

10ème Congrès Français d'Acoustique

Lyon, 12-16 Avril 2010

Etudes de l'intégration visuo-auditive dans un contexte de réalité virtuelle

Khoa-Van Nguyen¹, Clara Suied², Isabelle Viaud-Delmon¹, Olivier Warusfel¹

¹ Ircam - CNRS UMR 9912, Paris, France

² Laboratoire Psychologie de la Perception,

CNRS & Université Paris Descartes & Département d'Etudes Cognitives,

Ecole Normale Supérieure, Paris, France

La notion de fenêtre spatiale d'intégration visuo-auditive désigne les conditions suivant lesquelles un stimulus auditif et un stimulus visuel sont perçus sous forme d'un percept unique bien que leurs positions spatiales diffèrent objectivement. Cette question a fait l'objet de nombreuses recherches en neurosciences. Cependant, la plupart d'entre elles reposent sur des conditions expérimentales de laboratoire extrêmement simplifiées et contrôlées. L'utilisation de la réalité virtuelle permet d'enrichir le contexte expérimental et déterminer si les observations relevées sont maintenues dans des environnements plus complexes.

Les expériences présentées dans ce papier sont menées dans un environnement de réalité virtuelle utilisant un écran à stéréoscopie passive et un rendu binaural individualisé. Ces expériences sont dérivées du paradigme du point d'alignement spatial subjectif (PSSA) qui consiste à émettre un flux auditif selon une direction fixe tandis qu'un stimulus visuel synchrone se déplace latéralement. La tâche du participant est d'indiquer l'instant dès lequel il perçoit les deux stimuli alignés et qui définit le PSSA.

Nous avons repris ce paradigme avec des objets complexes afin d'étudier l'effet d'attraction visuo-auditif lorsque les stimuli sont plus complexes, physiquement et sémantiquement. Nous avons montré que l'effet principal d'attraction du son par une image est toujours présent, mais son amplitude est plus faible. Nos premiers résultats soulignent l'importance de considérer des objets multisensoriels complexes, ici pour l'étude de la localisation spatiale.

1 Introduction

Notre perception du monde est multisensorielle : c'est le résultat de l'intégration par le cerveau de différentes informations sensorielles. Cette intégration multisensorielle modifie particulièrement la perception unimodale. Par exemple la vision influe sur l'audition spatiale, alors que l'audition influe sur la vision temporelle. L'intégration multisensorielle s'établit sur des facteurs aussi bien structurels (alignement spatial ou temporel entre un son et une image) que cognitifs (cohérence sémantique entre le son et l'image) [3] [13].

Les précédentes études de la littérature portant sur l'intégration visuo-auditive ont généralement été menées avec des stimuli simples (flash, diode, clic ou bruit) [12] [2] [1]. Seules quelques unes ont abordées l'intégration multisensorielle avec des stimuli complexes [11] [6] [14]. Dans ces conditions, les expériences n'avaient pas la richesse des objets naturels et les aspects cognitifs de cohérence sémantique entre les informations auditives et visuelles n'étaient pas considérés.

De son côté, la réalité virtuelle (RV) permet à la fois d'utiliser des stimuli plus riches et de les contrôler de manière plus flexible. Elle permet d'enrichir les expériences d'intégration visuo-auditive en terme de complexité et de réalisme. Il est en même temps

important de savoir comment les deux canaux de rendu doivent interagir et l'effet produit sur la perception. Pour cela, il est nécessaire de tester ces effets dans des conditions proches de l'utilisation habituelle de la RV, i.e. avec des stimuli complexes.

Nous proposons d'étudier les phénomènes d'intégration visuo-auditive dans un contexte de réalité virtuelle. Pour cela, nous avons choisi de nous intéresser au paradigme du point d'alignement spatial subjectif (PSSA) entre un stimulus visuel mobile et un stimulus auditif statique. Ce paradigme permet d'étudier l'effet ventriloque, effet selon lequel deux flux auditifs et visuels spatialement discordants peuvent résulter en un percept unique. Dans un contexte de RV, il est intéressant d'étudier le paradigme du PSSA puisque le stimulus visuel est en mouvement. Le paradigme est plus proche d'un scénario classique de RV que d'autres paradigmes étudiant l'intégration visuo-auditive avec des stimuli statiques [4].

Dans une précédente expérience [10], nous avons simplement dérivé une expérience de Lewald [7] sur le PSSA dans un environnement de réalité virtuelle, en utilisant des stimuli simples. Nous avons reproduit les résultats obtenus [7], en validant ainsi notre setup de réalité virtuelle. Nous avons également montré que la direction de mouvement du stimulus visuel

(du centre vers la périphérie ou vice versa) avait un effet significatif sur la détermination du PSSA et ce indépendamment du stimulus auditif utilisé (bruit ou sinus), de l'hémi-espace (droit ou gauche) ainsi que du déphasage entre les flux auditifs et visuels. La direction du mouvement du stimulus visuel entraînait un alignement spatial subjectif anticipé par rapport à la coïncidence objective entre le stimulus visuel et auditif, révélant ainsi un effet ventriloque. Celui-ci semblait plus prononcé lorsque le stimulus était situé en périphérie du stimulus auditif. Par ailleurs, nous avons constaté une meilleure précision des réponses lorsque le stimulus était un bruit blanc, soulignant ainsi une influence des caractéristiques sonores pour cette tâche.

A partir de ces résultats, nous avons voulu tirer profit des avantages de la réalité virtuelle pour enrichir et complexifier les scénarii de nos expériences sur le PSSA et observer d'éventuelles modifications. Le premier objectif de cette étude est de déterminer si l'utilisation de stimuli complexes pouvait changer les résultats obtenus avec des stimuli simples. Dans la première expérience, nous réévaluons le PSSA avec des stimuli complexes et porteurs d'une signification sémantique. La première hypothèse est que l'effet d'attraction augmente puisque les stimuli sont complexes : leur cohérence sémantique s'ajoute à leur cohérence structurelle (spatiale en l'occurrence). Au contraire, la deuxième hypothèse est que l'effet d'attraction diminue car le stimulus sonore complexe utilisé est plus facilement localisable et donc moins sujet à l'effet ventriloque. Puis dans la seconde expérience, nous évaluons un PSSA "inversé" entre des stimuli auditifs mobiles et des stimuli visuels statiques. Nous voulons savoir si, pour le PSSA, les modalités sont interchangeables, i.e un son mobile peut-il attirer une image statique.

2 Point d'alignement spatial subjectif avec des stimuli complexes.

2.1 Matériels et Méthodes

2.1.1 Participants

Dix sujets (4 femmes et 6 hommes âgés de 19 à 55 ans) ont participé aux 2 expériences. Ils étaient tous droitiers, n'avaient aucun problèmes auditifs et avaient une vue normale ou corrigée. Tous les sujets avaient leur HRTFs individuelles, mesurées auparavant en chambre anéchoïque. Deux des sujets connaissaient les buts de l'expérience.

2.1.2 Environnement matériel et logiciel

L'expérience s'est déroulée dans un studio de l'IRCAM, insonorisé et dans l'obscurité, à l'aide de notre environnement de réalité virtuelle. La machine de rendu disposait d'un processeur AMD Dual Core Opteron 275 (2.2 GHz) et de deux cartes graphiques GeForce 7950 GX2 M500.

La scène visuelle était présentée sur un écran de $300 \times 225 \text{ cm}^2$ à stéréoscopie passive. L'image stéréosco-

pique était projetée par deux projecteurs arrière *F20 SX+ ProjectionDesign* équipés de lentilles grand angle et de filtres à polarisation circulaire. Les participants devaient porter des lunettes 3D et maintenir leur tête fixe à l'aide d'une mentonnière. Sur le plan visuel, les scènes virtuelles tridimensionnelles ont été réalisées grâce à *OgreVR*, une interface de programmation (API) développée par l'*INRIA* permettant de créer des scènes de réalité virtuelle. *OgreVR* est la combinaison du moteur de rendu 3D *Ogre* [9] et de la librairie *VRPN* [5]. Cette dernière est une interface transparente entre une application et la plupart des outils utilisés dans un environnement virtuel tels que les systèmes de suivi de position, joysticks etc. Ces deux librairies C++ sont open source.

Pour le rendu audio tridimensionnel, les stimuli binauraux (échantillonnés à 44.1 kHz) étaient préenregistrés avec MaxMsp, avec des HRTFs individualisées. En effet, puisque les sujets avaient la tête fixe, le suivi dynamique n'était pas nécessaire. Les stimuli binauraux étaient donc envoyés directement de l'application à la sortie stéréo d'une carte son *RME Fireface* et diffusés à travers un casque d'écoute *Sennheiser HD650*.

Le sujet était assis à 1,5 m de l'écran décrivant ainsi d'un champ de vision de $\pm 45^\circ$. Afin d'accomplir la tâche d'alignement spatial, il devait indiquer le point d'alignement spatial subjectif en appuyant sur la barre espace du clavier.

2.2 Expérience A : PSSA entre un stimulus visuel mobile et un stimulus auditif statique

2.2.1 Stimuli

Le stimulus visuel était un hélicoptère vert virtuel de 50 cm de long projeté de manière stéréoscopique au niveau des yeux. Il se déplaçait sur un segment circulaire de 100 cm de rayon autour du sujet. Une sphère verte de rayon 5 cm située au même niveau que l'hélicoptère servait à indiquer le point de départ de celui-ci. L'arrière plan était uniformément bleu.

Le stimulus sonore était un son de rotor avec 10 ms de temps de montée et de descente, d'une durée totale de 12 s. Il était filtré par des HRTFs individuelles afin d'être spatialisé aux différents azimuts et l'élévation 0° . Les HRTFs étaient égalisées champ diffus et il n'y a pas eu d'égalisation individuelle du casque. Le niveau sonore du rotor était de 70 dB SPL.

2.2.2 Procédure

Chaque séquence débutait par l'apparition de la sphère verte pour indiquer la position de départ de l'hélicoptère. Après 1000 ms, l'hélicoptère et le son de rotor associé apparaissaient. L'hélicoptère évoluait horizontalement à la vitesse de $2.5^\circ/\text{sec}$ pendant 12 sec. Pendant cette durée, le sujet devait fixer l'axe du rotor et appuyer sur la touche espace du clavier dès qu'il percevait l'axe du rotor aligné spatialement avec le son du rotor. Une fois la touche appuyée, l'azimut

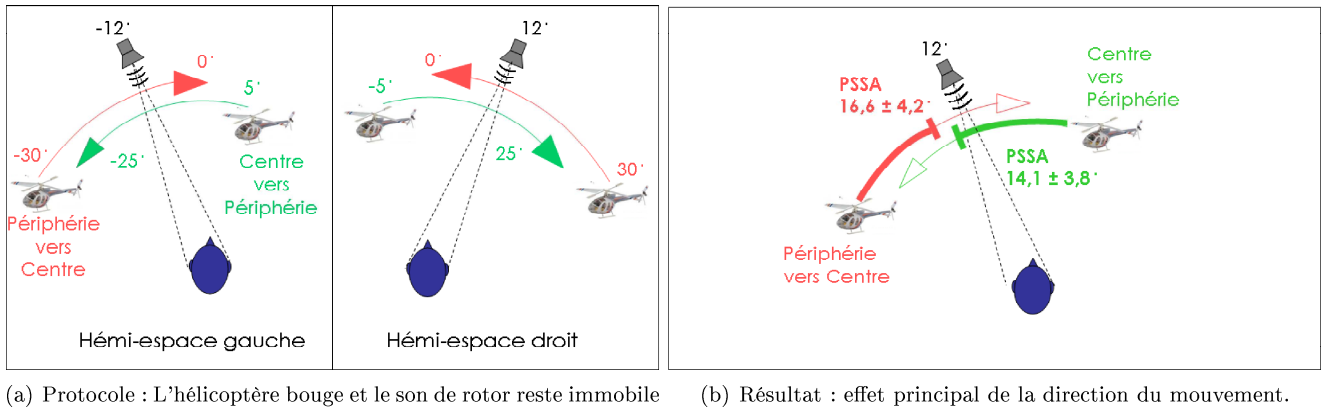


FIGURE 1 – PSSA avec des stimuli complexes

de l'hélicoptère était enregistré (le PSSA), l'hélicoptère devenait alors rouge pour signaler l'acquisition de la réponse tout en continuant son trajet jusqu'à la fin des 12 s. La séquence suivante commençait 1000 ms après la disparition de l'hélicoptère et du son de rotor.

Le PSSA a été étudié par rapport à 2 facteurs : l'hémi-espace de présentation du son de rotor statique et la direction du mouvement de l'hélicoptère croisant la position du son de rotor. Lorsque le stimulus sonore était dans l'hémi-espace droit (à l'azimut 12° par rapport au plan médian), le stimulus visuel parcourait un angle compris entre -5° sur la gauche du sujet et 25° sur sa droite ou un angle compris entre 30° et 0° . Les différentes conditions spatiales pour l'hémi-espace gauche sont construites symétriquement par rapport au plan médian.

Chacune des 4 conditions spatiales a été répétée 12 fois. Au total, 48 séquences ont été testées dans un ordre pseudo-aléatoire pendant 12 min après une session d'apprentissage de 3 min.

2.2.3 Analyses

Pour les deux expériences, les valeurs de PSSA sont d'abord symétrisées afin d'étudier les effets potentiels de l'hémi-espace de présentation des stimuli. L'analyse a été faite sur la moyenne des PSSA sur toutes les répétitions d'une condition donnée. Une ANOVA à mesures répétées est menée avec la moyenne des PSSA décrite ci-dessus comme variable dépendante. Les facteurs intra-sujets sont l'hémi-espace (droit ou gauche) et la direction du mouvement (de la périphérie vers le centre ou du centre vers la périphérie) et le facteur inter-sujets est l'ordre des expériences (A puis B ou B puis A).

2.2.4 Résultats

L'Anova a révélé un effet principal de la direction du mouvement ($F(1,8) = 5.5629$, $p < 0.05$) mais pas d'effet de l'hémi-espace ni de l'ordre de passage des expériences. Les moyennes de PSSA sont significativement différentes ($T = 2.4$, $p < 0.003$ pour le mouvement du centre vers la périphérie et $T = 4.8$, $p < 0.001$ pour l'inverse) de la position objective de la source sonore.

La figure 1(b) montre le PSSA moyenné sur l'hémi-espace et les répétitions en fonction de la direction du mouvement. Comme pour l'expérience que nous avons menée avec des stimuli simples, nous observons l'effet de la direction du mouvement sur la détermination du PSSA. Le stimulus visuel attire le stimulus auditif. L'effet d'attraction est dans les deux cas vers la périphérie.

2.3 Expérience B : PSSA entre un stimulus visuel statique et un stimulus auditif mobile

2.3.1 Stimuli

Nous avons utilisé les mêmes stimuli visuels et sonores dans cette expérience. Il s'agit donc toujours de l'hélicoptère et du son de rotor associé et seules les conditions spatiales ont changé.

L'hélicoptère se trouve cette fois-ci à 2 positions fixes. Le son de rotor est filtré par les HRTFs individuelles du sujet de manière à évoluer sur différentes trajectoires circulaires autour du sujet à l'élévation 0° .

2.3.2 Procédure

La procédure et les conditions spatiales sont identiques à la procédure 2.2.2. Il suffit de considérer que dans ce cas, c'est le son du rotor mobile qui croise la direction de l'hélicoptère statique (cf fig 2(a)).

2.3.3 Résultats

L'anova a révélé pour cette expérience un effet marginal de la direction du mouvement ($F(1,8) = 4.25$, $p < 0.07$), mais pas d'effets de l'hémi-espace ni de l'ordre dans lequel les expériences ont été menées. Seule la moyenne de PSSA est lorsque le son se déplace du centre vers la périphérie est significativement différente ($T = 5.46$, $p < 0.001$) de la position objective de la source sonore.

De même, la figure 2(b) montre le PSSA moyenné sur l'hémi-espace et les répétitions, en fonction de la direction du mouvement. Elle illustre l'effet de la direction du mouvement sur la détermination du PSSA. Cette fois-ci, au contraire de la figure 1(b), l'effet d'attraction semble plus fort vers le centre.

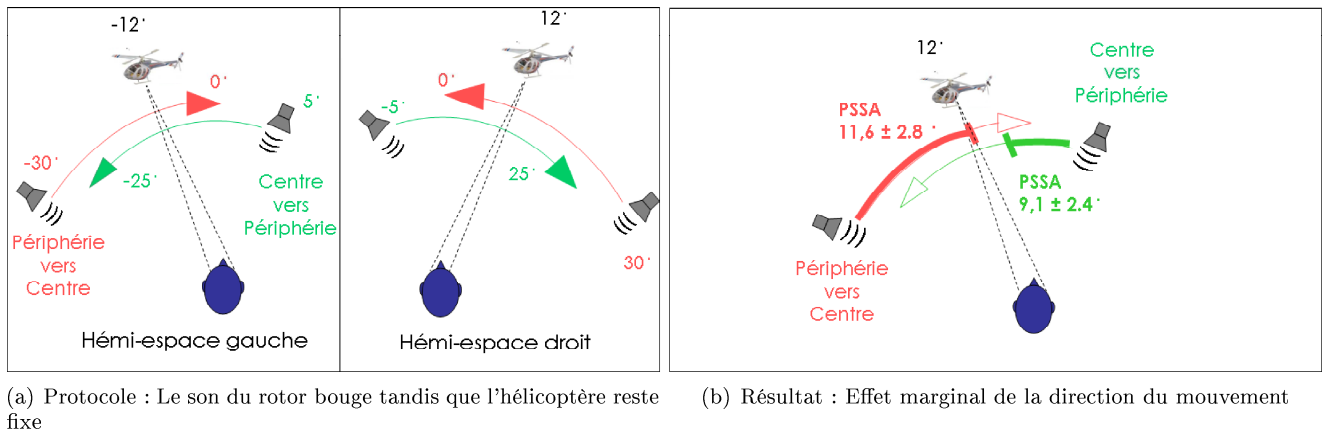


FIGURE 2 – PSSA avec des stimuli complexes et inversion de la modalité en mouvement.

2.4 Discussion

En utilisant des stimuli complexes en réalité virtuelle, nous avons pu confirmer et étendre des résultats obtenus dans la littérature avec des stimuli simples. Un stimulus visuel en mouvement attire la localisation d'un son statique. Cependant l'effet d'attraction semble plus faible que pour des stimuli simples (flash et bips). Une de nos hypothèses était que l'utilisation de stimuli complexes aurait plutôt favorisé la fusion des flux auditifs et visuels, et aurait renforcé l'effet d'attraction. Il est probable que les caractéristiques acoustiques et visuelles plus complexes des stimuli soient à l'origine de la réduction de l'effet d'attraction. En effet, le son de rotor comportant de nombreuses modulations, sa localisation était plus précise [8].

Lorsque un flux auditif est mobile, nous avons de nouveau observé un effet de la direction du mouvement sur le PSSA, légèrement moins fort que pour l'expérience inverse. L'effet d'attraction s'est réduit dans les mêmes proportions que pour le cas où le flux visuel était dynamique. En revanche, dans ce cas, l'effet d'attraction semble légèrement plus fort vers le centre.

3 Conclusion

L'utilisation de stimuli complexes en réalité virtuelle a permis d'affiner les résultats existants [10]. Il y a bien un effet d'attraction d'un son par un stimulus visuel en mouvement. Cependant, celui-ci est beaucoup plus faible que celui observé avec des stimuli simples. Ce résultat illustre l'importance d'étudier la localisation auditive d'une part dans un contexte multisensoriel (avec des sources visuelles) et d'autre part avec des stimuli complexes.

Remerciements

Ces recherches ont été financées par le projet européen Crossmod (FET Open) N° IST-014891-2, <http://www.crossmod.org>.

Références

- [1] P. Bertelson and M. Radeau. Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Percept Psychophys*, 29(6) :578–84, 1981.
- [2] M. O. Ernst and H. Bulthoff. Merging the sense into a robust percept. *Trends in Cognitive Sciences*, 8(4) :162–169, 2004.
- [3] P. Fuchs, G. Moreau, A. Berthoz, J.L. Vercher, D. Amarantini, F. Multon, G. Rao, and F. Aubert. Le traité de la réalité virtuelle (Volume I, L'homme et l'environnement virtuel). *Sciences mathématiques et informatique*.
- [4] W. D. Hairston, M. T. Wallace, J. W. Vaughan, B. E. Stein, J. L. Norris, and J. A. Schirillon. Visual localization ability influences cross-modal bias. *Journal of Cognitive Neuroscience*, 15(1) :20–29, 2003.
- [5] T.C. Hudson, A. Seeger, H. Weber, J. Juliano, and A.T. Helser. VRPN : a device-independent, network-transparent VR peripheral system. *Proceedings of the ACM symposium on Virtual reality software and technology*, pages 55–61, 2001. <http://www.cs.unc.edu/Research/vrpn/>.
- [6] C.E. Jack and W.R. Thurlow. Effects of degree of visual association and angle of displacement on the "ventriloquism" effect. *Perceptual and Motor Skills*, 37(3) :967–979, 1973.
- [7] J. Lewald, W. H. Ehrenstein, and R. Guski. Spatio-temporal constraints for auditory-visual integration. *Behavioral Brain Research*, 121 :69–79, 2001.
- [8] J.C. Middlebrooks and D.M. Green. Sound localization by human listeners. *Annual Review of Psychology*, 42(1) :135–159, 1991.
- [9] I. Milne and G. Rowe. OGRE-3D Program Visualization for C++'. *Proceedings of the 3rd Annual LTSN-ICS Conference*, 2002. <http://www.ogre3d.org/>.
- [10] K.V. Nguyen, M. Dellepiane, I. Viaud-Delmon, and O. Warusfel. Investigation of auditory-visual integration in VR environments. In *8th International Multisensory Research Forum*, 2007.
- [11] M. Radeau and P. Bertelson. Adaptation to auditory-visual discordance and ventriloquism in

semirealistic situations. *Perception & Psychophysics*, 22(2) :137–146, 1977.

- [12] G.H. Recanzone. Interactions of auditory and visual stimuli in space and time. *Hearing research*, 2009.
- [13] C. Spence. Audiovisual multisensory integration. *Acoustical Science and Technology*, 28(2) :61–70, 2007.
- [14] C. Suied, N. Bonneel, and I. Viaud-Delmon. Integration of auditory and visual information in the recognition of realistic objects. *Experimental Brain Research*, 194(1) :91–102, 2009.