle des mouvements de tête pour améliorer l'externalisation. À 0°, les mouvements de sources amélioraient aussi l'externalisation, mais dans une moindre mesure que les mouvements de tête. Ces résultats suggèrent tout de même qu'une rotation automatique d'une scène sonore binaurale peut être une alternative intéressante pour améliorer l'externalisation quand il n'est pas possible d'avoir accès à un head-tracker.

Remerciements

Les auteurs souhaiteraient remercier Alexis Baskind, Thibaut Carpentier, Vincent Koehl, Julian Palacino, Mathieu Paquier, Claire Voirin, Olivier Warusfel et tous les sujets qui ont participé aux tests. Cette étude a été financée en partie par le projet FUI BiLi (« Binaural Listening », www.bili-project.org, FUI-AAP14).

Références

- [1] E. M. Wenzel, M. Arruda, D. J. Kistler, F. L. Wightman, Localization using nonindividualized head-related transfer functions, *J. Acoust. Soc. Am.* **94**, 111–123 (1993).
- [2] S. M. Kim, W. Choi, On the externalization of virtual sound images in headphone reproduction : A Wiener filter approach, *J. Acoust. Soc. Am.* **117**, 3657–3665 (2005).
- [3] D. R. Begault, E. M. Wenzel, Headphone localization of speech, *Hum. Fac. Erg. Soc.* **35**, 361–376 (1993).
- [4] D. R. Begault, E. M. Wenzel, M. R. Anderson, Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source, *J. Audio Eng. Soc.* **49**, 904–916 (2001).
- [5] F. L. Wightman, D. J. Kistler, Resolution of front-back ambiguity in spatial hearing by listener and source movement, *J. Acoust. Soc. Am.* **105**, 2841–2853 (1999).
- [6] C. Mendonça, G. Campos, P. Dias, J. A. Santos, Learning auditory space : generalization and long-term effects, *PloS One* **8**, e77900 (2013).
- [7] N. Sakamoto, T. Gotoh, Y. Kimura, On “out-of-head localization” in headphone listening, *J. Audio Eng. Soc.* **24**, 710–716 (1976).
- [8] E. Hendrickx, P. Stitt, J.-C. Messonnier, J.-M. Lyzwa, B. F. G. Katz, C. de Boishéraud, Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis, *J. Acoust. Soc. Am.* **141**, 3678–3688 (2017).
- [9] J. M. Loomis, C. Hebert, J. G. Cicinelli, Active localization of virtual sounds, *J. Acoust. Soc. Am.* **88**, 1757–1764 (1990).
- [10] J. I. Kawaura, Y. Suzuki, F. Asano, T. Sone, Sound localization in headphone reproduction by simulating transfer functions from the sound source to the external ear, *J. Acoust. Soc. Jpn.* **12**, 203–216 (1991).
- [11] W. O. Brimijoin, A. W. Boyd, M. A. Akeroyd, The contribution of head movement to the externalization and internalization of sounds, *PloS One* **8**, e83068 (2013).
- [12] E. M. Wenzel, The relative contribution of interaural time and magnitude cues to dynamic sound localization, *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 80–83 (1995).
- [13] M. Williams, Microphone arrays for natural multiphony, *Proc. 91st Conv. Audio Eng. Soc.*, paper no. 3157 (1991).
- [14] A. Baskind, T. Carpentier, M. Noisternig, O. Warusfel, J.-M. Lyzwa, Binaural and transaural spatialization techniques in multichannel 5.1 production, *27th Tonmeistertagung, VDT Intl. Conv.* (2012).
- [15] A. Baskind, Hedrot, an open-source head tracker (2016). Disponible sur <https://abaskind.github.io/hedrot/>.
- [16] D. S. Brungart, A. J. Kordik, B. D. Simpson, Effects of headtracker latency in virtual audio displays, *J. Audio Eng. Soc.* **54**, 32–44 (2006).
- [17] O. Warusfel, Listen HRTF database (2003). Disponible sur <http://recherche.ircam.fr/equipements/salles/listen/>.
- [18] G. Wersényi, Effect of emulated head-tracking for reducing localization errors in virtual audio simulation, *IEEE Trans. on audio, speech, and language processing* **17**, 247–252 (2009).
- [19] C. Faller, F. Menzer, C. Tournery, Binaural audio with relative and pseudo head tracking, *Proc. 138th Conv. Audio Eng. Soc.*, paper no. 9223 (2015).
- [20] S. Carlile, J. Leung, The perception of auditory motion, *Trends hear.* **20**, 1–19 (2016).