



Influence des visuels sur l'évaluation subjective d'auralisations dans un contexte de réalité virtuelle

David Thery, Barteld N J Postma, Brian F. G. Katz

► To cite this version:

David Thery, Barteld N J Postma, Brian F. G. Katz. Influence des visuels sur l'évaluation subjective d'auralisations dans un contexte de réalité virtuelle. 29ème conférence francophone sur l'Interaction Homme-Machine, AFIHM, Aug 2017, Poitiers, France. 10 p., 10.1145/3132129.3132164 . hal-01578352

HAL Id: hal-01578352

<https://hal.science/hal-01578352>

Submitted on 29 Aug 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.

Influence des visuels sur l'évaluation subjective d'auralisations dans un contexte de réalité virtuelle

David Thery

LIMSI, Université Paris-Sud, CNRS
Orsay 91400, FRANCE
david.thery@limsi.fr

Barteld N.J. Postma

LIMSI, Université Paris-Sud, CNRS
Orsay 91400, FRANCE
postma@limsi.fr

Brian F.G. Katz

Institut Jean Le Rond d'Alembert
Sorbonne Universités, UPMC, CNRS
Paris 75005, FRANCE
brian.katz@upmc.fr

Abstract

Nous étudions ici l'impact de la composante visuelle sur l'évaluation subjective d'auralisations, permettant également d'évaluer l'impact du support de rendu visuel (comparaison entre CAVE-portable ou Head Mounted Display - HMD). Une visualisation 3D du Théâtre de l'Athénée incorporant une performance virtuelle de deux acteurs jouant sur scène, et incluant un rendu audio binaural de la scène, avec suivi de mouvements de la tête du sujet ont permis cette évaluation.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

IHM'17, August 29-September 1, 2017, Poitiers, France

© 2017 Copyright is held by the owner/author(s).

ACM ISBN 978-1-4503-5109-6/17/08. <https://doi.org/10.1145/3132129.3132164>

CCS Concepts

• **Human-centered computing** → **Virtual reality**; *Sound-based input / output*; • **Information systems** → *Speech / audio search*;

Keywords

Réalité virtuelle; Audio 3D; Évaluation subjective, Auralisation, Multimodal

ACM Reference Format :

D. Thery, B.N.J. Postma and B.F.G. Katz 2017. Influence des visuels sur l'évaluation subjective d'auralisations dans un contexte de réalité visuelle. In *Proceedings of 29ème Conférence Francophone sur l'Interaction Homme-Machine, Poitiers Futuroscope, FR Août 2017 (IHM 2017)*, 6 pages.

DOI : 10.1145/3132129.3132164

1 INTRODUCTION

L'auralisation, de manière analogue à la visualisation, est une technique permettant de rendre audible, par modélisation physique ou mathématique, le champ sonore d'une source dans un espace, de manière à simuler l'écoute binaurale à une position donnée dans l'espace modélisé [12].

Dans un objectif d'amélioration des auralisations et plus généralement des représentations virtuelles les incluant, Postma et Katz ont étudié la fidélité subjective de ces auralisations [2, 4]. Ils présentent respectivement une procédure méthodique de calibration d'auralisations simulées (basée sur une minimisation des écarts entre paramètres simulés et mesurés – ces paramètres, incluant des temps de réverbération tels que RT20, EDT ou encore des indices de clarté comme C50 ou C80 peuvent en effet être extraits de mesures réelles de réponses impulsionnelles dans un espace, ou calculés grâce au logiciel de modélisation cité ci-dessus [11]) et une validation perceptive de cette procédure (basée sur un test d'écoute de comparaison entre auralisations simulées et mesurées des attributs subjectifs suivants : Réverbérance, Clarté, Distance, Balance Tonale, Coloration, Plausibilité, Largeur apparente de source (ASW), Enveloppement de l'auditeur). Par la suite, Postma et Katz [5] ont également montré que l'ajout de directivité dynamique de la voix dans les auralisations mène à une amélioration des jugements de la plausibilité de ces dernières. Enfin, précédemment à la présente étude, Postma et Katz (publication soumise) ont montré que l'ajout de visuels avait un impact sur l'évaluation acoustique de certains attributs : les sujets devaient évaluer la même scène sous deux conditions : 1) auralisations seulement et 2) auralisations avec visualisation du théâtre de l'Athénée. Sur les 24 participants au total, les participants ont été classés dans trois groupes différents : 1) les participants qui jugeaient la scène acoustiquement plus éloignée avec l'augmentation de la distance visuelle, 2) les participants qui jugeaient la scène acoustiquement plus forte avec une augmentation de la distance visuelle

source-récepteur, et 3) ceux dont le jugement acoustique n'était pas influencé par l'ajout de visuels. De plus, les résultats ont également montré des différences significatives de jugement sur la largeur apparente de la source entre les conditions [seulement audio] et [audio-visuel].

Un aspect lié à cet objectif d'amélioration est donc l'utilisation des auralisations dans des environnements (audio-visuels) virtuels. Pour des applications architecturales par exemple, lors de visites virtuelles, l'information auditive aide à donner du sens à l'information visuelle et à l'impression générale [16]. Il est maintenant connu que les informations perçues par une modalité sensorielle peuvent affecter la perception et le traitement d'une autre modalité sensorielle [8].

Cependant, peu d'études se sont intéressées à l'effet des visuels sur l'évaluation subjective acoustique d'auralisations. Larsson et al [13] ont réalisé des tests d'écoute sous les conditions suivantes : 1) Auralisations seules, 2) Auralisations accompagnées de photos, 3) Auralisations accompagnées d'un modèle visuel virtuel sous HMD, et 4) Auralisations en situations réelle (dans la salle). Les résultats ont montré que les conditions réelles et virtuelles (3 et 4) étaient jugées significativement plus larges (ASW) que dans les conditions 1 et 2 (audio seul et photos). De plus, la condition « situation réelle » a conduit à des jugements de taille de la salle et de distance significativement plus petites et plus courtes.

La perception de certains attributs acoustiques comme la localisation ou la distance ont été en revanche largement étudiés. Une bonne synthèse, incluant le rôle de la vision est faite par Calcagno et al [15]. Au-delà de rappeler les indices essentiels liés à la perception de la distance, bien connus en psycho-acoustique (Intensité, Ratio son direct/son réverbéré, contenu spectral -type de source-), ils rappellent également l'effet de « proximité de l'image » désigné par [1] : la distance acoustique est perçue plus faible avec la vision d'une source visuelle (enceinte) proche que sans (dans l'obscurité). Zahorik [14] a réitéré ces expériences dans un environnement réverbérant et montre que la vision d'une source améliore la précision du jugement acoustique de la distance, et réduit la variabilité des résultats. Aucune étude n'a cependant étudié l'influence du support de RV sur l'évaluation de la composante auditive.

L'objectif de cette étude est donc d'étudier l'influence des visuels sur l'évaluation acoustique subjective des auralisations, incluant la plausibilité de la scène, la distance, l'intensité, ou des paramètres davantage liés à l'espace comme la largeur apparente de source, la réverbérance ou encore l'enveloppement. La section Framework décrit les différents éléments de ce dernier pour réaliser l'environnement virtuel ainsi que le protocole expérimental. La section Résultats présente les résultats des deux expériences, discutés dans la partie Discussion.

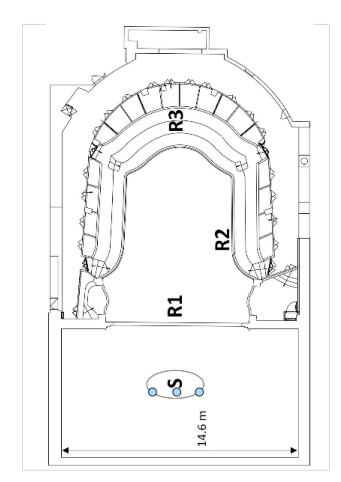


Figure 1: Plan du Théâtre représentant la zone des positions de la source (acteurs sur scène) et des positions d'écoute utilisées dans le théâtre. Les 3 positions de calcul des réponses acoustique pour les acteurs sont indiquées dans la zone S.



Figure 2: Photo du rendu visuel de la CAVE-portable du théâtre et les acteurs.

2 FRAMEWORK

L'architecture globale du système, incluant des échanges entre BlenderVR [6], MaxMSP¹, et un serveur VRPN (virtual reality peripheral network²) envoyant les informations de tracking est décrite ci-dessous. Une description détaillée est disponible dans Poirier-Quinot et al. [7].

2.1 Audio

Le modèle acoustique géométrique (GA) du Théâtre de l'Athénée a été créé et calibré grâce au logiciel CATT-Acoustics. La procédure de calibration [2] consiste à minimiser l'erreur moyenne et la variance sur toutes les paires sources-récepteurs par rapport aux mesures réelles, pour les paramètres objectifs temps de décroissance précoce (EDT), temps de réverbération (T20) et des indices de clarté (C50 & C80), tous définis en ISO 3382 [11]. De plus, cette calibration a ensuite été validée perceptivement lors de tests d'écoute évaluant des auralisations réalisées grâce à ce modèle.

Les auralisations sont générées de la manière suivante : des RIRs (réponses impulsionnelles de la salle) Ambisoniques du 2ème ordre sont générées grâce au modèle (utilisant la méthode de tracé de rayons), puis convoluées avec un signal mono et « anéchoïque » d'enregistrements d'un extrait d'une pièce de théâtre comportant 2 acteurs (« Ubu Roi » d'Alfred Jarry, réalisé au Théâtre de la Reine Blanche à Paris, en utilisant 2 micros cravates sans fil). Au total, 3 positions différentes d'auralisations ont ainsi été générées pour les tests (Fig. 1 : positions **R1**, **R2**, **R3**). De plus, la directivité des voix des acteurs a été incluse (voir [5] pour les détails), ainsi qu'un post-traitement permettant un tracking de la position des acteurs. En effet la scène, composée de deux acteurs, restait dans la zone ellipsoïdale **S** délimitée sur la Fig. 1. L'homme restait assis sur une chaise au centre de la zone **S**. Le déplacement ne concernait que la femme qui allait de gauche à droite autour de l'homme, et ne se déplaçait en profondeur que de manière négligeable. Le tracking audio a été réalisé en interpolant selon 3 réponses impulsionnelles situées sur une ligne (représentées par les points bleus sur la Fig. 1, donc pas de tracking audio en profondeur), selon la position de l'actrice. En conséquence, il était imposé que la distance restait constante.

Un système de head-tracking alimentant un rendu binaural dynamique utilisant l'IRCAM Spat³ en choisissant l'HRTF du KEMAR (filtres acoustique spatiaux d'une tête artificielle appliqués aux flux audio pour reproduire les filtrages effectués par le buste, la tête, et la forme du pavillon) génère un rendu audio 3D au casque et permet également de suivre l'orientation du sujet lors des expériences.

2.2 Visuel

Le modèle visuel du théâtre a été porté et modifié dans BlenderVR, puis rendu respectivement sur une CAVE-portable (Fig. 2), ou sous HMD (Oculus DK2). Le rendu est adaptatif, en prenant en compte le tracking de la position et l'orientation de la tête du sujet lors des expériences.

Pour animer la scène visuelle, un enregistrement vidéo réel de la pièce « Ubu Roi » a également été réalisé à l'aide d'un capteur Kinect2, permettant d'avoir des vidéos RGB et Carte de Profondeur. Ces deux vidéos combinées ont permis de générer un nuage de points (512×424 pixels) permettant d'avoir une représentation visuelle de la scène assez réaliste, malgré la faible qualité graphique du rendu. Le nuage de points a été rendu grâce à une texture GLSL⁴, rendant les pixels de l'image RGB en 3 dimensions dans la scène virtuelle comme des points distincts. Ces points reflètent bien la lumière, mais ne produisent en revanche pas d'ombre. La vidéo « Carte de Profondeur » a été utilisée pour définir la position spatiale de chaque pixel dans la scène, et la vidéo RGB pour définir le couleur. De plus, les éléments alentours de l'estrade permettaient de bien localiser les acteurs sur la scène.

La CAVE-portable est constituée d'un projecteur unique (sans stéréoscopie, résolution 1280×1024) et de 3 écrans de 2,28 m de côté, disposés en forme de « U », le sujet étant placé à l'entrée au milieu de ce « U », lui donnant un angle de vue de 180°. Ainsi, malgré l'absence de stéréoscopie, les sujets percevaient une sensation d'immersion et de profondeur du théâtre grâce à la taille de l'image, les indices géométriques et les indices dynamiques. De plus, pour éviter des mouvements contre-intuitifs liés à l'absence de stéréoscopie, les objets situés entre l'écran et le sujet (e.g. certains sièges) étaient « clippés ». On peut noter que la position **R1** était à environ 4 m des acteurs dans le monde virtuel.

Le HMD est lui relié à un ordinateur qui génère un flux vidéo constitué de paires d'images animées (destinées à chaque oeil) à un taux de rafraîchissement de 60 Hz, et donnant un angle de vue d'environ 100°. La technologie de l'écran est AMOLED⁵, et sa résolution est de 960×1080 pixels par oeil (« moitié » de « full HD »). Les images sont générées de manière à simuler les points de vue respectifs de l'oeil gauche et droit, permettant ainsi à l'utilisateur de voir en relief.



Figure 3: Interface physique.

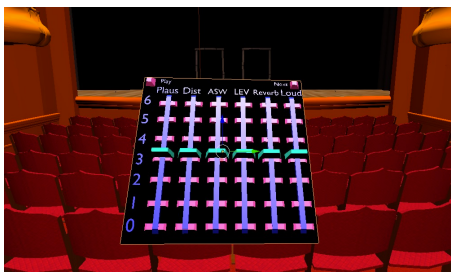


Figure 4: Avatar virtuel de l'interface (vu sous HMD).

2.3 Protocole expérimental

Cette étude comprend deux expériences espacées dans le temps de quelques semaines, et différant uniquement par le support visuel de RV utilisé : CAVE-portable ou HMD.

Ainsi, au total, les 14 mêmes sujets ont participé (ratio homme/femme = 10/4 ; moyenne d'âge = 31.9, écart-type : 6.4). L'évaluation de ces critères perceptifs nécessite une écoute « avertie ». Ainsi, la condition pour pouvoir participer était de posséder une expérience significative en musique ou en acoustique des salles, et de ne présenter aucune perte auditive. Le sujet était successivement virtuellement placé à trois différentes positions visuelles dans le théâtre et écoutait les trois différentes auralisations correspondant à ces différentes positions (Fig. 1 : **R1**, **R2**, **R3**). Au total, le sujet passait donc les 9 combinaisons possibles (3 positions visuelles × 3 auralisations, ce qui donnait 3 conditions cohérentes et 6 conditions incohérentes), répétées 3 fois (27 évaluations au total). Une première phase de familiarisation permettant de vérifier que le sujet comprenait la tâche était donc suivie des 27 configurations de l'expérience.



Figure 5: Sujet sous HMD.

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

z représente la réponse normalisée, en fonction de x , la réponse, μ , la moyenne de l'ensemble des notes des participants pour l'attribut considéré et σ , l'écart-type correspondant.

A chaque fois, le sujet évaluait la scène sur les attributs perceptifs *acoustique* suivants, sur une échelle de 0 à 6 :

- Plausibilité : la scène vous paraît-elle réaliste au regard de votre position dans le théâtre ?
- Distance : la distance acoustique des acteurs depuis votre position.
- Intensité : l'intensité perçue du son.
- Largeur apparente de la source (ASW) : l'étendue de l'image acoustique sur un plan horizontal. La source peut être étroite ou au contraire perçue très large.
- Enveloppement (LEV) : la sensation d'être entouré(e) par le son.
- Reverberance (Rev) : la chute du son après son extinction.

L'interface utilisée pour l'évaluation dans les deux expériences comporte des sliders (Behringer BCF2000, voir Fig. 3). Pour garder une interface *unique* dans les deux conditions une représentation virtuelle, animée par les positions des sliders réels, a été réalisée, et affichée dans le théâtre virtuel (voir Fig. 4). En plus du retour visuel sur la position actuelle des sliders, des repères tactiles entre chaque slider sur l'interface aidaient le sujet à se repérer sur l'interface. Sous HMD, les mains de l'utilisateur n'étaient en effet pas représentées virtuellement, ce qui pouvait nécessiter quelques secondes de prise en main (Fig. 5).

3 RÉSULTATS

3.1 Normalisation des résultats et statistiques appliquées

Les réponses allant de 0 à 6 ont été normalisées, de manière à pouvoir comparer les résultats entre participants. Les caractéristiques des sujets étant connus, une normalisation standard a été appliquée (Eq. 1). Une rm-ANOVA (*repeated measures*, via MATLAB⁶) suivie d'un *t*-test de comparaison multiple a été indépendamment appliquée pour chaque attribut, dont les résultats sont listés pour chaque attribut sur la Table 1.

3.2 Configurations audio-visuelles cohérentes

La 1ère étape consiste à comparer les résultats obtenus avec les deux modes de rendu, uniquement pour les configurations audio-visuelle cohérentes. On observe sur la Fig. 6 (respectivement gauche et droite : CAVE-portable et HMD) des tendances similaires. En effet, dans les deux cas :

- Aucune tendance n'est observée en ce qui concerne la plausibilité.
- La configuration audio-visuelle AV1 (juste devant la scène) est perçue plus proche, que AV2 et AV3 (position latérale et au 1er balcon).
- L'intensité est perçue plus forte devant la scène en configuration AV1 qu'en AV2 et AV3. De plus, AV3 est perçue significativement plus forte en HMD qu'en CAVE-p ($F = 6.54, p = 0.012$).
- La source est perçue plus étroite en configurations AV2 et AV3 qu'en AV1 devant la scène.

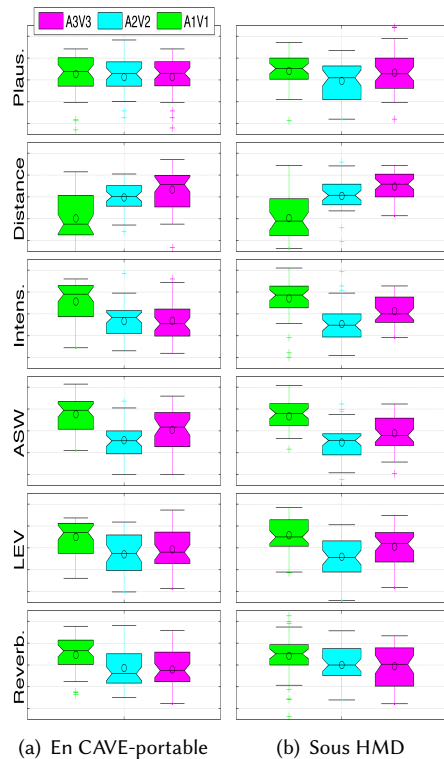


Figure 6: Configurations audio-visuelle cohérentes

Description du diagramme à boîte : la boîte contient 50% des réponses (intervalle de confiance), les limites de la boîte représentent les 1ère et 3ème quartiles (25 et 75%), (+) les données aberrantes, (—) la médiane et (o) la valeur moyenne.

- De plus, la configuration AV1 est également perçue plus enveloppante et plus réverbérante qu'en AV2 et AV3.

3.3 Configurations audio-visuelle incohérentes

La 2ème étape consiste à analyser l'ensemble des configurations, comprenant les configurations audio-visuelle incohérentes (par exemple une position visuelle V1 avec l'auralisation A3) pour étudier l'influence d'une modalité sur une autre. D'autres études ont évalué l'influence d'une modalité sur une autre de cette manière, comme par exemple ces recherches montrant l'influence des visuels sur la perception auditive et vocale [9, 10]. La Fig. 7 montre ainsi les résultats combinés pour toutes les positions visuelles (Fig. 7(a)) et toutes les auralisations (Fig. 7(b)).

3.3.1 Plausibilité (1ère ligne de Fig. 7) La plausibilité étant notée par rapport à la position visuelle dans le théâtre, il convient d'en étudier l'effet par position visuelle. Dans les deux expériences, on observe pour la position visuelle V1 que l'auralisation R1 est jugée plus plausible que les auralisations A2 et A3 ($F = 6.04/p = 0.0032$ et $F = 21.18/p < 1e^{-2}$ resp. pour CAVE-p et HMD). Pour les positions V2 et V3, aucune différence significative n'a été relevée. De plus, aucune différence significative de jugement de plausibilité n'apparaît entre les deux supports visuels.

3.3.2 Effet de la modalité auditive (résultats combinés sur toutes les positions visuelles, (Colonne droite des Fig. 7 a) et b)))

- Distance : Dans les deux cas (CAVE-p et HMD), l'auralisation A1 est perçue plus proche que les auralisations A2 et A3, quelque soit la position visuelle.
- Intensité : Dans les deux cas, A1 est jugée plus forte que A2 et A3.
- ASW : Dans les deux cas, l'auralisation A1 est perçue plus large que les auralisations A2 et A3.
- LEV : Dans les deux cas, l'auralisation A1 est jugée plus enveloppante que A2 et A3.
- Rev : En CAVE-p, A1 est jugée plus réverbérante que A2 et A3, alors qu'avec HMD, A2 est jugée moins réverbérante que A1 et A3.

3.3.3 Effet de la modalité visuelle (résultats combinés sur toutes les auralisations, (Colonne gauche des Fig. 7a et Fig. 7b))

- Distance : Pour les deux supports CAVE-p et HMD, l'ensemble des auralisations sont jugées plus proches à la position visuelle V1 qu'aux positions V2 et V3.
- Intensité : Aucun effet significatif du visuel.
- ASW : Aucun effet significatif, jugée identique en V1, V2 et V3.
- LEV : Dans les deux cas, les auralisations sont perçues plus enveloppantes à la position visuelle V3 qu'en V2 et V3.
- Rev : Aucun effet significatif du visuel.

Table 1: rm-ANOVA F/p -value pour les Fig. 7a et Fig. 7b. $\epsilon = p < 0.01$.

Attr	CAVE-p			
	Vis		Aur	
	F	p	F	p
Dist	13.4	ϵ	20.8	ϵ
Inten	0.39	0.7	52.5	ϵ
ASW	0.32	0.7	73.9	ϵ
LEV	4.1	0.02	56	ϵ
Rev	1.6	0.2	14	ϵ

Attr	HMD			
	Vis		Aur	
	F	p	F	p
Dist	22	ϵ	21.5	ϵ
Inten	0.07	0.9	38	ϵ
ASW	2.0	0.13	48	ϵ
LEV	2.2	0.11	51	ϵ
Rev	4.	0.02	12	ϵ

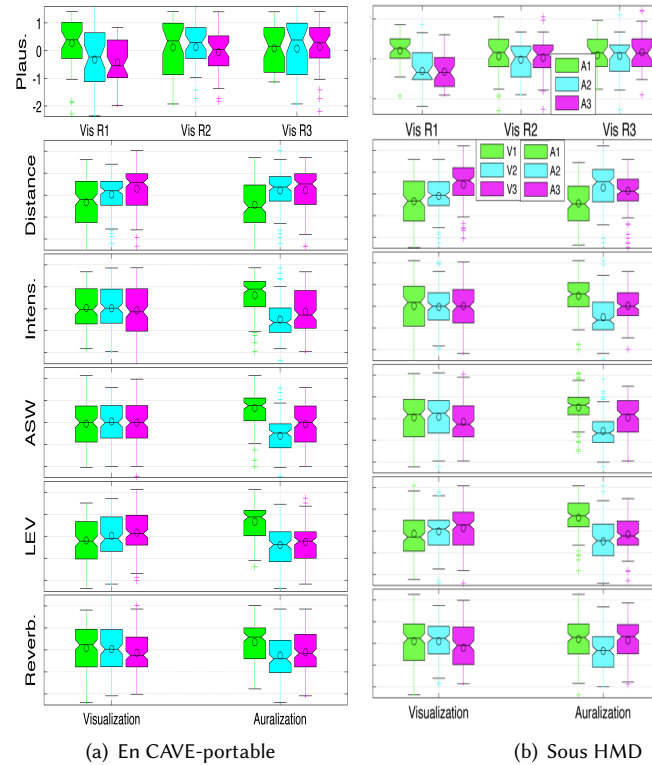


Figure 7: Résultats combinés. Colonne de gauche : visualisations R1, R2 et R3 (combinant toutes les auralisations); Colonne de droite : auralisations R1, R2 et R3 (combinant toutes les positions visuelles).

4 DISCUSSION

La plausibilité de la scène est jugée globalement identique sous HMD et en CAVE-p. Toutefois, des commentaires informels post-expérience des participants notaient une meilleure « impression d'immersion » sous HMD, qu'ils trouvaient liée à la plausibilité de la scène.

La perception de la distance acoustique augmente en augmentant la distance visuelle, quelle que soit l'auralisation écoutée. Ceci est en accord avec [13, 15]. Cependant, un effet semblable sur l'intensité apparaissait dans ces études, non répété ici. Ces résultats s'ajoutent aux effets bénéfiques sur la plausibilité de l'ajout de la directivité dynamique dans les auralisations [3].

5 CONCLUSION

Cette étude a permis d'étudier l'influence des visuels et du support de rendu visuel (CAVE-p et HMD) sur l'évaluation subjective acoustique d'auralisations simulées d'une scène de théâtre virtuelle incluant deux acteurs.

Une interface d'évaluation « universelle » a été conçue et utilisée dans ces 2 expériences (cette dernière avec son avatar virtuel lors de l'expérience sous HMD). Deux principales conclusions peuvent être tirées de cette étude :

- Le jugement acoustique de la distance augmente avec la distance visuelle (en accord avec les résultats précédant sur le CAVE-p seulement [4]).
- Sur les critères perceptifs considérés, les résultats montrent qu'aucune différence significative n'a été relevée entre les deux supports visuels utilisés (CAVE-portable ou HMD).

La validation perceptive des auralisations et l'indépendance du support de rendu visuel (une autre expérience ajoutant un 3ème mode de rendu - grande CAVE haute-définition est en cours) permettra d'étendre ce type d'expérimentation à d'autres types de stimuli, dans un objectif d'amélioration de la perception des environnements (audio-visuels) virtuels.

Remerciements

Nous remercions Francis Cousson et Dominique Lemaire, directeur technique du Théâtre de l'Athénée, de nous avoir fourni le modèle graphique. Nous remercions aussi tous les sujets qui ont passé notre expérience.

Références

- [1] Gardner M B. 1968. Proximity image effect in sound localization. *J. Acoust. Soc. Am.* 43 (1968), 163. <https://doi.org/10.1121/1.1910747>
- [2] Postma B and Katz B. 2015. Creation and calibration method of virtual acoustic models for historic auralizations. *J. Virtual Reality* 19 (2015), 161–180. <https://doi.org/10.1007/s10055-015-0275-3>
- [3] Postma B and Katz B. 2016. Dynamic voice directivity in room acoustic simulation auralizations. In *German Annual Conf. on Acoustics (DAGA)*. 352–355.
- [4] Postma B and Katz B. 2016. Perceptive and objective evaluation of calibrated room acoustic simulation auralizations. *J. Acoust. Soc. Am.* 140 (2016), 4326–4337. <https://doi.org/10.1121/14971722>
- [5] Postma B, Demontis H, and Katz B. 2017. Subjective evaluation of dynamic voice directivity for auralizations. *Acta Acustica united with Acustica* 103 (2017), 181–184. <https://doi.org/10.3813/AAA.919045>

Notes

¹MaxMSP 6.1, <https://cycling74.com/products/max/>

²<https://github.com/vrpn/vrpn/wiki>

³<http://forumnet.ircam.fr/fr/produit/spat/>

⁴https://en.wikibooks.org/wiki/GLSL_Programming/Blender

⁵<https://en.wikipedia.org/wiki/AMOLED>

⁶<https://www.mathworks.com/products/matlab.html>

- [6] Katz BFG, Felinto D, Touraine D, Poirier-Quinot D, and Bourdot P. 2015. BlenderVR : Open-source framework for interactive and immersive VR. In *IEEE Virtual Reality, VR*. 203–204. <https://doi.org/10.1109/VR.2015.7223366>
- [7] Poirier-Quinot D, Postma B, and Katz B. 2016. Augmented Auralizations : Complimenting auralizations with immersive virtual reality technologies. In *Intl. Sym. on Music and Room Acoustics (ISMRA)*. ISMRA, 1–10.
- [8] Stein B E, Meredith M A, Honeycutt W S, and McDade L. 1989. Behavioral indices of multisensory integration : orientation of visual cues is affected by auditory stimuli. *J. of Cognitive Neuroscience* 1 (1989), 12–24.
- [9] McGurk H and MacDonald J. 1976. Hearing leaps and seeing voices. *Nature* 264 (1976), 746–748.
- [10] Howard I and Templeton W. 1966. *Human spatial orientation*. London : Wiley.
- [11] ISO 3382-1 :2009. 2009. *Acoustics - Measurement of room acoustic parameters -I : Performance spaces*. International Standards Organization, ISO Standard.
- [12] Kleiner M, Dalenbäck B-I, and Svensson P. 1993. Auralization - An Overview. *J. Audio Eng. Soc.* 41, 11 (1993), 861–875.
- [13] Larssön P, Västfjäll, and Kleiner M. 2001. Ecological acoustics and the multi-modal perception of rooms : real and unreal experiences of auditory-visual virtual environments. In *Proc. Intl. Conf. on Auditory Display (ICAD)*, Vol. 1. 245–249.
- [14] Zahorik P. 2002. Assessing auditory distance perception using virtual acoustics. *J. Acoust. Soc. Am* 111 (2002), 1832–1846. <https://doi.org/10.3813/AAA.919045>
- [15] Calcagno E R, Abregú E L, Eguia M C, and Vergara R. 2012. The role of vision in auditory distance perception. *Perception* 41, 17 (2012), 175–192. <https://doi.org/10.1069/p7153>
- [16] M Vorländer. 2014. Virtual Acoustics. *Archives of Acoustics* 39-3 (2014), 307–318. <https://doi.org/10.2478/aoa-2014-0036>