CFA/VISHNO 2016

Effet ventriloque pour des sources sonores variant simultanément en azimut et élévation

E. Hendrickx[a], M. Paquier[b], V. Koehl[b] et J. Palacino[b]
[a]Conservatoire National Supérieur de Musique et de Danse de Paris, 209 avenue Jean Jaurès, 75019 Paris, France
[b]UBO, Lab-STICC UMR CNRS 6285, 6 avenue Victor Le Gorgeu, CS 93837, 29238 Brest Cedex 3, France
mathieu.paquier@univ-brest.fr

Lorsque l'on présente à un sujet des stimuli audio et visuel temporellement coïncidents mais spatialement disparates, les sujets perçoivent parfois le stimulus sonore au même endroit que le stimulus visuel. On appelle ce phénomène l'effet ventriloque car il rappelle l'illusion créée par le ventriloque lorsque sa voix semble plutôt provenir de sa marionnette que de sa propre bouche. Cet effet a été très largement étudié en azimut, mais beaucoup moins en élévation. Dans cette étude, des séquences montrant un homme en train de parler ont été présentées à des sujets. La voix de l'homme pouvait être reproduite sur différents haut-parleurs, qui créaient des disparités plus ou moins grandes en azimut et en élévation entre le son et l'image. Pour chaque présentation, les sujets devaient indiquer s'ils avaient perçu ou non la voix dans la même direction que la bouche de l'acteur. Les résultats ont montré que l'effet ventriloque fonctionnait à des angles plus larges en élévation qu'en azimut.

1 Introduction

L'effet ventriloque est souvent étudié à l'aide d'une tâche de discrimination, dans laquelle les sujets doivent indiquer si le stimulus visuel et le stimulus sonore « fusionnent » entre eux ou, dans une formulation légèrement différente, si ils perçoivent que les stimuli sonore et visuel proviennent ou non de la même direction. Les études résument en général les performances des sujets en indiquant le « seuil à 50% » (qui correspond à l'écart angulaire entre stimuli sonore et visuel associés pour lequel les sujets trouvent que les stimuli fusionnent - ou perçoivent que les stimuli proviennent du même endroit - une fois sur deux).

Les nombreuses études sur l'effet ventriloque en azimut ont permis de dégager plusieurs facteurs déterminant son efficacité (voir [1] pour une présentation détaillée de la littérature) :

- la disparité spatiale (l'effet décroît lorsque l'écart angulaire entre son et image augmente) ;
- la disparité temporelle (l'effet fonctionne mieux si le son et l'image sont synchrones) ;
- l'expérience du sujet (l'effet fonctionne mieux avec des sujets naïfs qu'avec des sujets experts) ;
- le « réalisme » de la combinaison son-image (l'effet fonctionne d'autant mieux que la combinaison son-image est réaliste et convaincante).

Il a également été montré que l'effet ventriloque dépendait de la précision spatiale du système auditif (l'effet fonctionne d'autant mieux que la précision de localisation est faible) [2]. Comme les performances de localisation sont moins bonnes dans le plan vertical que dans le plan horizontal [3], l'effet ventriloque devrait donc mieux fonctionner en élévation qu'en azimut.

Les rares études ayant exploré l'effet ventriloque en élévation sont contradictoires : d'un côté, l'étude de Thurlow et Jack [4] suggère que l'effet ventriloque fonctionne en effet mieux dans le plan vertical que dans le plan horizontal (cependant, Thurlow et Jack n'ont pas mesuré de « seuils à 50% » et leurs résultats ne peuvent donc pas être comparés avec la littérature) ; d'un autre côté, Werner *et al.* [5] ont bel et bien mesuré des seuils à 50% dans le plan médian, mais les valeurs obtenues sont semblables à celles obtenues en azimut par d'autres études (entre 8° et 10°). Werner *et al.* concluent que la magnitude de l'effet ventriloque est similaire en élévation et en azimut. Nous pensons que cette similarité est plutôt due à la spécificité des conditions expérimentales de Werner *et al.*, et qu'il est nécessaire de comparer des seuils mesurés en azimut et en élévation dans les mêmes conditions expérimentales pour pouvoir véritablement conclure.

Plusieurs études ont également suggéré que l'effet ventriloque fonctionnait mieux si le sujet prêtait moins attention à la position de la source sonore. Par exemple, lorsqu'une personne parle « dans la vraie vie », l'attention est plutôt focalisée sur le contenu sémantique, et l'individu n'accorde vraisemblablement que très peu d'importance à la position spatiale de la voix.

Le but de cette expérience est de comparer la force de l'effet ventriloque en azimut et en élévation dans des conditions « réalistes ». Nous formulons les hypothèses que :

- l'effet ventriloque est plus efficace en élévation qu'en azimut. Nous avons ainsi mesuré des seuils à 50% pour des stimuli sonores variant à la fois en azimut et en élévation, afin que l'effet ventriloque dans les deux dimensions puisse être comparé directement ;
- l'effet ventriloque peut fonctionner à des angles encore plus larges si l'attention du sujet est focalisée sur le contenu sémantique des stimuli. Nous avons donc interrogé les sujets sur le contenu des stimuli, afin qu'ils ne soient pas uniquement concentrés sur les positions relatives des sources sonore et visuel.

2 Matériel et méthode

2.1 Stimuli

Les séquences utilisées dans ce test montraient un jeune homme sur fond noir prononçant des phrases de 5 secondes. Les séquences avaient été filmées en 3D-stéréoscopique à l'aide d'une caméra Panasonic AG-3DP1 et étaient projetées sur un écran face au sujet. Les phrases étaient construites sur le modèle suivant : « Je m'appelle {nom}, ma couleur préférée est le {couleur} et j'habite à {ville} ». Il y avait trois noms possibles (Antoine, Clément, Pierre), trois couleurs possibles (rouge, vert, bleu) et trois villes possibles (Bordeaux, Lyon, Marseille). Avec toutes les combinaisons possibles de noms, couleurs et villes, le stimulus pouvait donc prendre $3\times3\times3 = 27$ formes différentes.

L'enregistrement des séquences a eu lieu dans une cabine de prise de son (acoustique mate) de l'Université de Brest, avec un microphone DPA 4006 placé 22 cm au-dessus de la bouche de l'acteur (pour que le microphone ne soit pas dans le champ de la caméra) et relié à une interface RME Fireface 800.

2.2 Système de reproduction

Le test subjectif s'est déroulé dans la salle 3D de l'université de Brest (acoustique mate). Les lumières avaient été éteintes pour minimiser l'influence d'éventuels indices visuels. Le sujet était assis au centre de la pièce.

Les images (25 i/s) étaient diffusées à l'aide d'un projecteur Epson EH-TW6000 sur un écran acoustiquement transparent, avec des lunettes 3D actives Epson ELPGS01.

La diffusion des stimuli et l'enregistrement des réponses des sujets étaient assurés par un logiciel programmé sous Max/MSP sur un ordinateur MacBook Pro relié à une interface RME MADIface USB.

Le système de diffusion sonore était composé de 28 enceintes Amadeus PMX4, alimentées par un convertisseur numérique-analogique D.O.Tec Andiamo 2.DA et des amplificateurs Audac DPA154. Chaque haut-parleur avait été filtré numériquement pour égaliser les réponses en fréquence. Pour chaque présentation, la voix de l'acteur était reproduite aléatoirement sur l'un des 28 haut-parleurs à un niveau sonore d'environ 65 dBA.

2.3 Placement des haut-parleurs

Plusieurs études ont obtenu une symétrie gauche-droite pour l'effet ventriloque [5] et il a donc été décidé de placer toutes les enceintes sur la droite du stimulus visuel.

Par contre, les résultats de Werner *et al.* [5] suggèrent que les seuils à 50% dans le plan médian ne sont pas les mêmes selon que l'écart angulaire entre son et image est positif (son au-dessus de l'image) ou négatif (son en-dessous de l'image). Cependant, pour que la durée du test reste raisonnable, nous avons décidé d'étudier uniquement des écarts angulaires positifs (son au-dessus de l'image).

Un système de coordonnées sphériques bi-dimensionnel à deux pôles [6], avec l'azimut et l'élévation représentés par θ et ϕ respectivement, a été utilisé pour décrire les positions des haut-parleurs et des « seuils à 50% » sur une sphère de diamètre 2,40 m centrée sur la tête du sujet.

Le stimulus visuel était projeté sur un écran droit devant le sujet, avec la bouche de l'acteur positionnée à azimut 0°, élévation 0° et 2,40 m de distance.

Le stimulus sonore pouvait être plus ou moins décalé par rapport au stimulus visuel le long de plusieurs arcs de cercles également centrés sur la tête du sujet. Les arcs de cercle pouvaient être plus ou moins inclinés par rapport au plan horizontal : l'angle que formaient au niveau de la bouche de l'acteur un arc de cercle et le plan horizontal est noté δ et appelé *orientation*. Pour que la durée du test reste raisonnable, 4 valeurs ont été retenues pour δ : 0° (décalage de la source sonore vers la droite), 45° (en diagonale), 67.5° (en diagonale) et 90° (vers le haut). Les 4 orientations δ sont représentées dans la Fig. 1. Pour chaque orientation δ :

- une *indication de fusion* correspond à une situation où le sujet indique que la voix et la bouche de l'acteur lui semblent provenir de la même direction ;
- l'angle au niveau de la tête du sujet entre le stimulus visuel (la bouche de l'acteur située droit devant le sujet) et le stimulus sonore (la voix de l'acteur) est appelé *écart angulaire* et noté Ψ ;
- la valeur de Ψ pour laquelle le pourcentage d'indications de fusion est égal à 50% (c'est-à-dire l'écart angulaire Ψ pour lequel la voix et la bouche semblent provenir de la même direction une fois sur deux) est appelé *seuil à 50%* et noté $\Psi_{50\%}$;
- La valeur de Ψ peut être décomposée en différences d'azimut et d'élévation entre le son et l'image : l'azimut et l'élévation correspondants sont notés θ et ϕ respectivement.
- Le seuil à 50% $\Psi_{50\%}$ peut être décomposé en différences d'azimut et d'élévation entre le son et l'image : l'azimut et l'élévation correspondants sont notés $\theta_{50\%}$ et $\phi_{50\%}$ respectivement.

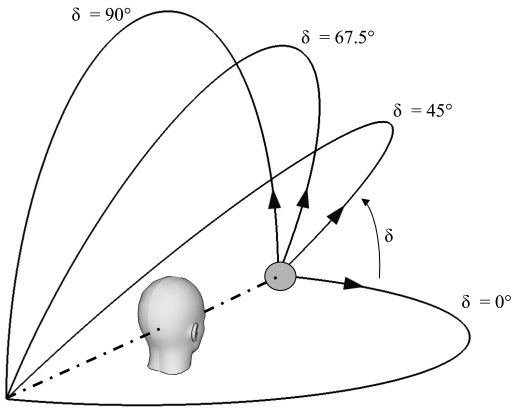

FIGURE 1 – Les 4 orientations δ, le long desquelles pouvait être décalé le stimulus sonore par rapport au stimulus visuel. Les 4 orientations étaient centrées sur la tête du sujet. Le stimulus visuel était toujours projeté à azimut 0°, élévation 0°, et est représenté sur la figure par un disque gris.

La Fig. 2 donne l'exemple d'une enceinte A positionnée le long de l'orientation δ 67.5° avec un écart angulaire Ψ = 36°.

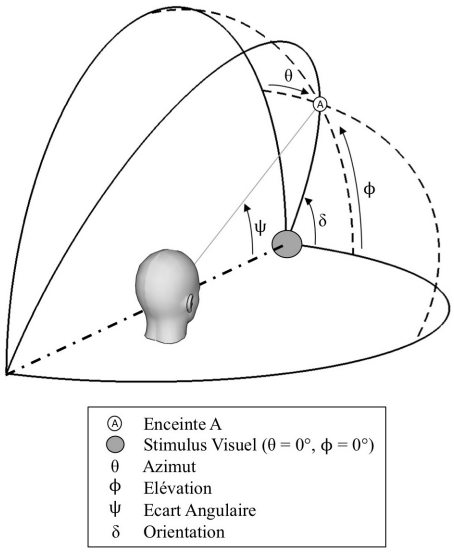

FIGURE 2 – Exemple d'une enceinte A d'orientation δ = 67.5° et d'écart angulaire Ψ = 36°.

Le but de cette expérience est de définir pour chaque orientation δ le seuil à 50% $\Psi_{50\%}$. En plaçant des enceintes le long de chacune des 4 orientations δ (7 enceintes par orientation), 4 fonctions psychométriques peuvent être

estimées, à partir desquelles les seuils à 50% et les pentes à 50% sont déterminés (la pente à 50% d'une fonction psychométrique correspond à la valeur de la pente au point où le pourcentage d'indications de fusion est égal à 50%).

Les valeurs d'orientation δ ont été déterminées à partir d'un test informel passé par les expérimentateurs, qui suggérait que le seuil à 50% $\Psi_{50\%}$ variait modérément entre $\delta = 0°$ et $\delta = 45°$ et substantiellement entre $\delta = 45°$ et $\delta = 90°$. Un test informel mené avec 6 sujets a également permis d'estimer grossièrement la forme des fonctions psychométriques. Le placement des enceintes a ainsi pu être optimisé en suivant les recommandations de Lam *et al.* [7].

2.4 Sujets et Protocole

8 sujets naïfs ont pris part à l'expérience (4 hommes, 4 femmes, âgés de 19 à 40 ans). Ils étaient rémunérés pour leur participation, et aucun d'entre eux n'avait participé à un test d'écoute auparavant.

Un premier test (tâche A « sans question sémantique ») a été mené, dans lequel les sujets devaient répondre après chaque présentation à la question : « la voix et la bouche de l'acteur semblent-elles provenir de la même direction ? ». Une fois qu'ils avaient donné leur réponse, le stimulus suivant était automatiquement lancé.

Un test supplémentaire a été mené (tâche B « avec questions sémantiques »), dans lequel les sujets devaient après chaque présentation rapporter le nom, la couleur favorite et le lieu d'habitation du personnage avant de donner leur réponse sur la cohérence audiovisuelle. En cas de mauvaises réponses, l'essai en question était répété un peu plus tard. Cette tâche supplémentaire a été conduite afin de vérifier si le fait d'attirer l'attention du sujet sur le contenu sémantique (comme dans la « vraie vie ») pouvait permettre à l'effet ventriloque de mieux fonctionner.

En accord avec les recommandations de Lam *et al.* [7], les sujets ont été interrogés 30 fois par enceinte pour chacune des deux tâches. L'ordre de diffusion des enceintes était aléatoire et différent pour chaque sujet. Pour chaque essai, un stimulus était choisi aléatoirement parmi les 27 combinaisons possibles de noms, couleurs et villes. Chaque tâche était divisée en deux sessions d'environ 1 heure, et tous les sujets ont passé les deux tâches sur 4 jours différents. Les sujets A, B, C et D ont commencé par la tâche A, tandis que les sujets E, F, G et H ont commencé par la tâche B.

3 Résultats

Pour estimer les fonctions psychométriques, une approche non-paramétrique basée sur un ajustement linéaire local (*local linear fitting* en anglais) a été utilisée [8].

Comme le taux d'erreur était extrêmement bas (inférieur à 1% pour chaque sujet) durant la tâche B, il a été décidé d'ignorer les essais pour lesquels les sujets avaient donné de mauvaises réponses aux questions sémantiques.

3.1 Influence de l'orientation δ

La Fig. 3 montre les seuils à 50% $\Psi_{50\%}$ en fonction de l'orientation δ, pour chaque sujet, durant la tâche A (sans question sémantique). $\Psi_{50\%}$ ne pouvait pas toujours être déterminé à $\delta = 90°$: pour les sujets B et F par exemple, le pourcentage d'indications de fusion était toujours supérieur à 50%, qu'importe l'enceinte (même pour l'enceinte derrière le sujet, à $\Psi = 137°$, le pourcentage d'indications de fusion était égal à 85% et 77% respectivement). La valeur 137° a été retenue pour la figure, mais il est probable que le pourcentage d'indications de fusion ait été supérieur à 50% à des valeurs d'écart angulaire plus grandes, peut-être même dans tout le plan médian.

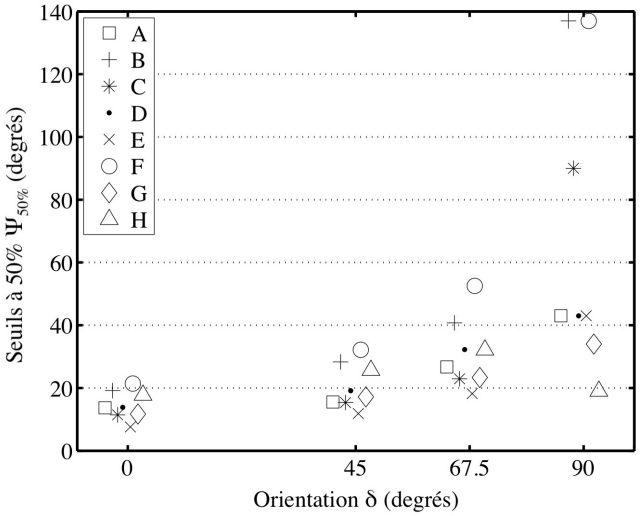

FIGURE 3 – Seuils à 50% $\Psi_{50\%}$ pour chaque sujet en fonction de l'orientation δ. Tâche A.

Il a donc été décidé de ne pas utiliser les seuils à 50% pour les données obtenues à 90° d'orientation δ, mais plutôt d'indiquer la valeur maximale d'écart angulaire Ψ pour laquelle nous avions pu observer un pourcentage d'indications de fusion supérieur à 50%. Comme les fonctions psychométriques étaient décroissantes et monotones pour tous les sujets, il est probable que les véritables seuils à 50% aient été en fait à des écarts angulaires plus importants que ceux indiqués.

Sur la Fig. 3, les seuils à 50% sont déjà relativement dispersés à 0° d'orientation δ, allant de 7° (sujet E) à 21° (sujet F) : certains sujets peuvent donc tolérer des écarts angulaires jusqu'à trois fois supérieurs que d'autres sujets.

Lorsque l'orientation δ augmente de 0° à 45°, tous les sujets montrent à peu près la même tendance, avec une augmentation légère des seuils (moyenne = +6°).

Lorsque l'orientation δ augmente de 45° à 67,5°, les seuils à 50% augmentent plus rapidement, et ce pour tous les sujets (moyenne = +10°, ce qui veut dire que l'augmentation moyenne du seuil est deux fois plus grande qu'entre $\delta = 0°$ et $\delta = 45°$, quand bien même la différence d'orientation δ est deux fois plus petite). Cependant, la vitesse d'augmentation du seuil varie significativement d'un sujet à l'autre : par exemple, elle est de +20° pour le sujet F, alors qu'elle n'est que de +6° pour le sujet G. Ces différentes vitesses accentuent la variabilité inter-sujet des seuils à 50% : à 67,5° d'orientation δ, deux sujets (B et F) présentent de bien plus grands seuils (41° et 53° respectivement) que les autres sujets (dont les seuils à 50% varient de 18° à 32°).

Lorsque l'orientation δ augmente de 67,5° à 90°, les seuils à 50% augmentent encore plus vite pour la plupart des sujets, avec des différences encore plus marquées entre les vitesses d'augmentation : +96° pour le sujet B, mais seulement +12° pour le sujet G. Cela entraine une dispersion

importante des seuils à 50% à 90° d'orientation δ : de 19° pour le sujet H jusqu'à 137° pour les sujets B et F. Le sujet H est l'unique cas de figure où, étrangement, le seuil à 50% décroit de 9.5°. Il est à noter que les sujets A, D et E présentent le même seuil à 90° d'orientation δ, égal à 43°. Cependant, cette égalité est vraisemblablement due à notre définition particulière du « seuil » pour cette orientation, et les véritables seuils à 50% sont probablement différents d'un sujet à l'autre et éparpillés entre 43° (cinquième enceinte de l'orientation) et 90° (sixième enceinte de l'orientation).

Une tendance similaire a été obtenue pour la tâche B avec questions sémantiques (cf. Fig. 4), bien que la différence de seuils à 50% entre δ = 67.5° et δ = 90° soit moins marquée. À nouveau, certains sujets (A, C, D et E) présentent le même seuil à 90° d'orientation δ, égal à 34°, mais les véritables seuils à 50% sont probablement différents d'un sujet à l'autre et éparpillés entre 34° (quatrième enceinte de l'orientation) et 43° (cinquième enceinte de l'orientation).

En résumé, les résultats des deux tâches A et B montrent que les seuils à 50% augmentent strictement lorsque l'orientation δ augmente de 0° à 90° (sauf pour un sujet), et que cette augmentation est plus ou moins rapide selon le sujet, ce qui a pour effet d'élargir la variabilité inter-sujet au fur et à mesure que l'orientation δ augmente de 0° à 90°. À 0° d'orientation δ, les seuils à 50% sont déjà dispersés (de 7° à 21°), mais cette dispersion est finalement modérée par rapport à celle observée à 90° d'orientation δ, où les seuils à 50% vont de 19° jusqu'à 137°.

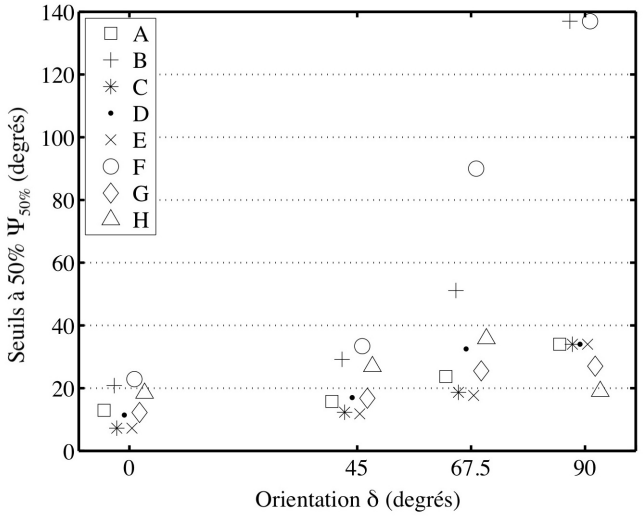

FIGURE 4 – Seuils à 50% $\Psi_{50\%}$ pour chaque sujet en fonction de l'orientation δ. Tâche B

3.2 Influence de la tâche

Des tests de Wilcoxon ont été appliqués aux données pour déterminer si la tâche (« sans question sémantique » - tâche A vs. « avec questions sémantiques » - tâche B) avait eu un impact significatif sur les réponses des sujets :

- un premier test a comparé les seuils à 50% et les pentes obtenues aux orientations δ = {0°, 45°, 67.5°}. L'influence de la tâche s'est révélée non significative (p = 0.976 pour les seuils à 50% et p = 0.224 pour les pentes).
- un second test de Wilcoxon a été mené sur les résultats obtenus à 90° d'orientation δ. Le test comparait les pourcentages d'indications de fusion obtenus pour chaque enceinte, en intégrant les résultats obtenus pour tous les sujets et toutes les enceintes de l'orientation. L'influence de la tâche s'est révélée significative (p = 0.011), avec plus d'indications de fusion lors de la tâche A « sans question sémantique » (64% en moyenne) que lors de la tâche B « avec questions sémantiques » (59% en moyenne).

3.3 Influence de l'azimut et de l'élévation

La Fig. 5 montre les seuils à 50% $\Psi_{50\%}$ décomposés en azimut $\theta_{50\%}$ et en élévation $\phi_{50\%}$ pour la tâche A.

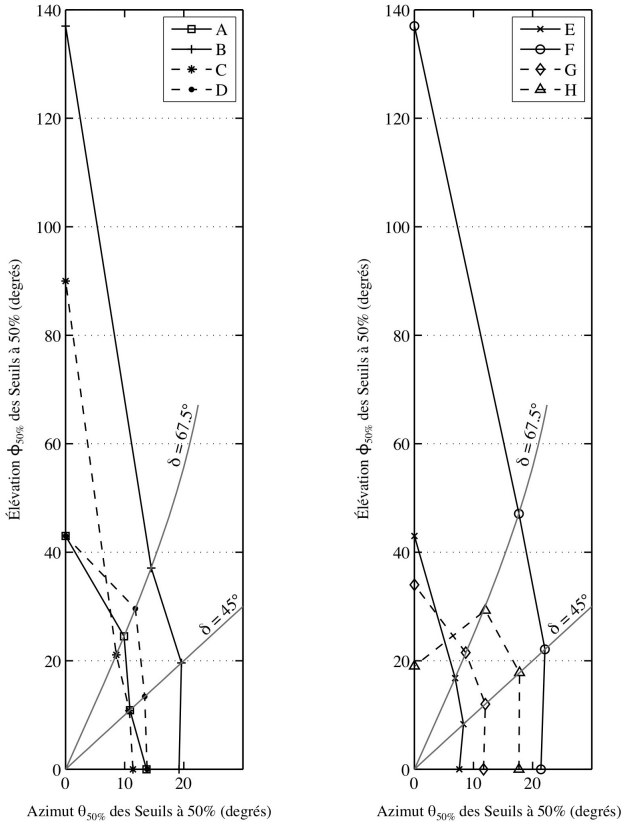

FIGURE 5 – Azimut $\theta_{50\%}$ et élévation $\phi_{50\%}$ des seuils à 50% pour les sujets A, B, C et D (figure de gauche) et les sujets E, F, G et H (figure de droite). Tâche A. Pour plus de lisibilité, les résultats ont été répartis sur deux diagrammes différents. L'axe des abscisses correspond à 0° d'orientation δ tandis que l'axe des ordonnées correspond à 90° d'orientation δ.

Au fur et à mesure que l'orientation δ augmente, les variations d'azimut sont modérées comparées aux variations d'élévation. Un test de Wilcoxon, intégrant les résultats des deux tâches A et B, montre même qu'il n'y a pas de différence significative d'azimut entre les orientations δ = 0° et δ = 45° (p = 0.820). À une certaine orientation comprise entre δ = 45° et δ = 67.5°, l'azimut du seuil à 50% $\theta_{50\%}$ commence à décroître. Cependant, comme on peut le voir sur la Fig. 5, cette décroissance est faible comparée à l'augmentation simultanée de l'élévation $\phi_{50\%}$.

Ainsi, l'efficacité de l'effet ventriloque dépend uniquement de la différence d'azimut entre les stimuli visuel et sonore sur une large amplitude d'orientations (de

δ = 0° à au moins δ = 45°). Au-delà de 45°, la différence d'élévation doit également être prise en compte, mais les variations d'élévation ont bien moins d'impact sur l'effet ventriloque que les variations d'azimut.

4 Discussion

4.1 L'effet ventriloque fonctionne mieux en azimut qu'en élévation

Les résultats de la présente étude confortent l'hypothèse que l'effet ventriloque fonctionne mieux dans le plan vertical que dans le plan horizontal : les seuils à 50% sont en moyenne 4 fois plus grands à 90° d'orientation δ (plan vertical) qu'à 0° d'orientation δ (plan horizontal). Pour deux sujets, il est même probable que l'effet ventriloque ait fonctionné dans la totalité du plan médian, puisque les pourcentages d'indications de fusion étaient de 85% et 77% même quand le stimulus sonore était diffusé dans leur dos (Ψ = 137° à 90° d'orientation δ). Ces résultats sont en accord avec les études précédentes de Thurlow et Jack [4], ainsi qu'avec les performances de localisation du système auditif : comme la précision spatiale est moins bonne dans le plan vertical que dans le plan horizontal, l'influence de la position du stimulus sonore décroît et de plus larges écarts angulaires sont ainsi tolérés.

Cependant, une importante variabilité inter-sujet a pu être constatée, puisque les seuils à 50% pouvaient être de 1.1 jusqu'à 8 fois plus grands dans le plan vertical que dans le plan horizontal selon le sujet.

Pour des directions « obliques » (à 45° ou 67,5° d'orientation δ), les stimuli sonore et visuel présentaient à la fois des différences d'azimut et d'élévation. Cependant, les résultats montrent que les variations d'élévation avaient très peu d'impact sur l'effet ventriloque et que les seuils à 50% dépendaient principalement des différences d'azimut.

4.2 Les fluctuations d'attention influencent l'effet ventriloque, mais uniquement dans le plan médian

Nous avions également formulé l'hypothèse que focaliser l'attention du sujet sur le contenu sémantique des stimuli permettrait à l'effet ventriloque de fonctionner à des écarts angulaires plus importants entre son et image. Cependant, la plupart des sujets ont rapporté que la mémorisation des noms, couleurs favorites et lieux d'habitations du personnage était une tâche simple qui n'avait pas détourné leur attention des disparités spatiales. Les résultats ont montré qu'il n'y a effectivement pas eu d'influence significative de la tâche aux orientations $\delta = \{0°, 45°, 67.5°\}$.

Cependant, l'effet ventriloque a mieux fonctionné durant la tâche A « sans question sémantique » que durant la tâche B « avec questions sémantiques » à 90° d'orientation δ. Ce phénomène ne peut résulter d'un effet d'apprentissage puisque l'ordre des tâches n'était pas le même pour tous les sujets. Forcer les sujets à se concentrer sur le contenu sémantique durant la tâche B a peut-être maintenu leur niveau de stimulation à un plus haut degré, les rendant ainsi plus discriminants sur la durée par rapport à la tâche A.

Même si ces résultats contredisent notre hypothèse de départ, ils montrent tout de même que des fluctuations d'attention du sujet peuvent avoir une influence significative sur l'effet ventriloque. À 0°, 45° et 67,5° d'orientation δ, ces fluctuations sont négligeables, mais ne le sont plus à 90° d'orientation δ.

4.3 Une pondération différente des facteurs déterminant l'effet ventriloque pourrait expliquer pourquoi la variabilité inter-sujet augmente au fur et à mesure que l'orientation δ augmente

Une hypothèse, très semblable au modèle proposé par Thurlow et Jack [4], pourrait expliquer les tendances observées. Le fait qu'un sujet « fusionne » un stimulus sonore avec un stimulus visuel est une décision complexe qui repose sur plusieurs facteurs tels que la position du stimulus sonore par rapport au stimulus visuel, à quel point le sujet estime que les deux stimuli « vont bien ensemble » (réalisme et crédibilité de la combinaison son-image) et à quel point le sujet prête attention à la position de la source sonore. L'influence de ces facteurs est plus ou moins pondérée en fonction de la situation. Par exemple, si la précision de localisation auditive est faible, le sujet hésitera parmi un nombre important de directions pour la localisation du son. Si la combinaison son-image est très convaincante, alors le sujet présumera que la direction la plus probable pour le stimulus sonore est celle de la source visuelle. Ainsi, l'influence du facteur « position de la source sonore » décroît au profit du facteur « réalisme de la combinaison son-image ».

Tandis que certains facteurs sont relativement constants d'un sujet à l'autre, d'autres facteurs présentent une grande variabilité inter-sujet :

- des études ont montré que les performances de localisation étaient comparables d'un sujet à l'autre [3] ;
- la « présomption d'unité » (i.e. à quel point un sujet estime qu'un son et un image vont « bien ensemble » et qui est en grande partie liée au réalisme de la combinaison son-image) dépend étroitement de l'expérience du sujet et de son passé avec des situations semblables [9] ;
- l'attention prêtée aux informations du système auditif peut fortement varier d'un sujet à l'autre [10].

Ainsi, si la pondération associée à un facteur hautement subjectif (tel que le réalisme de la combinaison son-image ou l'attention du sujet) augmente, alors nous pouvons supposer que la variabilité inter-sujet augmentera également.

Dans le plan horizontal, les performances de localisation sont bonnes. Ainsi, tant qu'il y a un minimum de différences azimutales entre le son et l'image, l'effet ventriloque est fortement influencé par la position horizontale de la source sonore par rapport à la source visuelle. Ceci a plusieurs conséquences :

- l'effet ventriloque est limité ;
- une variation d'élévation de la source sonore n'a pas d'effet prononcé car le manque de précision en localisation verticale fait de cette variation un

indice spatial négligeable par rapport à la différence d'azimut ;

- l'influence des autres facteurs tels que le réalisme de la combinaison son-image ou l'attention du sujet est réduite. Ainsi, la variabilité inter-sujet observée est modérée, et les fluctuations d'attention (tâche A vs. tâche B) sont négligeables ;

Cependant, au fur et à mesure que l'orientation δ augmente, il y a de plus en plus de différences d'élévation et de moins en moins de différences d'azimut entre les stimuli visuel et sonore. Comme la localisation est moins précise en élévation, l'influence de la position du stimulus sonore décroît, ce qui a deux conséquences :

- l'effet ventriloque fonctionne à des écarts angulaires plus larges ;
- l'influence des autres facteurs, tels que le réalisme de la combinaison son-image ou les fluctuations d'attention du sujet, augmente. Comme ces facteurs sont très subjectifs, une variabilité inter-sujet plus grande est observée. Une telle hypothèse expliquerait également pourquoi le facteur « attention » (tâche A vs. tâche B) n'est devenu significatif qu'à 90° d'orientation δ.

5 Conclusion

Les résultats ont montré que l'effet ventriloque fonctionnait bien mieux en élévation qu'en azimut (en accord avec les performances de localisation du système auditif) et pouvait fonctionner à des écarts angulaires très élevés (certains sujets continuaient de percevoir la voix du personnage sur sa bouche même lorsque le son était diffusé dans leur dos). Cependant, la variabilité inter-sujet était plus grande en élévation qu'en azimut.

Dans une autre tâche, les sujets devaient répondre à des questions sur le contenu sémantique des stimuli avant de donner leur réponse sur l'effet ventriloque. Tant qu'il y avait un minimum de différence en azimut entre les stimuli sonore et visuel, le fait de poser des questions sémantiques n'avait aucun effet sur les réponses de fusion image/son des sujets. Dans le plan médian (où il n'y a aucune différence d'azimut), l'effet était significatif mais contraire à notre hypothèse de départ. En effet, l'effet ventriloque fonctionnait moins bien lorsque le sujet devait se focaliser sur le contenu sémantique.

Les résultats suggèrent que :

- en azimut, l'efficacité de l'effet ventriloque dépend principalement de la position horizontale du stimulus sonore par rapport à la position du stimulus visuel. Comme les performances de localisation en azimut sont précises et comparables d'un individu à l'autre, l'effet ventriloque fonctionne pour des écarts angulaires limités et la variabilité inter-sujet est modérée par rapport à celle observée en élévation ;
- en élévation, la localisation auditive n'est pas précise, et l'influence de la position du stimulus sonore décroît substantiellement au profit d'autres facteurs à variabilité inter-individuelle élevée tels que l'attention du sujet et le réalisme de la combinaison son-image. Ainsi, des seuils plus importants sont obtenus (surtout si la combinaison son-image est convaincante) et la variabilité inter-sujet augmente.

Remerciements

Les auteurs souhaiteraient remercier Pierre Souchar, Baptiste Le Deun, Vincent Mazo et tous les sujets qui ont participé aux tests subjectifs. Cette étude s'inscrit dans le projet « Cross Channel Film Lab 2 », sélectionné dans le cadre du programme européen de coopération transfrontalière INTERREG IV A France (Manche) - Angleterre, cofinancé par le FEDER. L'étude a également été soutenue par l'Agence Nationale de la Recherche dans le cadre du projet EDISON 3D (ANR-13-CORD-0008-02).

Références

[1] E. Hendrickx, Influence de la stéréoscopie sur la perception du son - Cas des mixages sonores pour le cinéma en relief, Thèse de doctorat, Université de Brest, France (2015).

[2] Alais, D. and Burr, D., The ventriloquist effect results from near-optimal bimodal integration, *Curr. Biol.* **14(3)**, 257-262 (2004)

[3] Makous, J. C. and Middlebrooks, J. C., Two-dimensional sound localization by human listeners, *J. Acoust. Soc. Am.* **87(5)**, 2188-2200 (1990)

[4] Thurlow, W. R. and Jack, C. E., Certain determinants of the "ventriloquism effect", *Percept. Mot. Skills* **36(3c)**, 1171-1184 (1973).

[5] Werner, S. and Liebetrau, J. and Sporer, T., Vertical Sound Source Localization Influenced by Visual Stimuli, *Signal Process. Res.* **2(2)**, 29-38 (2013).

[6] Middlebrooks, J. C. and Makous, J.C. and Green, D. M., Directional sensitivity of sound-pressure levels in the human ear canal, *J. Acoust. Soc. Am.* **86(1)**, 89-108 (1989).

[7] Lam, C. F. and Dubno, J. R. and Mills, J. H., Determination of optimal data placement for psychometric function estimation : a computer simulation, *J. Acoust. Soc. Am.* **106(4)**, 1969-1976 (1999).

[8] Zchaluk, K. and Foster, D. H., Model-free estimation of the psychometric function, *Percept. Psychophys.* **71(6)**, 1414-1425 (2009).

[9] Warren, D. H. and Welch, R. B. and McCarthy, T. J. , The role of visual-auditory compellingness in the ventriloquism effect : Implications for transitivity among the spatial senses, *Percept. Psychophys.* **30(6)**, 557-564 (1981)

[10] Giard, M. H. and Peronet, F., Auditory-visual integration during multimodal object recognition in humans : A behavioral and electrophysiological study, *J. Cognitive Neurosci.* **11**, 473-490 (1999)