

Les technologies du son 3D au service des aides aux non-voyants

Gaëtan Parseihian, Brian FG Katz

LIMSI - CNRS

BP 133

91403 Orsay CEDEX

Tél.: +(33) 1 69 85 80 80

Fax: +(33) 1 69 85 80 88

E-mail : gaetan.parseihian@limsi.fr



es dernières années, les techniques de génération de sons spatialisés ont atteint une certaine maturité permettant d'envisager leur usage dans des applications grand public. La synthèse binaurale, technique de restitution sur casque stéréophonique, est de plus en plus utilisée dans des applications mobiles ainsi que dans des systèmes commerciaux. Cette technique de spatialisation sonore est basée sur l'utilisation de fonctions de transfert relatives à la tête (HRTFs). Ces HRTFs contenant l'ensemble des transformations acoustiques entre la source et les conduits auditifs permettent de positionner perceptivement un signal sonore à un endroit quelconque de l'espace. Beaucoup étudiée dans la recherche, cette technique commence à être suffisamment éprouvée pour créer des espaces virtuels sonores permettant de suppléer à la vision. Une des applications la plus souvent citée est l'aide aux non-voyants (NV). Effectivement, la synthèse binaurale peut permettre de générer des informations spatiales sous la forme de son 3D d'une manière plus rapide et plus efficace que ce qui pourrait être fait avec des instructions vocales. Nous allons aborder, dans cet article, plusieurs manières d'utiliser la synthèse binaurale dans des systèmes d'aide aux NV. Ces aides ont pour objectif de restituer certaines fonctions assurées par le système visuel et dont l'absence peut engendrer un besoin chez les déficients visuels. Ce sont des moyens destinés à permettre à l'utilisateur de compenser l'absence de vision par l'intermédiaire de la modalité auditive.

Cet article dresse le bilan de plusieurs études de réalité virtuelle ou augmentée au service des NV, réalisées au LIMSI-CNRS, puis détaille plus spécifiquement les recherches réalisées pour la mise en place d'un dispositif de détection d'objets en champ proche.

Réalité Virtuelle/Augmentée pour les non-voyants (NV)

Se déplacer de manière autonome dans un environnement inconnu, rejoindre une destination en évitant les obstacles et en acquérant suffisamment d'informations pour se construire une représentation mentale de l'espace est une tâche très compliquée pour les NV. Ceci est dû au fait qu'ils rencontrent de nombreux problèmes comme garder une orientation correcte, connaître les noms des rues et la configuration des carrefours... Cette diminution du sentiment de sécurité rend difficile la découverte d'un nouvel itinéraire sans l'aide d'une tierce personne.

Plusieurs études ont exploré l'utilisation de la réalité virtuelle ou augmentée pour l'aide aux NV. Afonso et al. [1] ont exploré les capacités de cognition spatiale entre des sujets NV de naissance, NV tardifs et voyants privés de vision. Dans cette étude, un système de réalité virtuelle a été employé pour créer des scènes spatiales dans une pièce de 12 m².

Les résultats ont montré (contrairement à ce qui est établi) que la vision n'est pas nécessaire pour la création d'une carte cognitive et que l'exploration réelle d'une scène sonore virtuelle est suffisante pour étudier nos capacités de cognition spatiale.

Plus récemment, Katz et Picinali [7,8] ont étudié le réalisme d'un dispositif de réalité virtuelle en termes de création d'espaces architecturaux accessibles aux NV. Dans cette étude, les auteurs ont comparé les capacités d'exploration d'un bâtiment par des NV en réel (déplacement physique dans l'environnement) et en virtuel (déplacement dans environnement sonore binaural avec un modèle acoustique de chacune des pièces).

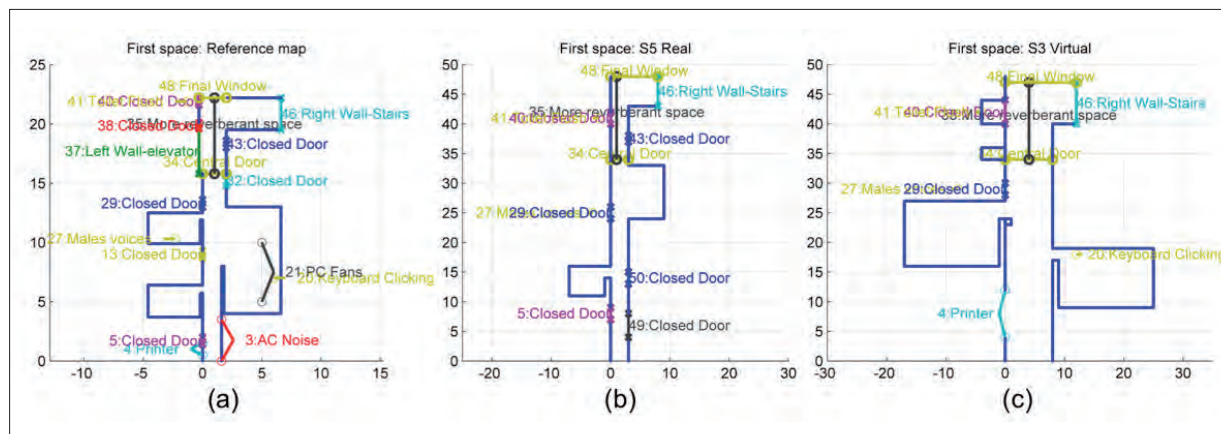


Fig. 1 : (a) Carte de l'environnement réel (unité en mètre). Exemples de cartes reconstruites après l'exploration réelle (b) et après l'exploration virtuelle (c), (unité en briques de LEGO®)

Les résultats ont mis en évidence des performances de compréhension de l'espace similaires pour ces deux conditions d'exploration active (Figure 1). Ce type de système pourrait donc être très utile pour l'apprentissage de lieux inconnus. Il permettrait aux utilisateurs de visiter virtuellement un bâtiment avec leur ordinateur, leur permettant ainsi de s'y déplacer en toute sérénité lors de leur déplacement réel.

Pour aider les NV dans leur vie quotidienne, en mobilité, la combinaison de la réalité augmentée auditive et des systèmes GPS («*Global Positioning System*») semble être une bonne combinaison. L'idée d'utiliser le GPS pour l'aide à la navigation des NV n'est pas nouvelle. De nombreux systèmes commerciaux (Kapten, Trekker, Blind navigator, etc.) basés sur le principe de géolocalisation ont été développés. Ces systèmes emploient un retour sonore monaural, souvent via un moteur de synthèse vocale. Il est cependant difficile de se focaliser sur les instructions verbales pendant un déplacement tout en effectuant simultanément une autre tâche, comme le contournement d'un obstacle, la détection de dangers potentiels ou simplement la discussion avec une autre personne. Plusieurs projets de recherche ont mis en évidence les avantages à utiliser les capacités de perception sonore spatiale de l'homme en transmettant les informations avec des sons 3D virtuels. Loomis [9] a démontré la supériorité du son 3D sur la synthèse vocale, au niveau de la charge cognitive et de la précision d'orientation pour les tâches de déplacement. Wilson [13] a quant à elle étudié les performances de navigation en fonction du type de son et montré la possibilité de guider avec des sons 3D diffusés sur casques osseux.

Récemment, le projet ANR-NAVIG [7] a mis en place un système d'aide à la navigation novateur par son intégration d'un système de reconnaissance d'image pour affiner la précision de localisation par rapport au GPS standard. Sur les aspects auditifs, il est important de transmettre les informations de manière ergonomique. Pour une tâche de navigation en environnement urbain, il faut, par exemple, s'orienter vers des sons suffisamment courts pour ne pas gêner la prise d'informations sonores réelles. Dans le projet NAVIG, différents types d'informations

organisées en catégories et sous-catégories doivent être transmises sous la forme de balises sonores spatialisées. Pour répondre aux attentes des utilisateurs en termes de personnalisation des sons, Parseihian et Katz [10] ont développé le concept de «morphocons». Ces notifications audio permettent de construire un vocabulaire sonore basé sur des variations morphologiques de paramètres acoustiques (fréquence, tempo, intensité, etc.). Ces signaux auditifs présentent l'avantage d'être indépendants du type de son. Ils permettent à l'utilisateur de changer de palette sonore sans introduire d'apprentissage supplémentaire. À partir de ce concept, un vocabulaire a été créé pour informer des différentes catégories et sous-catégories de messages. Trois palettes sonores différentes, basées sur ce vocabulaire, ont été construites. Leur ressemblance morphologique a été validée à partir d'un test de reconnaissance diffusé sur le web qui a permis de montrer que les formes sonores sont reconnues par l'utilisateur indépendamment du son qui les transmet. Il est donc possible d'utiliser le son 3D pour guider de manière efficace les NV en extérieur et l'utilisation d'un vocabulaire sonore personnalisable permet la mise en place d'un système agréable et non invasif.

Guidage en champ proche

Saisir un objet est une tâche compliquée pour les NV. Elle leur demande beaucoup de temps et une bonne mémoire spatiale. Plusieurs types de dispositifs ont vu le jour pour pallier ce problème. Basés sur le principe de la substitution sensorielle [2] ou de l'augmentation sensorielle [8], ces dispositifs visent à transmettre à l'utilisateur la position des objets situés directement devant lui. Ils ne permettent cependant pas de connaître la nature des objets détectés et transmettent un nombre trop important d'informations.

Pour effectuer ce type de tâche, nous proposons, dans le cadre du projet NAVIG, la combinaison d'un algorithme de reconnaissance visuelle permettant de détecter la position d'une cible définie (avec une caméra stéréoscopique placée sur la tête de l'utilisateur) à un moteur de spatialisation binaurale permettant de placer un son virtuel à l'emplacement physique de la cible à attraper.

Avec ce système, l'utilisateur peut demander à trouver sa tasse à café, une fois que celle-ci a été identifiée et localisée, l'utilisateur est guidé vers sa position avec des indications sonores spatialisées. En dehors des considérations techniques du système de reconnaissance d'image, cette application soulève des questions fondamentales sur notre capacité à localiser des sons en champ proche (<1,5 m), à attraper des objets sonores ainsi que sur la capacité de notre moteur de synthèse binaurale à reproduire les indices nécessaires à la localisation dans cette zone. Les capacités du système auditif sont moins connues dans cette zone et mènent en général à des performances plus faibles [3]. Nous souhaitons donc aussi étudier la possibilité d'utiliser un système de réalité augmentée à accentuer ou suppléer les indices de localisation naturels pour obtenir des performances supérieures à la situation réelle.

Afin de mettre en place un système qui transmette des positions spatiales de manière efficace, nous avons, dans un premier temps, étudié les performances de localisation et de saisie dans l'espace péripersonnel d'objets sonores réels. Plusieurs études ont permis de quantifier la précision qu'il est possible d'atteindre dans cette zone, d'étudier l'influence du type de stimulus utilisé, ainsi que les différences de performances entre voyants et NV ou les différences de pointages entre la main préférée et la main secondaire. Les résultats mettent en évidence une grande variabilité des réponses dans la zone étudiée. En général, les performances directionnelles sont meilleures dans la zone et les performances en distance bien que faibles sont meilleures sur les côtés. La répétition du stimulus permet d'augmenter la précision en azimuth et permet de diminuer sa longueur sans affecter la précision en distance. Les faibles différences de performances entre voyants et NV ainsi qu'entre main préférée et main secondaire permettent de s'affranchir de ce type de considérations dans la suite des investigations.

Une étude similaire a ensuite été réalisée pour quantifier les dégradations de performances induites par la synthèse binaurale et pour étudier les stratégies permettant d'améliorer la perception des indices de distance. Nous allons, après avoir décrit le dispositif utilisé et le protocole du test, comparer les performances obtenues avec des sons virtuels à celles qui sont obtenues précédemment avec des sons réels. Après avoir décrit le concept et la mise en place de stratégies de sonification de la distance, nous étudierons les améliorations qu'entraînent deux métaphores que nous avons créées.

Dispositif expérimental

L'expérience a été réalisée avec une plateforme expérimentale équipée de sources placées sur un plateau situé devant le sujet (Figure 2).

16 sujets adultes naïfs rémunérés ont participé à l'expérience. Équipés d'un casque stéréo ouvert (modèle Sennheiser HD570), ils devaient pointer avec leur main dominante vers des stimuli (3 bursts de 40 ms) spatialisés avec des HRTFs non-individuelles mesurées sur un mannequin KEMAR. Afin d'adapter les sujets à la synthèse binaurale avec des HRTFs non-individuelles, l'expérience était précédée d'une phase d'adaptation perceptive réalisée en trois séances de 12 minutes dans une plateforme audio-kinesthésique selon la méthode décrite par [11].

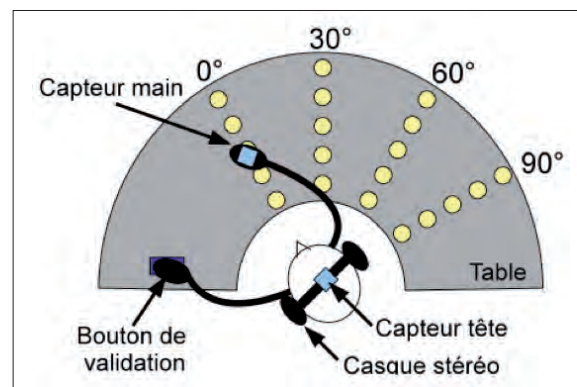


Fig. 2 : Schéma du dispositif expérimental avec le placement de chacune des sources

L'expérience était ensuite divisée en trois blocs correspondants à l'évaluation des performances avec les sons virtuels et à l'évaluation des deux métaphores de sonification mises en place. Pour chaque condition, les sujets devaient localiser 20 positions répétées quatre fois, réparties sur 4 azimuths (0, 30, 60 et 90°) et 5 distances.

Les dégradations dues à la synthèse binaurale

La figure 3 représente les positions moyennes perçues sur le plateau, reliées aux positions cible pour les sons réels (à gauche) et les sons virtuels (à droite).

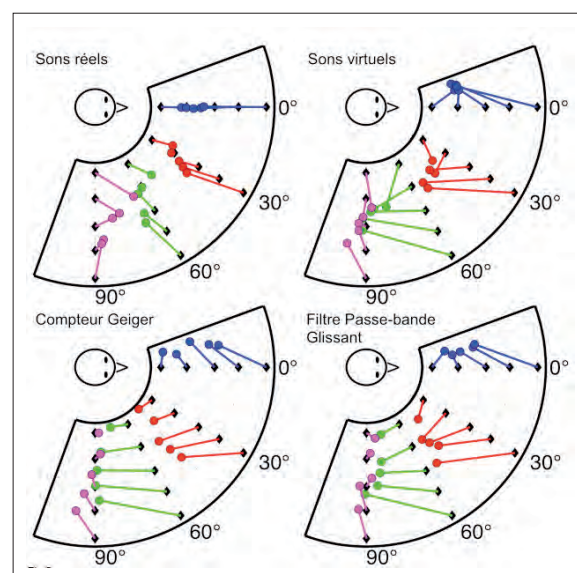


Fig. 3 : Représentation des positions moyennes pointées sur le plateau pour les sources réelles (en haut à gauche), virtuelles (en haut à droite) et pour les métaphores : Compteur Geiger (en bas à gauche) et Filtre Passe-bande Glissant (en bas à droite). Chaque position moyenne perçue est reliée à la position de la source correspondante (losange)

Dans le cas des sons réels, la moyenne d'erreur absolue globale en azimuth est de $9,3^\circ \pm 9,4^\circ$ alors que dans le cas des sons virtuels, elle est de $23,7^\circ \pm 17,5^\circ$. Cette dégradation des résultats n'est pas surprenante sachant que les HRTFs du mannequin KEMAR n'étaient pas individuelles et que les mesures étaient réalisées tous les 15° en azimuth.

Au niveau de la distance, l'erreur globale relative qui était d'environ 11,0% par rapport aux distances à localiser pour les sons réels passe à 13,5% pour les sons virtuels. Une régression linéaire sur l'ensemble des résultats en distance de la localisation des sons virtuels, donne un coefficient de régression de $0,16 \pm 0,09$ avec une qualité d'ajustement de 0,65 (contre $0,29 \pm 0,11$ et 0,95 pour les sons réels). Ce résultat montre qu'avec les sons virtuels, la distance n'est quasiment pas perçue par les sujets. Ce constat est confirmé par l'analyse de la position moyenne perçue sur le plateau représentée pour les deux conditions (Figure 3).

Amélioration des indices de perception de la distance

Dans un contexte de guidage en champ proche, la perception de la distance est très limitée comparée à la précision requise. Il est donc nécessaire d'utiliser d'autres indices acoustiques pour la perception de la distance. Notre approche consiste à simuler de nouveaux indices de distance en utilisant la sonification par *mapping* de paramètre (méthode qui consiste à représenter les variations des données à sonifier par des variations de paramètres acoustiques [4]). Avec ce concept, la distance est reliée à un ou plusieurs paramètres d'un effet audio appliqué au son. Celui-ci devient donc dépendant de la distance. Cette méthode permet la mise en place de plusieurs métaphores de sonification de la distance tout en laissant à l'utilisateur la possibilité de personnaliser les sons de l'interface. Elle a, en plus, l'avantage, une fois que la métaphore est comprise et apprise, de permettre à l'utilisateur de changer de son sans avoir à réapprendre le *mapping* de la sonification (les variations acoustiques dues au *mapping* de l'effet restent les mêmes). Deux métaphores de distance ont été développées à partir de ce concept [10,11]. La première, appelée «Compteur Geiger (CG)» consiste à répéter trois fois le stimulus et à faire varier l'intervalle temporel entre chaque répétition proportionnellement à la distance. La deuxième, appelée «Filtre Passe-bande Glissant (FPG)» consiste à appliquer au stimulus un filtre passe-bande dont la fréquence centrale et la bande passante, varient en fonction du temps. La fréquence centrale initiale du filtre est fixe alors que la fréquence finale augmente proportionnellement à la distance. Le choix de la polarité du *mapping* permet d'associer la distance physique à une «distance fréquentielle».

Les performances obtenues avec ces deux métaphores de sonification sont représentées dans la figure 3 avec les positions moyennes pointées sur le plateau. On remarque que ces deux métaphores n'ont pas affecté la perception en azimut et qu'elles ont permis d'augmenter significativement la perception en distance. La régression linéaire sur l'ensemble des résultats en distance donne un coefficient directeur de $0,64 \pm 0,29$ pour le CG (avec un coefficient de qualité r^2 de $0,96 \pm 0,05$) et de $0,50 \pm 0,20$ ($r^2 = 0,92 \pm 0,12$) pour le FPG contre $0,14 \pm 0,09$ ($r^2 = 0,66 \pm 0,26$) pour les sons virtuels sans sonification. Une ANOVA à mesures répétées, effectuée sur la moyenne des erreurs de distance, met en évidence un effet significatif du type de rendu ($F(2,36) = 19,76$, $p < 0,001$), avec une différence significative entre les conditions sans métaphore et GC ($p = 6,10^{-5}$) ainsi qu'entre les conditions sans métaphore et FPG ($p = 2.10^{-4}$).

Conclusion

Dans cet article, nous avons abordé plusieurs manières d'utiliser le son 3D pour l'aide aux non-voyants. Les travaux décrits montrent que les outils de génération d'environnement sonore virtuel ou augmenté décrivant acoustiquement un espace intérieur ou extérieur sont techniquement matures pour être utilisés comme des outils de recherche sur la cognition spatiale ou pour être utilisés dans des systèmes d'aide aux non-voyants. Dans certains cas comme le guidage en champ proche, il est néanmoins nécessaire d'augmenter les indices de localisation par l'utilisation de la sonification pour atteindre des performances acceptables. Ces différents outils nécessitent cependant une bonne connaissance de l'environnement (intérieur ou extérieur) qui est difficile à obtenir. Si un bon nombre d'informations utiles à la navigation en extérieure sont disponibles dans les services de la voirie des villes, leur récolte et leur implémentation dans un dispositif portable reste aujourd'hui un frein à la mise en place de ce type de système.

Remerciements

Ce travail a été soutenu en partie par l'Agence Nationale de Recherche (ANR) au travers du programme TecSan (projet NAVIG n° ANR-08-TECS-011) et la Région Midi-Pyrénées par le biais du programme APRRTT.

Références bibliographiques

- [1] Afonso A., Blum A., Katz B., Tarroux P., Borst G., and Denis M., Structural properties of spatial representations in blind people : Scanning images constructed from haptic exploration or from locomotion in a 3-D audio virtual environment. *Memory & Cognition*, 38:591–604, 2010.
- [2] Bach-y Rita P., Kaczmarek K. A., Tyler M. E., and Garcia-Lara J., Form perception with a 49-point electrostatic stimulus array on the tongue: a technical note. *Journal of Rehabilitation Research & Development*, 35(4):427, 1998.
- [3] Brungart D., Durlach N., and Rabinowitz W., Auditory localization of nearby sources. II. Localization of a broadband source. *J. Acoust. Soc. Am.*, 106(4):1956–1968, 1999.
- [4] Grond F. and Berger J., Parameter mapping sonification. In Hermann T., Hunt A., and Neuhoff J. G., editors, *The Sonification Handbook*. Logos Publishing House, 2011.
- [5] Katz B. and Picinali L., Advances in Sound Localization, chapter Spatial Audio Applied to Research with the Blind, pages 225–250. InTech. ISBN 978-953-307-224-1, 2011a.
- [6] Katz B. and Picinali L., Exploration of virtual acoustic room simulations by the visually impaired. In 1st International Seminar on Virtual Acoustics (ISVA), pages 169–177, 2011b.
- [7] Katz B., F. Kammoun S., Parseihian G., Gutierrez O., Brilhaut A., Auvray M., Truillet P., Denis M., Thorpe S. and Jouffrais C., NAVIG: augmented reality guidance system for the visually impaired. *Virtual Reality*, 16(4):253–269, 2012.
- [8] Kawai Y., Kobayashi M., Minagawa H., Miyakawa M., and Tomita F., A support system for visually impaired persons using three-dimensional virtual sound. In Int. Conf. Computers Helping People with special Needs (ICCHP 2000), 2000.
- [9] Loomis J. M., Golledge R. G., Klatzky R. L., Speigle J. M. and Tietz J., Personal guidance system for the visually impaired. In Proceedings of the first annual ACM conference on Assistive technologies, Assets '94, pages 85–91. ACM, 1994.
- [10] Parseihian G., Sonification binaurale pour l'aide à la navigation. PhD thesis, Université Paris 6. [Parseihian et al., 2012] Parseihian G., Conan S., and Katz B. (2012). Sound effect metaphors for near field distance sonification. In Proceedings of the 18th international conference on Auditory display (ICAD 2012), 2012.
- [11] Parseihian G. and Katz B., Morphocons: A new sonification concept based on morphological earcons. *J. Audio Eng. Soc.*, 60(6):409–418, 2012a.
- [12] Parseihian G. and Katz B., Rapid head-related transfer function adaptation using a virtual auditory environment. *J. Acoust. Soc. Am.*, 131(4), 2012b.
- [13] Wilson J., Walker B., Lindsay J., Cambias C. and Dellaert F., Swan: System for wearable audio navigation. In Proceedings of the 11th IEEE International Symposium on Wearable Computers, pages 1–8, Washington, DC, USA. IEEE Computer Society, 2007.