Techniques binaurales en contexte multimodal : aspects psychoacoustiques

Alan Blum E-mail : alan.blum@limsi.fr - Web site : www.limsi.fr, www.omnihead.com

Groupe « Perception Située » - Bâtiment 508, **LIMSI-CNRS**, BP133, 91402 Orsay Cedex 03 France

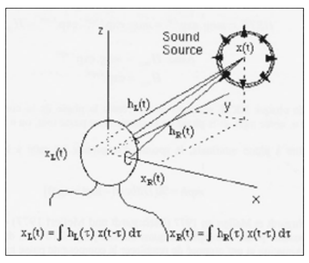
[1] INTRODUCTION AUX TECHNIQUES BINAURALES

Création d'un espace sonore tridimensionnel grâce au rendu binaural :

- Une technique élégante permettant un **rendu sonore 3D** avec seulement deux canaux audio : restitution sur **casque stéréophonique** standard.
- Utilisation du « décodage naturel » par le système auditif des indices acoustiques contenus dans les **HRTF** (Head Related Transfer Function - voir figures ci-contre).
- Principe applicable en **prise de son** pour la captation d'une scène sonore complexe comme en synthèse sonore pour la **spatialisation artificielle** d'un son.

Principales problématiques liées :

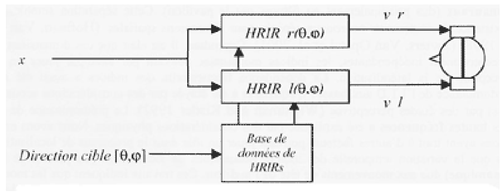
- Problème de l'**individualisation** du rendu binaural : Les HRTF dépendent de la morphologie de l'auditeur et sont par conséquent propres à l'individu.
- Interaction avec d'autres entrées sensorielles, notamment la **vision** et les informations proprioceptives (importance des **mouvements de tête** en localisation auditive).
- Aspects techniques (optimisation de l'implémentation - synthèse coûteuse en calcul, gestion d'effet de salle, dispositif de « tracking », intégration dans des plateformes multimodales).
- Evaluation des systèmes de rendu (test de localisation, questionnaire).



Ci-contre à droite : création d'une « source sonore virtuelle » par filtrage d'un signal audio monophonique avec les HRTF correspondant à la position voulue en azimut (θ) et élévation (φ) - coordonnées polaires.

Les HRTF sont les fonctions de transfert acoustiques entre une source sonore en un point de l'espace et chacun des tympans d'un auditeur.

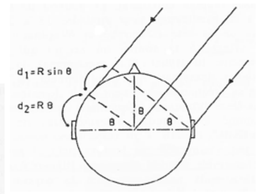
On peut en effectuer la mesure en chambre sourde en discrétisant la sphère auditive en un certain nombre de points où l'on enregistre la réponse impulsionnelle correspondante (HRIR pour Head Related Impulse Response - passage aux HRTF par Transformée de Fourier).



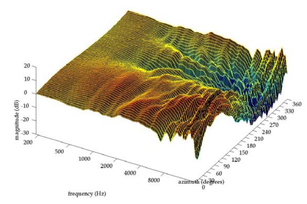
Décomposition des fonctions de transfert en sous-indices :

> **Indices interauraux** de temps (ITD) et de niveau (ILD)

> **Indices spectraux**



Principalement utilisés pour la **latéralisation** des sources



Principalement liés à la **sensation d'élévation**

[2] « ECOUTER AVEC LES OREILLES DE QUELQU'UN D'AUTRE »

Réalisé à l'IRCAM

avec Brian Katz et Olivier Warusfel

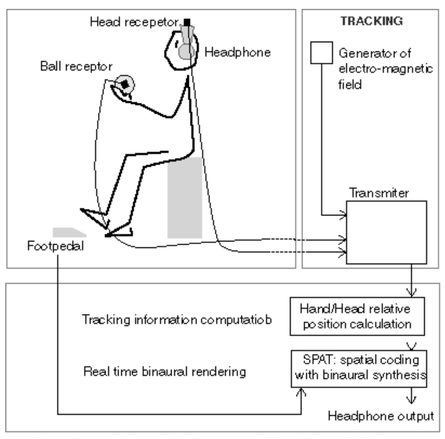
dans le cadre d'un DEA d'Acoustique à l'Université d'Aix-Marseille II

Contexte et Objectif :

Plusieurs approches existent pour essayer de résoudre le problème de l'individualisation de la synthèse binaurale, l'objectif étant de réussir à générer des HRTF personnalisées pour l'auditeur selon des critères plus simples que le passage par leur mesure directe, opération complexe et coûteuse. Dans l'étude présentée ici le problème est pris en sens inverse, la question étant de **savoir si un individu peut s'adapter à des HRTF qui ne sont pas les siennes**.

Dans une expérience par Hofman et al., les pavillons d'oreille des participants sont déformés à l'aide d'inserts. Les performances de localisation auditive sont alors très réduites mais reviennent progressivement à la normale avec le temps. Cette étude essaye d'accélérer une telle adaptation dans un contexte de spatialisation artificielle, en s'appuyant sur une idée de **calibrage du système auditif via un recoupement avec d'autres entrées sensorielles**.

Un dispositif d'entraînement est élaboré grâce à la mise en œuvre d'**une plateforme interactive auditivo-kinesthésique.** Le rendu sonore se fait avec des HRTF « étrangères » où la partie spectrale est non-individuelle mais l'ITD est personnalisé. Les participants ne sont pas au courant de l'objectif de l'expérience et se voient proposer une tâche d'exploration de la sphère auditive via un jeu de recherche de son d'animaux dans l'espace sonore. Une balle tenue dans leur main émet un son spatialisé alternatif (tic-tac) dont le rythme accélère à la proximité des cibles cherchées (type compteur Geiger).



Structure de la plateforme auditivo-kinesthésique

On calcule la position de la balle dans la main de l'auditeur par rapport au centre de sa tête grâce à deux capteurs de position à 6 degrés de liberté.

On spatialise un son à la position de la balle que l'auditeur peut manipuler. Celui-ci à donc la possibilité de contrôler le déplacement d'une source sonore virtuelle tout autour de lui.

A tout moment **les informations proprioceptives** (connaissance de la position de sa main) **sont synchrones avec la perception auditive**.

Méthode :

Toute l'expérience se fait en absence d'information visuelle (les participants ont les yeux bandés).

Paradigme en 3 phases :

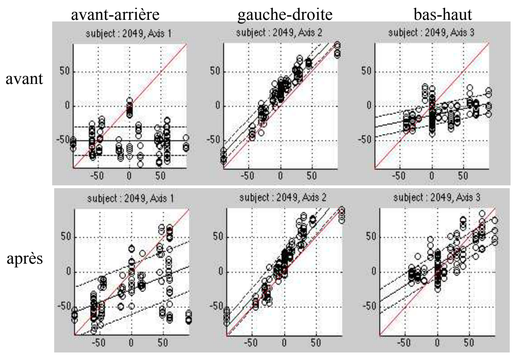
1. Evaluation des performances initiales (test de localisation)
2. Entraînement - jeu d'exploration (**durée limitée** à 12 min)
3. Deuxième évaluation des performances (même test qu'en 1.)

2 groupes de sujets :

- un groupe *test* avec des HRTF non-individuelles
- un groupe *contrôle* avec des HRTF individuelles
- 5 participants dans chaque groupe

Résultats :

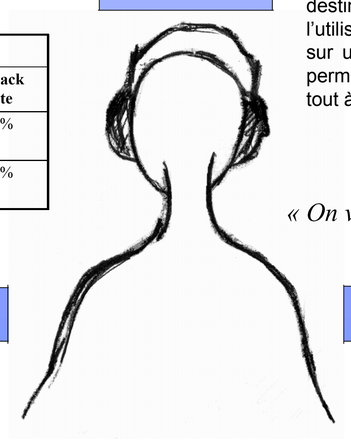
Les résultats montrent une amélioration des performances pour le pourcentage de confusions avant/arrière (artefact très courant en écoute binaurale sans mouvement de la tête) ainsi que pour la moyenne des erreurs angulaires en élévation. Comme attendu, aucune amélioration n'est constatée pour la latéralisation puisque les indices interauraux de temps (ITD) étaient individualisés. De plus, la progression du groupe *test* est significativement plus importante que celle du groupe *contrôle* (p = 0.034 pour l'erreur angulaire sphérique ; p = 0.038 pour l'erreur angulaire haut-bas).



Ces résultats semblent indiquer que l'amélioration est bien due à une adaptation des participants du groupe *test* aux HRTF non-individuelles et non à une simple habitude de la tâche de localisation.

Progression results summary — mean errors (before – after)	Spherical angle	Left/right angle	Up/down angle	Front/back conf. rate
Non-individual HRTF - test group -	6.1 °	2.8 °	5.9 °	9.3 %
Individual HRTF - control group -	0.4 °	0.7 °	0.1 °	2.2 %

Moyenne des erreurs angulaires avant-après selon différents axes d'observation.



« On vient au monde par les oreilles »

[3] DIFFUSION POUR UNE LARGE AUDIENCE DANS UN THEATRE

Réalisé par OmniHead

avec Antoine Gonot et Noël Chateau de FTR&D Lannion

Partenaires :
- Diacom Video
- Le Quartz
- France Culture
- La LILANN
- Carré Magique

Contexte et objectif :

L'association OmniHead s'intéresse depuis quelques années à la prise de son binaurale et à son utilisation dans des productions audio. Une première expérience de spectacle impliquant une diffusion binaurale pour un large **public** fut possible à l'occasion du festival de la radio et de l'écoute organisé par l'association Longueur d'onde à Brest. Une salle de théâtre (au Quartz de Brest) est entièrement équipée par plus de 300 casques sans fil pour l'enregistrement d'une fiction radiophonique par France Culture. L'intervention d'OmniHead consiste en une première partie d'environ 15 min mélangeant la diffusion d'une bande son préenregistrée à une captation en direct, et le jeu de 8 acteurs de la LILANN (Ligue d'Improvisation Lannionaise) se déplaçant dans toute la salle. A la fin de cet événement, un **questionnaire** est proposé aux spectateurs pour recueillir leurs impressions et évaluer l'impact de la spatialisation sonore dans ce type de condition.

Le but est de voir si le son 3D est correctement perçu en dehors d'un contexte très spécifique, sachant que l'enregistrement binaural peut paraître à ce point naturel que certains auditeurs ne comprennent pas ce qu'il contient de spécial. Comment est ressenti le port d'un casque dans une salle de spectacle ? Peut-il générer une gêne ou une sensation d'isolement ? L'immersion procurée par le rendu binaural peut-elle nuire à l'interaction avec le reste de la salle et les acteurs ? Avec une technique de prise de son centrée sur un point de vue (ou plutôt d'écoute), un spectateur excentré ressent-il des incohérences puisque son casque pourra lui faire entendre un son semblant provenir de derrière un mur par exemple ? Comment s'opère le mixage entre des sons provenant du "direct" et des sons préenregistrés ne présentant pas forcément lien de cause à effet avec une source émettrice identifiable ?

Prise de son autour du mannequin KEMAR pour la constitution de la bande son binaurale diffusée au cours du spectacle.

Le questionnaire comprenait 9 questions s'attachant à ces interrogations, avec à chaque fois 4 cases au choix à cocher en réponse à une affirmation de « Tout à fait » à « Pas du tout ».

Une autre partie avait pour rôle de réaliser un profil de l'auditeur selon une série de critères principalement liés aux habitudes d'écoutes :

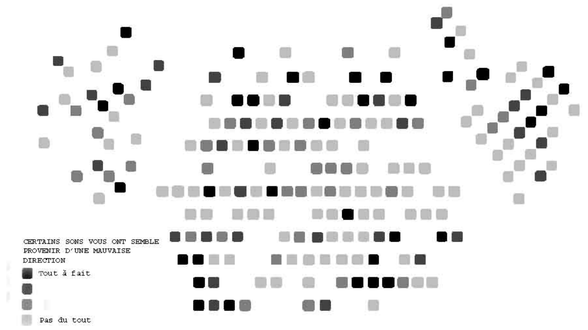
- Tranche d'âge.
- Fréquentation des théâtres.
- Heures d'écoute hebdomadaire de musique et de radio.
- Pratique d'un instrument ou non.
- Qualité de matériel audio possédé.
- Habitude d'écoute au casque.
- Habitude d'écoute de son spatialisé (système 5.1).

Résultats :

Profil moyen des participants :							
Age: 35-45	Radio: 3h-10h	Musique: 3h-10h	Th%tre: Peu	Instrument: Peu	Syst, me: minicha"ne	Support: CD	Casque: Peu

Une question à réponse binaire s'assurait également de savoir si le spectateur avait ressenti l'effet 3D dans le son ou non. Sur **215 personnes** ayant répondu au questionnaire, 212 ont ressenti cet effet, soit 98,6 % des participants.

L'analyse des données fait ressortir en premier lieu des réponses **extrêmement homogènes**. La nouveauté de l'événement, et donc l'absence de référence disponible, pourrait faire partie des explications. En excluant une question dont le détail des réponses est donné sur la figure ci-contre, on observe des réponses unanimes à 70 % en moyenne, avec un écart-type très petit de 1,3 %, allant toujours dans le sens d'une appréciation positive des points abordés. On peut estimer que les questions posées étaient assez « prudentes » car destinées à s'assurer de certaines bases concernant l'utilisation d'une spatialisation sonore individuelle sur une large audience. Elles semblent néanmoins permettre de conclure qu'une diffusion binaurale est tout à fait envisageable dans ce type de contexte.



La figure ci-dessus montre les réponses en niveau de gris à la question « Certains sons vous ont semblé provenir d'une mauvaise direction ? » en fonction de la position dans la salle. Les valeurs sont plus réparties que pour les autres questions ce qui semble indiquer que la spatialisation du son était parfois perçue comme surprenante. En revanche, **cette appréciation est indépendante de la position dans la salle** (test du chi-2), même en prenant les places situées aux extrémités de la salle qui pouvaient a priori être considérées comme défavorisées.

[4] ETUDE SUR LA COGNITION SPATIALE EN REALITE VIRTUELLE IMMERSIVE AUDIO

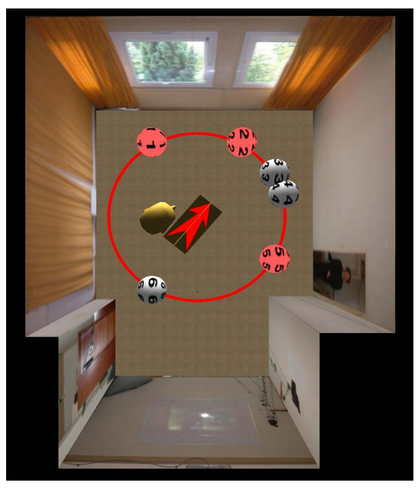
Réalisé au LIMSI-CNRS avec Amandine Afonso, Brian Katz, Christian Jacquemin et Michel Denis

Contexte et Objectif :

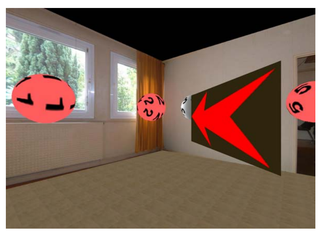
La réalité virtuelle offre des possibilités très riches pour la **psychologie expérimentale** en permettant l'élaboration d'environnements complexes, incluant plusieurs modalités sensorielles tout en restant facilement manipulable.

La plateforme élaborée ici intègre des outils de modélisation géométrique 3D, de rendu visuel, de rendu sonore 3D, et des périphériques de captation de mouvement. Le but de sa mise en œuvre est une expérimentation sur la **représentation cognitive de l'espace** dans le cadre de la thèse d'Amandine Afonso, où le participant à la possibilité de se déplacer et d'interagir au sein d'une scène sonore 3D.

L'étude cherche à observer l'**influence de la vision** dans la construction de représentations mentales de l'espace. Elle implique trois populations de participants : non-voyants de naissance, non-voyants tardifs et voyants privés de vision par un bandeau sur les yeux (54 personnes au total). L'hypothèse évaluée est celle de stratégies différentes, selon que la personne bénéficie de la vue (représentation mentale de nature iconique) ou en est définitivement privée (**mémorisation de contingences sensorimotrices**).

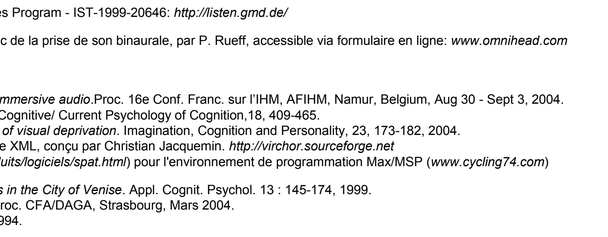


L'image ci-contre représente la scène expérimentale vue de haut. L'expérience à lieu dans une salle de 3x4 mètres environ, dans laquelle le participant (tête en jaune) peut se déplacer. Un son de vent spatialisé prévient de la présence d'un mur si l'on vient à s'en approcher de trop près. La consigne est d'**apprendre la configuration d'une scène sonore** constituée de 6 sources (sphères numérotées) disposées sur un cercle de 1,5 mètre de diamètre.



Deux modes d'apprentissage sont contrastés, celui-ci pouvant se faire soit via une **description verbale** avec un codage horaire, soit via une **exploration active** de l'environnement. Après cette phase, le participant est amené à interagir avec la scène à l'aide d'un pointeur (flèche rouge). Dans un premier temps, il lui est demandé de reconstruire la scène, ensuite il s'agit de détecter des modifications d'ordre topologique et/ou métrique de l'environnement.

L'expérimentateur dispose d'un retour visuel grâce au rendu graphique du modèle géométrique (on voit ici à gauche le point de vue subjectif correspondant à la position du participant dans la scène - une sphère est rouge quand elle émet du son). Tous les mouvements et actions sont également enregistrés, ce qui permet de rejouer la scène ultérieurement. Cette fonctionnalité doit être utile à l'analyse des résultats pour laquelle on désire observer les stratégies adoptées par les sujets pour s'orienter dans l'espace.



Le dispositif utilise une **architecture distribuée** sur plusieurs machines afin de répartir les tâches et leur coût de calcul élevé pour assurer le rendu « temps réel ». Les communications se font avec le format OSC (Open Sound Control) et le protocole réseau UDP.

Le rendu sonore binaural est réalisé sous *Max/MSP* avec le *Spat*. Une phase préliminaire permettait aux participants de choisir des HRTF leur convenant au mieux dans une base de données de 60 têtes mesurées.

Le modèle géométrique de la scène et le rendu graphique sont effectués grâce au logiciel *Virtual Choreographer*.

[5] PERSPECTIVES

Thèse au LIMSI-CNRS dans le cadre d'un projet européen STREP (Specific Targeted Research Project) nommé « Wayfinding » sous la direction de Brian Katz et de Michel Denis. Thématiques :

- Navigation au sein d'environnements virtuels audio.
- Acquisition de scènes spatiales via la modalité auditive.
- Travail sur la perception de l'espace chez les personnes non-voyantes.

Sous-projets de recherche :

- Influence des micro-mouvements de la tête dans la localisation auditive. Anticipation des mouvements par synchronisation d'événements visuels et sonores.
- Recherche des paramètres acoustiques utiles au repérage spatial dans une scène sonore.
- Réitération de l'expérience de diffusion binaurale sur une large audience avec un nouveau spectacle en salle, de plus grande durée.
- Poursuite de l'étude sur l'adaptation à des HRTF non-individuelles (expérience avec plus de sujets, investigation sur des aspects temporels de l'adaptation, autres méthodes d'entraînement).

BIBLIOGRAPHIE

[1]
- J. Blauert. *Spatial Hearing, the psychophysics of human sound localization*. MIT Press, Cambridge, 1996.
- D.R. Begault. *3-D Sound for Virtual Reality and Multimedia*. Academic Press, Cambridge, MA, 1994.
- B.F.G. Katz. *Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation*. J Acoust Soc Am, vol. 110(5): 2240-2248, 2001.
- R. Algazi, C.Avendano & R.O. Duda. *Estimation of a spherical-head model from anthropometry*. J. Audio Eng. Soc., vol. 49, no 6, pp 472-479, 2001.
- V. Larcher. *Techniques de spatialisation des sons pour la réalité virtuelle*. PhD Thesis, Université Pierre & Marie Curie, IRCAM, Paris, 2001.
- F.L. Wigthman & D.J. Kistler. *Headphone simulation of free-field listening. II: Psychophysical validation*. J Acoust Soc Am, 85: 868–878, 1989.
- F.L. Wightman & D.J. Kistler. *Resolution of front-back ambiguity in spatial hearing by listener and source movement*. J Acoust Soc Am, 105(5): 2841–2853, 1999.

[2]
- A. Blum, B.F.G. Katz, & O. Warusfel. *Eliciting adaptation to non-individual HRTF spectral cues with multi-modal training*. Proc. CFA/DAGA, Strasbourg, Mars 2004.
- P. Hofman, J.A.G. Riswick, & A.J. *Van Opstal. Relearning sound localization with new ears*. Nature Neuroscience, 1(5), 1998.
- P. Zahorik, C. Tam, K. Wang, P. Bangayan, & V. Sundareswaran. *Localization accuracy in 3-D sound displays : the role of visual-feedback training*. Adv. Displays Consortium: ARL's 5th Fed. Lab. Ann. Symp. 17–22, 2001.
- G.H. Recanzone. *Rapidly induced auditory plasticity : The ventriloquism aftereffect*. Proc. Natl. Acad. Sci. v.95, 869–875, Feb. 1998.
- J. Lewald. *Vertical sound localization in blind humans*. Neuropsychologia, 40 :1868–1872, 2002.
- LISTEN HRTF Database : *http://www.ircam.fr/equipes/salles/listen/* Listen Project – Information Society Technologies Program - IST-1999-20646: *http://listen.gmd.de/*

[3]
- OmniHead - Groupe de recherche sur les techniques de reproduction de l'espace sonore tridimensionnel. Livre Blanc de la prise de son binaurale, par P. Rueff, accessible via formulaire en ligne: *www.omnihead.com*
- Festival de la radio et de l'écoute par l'association Longueur d'ondes. *http://longueur.ondes.free.fr/*
- M. Chion. *Le son au cinéma*. Editions de l'Etoile / Cahiers du cinéma, Collection Essais, Paris, 1994.

[4]
- A. Blum, A. Afonso, B.F.G. Katz, & C. Jacquemin. *Expérimentation sur la perception de l'espace en réalité virtuelle immersive audio*.Proc. 16e Conf. Franc. sur l'IHM, AFIHM, Namur, Belgium, Aug 30 - Sept 3, 2004.
- Denis, M. & Kosslyn, S. M. (1999). *Scanning visual mental images: A window on the mind*. Cahiers de Psychologie Cognitive/ Current Psychology of Cognition,18, 409-465.
- A. Afonso, F. Gaunet & M. Denis. *The mental comparison of distances in a verbally described spatial layout: Effects of visual deprivation*. Imagination, Cognition and Personality, 23, 173-182, 2004.
- *Virtual Choreographer* : Outil de modélisation et d'animation d'objets graphiques et sonores, avec un langage de type XML, conçu par Christian Jacquemin. *http://virchor.sourceforge.net*
- *Spat* : Librairies pour la création d'effet de salle et la spatialisation sonore développée à l'IRCAM (*www.ircam.fr/produits/logiciels/spat.html*) pour l'environnement de programmation Max/MSP (*www.cycling74.com*)

[5]
- M. Denis, F. Pazzaglia, C. Cornoldi & L. Bertolo. *Spatial Discourse an d Navigation : An Analysis of Route Directions in the City of Venise*. Appl. Cognit. Psychol. 13 : 145-174, 1999.
- I. Viaud-Delmon, L. Sarlat & O. Warusfel. *Virtual Ventriloquism: Localization of Auditory Sources in Virtual Reality*. Proc. CFA/DAGA, Strasbourg, Mars 2004.
- S. McAdams & E. Bigand. *Penser les sons. Psychologie cognitive de l'audition*. Presses Universitaires de France. 1994.