



HAL
open science

Perception des variations de la réverbération en fonction de l'orientation d'une source en réalité virtuelle

Pablo Abehsera-Morell, David Poirier-Quinot, Brian F G Katz

► To cite this version:

Pablo Abehsera-Morell, David Poirier-Quinot, Brian F G Katz. Perception des variations de la réverbération en fonction de l'orientation d'une source en réalité virtuelle. Journées Françaises de la Réalité Étendue (jf.XR), AFXR, Oct 2024, Évry, France. <hal-05579536>

HAL Id: hal-05579536

<https://hal.science/hal-05579536v1>

Submitted on 3 Apr 2026

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.



Distributed under a Creative Commons CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License

Perception des variations de la réverbération en fonction de l'orientation d'une source en réalité virtuelle

Perception of changes in source orientation-dependent reverberation in virtual reality

Pablo Abehsera-Morell, David Poirier-Quinot, and Brian F.G. Katz

English Abstract—In extended reality experiences, dynamic rendering updates the auralisation as changes occur in the scene. In this study, we are interested in how the reverberant part of a directional, vocal sound source is perceived as it rotates and whether it contributes to the perceived naturalness of the scene's acoustics. An experiment was conducted where participants were immersed in a virtual scene displaying a virtual person talking. The person's voice was auralised with a range of dynamic rendering approaches. Participants were asked to rate the perceived naturalness of the acoustics of the scene when listening to the different auralisations. The results of the perceptual test provide insights into how the changes in the reverberation pattern were perceived and how those changes affected perceived naturalness.

1 INTRODUCTION

Dans le cadre de la création d'expériences en Réalité Étendue (XR), l'auralisation est utilisée pour attribuer les caractéristiques acoustiques souhaitées aux sources sonores virtuelles. Pour qu'une source virtuelle soit perçue comme plausible, ses caractéristiques acoustiques doivent correspondre à celles qu'elle aurait si elle était présente physiquement dans l'espace où elle est restituée [4], [5].

Le rendu dynamique est utilisé pour mettre à jour l'auralisation lorsque des changements, tels que le déplacement de la source ou de l'auditeur, se produisent. Les changements d'orientation peuvent être représentés en appliquant le rendu de la directivité.

Dans un environnement réverbérant, la directivité d'une source influe sur l'intensité de chacune des réflexions qui arrivent aux oreilles d'un auditeur à partir de différentes directions. Ce phénomène se traduit par des différences de signaux spectraux et spatiaux audibles [3].

Il est généralement admis que l'intégration du rendu de la directivité dans l'auralisation permet d'accroître la plausibilité de la source virtuelle, dans un contexte audio uniquement [7] ou audio-visuel [2].

Dans cette étude, on propose d'évaluer comment différents degrés de rendu de la directivité sont

perçus par des auditeurs lorsque la source virtuelle est visible. Une auralisation omnidirectionnelle (ou statique) d'une source qui tourne visiblement pourrait être perçue comme moins naturelle qu'une auralisation avec traitement de la directivité [6], car l'acoustique ne serait pas en accord avec l'image. À contrario, regarder la source tourner pourrait tromper l'auditeur en lui faisant percevoir un changement acoustique qui n'a pas lieu. Il serait pertinent qui plus est d'observer comment une mise en œuvre partielle de la directivité, ou une mise en œuvre complètement dynamique mais partiellement erronée, se comparerait à chaque extrême.

Une expérience a été menée pour tenter de répondre à cette question. Des participants ont été immergés dans des scènes virtuelles, à l'aide d'un casque de réalité virtuelle et d'un casque audio, et ont évalué le degré de naturel d'une série d'auralisations d'une source virtuelle visible. La source était toujours l'enregistrement d'une personne, qui racontait une histoire ou faisait d'autres formes de vocalisations comme chanter, siffler ou faire de la beat-box, tout en se tournant dans différentes directions.

Les résultats de cette expérience pourraient aider à déterminer le degré de traitement de la directivité à utiliser dans l'auralisation binaurale de sources sonores virtuelles visibles dans des applications de réalité étendue.

2 MÉTHODES

2.1 Environnements acoustiques et visuels

Des réponses impulsionnelles binaurales (BRIRs) ont été capturées dans une salle de (l 12 m, p 9 m, h 3 m; $T30_{mid} = 0.47$ s) à l'aide d'un haut-parleur directionnel

- Pablo Abehsera-Morell: Sorbonne Université, CNRS, Institut Jean Le Rond d'Alembert, UMR 7190, Paris, France
E-mail: pablo.abehsera_morell@sorbonne-universite.fr.
- David Poirier-Quinot: Sorbonne Université, CNRS, Institut Jean Le Rond d'Alembert, UMR 7190, Paris, France
E-mail: david.poirier-quinot@sorbonne-universite.fr.
- Brian F.G. Katz: Sorbonne Université, CNRS, Institut Jean Le Rond d'Alembert, UMR 7190, Paris, France
E-mail: brian.katz@sorbonne-universite.fr.



Fig. 1. Système de mesure dans la salle. **Pos_{center}** (G) et **Pos_{wall}** (D).

(Genelec 8331A) et d'une tête binaurale (Neumann KU-80) équipée de microphones (DPA 4060). Deux configurations ont été mesurées : l'une au centre de la pièce (**Pos_{center}**) et l'autre à 1 m d'un mur avec fenêtres (**Pos_{wall}**), la source et le récepteur étant séparés de 2 m, à 1.5 m de hauteur. Pour chaque configuration, des BRIRs ont été capturées pour 36 orientations de la source, de 0° à 350° par pas de 10° . Une caméra 360° (Ricoh Theta X) a été utilisée pour filmer la pièce pour chaque position de récepteur.

2.2 Sources acoustiques et visuelles

Des enregistrements de deux orateurs ont été réalisés dans une chambre anéchoïque avec un micro cravate (DPA 4088) et la caméra 360° placée à une distance de 2 m. Les orateurs parlaient ou vocalisaient en tournant la tête, et des enregistrements de 12 s ont été capturés sur fond vert, puis superposés dans la capture visuelle de la salle après suppression du fond faite via Adobe Premiere Pro. Les données d'orientation des orateurs ont été générées via Blender, synchronisées avec un avatar reproduisant leurs mouvements en temps réel. Les animations, enregistrées à 30 Hz, ont été exportées en fichiers texte.

2.3 Méthodes d'auralisation

Quatre stratégies d'auralisation des sources, ou conditions acoustiques, ont été évaluées : (1) **Ac_{real}**, avec les BRIRs mesurées et convolution dynamique selon l'orientation du locuteur; (2) **Ac_{direct}**, où la convolution dynamique apportait directivité correcte dans le champ direct mais réverbération identique pour toutes les orientations; (3) **Ac_{static}**, avec une seule BRIR (fixe à 0°); et (4) **Ac_{invert}**, où la convolution dynamique était réalisée mais en utilisant une orientation de source inversée à 180° . Neuf clips ont été générés : 3 de parole naturelle (**Stim_{natural}**), 3 forcés (**Stim_{forced}**: parole intense, chant, beat-box), et 3 d'entraînement (**Stim_{training}**: vocalisations arbitraires). Ces stimuli permettent d'évaluer les conditions acoustiques dans divers scénarios, pour différentes orientations de la source.

2.4 Expérience perceptive

Un test d'écoute a été réalisé via une application développée sous Unity. Les participants, portant un

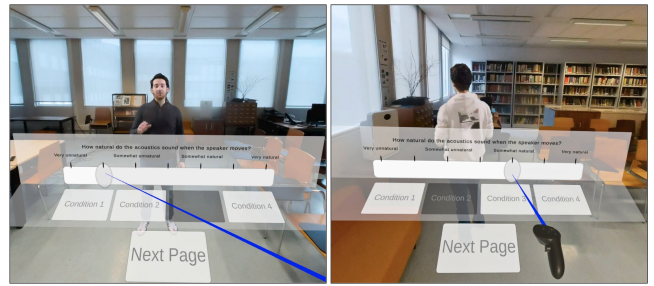


Fig. 2. Perspective des participants pour **Pos_{center}** (G) et **Pos_{wall}** (D).

casque de réalité virtuelle (Meta Quest 3) et des écouteurs (Sennheiser HD-600), étaient placés dans un studio acoustiquement semi-isolé. Dans l'application de réalité virtuelle, ils se trouvaient placés dans la salle virtualisée à l'une des deux positions capturées, ne pouvant pas se déplacer mais uniquement tourner la tête. Ils étaient invités à rester debout et regarder en direction du personnage en face d'eux. Le signal audio était coupé s'ils tournaient la tête de plus de 10° . Une interface utilisateur virtuelle placée devant eux permettait aux participants de sélectionner et redémarrer les lectures à l'aide d'un pointeur laser, avec une échelle de Likert à 7 points pour évaluer le naturel de l'acoustique.

Le test comportait 15 scènes : 3 scènes d'entraînement avec les stimuli **Stim_{training}**, et 12 scènes avec les 6 stimuli **Stim_{natural}** et **Stim_{forced}**, aléatoirement répartis entre les deux positions **Pos_{center}** et **Pos_{wall}**. Dans chaque scène, les participants devaient évaluer chacune des 4 méthodes d'auralisation (sans savoir laquelle était laquelle) pour la combinaison donnée de stimulus et position dans la pièce. Une pause était proposée à mi-parcours. Le test était suivi d'un entretien pour recueillir les impressions des participants.

17 participants avec une audition normale (4 femmes, 12 hommes, 1 non-binaire) et d'âge moyen 28.2 ± 3.5 ans, ont pris part à l'expérience. 8 participants ont rapportés travailler dans un domaine lié à l'écoute critique en RV, jugés comme "experts" dans l'analyse. Le test a duré en moyenne 27 ± 7 minutes et chaque participant a été rémunéré 15€.

3 RÉSULTATS

Pour évaluer l'impact de (1) la condition acoustique, (2) la position dans la salle et (3) le stimulus sur les évaluations des participants, des analyses de variance (ANOVA) ont été réalisées. Les interactions de premier ordre entre ces facteurs ont également été analysées. Une différence significative est reportée pour des p -values inférieures à 0,05, et la notation $p < \varepsilon$ est utilisée pour indiquer des p -values inférieures à 10^{-3} . Les comparaisons post-hoc pour les facteurs significatifs ont été effectuées avec des p -values ajustées

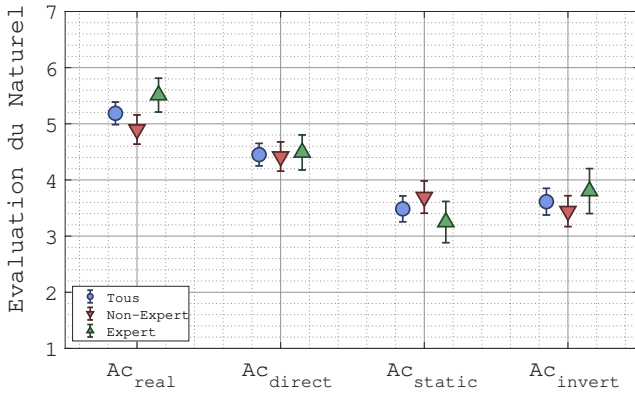


Fig. 3. Impact global de la condition acoustique sur les évaluations des participants. Moyenne et intervalles de confiance (IC) à 95 %.

de Tukey-Kramer ou de Wilcoxon pour comparer des échantillons déséquilibrés.

3.1 Impact direct des facteurs étudiés

La condition acoustique a influencé de façon significative les évaluations ($F = 51.7, p < \epsilon$), résultat illustré à la Figure 3. Ac_{real} a été mieux notée qu' Ac_{direct} (5.2 ± 0.1 vs $4.5 \pm 0.1, p < \epsilon$), elle-même mieux notée qu' Ac_{static} (4.5 ± 0.1 vs $3.5 \pm 0.1, p < \epsilon$) et Ac_{invert} (4.5 ± 0.1 vs $3.6 \pm 0.1, p < \epsilon$). Aucune différence n'a été observée entre Ac_{static} et Ac_{invert} ou entre Pos_{center} et Pos_{wall} . Le facteur stimuli a eu un impact significatif ($F = 3.1, p = 0.008$), avec deux $Stim_{natural}$ mieux notés que la beat-box de $Stim_{forced}$ (4.4 ± 0.1 et 4.3 ± 0.1 vs $3.7 \pm 0.1, p = 0.002$ et $p = 0.013$) et le chant mieux noté que la beat-box (4.4 ± 0.1 vs $3.7 \pm 0.1, p = 0.003$).

3.2 Autres interactions

L'analyse de variance n'a montré aucune interaction significative entre les facteurs acoustique, position, stimulus et catégorie de stimuli. Comme le montre la Figure 4, Ac_{invert} a été noté plus bas à Pos_{wall} qu'à Pos_{center} (4.0 ± 0.2 vs $3.3 \pm 0.2, p = 0.002$), sans différence pour les autres conditions. Les experts ont noté Ac_{invert} au-dessus d' Ac_{static} (3.8 ± 0.2 vs $3.2 \pm 0.2, p = 0.043$) et ont attribué à Ac_{invert} des notes similaires à celles d' Ac_{direct} à Pos_{center} (4.7 ± 0.2 et 4.2 ± 0.3 vs $3.1 \pm 0.3, p < \epsilon$). Ce changement n'a pas été observé à Pos_{wall} .

3.3 Entretiens

Les participants se sentaient le plus confiants en évaluant les auralisations des $Stim_{natural}$, les non-experts étant habitués à entendre des conversations dans la vraie salle, ce qui les aidait à établir une référence interne.

La beat-box $Stim_{forced}$ était en revanche jugé difficile à évaluer en raison de l'absence de référence interne.

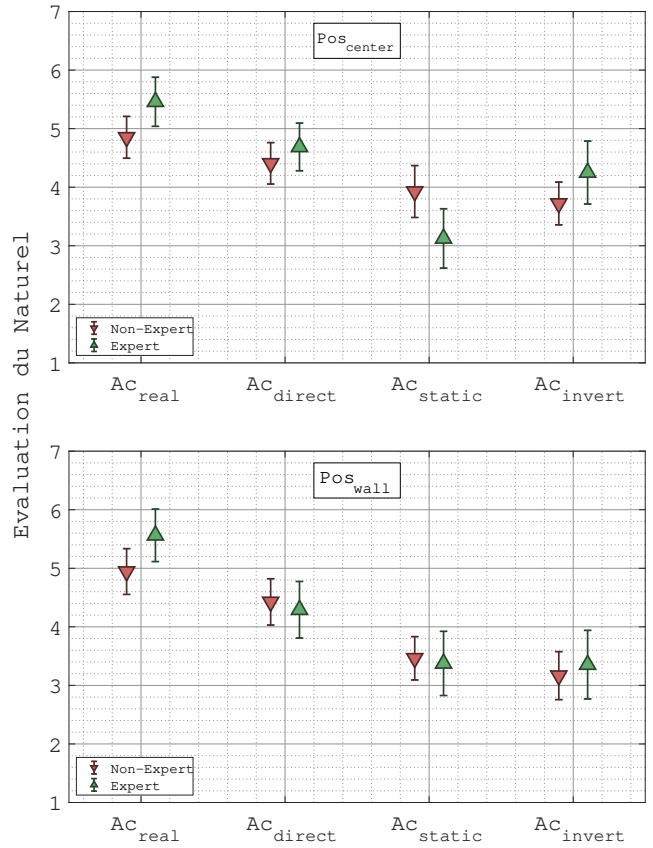


Fig. 4. Impact de la condition acoustique sur les évaluations des participants pour les deux positions dans la pièce. Moyenne et IC à 95 %.

Le mouvement des locuteurs influençait aussi leur jugement. Les participants trouvaient qu'une rotation progressive était plus facile à apprécier qu'un mouvement brusque. Environ la moitié des participants se sont sentis plus à l'aise pour évaluer les conditions acoustiques à Pos_{wall} , tandis que d'autres estimaient que la position n'affectait pas leur jugement.

La majorité des participants ont noté qu'une condition était toujours statique (ou omnidirectionnelle) et l'ont souvent jugée moins naturelle, bien que certains préféraient cette condition, trouvant les changements de réverbération dans les conditions dynamiques trop rigoureux et irréalistes.

4 DISCUSSION

L'impact des conditions acoustiques sur les évaluations ($Ac_{real} > Ac_{direct} > Ac_{static}$) était attendu, le réalisme étant affecté par la dégradation de la directivité. Cela concorde avec la littérature existante, qui indique que le rendu de la directivité contribue à des auralisations plus réalistes. La présence d'une référence visuelle n'a pas masqué les différences acoustiques entre Ac_{real} et les conditions dégradées.

Les conditions Ac_{invert} et Ac_{static} ont été notées de manière similaire, suggérant qu'un changement

acoustique lors de la rotation d'une source visible n'est pas suffisant sans congruence visuelle. Cependant, les experts ont noté Ac_{invert} plus haut que Ac_{static} , et au niveau de Pos_{center} , Ac_{invert} a été noté aussi haut qu' Ac_{direct} , indiquant qu'un rendu dynamique mais incorrect de la directivité peut parfois sembler plus naturel qu'un rendu sans directivité.

Les deux groupes de participants ont noté Ac_{invert} plus bas à Pos_{wall} qu'à Pos_{center} , probablement en raison des différences acoustiques plus marquées à Pos_{wall} à cause du mur proche. Ces résultats mettent en évidence l'importance d'un rendu correct de la directivité lorsque l'environnement acoustique engendre des différences de signaux acoustiques fortes.

Enfin, l'impact des stimuli sur la capacité des participants à discerner les conditions acoustiques était plus faible que prévu. Bien que l'on s'attendait à ce que $Stim_{forced}$ révèle mieux les différences, les interviews ont suggéré que l'évaluation des $Stim_{natural}$ était plus facile. D'autres facteurs, tels que la nature des changements d'orientation, ont pu influencer la perception du naturel, Ac_{real} étant perçu comme plus naturel lors de changements d'orientation cohérents et progressifs.

5 TRAVAUX À VENIR

Une analyse plus détaillée des résultats des tests perceptifs est actuellement en cours, dans le but de trouver des relations entre des différences constatées dans l'analyse physique des réponses impulsionnelles utilisées et les résultats de l'expérience.

Concernant des études futures, une prochaine étape serait d'explorer l'impact du rendu de la directivité sur le réalisme perçu d'une source virtuelle visible, en mettant l'accent sur les réflexions précoces. Nos résultats montrent que les participants peuvent distinguer un rendu correct de la directivité des approches simplifiées; toutefois, des recherches récentes s'intéressent à la mesure dans laquelle il est possible de simplifier le rendu tout en restant indiscernable d'un point de vue perceptif de la mesure de référence [1].

Une autre voie serait de rendre la tâche plus écologique. Bien que les résultats soient utiles, évaluer l'impact du rendu de la directivité sur le réalisme des expériences en XR sans que les participants se concentrent sur l'acoustique, en adoptant des paradigmes comme ceux de l'étude conduit en [8], serait intéressant. Par exemple, un dispositif où deux participants dialoguent pourrait mieux illustrer l'importance de la directivité pour des interactions naturelles en réalité étendue.

6 CONCLUSION

Cet article présente les résultats d'une expérience visant à évaluer l'impact du rendu de la réverbération lié à la directivité sur le naturel perçu d'une source audio virtuelle visible. Les participants étaient

immergés dans une scène en RV où une personne leur parlait en tournant dans différentes directions. Plusieurs stratégies d'auralisation ont été évaluées. Les auralisations avec un rendu de directivité réaliste dans la réverbération étaient généralement perçues comme plus naturelles que celles avec une réverbération statique ou une directivité partiellement incorrecte. Cependant, une mise en œuvre partielle de la directivité était mieux notée qu'aucune mise en œuvre, bien qu'une mise en œuvre complète mais partiellement erronée puisse varier selon la position dans la pièce.

D'autres facteurs, comme l'expertise d'écoute et la nature du mouvement de rotation de la source, ont également joué un rôle important. Ainsi, il en ressort que la présence d'un stimulus visuel permettait aux auralisations de bénéficier d'un rendu de directivité correct et congruent. Plutôt que de tromper les participants en leur faisant percevoir un rendu de directivité erroné comme correct, le visuel mettait en évidence les incongruences causées par des auralisations ne reflétant pas correctement l'affichage.

7 REMERCIEMENTS

Nous tenons à remercier les participants qui ont pris part à l'expérience d'écoute. Ce travail a été réalisé dans le cadre du projet SONICOM (www.sonicom.eu) qui a reçu un financement du programme de recherche et d'innovation Horizon 2020 de l'Union européenne dans le cadre de la convention de subvention n° 101017743.

REFERENCES

- [1] P. Abehsera-Morell, D. Poirier-Quinot, and B. F. G. Katz. Projecting source directivity variations onto an existing binaural room impulse response. In *Aud. Eng. Soc. Int. Conf. on Spatial and Immersive Aud.*, number 37, Huddersfield, UK, 2023.
- [2] M. Blau, A. Budnik, M. Fallahi, H. Steffens, S. D. Ewert, and S. Van de Par. Toward realistic binaural auralizations—perceptual comparison between measurement and simulation-based auralizations and the real room for a classroom scenario. *Acta Acustica*, 5(8), 2021.
- [3] J. Bradley, H. Sato, and M. Picard. On the importance of early reflections for speech in rooms. *J. of the Acoust. Soc. of America*, 113:3233–3244, 07 2003.
- [4] K. Brandenburg, F. Klein, A. Neidhardt, U. Sloma, and S. Werner. Creating Auditory Illusions with Binaural Technology. In J. Blauert and J. Braasch, editors, *The Technology of Binaural Understanding*, pages 623–663. Springer International Publishing, 2020.
- [5] A. Neidhardt, C. Schneiderwind, and F. Klein. Perceptual Matching of Room Acoustics for Auditory Augmented Reality in Small Rooms - Literature Review and Theoretical Framework. *Trends in Hearing*, 26, 2022.
- [6] B. N. J. Postma and B. F. G. Katz. Influence of visual rendering on the acoustic judgements of a theater auralization. In *Proc. Mtgs. Acoust.*, number 015008, Boston, Massachusetts, 2017.
- [7] U. Sloma, S. Werner, F. Klein, and T. Pappachan. Synthesis of binaural room impulse responses for different listening positions considering the source directivity. In *147th Aud. Eng. Soc. Conv.*, number 10237, 10 2019.
- [8] J. Wendt, B. Weyers, J. Stienen, A. Bönsch, M. Vorländer, and T. W. Kuhlen. Influence of Directivity on the Perception of Embodied Conversational Agents' Speech. In *Proc. of the 19th ACM Int. Conf. on Intell. Virtual Agents*, pages 130–132, Paris France, July 2019. ACM.