

Le « head-tracker », une solution pour l'écoute 3D au casque ?

1 Introduction

La synthèse binaurale permet de recréer une scène sonore 3D réaliste à l'aide d'un casque audio. Ainsi, en fermant les yeux, l'utilisateur pourra percevoir un violon en face de lui, une clarinette sur sa gauche, ou encore un avion volant au-dessus de sa tête (contrairement à l'écoute stéréophonique au casque traditionnelle, dans laquelle les sources sonores ne peuvent être perçues qu'à l'intérieur de la tête, étalées le long d'une ligne fictive entre les deux oreilles).

Idéalement, la synthèse binaurale doit être *individualisée*, c'est-à-dire parfaitement adaptée à la morphologie de l'utilisateur. Une telle individualisation nécessite la mesure de ce qu'on appelle les HRTFs (*Head-Related Transfer Functions*, ou fonctions de transfert de la tête), sorte de scanner permettant d'établir comment le corps et le visage d'un individu agit sur le son selon sa provenance. Il s'agit d'une opération coûteuse et délicate (nécessitant une chambre anéchoïque, du personnel qualifié et du matériel spécialisé), inenvisageable à grande échelle.

Une solution plus simple consiste à écouter du binaural « à travers les oreilles » d'un autre individu, dont les HRTFs ont déjà été mesurées. On parle alors de synthèse binaurale *non-individualisée*. Plusieurs études ont cependant montré qu'une telle synthèse pouvait provoquer un certain nombre d'artefacts, notamment l'internalisation des sources sonores [1, 2], 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 précédentes ont envisagé l'utilisation d'un head-tracker¹ pour remédier à ce problème, mais sans aboutir à des résultats convaincants (voir [3] pour une analyse critique de ces études).

Depuis (date ??), l'équipe du Service Audiovisuel du Conservatoire de Paris explore la synthèse binaurale non-individualisée (ici 2-3 phrases sur les diverses productions, webTV, etc. que vous avez effectuées en binaural jusqu'à maintenant). Ces années de pratique ont convaincu l'équipe que le head-tracker pouvait efficacement résoudre les problèmes d'internalisation du son. Sous l'impulsion du projet Bili², le Conservatoire de Paris a décidé de mettre en place plusieurs tests perceptifs pour vérifier scientifiquement cette hypothèse, qui ont fait l'objet de deux publications dans des revues scientifiques [3, 4].

2 Expérience I

2.1 Protocole

Une première expérience, sans head-tracker, a été mise en place au lycée Charlemagne (Paris IV), pour prendre la

1. Le head-tracker est un capteur posé sur la tête de l'utilisateur. Directement relié au module de synthèse binaurale, il permet de prendre en compte les mouvements de l'utilisateur en temps réel. Ainsi, si l'utilisateur tourne la tête vers la droite, le violon qui se trouvait devant lui se retrouvera sur sa gauche, etc.

2. Projet FUI (Fonds unique interministériel) de recherche autour du binaural, regroupant divers partenaires tels que France Télévisions, Radio France, Orange, le LIMSI-CNRS ou encore l'IRCAM

mesure du problème d'internalisation du son.

Dans cette expérience, les sujets devaient écouter au casque des stimuli de 5 secondes : soit une voix d'homme récitant un poème, soit un morceau de maracas.

La synthèse binaurale permet de spatialiser la voix de l'homme ou le maracas n'importe où dans l'espace 3D autour du sujet. Pour cette première expérience, 8 directions différentes ont été choisies. Ainsi, le son de la voix d'homme ou du maracas provenaient soit de devant (0°), soit de derrière (180°), soit des côtés ($\pm 45^\circ$, $\pm 90^\circ$, $\pm 135^\circ$). Cette spatialisation du son a été effectuée à l'aide du moteur de synthèse binaurale *Bipan* [5], développé par Alexis Baskind en étroite collaboration avec le CNSMDP.

Après chaque présentation d'un stimulus, le sujet devait répondre sur un questionnaire papier s'il avait perçu le son à l'intérieur ou à l'extérieur de sa tête. Ensuite, le stimulus suivant était diffusé.

Le test perceptif a eu lieu dans une salle de classe du lycée Charlemagne. Les stimuli étaient présentés avec un casque Sennheiser Ultrason HFI580, à un niveau de diffusion confortable. Des distributeurs et amplificateurs Feichter Audio permettaient de faire passer le test à une trentaine de sujets en même temps.

98 sujets ont pris part au test (46 femmes et 52 hommes, âgés de 15 à 21 ans), tous élèves du lycée Charlemagne. 89% d'entre eux écoutaient de la musique au casque « souvent » voire « très souvent », tandis que 2% seulement affirmaient ne jamais écouter au casque. 89% des sujets n'avaient jamais écouté de contenus en binaural.

La moitié des sujets ont passé le test avec la voix d'homme, l'autre moitié avec le morceau de maracas. Le test durait 15 minutes.

2.2 Résultats

Une analyse statistique montre qu'il n'y a pas de différence significative entre les résultats du groupe « maracas » et les résultats du groupe « voix d'homme ». Ce résultat est intéressant car il suggère que même pour des stimuli très différents (ici, une voix d'homme et un maracas), les problèmes d'internalisation sont à peu près les mêmes.

La Fig. 1 montre les taux d'externalisation (c'est-à-dire le pourcentage de fois qu'un stimulus était perçu à l'extérieur de la tête) pour chaque direction, moyennés sur l'ensemble des sujets.

Les résultats montrent que, malgré l'absence de head-tracker, l'externalisation est élevée sur les côtés ($\pm 45^\circ$, $\pm 90^\circ$, $\pm 135^\circ$), avec des taux d'externalisation entre 74% et 81%. Par contre, la chute d'externalisation est vertigineuse à l'avant et à l'arrière : seulement 16% à 0° et 20% à 180° .

3 Expérience II

Une seconde expérience a été mise en place pour étudier l'influence du head-tracking sur l'externalisation. La première expérience suggère que c'est surtout dans le cas des sources frontales et arrière que l'on peut s'attendre à un effet bénéfique du head-tracker. Sur les côtés, comme

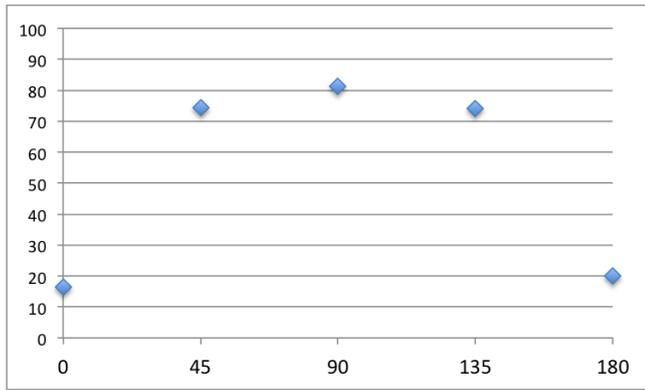


FIGURE 1 – Taux d’externalisation obtenus lors de l’Expérience I, pour chaque direction. Les résultats ont été repliés de part et d’autre du plan médian en raison de la symétrie gauche-droite des résultats.

les sujets externalisent très bien même sans head-tracker, la marge de progrès est bien moins importante.

3.1 Protocole

Comme la première expérience n’avait pas montré de différence significative entre la voix d’homme et le maracas, il a été décidé pour cette seconde expérience de ne garder que la voix d’homme.

Pour gagner en précision, 12 directions et non pas 8 ont été étudiées, par pas de 30° : 0°, ±30°, ±60°, ±90°, ±120°, ±150°, 180°. La spatialisation du son a de nouveau été effectuée avec le moteur binaural *Bipan*, ainsi que le système de head-tracking *Hedrot* [6], également développé par Alexis Baskind en étroite collaboration avec le CNSMDP.

Le test perceptif s’est déroulé dans une cabine insonorisée du Conservatoire de Paris. Les lumières étaient éteintes et le sujet était assis au centre de la pièce. Les stimuli étaient présentés avec un casque Sennheiser HD 600.

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 ±90°. Le head-tracker était soit actif, soit inactif. Le sujet devait donc évaluer 4 conditions différentes :

SØ : tête fixe, sans head-tracking.

ST : tête fixe, avec head-tracking³.

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

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

Une fois la présentation d’un stimulus terminée, le sujet devait rapporter à quel point la voix d’homme était externalisée, en utilisant une échelle à 6 points affichée

3. 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) pouvaient potentiellement améliorer l’externalisation [7].

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.

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

sur un écran d’ordinateur (voir Table 1). Une fois que le sujet avait répondu, le stimulus suivant était diffusé. Le test durait environ 4 h, divisées en 4 sessions d’1 h sur 4 jours différents.

3.2 Résultats

La Fig. 2 montre les notes d’externalisation pour chaque condition et chaque direction, moyennées sur l’ensemble des sujets et des HRTFs.

Il peut être observé que :

- Les résultats obtenus pour les conditions sans mouvement de tête sont très semblables, que le head-tracker 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 directions latérales (±60°, ±90°, ±120°), 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 directions frontales et arrières (0°, ±30°, ±150°, 180°) l’externalisation a été plus modérée pour les conditions sans mouvement de tête (**SØ** and **ST**). Aux directions ±30°, ±150° et 180°, les mouvements de tête avec head-tracking (**MT**) ont permis de maintenir une externalisation importante, comparable à celle obtenue pour les directions 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).

Dans cette deuxième expérience, les sujets étaient des ingénieurs du son professionnels, habitués à l’écoute de contenus binauraux. Une troisième expérience, non développée ici, a été mise en place pour vérifier si le head-tracking pouvait également améliorer l’externalisation chez des sujets n’étant pas des professionnels du son, et n’ayant même aucune expérience de l’écoute en binaural. Des résultats très semblables ont été observés (voir [4]).

4 Conclusion

Ces expériences ont montré que le head-tracking peut considérablement améliorer l’externalisation en écoute

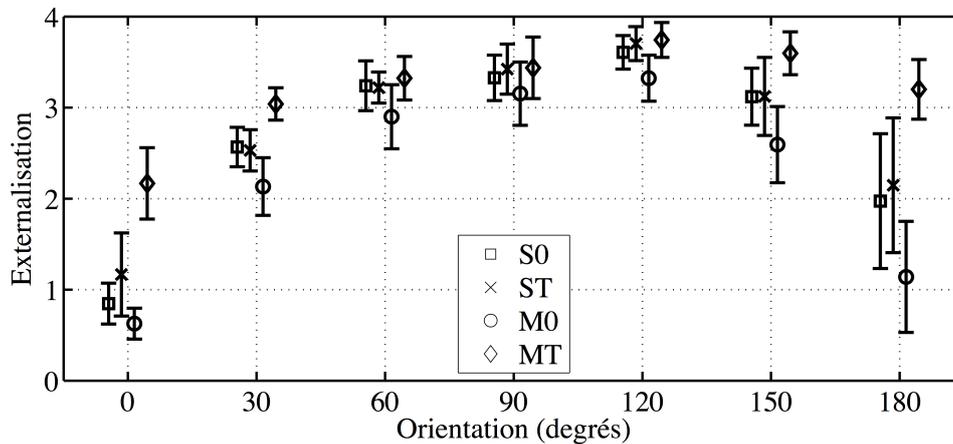


FIGURE 2 – Notes d’externalisation obtenues lors de l’Expérience I, pour chaque condition et chaque direction. Moyennes sur l’ensemble des sujets, avec intervalles de confiance à 95%. Les résultats ont été repliés de part et d’autre du plan médian en raison de la symétrie gauche-droite des résultats.

binaurale non-individualisée, surtout pour des sources sonores situées à l’avant ou à l’arrière. Cette amélioration est particulièrement importante par rapport à une situation où le sujet bouge la tête sans head-tracking (M0), une situation fréquente 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 S0, ce qui suggère que le head-tracking ne peut améliorer l’externalisation que si les mouvements de tête de l’utilisateur sont suffisamment amples.

L’amélioration de l’externalisation due au head-tracking a été observée aussi bien avec des sujets experts que naïfs. Le head-tracking apparaît donc comme un moyen efficace d’améliorer l’externalisation, indépendamment de l’expérience du sujet.

Remerciements

Cette étude a été financée en partie par le projet FUI BiLi (“Binaural Listening”, www.bili-project.org, FUI-AAP14).

Références

- [1] S. M. Kim and W. Choi, “On the externalization of virtual sound images in headphone reproduction : A Wiener filter approach,” *J. Acoust. Soc. Am.*, vol. 117, pp. 3657–3665, 2005. <https://doi.org/10.1121/1.1921548>.
- [2] D. R. Begault and E. M. Wenzel, “Headphone localization of speech,” *Hum. Fac. Erg. Soc.*, vol. 35, pp. 361–376, 1993. <https://doi.org/10.1177/001872089303500210>.
- [3] E. Hendrickx, P. Stitt, J.-C. Messonnier, J.-M. Lyzwa, B. F. G. Katz, and C. de Boishéraud, “Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis,” *J. Acoust. Soc. Am.*, vol. 141, pp. 3678–3688, 2017. <https://doi.org/10.1121/1.4978612>.
- [4] E. Hendrickx, P. Stitt, J.-C. Messonnier, J.-M. Lyzwa, B. F. G. Katz, and C. de Boishéraud,

“Improvement of externalization by listener and source movement using a “binauralized” microphone array,” *J. Audio Eng. Soc.*, vol. 65, pp. 589–599, 2017. <https://doi.org/10.17743/jaes.2017.0018>.

- [5] A. Baskind, T. Carpentier, M. Noisternig, O. Warusfel, and J.-M. Lyzwa, “Binaural and transaural spatialization techniques in multichannel 5.1 production,” in *27th Tonmeistertagung, VDT Intl. Conv.*, 2012.
- [6] A. Baskind, “Hedrot, an open-source head tracker,” 2016. <https://abaskind.github.io/hedrot/>.
- [7] G. Wersényi, “Effect of emulated head-tracking for reducing localization errors in virtual audio simulation,” in *IEEE Trans. on audio, speech, and language processing*, vol. 17, pp. 247–252, 2009. <https://doi.org/10.1109/TASL.2008.2006720>.