



Perception de la distance en condition unimodale et bimodale avec un système audio-visuel 3D utilisant la Wave Field Synthesis

S. Moulin^{a,b}, R. Nicol^a, L. Gros^a et P. Mamassian^b

^aOrange Labs, 2 Avenue Pierre Marzin, 22300 Lannion, France

^bLaboratoire des Systèmes Perceptifs - Ecole Normale Supérieure, 29 rue d'Ulm, 75005 Paris, France
samuel.moulin@orange.com

L'audition et la vision nous permettent au quotidien d'obtenir des informations spatiales et temporelles sur l'environnement qui nous entoure. Bien qu'elles soient captées par des organes différents, les informations visuelles et auditives interagissent et sont souvent interprétées comme un percept unique par notre cerveau. Ce phénomène d'intégration multimodale fait l'objet de nombreuses études dans le domaine de la psychophysique mais reste peu exploré dans les sciences appliquées.

Dans la perspective d'évaluer l'apport de l'audio 3D sur la vidéo 3D, cette étude porte sur la perception de la distance d'objets dans le cas d'une présentation unimodale (audio ou visuelle) et bimodale (audio-visuelle). Deux expériences sont menées pour la présentation unimodale. Dans la première, les participants doivent estimer la distance qui les sépare d'un objet sonore synthétisé par Wave Field Synthesis entre 1 et 10 mètres. Dans la seconde expérience, la cible à localiser est un objet visuel placé entre 2 et 7 mètres. Pour la présentation bimodale, les participants doivent localiser un objet à la fois sonore et visuel, dont la distance varie entre 2 et 7 mètres.

Le système de restitution audio-visuelle utilisé dans ces expériences est composé d'un vidéoprojecteur 3D passif couplé à une barre de 24 haut-parleurs utilisant la technologie Wave Field Synthesis. Les estimations de distance sont collectées suivant une méthode de report direct. Trois types de stimuli sonores et/ou visuels sont étudiés : signal de parole / personnage animé, sonnerie téléphonique / téléphone, salves de bruit blanc / cylindre.

Les résultats montrent une sous-estimation générale de la distance perçue des stimuli sonores, visuels, et audio-visuels. Cependant, des différences sont observées selon la modalité stimulée. Ces différences concernent à la fois le degré de compression de la distance perçue et la précision des jugements. Les résultats mettent également en évidence un effet du type de stimulus évalué.

1 Introduction

Le développement de la vidéo 3D a marqué le monde du divertissement numérique durant la dernière décennie. Il est désormais possible pour tout un chacun de visualiser des vidéos stéréoscopiques sur différents supports : au cinéma, à la télévision, dans les jeux-vidéos, etc. Les technologies et les moyens associés à l'image 3D ont nettement progressé mais qu'en est-il de l'évolution des technologies audio associées ? La plupart du temps, le son qui accompagne la vidéo 3D propose uniquement des effets de latéralisation droite/gauche, plus au moins étendus (stéréophonie, systèmes 5.1 surround). Il est pourtant naturel de s'interroger sur le besoin d'introduire des événements sonores correspondant à l'ajout d'une nouvelle dimension visuelle : la profondeur. Plusieurs technologies audio semblent pouvoir offrir une description sonore 3D de l'espace (technologies binaurales [1], Ambisonics [2], Wave Field Synthesis [3]). Le recourt à ces technologies pourrait potentiellement améliorer la qualité d'expérience de l'utilisateur, en termes de cohérence spatiale audio-visuelle, ou de sensation d'immersion par exemple. Avant de pouvoir étudier ces interactions multimodales complexes, il est nécessaire de comprendre comment un système audio-visuel peut restituer des informations spatiales relatives à la profondeur.

Cette étude porte sur la perception de la distance d'objets virtuels sonores, visuels mais aussi audio-visuels. Ce travail vise à évaluer :

- la capacité d'un système audio basé sur la WFS à restituer des événements sonores à différentes distances,
- la capacité d'un système 3D stéréoscopique à restituer des objets visuels à différentes distances,
- l'évolution de la perception de la distance en condition bimodale (audio-visuelle) et notamment l'apport potentiel des informations sonores.

Cet article présente dans un premier temps quelques rappels sur la perception de la distance. Les indices perceptifs auditifs et visuels ainsi que les méthodes d'estimation de la distance sont notamment abordés. Ensuite, trois expériences sont présentées sur la perception de la distance de stimuli sonores (expérience 1), visuels

(expérience 2), et audio-visuels (expérience 3). Les résultats sont finalement discutés avant de conclure.

2 Perception de la distance

2.1 Indices perceptifs sonores et visuels

La communauté scientifique s'accorde à dire que la perception de la distance de sources sonores statiques dépend principalement de quatre indices acoustiques [4,5]. Pour commencer, le niveau sonore et le rapport entre le champ direct et le champ réverbéré sont probablement les deux indices ayant le plus d'impact. Le niveau sonore est un indicateur « relatif » contrairement au rapport champ direct-sur-réverbéré qui peut fournir une information « absolue » sur la distance d'un objet sonore. Le troisième indice concerne les propriétés spectrales d'une source sonore. Du fait de l'absorption par l'air, les composantes hautes fréquences d'une source acoustique vont être atténuées progressivement avec la distance de propagation. Les différences binaurales constituent le quatrième indice acoustique capable de fournir une information sur la distance d'une source sonore, notamment pour les sources situées en champ proche (moins d'un mètre environ) [6]. D'autres indices « non-acoustiques » peuvent être utiles pour estimer la distance d'une source sonore. La vision par exemple donne des informations précieuses mais elle peut également biaiser la perception sonore. Cet effet a été mis en évidence pour des tâches de localisation en azimut et porte le nom d'effet ventriloque [7]. Des études mentionnent des effets similaires pour la localisation en distance [8,9]. La perception de la distance d'une source sonore peut également être influencée par la familiarité avec le stimulus sonore employé. D'après certaines études, l'utilisation de la parole [10] ou le fait de multiplier les expériences [11] pourraient améliorer la précision des jugements.

De nombreux indices permettent d'évaluer la distance d'objets visuels [12]. On peut distinguer les indices perceptifs monoculaires des indices binoculaires. En effet, la perception du relief ou de la distance d'un objet peut être réalisée par chaque œil, grâce à différents indices : la taille

relative des objets, l'occlusion, la hauteur dans le champ visuel, la perspective linéaire, le gradient de texture, l'ombrage, la perspective aérienne, ou encore l'accommodation. À ces indices monoculaires s'ajoutent des indices binoculaires comme la disparité binoculaire et la convergence par exemple. Il faut noter que dans le cas d'un objet (ou d'un observateur) en mouvement, des indices perceptifs dynamiques sont également pris en compte afin d'estimer une distance de manière plus précise. Les indices perceptifs visuels sont intégrés et pondérés différemment en fonction de la distance de l'objet cible.

2.2 Méthodes d'estimation de la distance

Il existe de nombreuses méthodologies d'estimation de la distance. Quatre types de protocoles peuvent être répertoriés [13] : l'estimation verbale, les actions guidées, les actions imaginées, les correspondances perceptuelles. Les méthodes basées sur la verbalisation se font à l'aide d'une échelle de mesure utilisant une unité familière (le mètre par exemple). Cette catégorie regroupe également les méthodes de « report direct » (sans échelle). Dans les protocoles basés sur les actions guidées, les sujets doivent effectuer une action pour estimer la distance de l'objet. Il peut s'agir de marcher directement jusqu'à la cible (« marche aveugle »), ou de suivre un chemin indirect (« triangulation »). Les actions imaginées peuvent être de différente nature (estimation du temps de marche, ou du nombre de pas à effectuer pour atteindre une cible). Dans les protocoles basés sur la correspondance perceptuelle, les sujets doivent estimer la distance d'un objet cible en effectuant une comparaison avec un objet témoin. Ces protocoles regroupent les « méthodes d'ajustement » et les méthodes utilisant le placement d'un objet à mi-distance (« *perceptual bisection* »).

2.3 Modélisation de la perception de la distance

Il y a quelques années, Zahorik a entrepris un travail bibliographique [5] regroupant plus de 80 articles sur la perception de la distance des sources sonores réelles. Les conditions expérimentales de ces études sont très variées en termes de stimuli, d'environnements expérimentaux, ou de méthodes d'estimation. Cependant, la plupart des résultats indiquent que les participants ont tendance à sur-estimer la distance des sources sonores proches (moins de 1,5m environ), alors que la distance des sources placées plus loin est sous-estimée. Zahorik remarque que la loi de puissance de Stevens [14] est une bonne approximation de la relation entre la distance perçue et la distance physique des sources sonores. Cette loi s'écrit sous la forme de la fonction $d_p = k(d_r)^a$ où d_p est la distance perçue, d_r est la distance physique, l'exposant « a » est le facteur de compression de la fonction, et « k » est une constante. Plusieurs études ont récemment été menées sur la perception de la distance des sources sonores dans les environnements virtuels. Ces études mettent en œuvre de nombreuses technologies audio 3D comme la technologie binaurale [9,15], ou encore les méthodes basées sur l'holophonie [16,17]. Ces études montrent que les performances de localisation auditive en distance sont comparables entre les environnements réels et virtuels. Les distances sont en règle générale sous-estimées et les résultats sont en accord avec le modèle de compression évoqué par Zahorik.

La littérature portant sur la localisation en distance d'objets visuels dans un environnement réel ou virtuel est relativement fournie. La distance d'un objet visuel est relativement bien perçue dans les environnements réels, mais est sous-estimée dans les environnements virtuels (casques de réalité virtuelle, CAVE, etc) [12,18]. La perception de la distance visuelle est généralement modélisée par la loi de Stevens comme dans le cas de l'audition [19].

3 Expériences

Notre étude porte sur la perception de la distance d'objets virtuels dans le cas d'une présentation unimodale (audio ou visuelle), ou bimodale (audio-visuelle). La restitution audio-visuelle est assurée par un vidéoprojecteur 3D couplé à un système sonore basé sur la Wave Field Synthesis. Dans la première expérience, les participants doivent estimer la distance qui les sépare d'un objet sonore virtuel. Dans la seconde expérience, la cible à localiser est un objet visuel. Enfin, dans la troisième expérience, les participants doivent localiser un objet audio-visuel.

3.1 Environnement expérimental

Les expériences se déroulent dans une salle de test d'Orange Labs. Sa superficie est d'environ 20 m² et le temps de réverbération (Tr_{60}) est d'environ 350 ms. Le bruit de fond est inférieur à 30 dB(A) pour la première expérience (Audio), et ne dépasse pas 35 dB(A) pour les deux autres expériences¹.

Un rideau acoustiquement transparent est placé devant les participants dans l'expérience Audio, afin de les empêcher de voir les haut-parleurs.

3.2 Restitution audio-visuelle 3D

La présentation des objets visuels est assurée grâce à un vidéoprojecteur 3D passif (LG CF3D). La distance de visualisation est fixée à 3 m. Une toile de projection métallisée permet de conserver la polarisation des images droite et gauche. L'image projetée mesure 2 m de base (16:9). Afin d'éviter un problème de gêne visuelle lors de la restitution, les objets virtuels doivent être placés dans une gamme de distance définie par la « boîte scénique ». Le DoF [20] (*Depth of Focus*) est fixé à 0,2 et assure une visualisation « confortable » si les objets virtuels sont placés dans la gamme de distance allant de 1,87 m à 7,5 m.

Les sources sonores virtuelles sont quant à elles synthétisées par Wave Field Synthesis grâce à une barre de 24 haut-parleurs Studer placée juste derrière l'écran de projection, à 1,1 m de hauteur. L'écart entre le centre de deux haut-parleurs consécutifs est de 9 cm. Chaque haut-parleur est contrôlé en gain et en retard dans Max/MSP. Les réponses en fréquence des haut-parleurs sont égalisées afin de supprimer les différences de coloration spectrale et les artefacts liés à la directivité naturelle des transducteurs. Le niveau sonore restitué correspond à celui d'un monopole émettant 73 dB(A) à 1 m, et respecte une loi d'atténuation de 6 dB par doublement de distance.

La Figure 1 propose une vue de dessus du système de restitution audio-visuel 3D utilisé.

¹ L'augmentation du bruit de fond est liée à la présence du vidéoprojecteur.

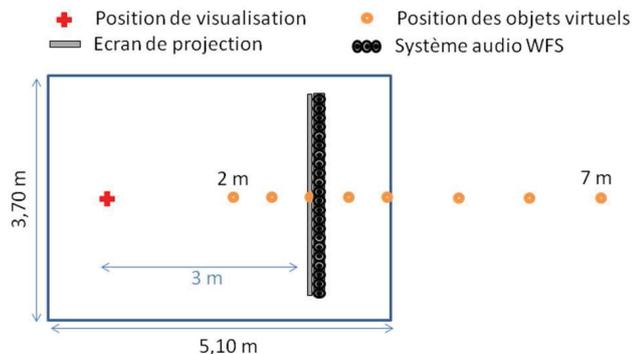


Figure 1 : Système de restitution audio-visuel 3D.

3.3 Stimuli

Trois stimuli audio et/ou visuels sont testés en fonction des expériences :

- femme assise / voix de femme (« Avatar/Voix »),
- cylindre / salves de bruit blanc (« Capsule/Bruit »),
- téléphone / sonnerie de téléphone (« Tel/Sonnerie »).

Ces couples de stimuli sont choisis pour leurs différences en termes de contenu spectral, mais permettent aussi d'explorer le potentiel effet de la familiarité avec le contenu. Un personnage qui parle par exemple, est supposé plus familier qu'un cylindre émettant des salves de bruit blanc. En effet, la taille d'une personne et le niveau sonore qu'elle peut générer en situation de conversation nous sont *a priori* connus.

3.4 Procédure de test

Le participant doit estimer la distance d'un objet cible synthétisé en face de lui. Ces objets sont unimodaux (sonores ou visuels) ou bimodaux (audio-visuels) en fonction des expériences. Le Tableau 1 répertorie les distances simulées en fonction de la modalité étudiée.

Tableau 1 : conditions de distance testées.

| Exp. | Modalité | Distances des objets (m) |
|------|----------------|-------------------------------------|
| 1 | Audio | 1, 1.5, 2, 2.5, 3, 3.5, 4, 5, 7, 10 |
| 2 | Visuelle | 2, 2.5, 3, 3.5, 4, 5, 6, 7 |
| 3 | Audio-Visuelle | 2, 2.5, 3, 3.5, 4, 5, 6, 7 |

La méthode d'estimation de la distance retenue est celle du report au clavier (la précision demandée aux participants est d'un chiffre après la virgule). Aucune limitation n'est imposée quant à la distance minimale ou maximale des objets virtuels. Le stimulus est présenté de manière continue jusqu'à ce que le sujet valide son jugement.

Au début de chaque expérience, une phase de familiarisation à la visualisation 3D et/ou à l'écoute du système WFS est proposée lors de laquelle tous les objets virtuels sont présentés à toutes les positions dans un ordre aléatoire. Une phase de familiarisation à la tâche est ensuite réalisée : le sujet est mis en situation de test. Il peut ainsi prendre en main l'interface de test et s'entraîner à estimer la distance des objets virtuels.

Lors de la phase de test, chacune des positions est présentée 5 fois. Le sujet doit donc estimer 50 conditions

par stimulus dans l'expérience portant sur la modalité audio seule, et 40 conditions dans les deux autres expériences. L'ordre de présentation des conditions est aléatoire. L'ordre de présentation des stimuli est alterné pour que chaque combinaison possible soit présentée au même nombre de participants.

Les trois expériences sont réalisées à des dates différentes avec, à chaque fois, des panels de 24 participants rémunérés.

4 Résultats

La moyenne d'âge des participants est d'environ 31 ans pour toutes les expériences. Les estimations de distance recueillies sont soumises à une analyse de variance.

4.1 Expérience 1 : stimuli sonores

Dans la première expérience, les participants doivent estimer la distance de sources sonores synthétisées par Wave Field Synthesis. Les sources sonores sont synthétisées entre 1 et 10 m. Une analyse de variance (ANOVA) est effectuée sur les notes des 24 participants. La Figure 2 représente les distances moyennes et les intervalles de confiance à 95% associés.

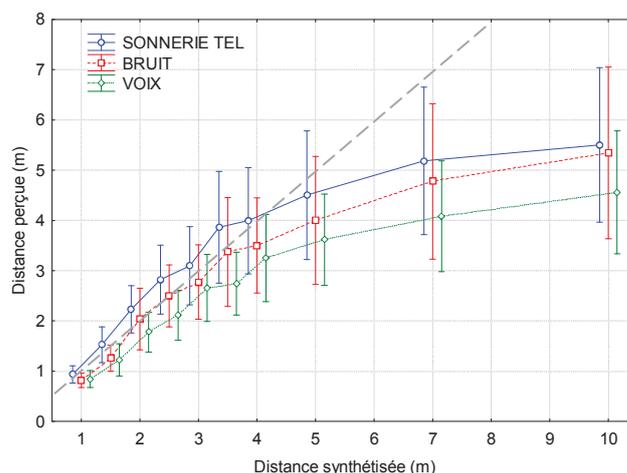


Figure 2 : moyennes et intervalles de confiance à 95% pour l'expérience 1 (stimuli sonores).

Pour commencer, les distances perçues sont en règle générale sous-estimées pour tous les stimuli à partir de 4 m. Ensuite, les intervalles de confiance (représentés par les barres verticales sur la Figure 2) indiquent que les jugements des participants sont plus précis pour les sources sonores les plus proches. Cette figure montre également que les différences perçues entre les stimuli s'amplifient avec la distance. Le stimulus « Voix » est généralement perçu comme étant plus proche que les stimuli « Bruit » et « Sonnerie Tel. ». De plus, la variabilité des résultats pour le signal de parole est moins importante, en particulier pour les distances de 5, 7, et 10 mètres. Cet effet du stimulus est confirmé par l'analyse ANOVA ($F(2,46)=6.13, p<0.005$). Un test de post-hoc (Bonferroni) révèle que cet effet est uniquement présent entre les stimuli « Voix » et « Sonnerie Tel. ». Il n'y a pas de différences statistiquement significatives entre les autres stimuli. Il a été montré que la perception de la distance d'un signal de parole peut être influencée par l'intonation du locuteur [21]. Les conclusions de ces travaux montrent notamment qu'une voix « parlée » est naturellement perçue comme étant plus

proche qu'une voix « criée ». Dans notre cas, le stimulus de parole utilisé est comparable à une voix « parlée ». De ce fait, pour les distances supérieures à quelques mètres, il est possible qu'une sous-estimation de la distance vienne s'ajouter au phénomène de compression décrit dans la partie 2.3, du fait de l'intonation. En effet, dans un environnement naturel, une personne placée à plusieurs mètres aurait plutôt tendance à crier pour se faire entendre. Concernant la variabilité des estimations, Coleman [10] a montré que l'utilisation de stimuli sonores familiers peut potentiellement augmenter la précision des jugements ; ce qui serait en accord avec les résultats présentés ici.

Il est intéressant de comparer ces résultats expérimentaux au modèle de perception de la distance régulièrement évoqué dans la littérature. Le Tableau 2 présente les valeurs des coefficients « k » et « a » du modèle de Stevens explicité dans le paragraphe 2.3, ainsi que les coefficients de détermination (R^2) associés qui caractérisent la pertinence du modèle.

Tableau 2 : coefficients des fonctions de compression et coefficients de détermination (R^2) de l'expérience 1.

| Stimulus | k | a | R^2 |
|---------------|------|------|-------|
| Sonnerie Tel. | 1,21 | 0,78 | 0,919 |
| Bruit | 1,03 | 0,82 | 0,935 |
| Voix | 1,00 | 0,75 | 0,945 |

Les valeurs présentées dans le Tableau 2 indiquent que les résultats expérimentaux sont comparables aux lois de compression évoquées dans la littérature. En effet, les valeurs des coefficients de détermination sont élevées ($R^2 > 0,91$) et témoignent donc d'une forte corrélation entre le modèle de prédiction et les valeurs expérimentales.

4.2 Expérience 2 : stimuli visuels

Dans la deuxième expérience, les participants doivent estimer la distance d'objets visuels présentés entre 2 et 7 m grâce à un vidéoprojecteur 3D. La Figure 3 représente les distances moyennes perçues et les intervalles de confiance à 95% associés.

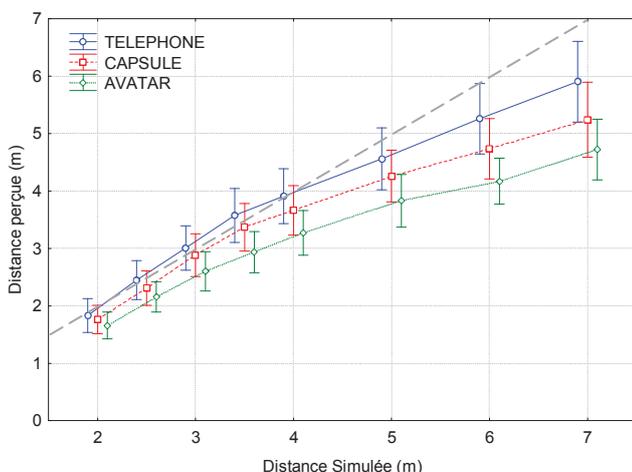


Figure 3 : moyennes et intervalles de confiance à 95% pour l'expérience 2 (stimuli visuels).

La Figure 3 montre que les participants sont capables d'estimer la distance des différentes conditions ($F(7,161)=193.15, p < 0.001$), mais que les distances des

objets visuels sont globalement sous-estimées, ce qui est en accord avec les résultats issus de la littérature [19]. Des différences sont constatées en fonction de l'objet à localiser. L'objet « Avatar » est perçu comme étant plus proche que « Capsule », alors que le stimulus « Téléphone » est perçu comme étant l'objet le plus lointain. Cette hiérarchie est respectée pour chaque distance même si les différences ne sont pas toujours statistiquement significatives (notamment entre les stimuli « Capsule » et « Téléphone »). Cet effet du stimulus est également révélé par l'ANOVA ($F(2,46)=16.88, p < 0.001$) et pourrait être expliqué par la taille relative des objets présentés [12].

Comme pour l'expérience 1, les intervalles de confiance montrent une augmentation de la variabilité des jugements en fonction de la distance. Le Tableau 3 présente les valeurs des coefficients « k » et « a » du modèle de Stevens utilisé précédemment.

Tableau 3 : coefficients des fonctions de compression et coefficients de détermination (R^2) de l'expérience 2.

| Stimulus | k | a | R^2 |
|-----------|------|------|-------|
| Téléphone | 1,07 | 0,90 | 0,981 |
| Capsule | 1,08 | 0,84 | 0,972 |
| Avatar | 1,03 | 0,80 | 0,982 |

Au vu des valeurs élevées du coefficient de détermination ($R^2 > 0,97$), la loi de Stevens semble être un bon modèle pour décrire la distance perçue des objets visuels. Les résultats mettent également en évidence les différences de compression de la distance entre les stimuli. La compression est par exemple plus importante pour le stimulus « Avatar » ($a=0,80$). Le facteur k est quant à lui relativement stable indépendamment du stimulus.

4.3 Expérience 3 : stimuli audio-visuels

Dans cette troisième expérience, les participants doivent estimer la distance d'objets virtuels audio-visuels, placés entre 2 et 7 m. La Figure 4 représente les distances moyennes perçues et les intervalles de confiance à 95% associés.

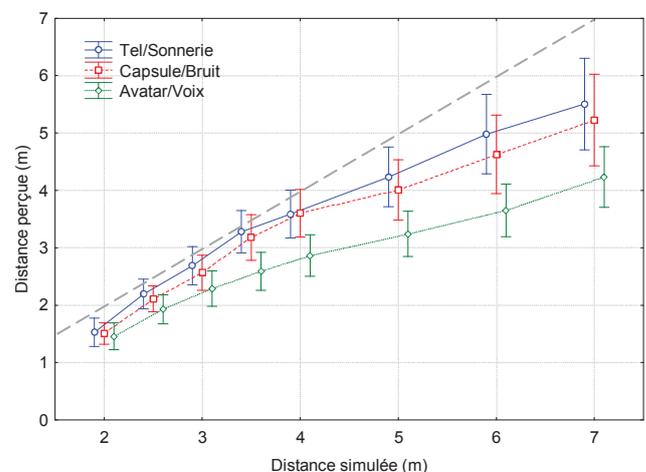


Figure 4 : moyennes et intervalles de confiance à 95% pour l'expérience 3 (stimuli audio-visuels).

Ici encore, bien que les participants arrivent à distinguer les différentes conditions de distance ($F(7,161)=118.42, p < 0.001$), les distances des objets audio-visuels sont sous-

estimées. La Figure 4 illustre également l'augmentation de la variabilité des réponses avec la distance des objets.

L'analyse ANOVA révèle un effet du stimulus ($F(2,46)=37.34, p<0.001$) principalement dû au stimulus « Avatar/Voix » qui est perçu comme étant plus proche que les autres stimuli. En effet, un test de post-hoc (Bonferroni) révèle qu'il n'y a pas de différences statistiquement significatives entre les deux autres stimuli. Le Tableau 4 présente les valeurs des coefficients « k » et « a » du modèle de Stevens dans le cas de l'expérience 3.

Tableau 4 : coefficients des fonctions de compression et coefficients de détermination (R^2) de l'expérience 3.

| Stimulus | k | a | R^2 |
|---------------|------|------|-------|
| Tel/Sonnerie | 0,87 | 0,98 | 0,974 |
| Capsule/Bruit | 0,87 | 0,95 | 0,968 |
| Avatar/Voix | 0,91 | 0,80 | 0,982 |

Comme précédemment, les valeurs élevées du coefficient de détermination indiquent que la loi de Stevens semble être une bonne approximation de la distance perçue des objets audio-visuels. La compression de la distance est plus importante pour le stimulus « Avatar/Voix » ($a=0,80$) et le facteur k est stable quel que soit le stimulus présenté.

4.4 Discussions

Il semble que les effets d'interaction entre modalité et distance ne dépendent pas du stimulus évalué. C'est pourquoi la Figure 5 représente les résultats des expériences 1, 2, et 3, moyennés sur tous les stimuli.

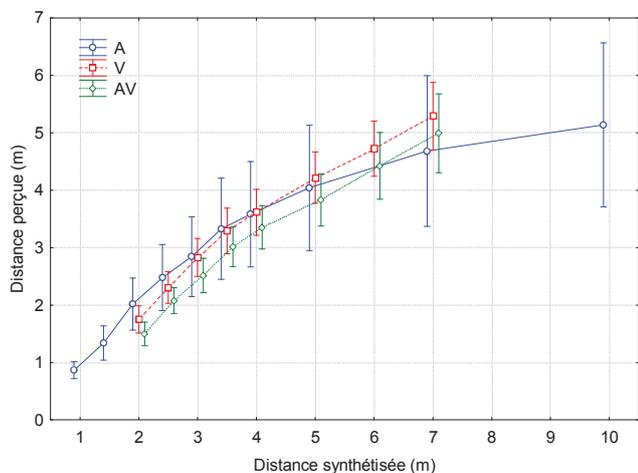


Figure 5: moyennes et intervalles de confiance à 95% des trois expériences (Audio, Vidéo, et Audio-Visuel)

Il est impossible de réaliser une analyse ANOVA pour comparer les résultats issus des trois expériences car les conditions de distances diffèrent suivant les expériences (cf. Tableau 1). Cependant, il est possible de commenter les tendances générales qui s'en dégagent. Les expériences 2 et 3 (Vidéo et Audio-Visuel) utilisent quant à elles les mêmes conditions de distance et seront donc comparées au moyen d'une analyse ANOVA. Enfin, les performances de prédiction de la loi de Stevens seront comparées à celles d'un modèle logarithmique.

4.4.1 Présentation sonore vs. visuelle

Il est intéressant de comparer les résultats issus des conditions unimodales (audition dans l'expérience 1 et vision dans l'expérience 2). Des différences importantes ressortent de la comparaison des courbes « A » et « V » de la Figure 5. Pour commencer, la variabilité est plus importante pour la modalité audio. En effet, les intervalles de confiance à 95% sont environ deux fois plus importants dans le cas de l'audition. Ensuite, le phénomène de compression de la perception de la distance semble plus marqué pour la modalité audio seule que pour la modalité visuelle seule. La comparaison des valeurs de l'exposant « a » dans les Tableaux 2 et 3 va dans ce sens : $a \in [0,75 ; 0,82]$ pour l'audition et $a \in [0,80 ; 0,90]$ pour la vision.

Ces résultats semblent confirmer que la perception de la distance d'un objet est plus performante (valeurs estimées et variabilité) avec la modalité visuelle qu'avec la modalité auditive seule.

4.4.2 Présentation visuelle vs. audio-visuelle

La Figure 5 met en évidence les similitudes entre les résultats de l'expérience 2 (Vidéo) et l'expérience 3 (Audio-Visuel). Il apparaît cependant que la distance perçue dans l'expérience 3 (présentation bimodale) a tendance à être sous-estimée par rapport aux jugements basés sur la perception visuelle seule. Une analyse ANOVA est réalisée sur les résultats de ces deux expériences (3 stimuli et 8 conditions de distance). Cette analyse révèle qu'il n'y a pas d'effet statistiquement significatif de la modalité entre ces deux expériences ($F(1, 23)=1.44, p>0.1$).

4.4.3 Loi de Stevens vs. modèle logarithmique

La distance perçue d'un objet sonore ou visuel est en générale décrite par la loi de puissance de Stevens (voir § 2.3). Les résultats expérimentaux présentés dans cette étude ont été comparés à deux lois de comportement : la loi de Stevens décrite par $d_p = k (d_r)^a$, et un modèle « logarithmique » du type $d_p = A \ln(d_r) + B$. Dans ces équations, a, k, A et B sont les paramètres des deux modèles de régression étudiés. Le Tableau 5 répertorie les valeurs des coefficients de détermination R^2 associés à chaque stimulus de chaque expérience de cette étude.

Tableau 5: coefficients de détermination (R^2) issus de la loi de Stevens et du modèle logarithmique (Tel/Sonnerie, Caps./Bruit, Avatar/Voix notés respectivement A, B, et C).

| Exp. | R^2 loi de Stevens A / B / C | R^2 modèle logarithmique A / B / C |
|------|-----------------------------------|---|
| 1 | 0,919 / 0,935 / 0,945 | 0,983 / 0,993 / 0,987 |
| 2 | 0,981 / 0,972 / 0,982 | 0,996 / 0,999 / 0,998 |
| 3 | 0,974 / 0,968 / 0,982 | 0,997 / 0,995 / 0,991 |

Les coefficients de détermination du modèle logarithmique sont toujours supérieurs, ce qui semble indiquer que ce modèle est plus adapté pour prédire la perception de la distance d'un objet virtuel (sonore, visuel, ou audio-visuel). À titre illustratif, la Figure 6 représente les valeurs expérimentales pour le stimulus « Capsule » (expérience 2), ainsi que les courbes de tendances des deux modèles comparés.

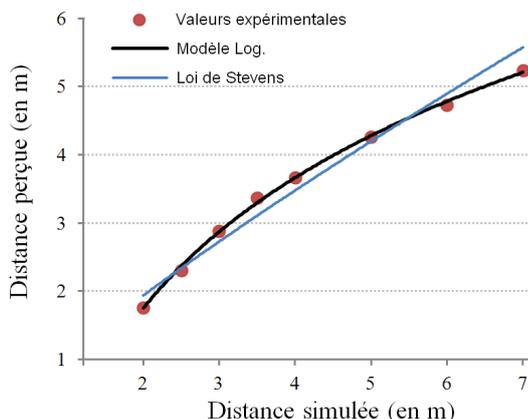


Figure 6 : distance perçue pour le stimulus « Capsule » de l'expérience 2 et courbes de tendances associées (R^2 Stevens =0,972, et R^2 Log.=0,999, cf. Tableau 5).

Cette figure illustre la précision des modèles comparés pour cette condition expérimentale particulière.

5 Conclusion

Cette étude présente trois expériences sur la perception de la distance d'objets virtuels dans le cas d'une présentation unimodale (audio ou visuelle), ou bimodale (audio-visuelle). La restitution des stimuli est assurée par un système audio-visuel 3D utilisant la Wave Field Synthesis. Les résultats montrent que la perception de la distance des stimuli sonores, visuels, et audio-visuels est généralement sous-estimée avec un phénomène de compression plus ou moins marqué suivant la modalité impliquée. Les distances évaluées par la modalité auditive semblent plus compressées et également plus variables. Les résultats issus des deux autres expériences (stimuli visuels et audio-visuels) montrent des comportements similaires en termes de variabilité mais aussi de niveau de compression de la perception. Un effet du stimulus a été mis en évidence dans chacune des expériences. En effet, le stimulus « Avatar » et/ou « Voix » est perçu comme étant le plus proche quelle que soit la modalité étudiée.

Les perspectives dégagées par ce travail sont nombreuses. Il serait par exemple intéressant d'explorer plus finement l'apport potentiel de la modalité auditive dans la présentation bimodale en utilisant une méthode d'estimation plus précise. Un autre axe d'étude porte sur l'évaluation de la fenêtre d'intégration audio-visuelle en profondeur. Il s'agit d'étudier la gamme de distance dans laquelle des stimuli audio et visuels spatialement incohérents sont perçus comme un percept unique.

Références

[1] Møller H., Fundamentals of binaural technology, *Applied Acoustics*, Vol. 36, 171-218. (1992).
 [2] Gerzon M.A., Periphony: With-eight sound reproduction, *J.Audio.Eng.Soc.*, 21(1), 2-10. (1973).
 [3] Berkhout A.J., De Vries D., and Vogel P., Acoustic control by Wave Field Synthesis, *J.Acoust.Soc.Am*, 93(5), 2764-2778. (1993).
 [4] Blauert J., Spatial Hearing, The Psychophysics of Human Sound Localization, MIT Press, Cambridge. (1997).

[5] Zahorik P., Brungart D.S., and Bronkhorst A.W., Auditory Distance Perception in Humans: A Summary of Past and Present Research, *Acta Acust. united with Acustica*, Vol. 91, 409-420. (2005).
 [6] Shinn-Cunningham B.G., Kopčo N., and Martin T., Localizing nearby sound sources in a classroom: Binaural room impulse responses, *J.Acoust.Soc.Am*, 117(5), 3100-3115. (2005).
 [7] Howard I.P., and Templeton W.B., Human spatial orientation, Wiley, New York. (1966).
 [8] Mershon D.H., Desaulniers D.H., Amerson T.L.J., and Kiefer S.A., Visual capture in auditory distance perception: Proximity image effect reconsidered, *J. Aud. Res.*, 20(2), 129-136. (1980).
 [9] Côté N., Khoel V., Paquier M., and Devillers F., Interaction between auditory and visual distance cues in virtual reality applications, *Proc. Forum Acusticum 2011*, 1275-1280, Aalborg. (2011).
 [10] Coleman P.D., Failure to localize the source distance of an unfamiliar sound, *J.Acoust.Soc.Am*, 34(3), 345-346. (1962).
 [11] Kopčo N., Schoolmaster M., and Shinn-Cunningham B.G., Learning to judge distance of nearby sounds in reverberant and anechoic environments, *Proc. Joint CFA/DAGA'04*, Strasbourg. (2004).
 [12] Cutting J.E., and Vishton P.M., Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. *Perception of space and motion*, Vol. 5, 69-117. (1995).
 [13] Klein E., Swan J.E., Schmidt G.S., Linvingston M.A., and Staadt O.G., Measurement protocols for medium-field distance perception in large-screen immersive displays, *PRO. IEEE Virtual Reality*, 107-113. (2009).
 [14] Stevens S.S., On the psychophysical law. *Psychological Review*, 64(3):153-181. (1957).
 [15] Zahorik P., Assessing auditory distance perception using virtual acoustics, *J.Acoust.Soc.Am*, 111(4), 1832-1846. (2002).
 [16] Kearney G., Gorzel M., Rice H., and Boland F., Distance perception in interactive virtual acoustic environments using first and higher order ambisonic sound fields, *Acta Acustica with Acustica*, 98(1), 61-71. (2012).
 [17] Rébillat M., Boutillon X., Corteel E., and Katz F.G.B., Audio, visual, and audio-visual egocentric distance perception by moving subjects in virtual environments, *ACM Trans. Appl. Percept.*, 9(4), 19:1-19:17. (2012).
 [18] Loomis J. M., and Knapp J. M., Visual perception of egocentric distance in real and virtual environments. In *Virtual and Adaptive Environments*. Lawrence Erlbaum Associates, Mahwah, NJ, USA. L. J. Hettinger and J. W. Haas, Eds., 21-46, (2003).
 [19] Teghtsoonian M., and Teghtsoonian R., Scaling apparent distance in natural indoor settings. *Psychonomic Science*, Vol. 16, 281-283. (1969).
 [20] Chen W., Caractérisation multidimensionnelle de la qualité d'expérience en télévision de la TV3D stéréoscopique, PhD thesis. (2012).
 [21] Brungart D.S., and Scott K.R., The effects of production and presentation level on the auditory distance perception of speech, *J.Acoust.Soc.Am*, 110(1), 425-440. (2001).