



HAL
open science

Environnement de simulation pour l'évaluation psychoacoustique des systèmes de prise et de restitution du son dans un contexte de téléconférence

M. Marin, A. Gilloire, J. Lacoume, O. Warusfel, J. Jot

► **To cite this version:**

M. Marin, A. Gilloire, J. Lacoume, O. Warusfel, J. Jot. Environnement de simulation pour l'évaluation psychoacoustique des systèmes de prise et de restitution du son dans un contexte de téléconférence. Journal de Physique IV Proceedings, 1994, 04 (C5), pp.C7-233-C7-236. 10.1051/jp4:1994546 . jpa-00253041

HAL Id: jpa-00253041

<https://hal.science/jpa-00253041v1>

Submitted on 4 Feb 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Environnement de simulation pour l'évaluation psychoacoustique des systèmes de prise et de restitution du son dans un contexte de téléconférence

M. MARIN, A. GILLOIRE, J.F. LACOUME*, O. WARUSFEL* et J.M. JOT**

Centre National d'Etudes des Télécommunications, LAA/TSS/CMC, Route de Trégastel, BP. 40, 22301 Lannion, France

** IRCAM, 1 place Stravinsky, 75004 Paris, France*

*** Espaces Nouveaux, 56 Boulevard Davout, 75020 Paris, France*

Abstract : A simulation environment is used for the perceptual evaluation of sound pickup and reproduction techniques in a teleconference context. A project, named "Spatialisateur", developed by Ircam and Espaces Nouveaux aims at synthesizing virtual room effects, controlling the acoustical quality and the localization aspects in real time. The prototype generates a localization effect by use of binaural reproduction techniques, for a single listener with headphones or a loudspeaker pair. Binaural synthesis relies on the use of head-related transfer functions (HRTF), which realize a listener-specific encoding of localization cues for any position of the sound source. Parametric modelling techniques are used in order to reduce the computational cost of the binaural filtering process. Psychoacoustical tests have been undertaken to validate this prototype by comparing, for various configurations, the localization judgements of subjects in actual stereophonic listening and in headphone simulation of the stereophonic reproduction.

1. INTRODUCTION

Le développement futur des systèmes de téléconférence est conditionné par l'amélioration de la qualité sonore de ces systèmes. En particulier, il est souhaitable de restituer la localisation auditive des locuteurs de la salle distante afin de se rapprocher des conditions naturelles d'écoute.

Pour apporter des solutions techniques appropriées, il est important de répondre aux exigences psycho-acoustiques régissant notre perception auditive. Cela implique de connaître et de contrôler ces paramètres subjectifs tout en les reliant à des critères objectifs. Un environnement de simulation psycho-expérimental reproduisant diverses configurations acoustiques de visioconférence permettrait, d'une part, de définir les différents attributs perceptifs pertinents pour la téléconférence et, d'autre part, d'étudier l'influence des paramètres objectifs liés aux systèmes de prise ou de restitution du son sur la localisation subjective.

Le projet "Spatialisateur", développé par l'Ircam et Espaces Nouveaux, permet de simuler, en temps réel, différentes qualités acoustiques et de gérer la localisation de sources virtuelles. Dans le cas d'une reproduction sur casque ou sur une paire d'enceintes, la localisation est restituée en utilisant les techniques de synthèse binaurale, nécessitant l'implémentation des HRTF par filtrage numérique. Ces techniques ne sont pas adaptées pour une diffusion destinée à une audience étendue : le champ sonore n'est recréé que pour un auditeur unique, et, de plus, les HRTF présentent une grande disparité inter-individuelle. Cependant, dans le cadre d'études psycho-expérimentales, elles permettent de simuler des environnements complexes sur un dispositif de reproduction réduit, à condition d'effectuer les mesures de HRTF pour chaque sujet.

La validation du "Spatialisateur", utilisé comme environnement de tests, a été réalisée en comparant les jugements de localisation donnés en situation d'écoute réelle d'un dispositif stéréophonique conventionnel et ceux obtenus par simulation sur casque de ce même dispositif.

2. SYNTHÈSE DE LA LOCALISATION ET MESURE DES HRTF

La synthèse binaurale repose sur l'hypothèse qu'une reproduction fidèle sera obtenue si on reconstruit le signal de pression capté par les tympans. La simulation d'un environnement acoustique donné peut être réalisée par convolution avec les réponses impulsionnelles mesurées au niveau du tympan dans ce même environnement. Pour ce qui est de la localisation, elle est restituée en utilisant des réponses (que l'on nomme alors HRTF), mesurées en champ libre dans le cas d'un événement sonore placé dans une position voulue [1]. Les signaux droit et gauche issus de cette convolution sont alors utilisables pour une écoute sur casque au même titre que des enregistrements par tête artificielle.

Il est reconnu que les performances de localisation réalisées avec des HRTF personnalisées sont meilleures que celles obtenues avec des HRTF universelles, mesurées sur une tête artificielle ou obtenues par moyennage de mesures effectuées sur différents sujets. Nous avons donc constitué notre propre base de données de HRTF, en les mesurant sur une vingtaine de sujets au CNET et à l'Ircam, afin de minimiser les biais des résultats des expériences psycho-acoustiques réalisées par la suite.

Pour chaque sujet, 48 mesures ont été réalisées pour différents azimuts et élévations : l'ensemble du plan horizontal a été exploré par pas de 15° pour les élévations 0° et 30°. Les réponses impulsionnelles ont été mesurées en chambre sourde (échantillonnées à 48 kHz) à partir d'un haut-parleur à deux voies concentriques, excité par des séquences pseudo-aléatoires. Le sujet était assis avec des microphones miniatures à électret directement insérés dans le conduit auditif [2]. Dans l'optique d'une reproduction ultérieure sur casque, il a été nécessaire de mesurer, pour chaque oreille, la fonction de transfert entre l'écouteur et le microphone placé à la même position dans le conduit que lors des mesures précédentes.

3. MODELISATION DES HRTF

Une implémentation en temps réel des filtres binauraux sous la forme de filtres FIR, par convolution directe avec les coefficients de la réponse impulsionnelle, nécessite une puissance de calcul élevée (une longueur de 5 ms requiert environ 250 coefficients à 44,1 kHz). L'objectif, pour diminuer le coût de calcul, est de réduire l'information contenue dans les HRTF sans perdre pour autant les caractéristiques indispensables à la localisation, de façon à ce qu'aucune différence perceptible ne soit décelée entre le modèle et son original (tant du point de vue du timbre que de la localisation). A cette fin, une modélisation paramétrique a été effectuée sur chaque réponse impulsionnelle, fournissant un filtre récursif équivalent. La modélisation s'appuie sur deux propriétés [3]:

- une propriété générale des filtres rationnels stables, qui permet de décomposer tout système en un filtre à phase minimale associé, en série, avec un filtre passe-tout modélisant l'excès de phase;
- une propriété particulière des HRTF, pour lesquelles l'excès de phase est linéaire en fonction de la fréquence (jusqu'au voisinage de 8 kHz), ce qui rend le filtre passe-tout équivalent à un retard pur. La fonction d'excès de phase est obtenue par la différence entre la phase et la phase minimale de la réponse et le retard pur est estimé par régression linéaire à partir de l'excès de phase (fig.1).

Avant de synthétiser, par une méthode de modélisation paramétrique, le filtre à phase minimale, deux traitements sont préalablement appliqués au module de la réponse en fréquence afin d'éliminer la structure fine des réponses tout en respectant la résolution fréquentielle de l'oreille [4]. Le premier consiste à effectuer un lissage perceptif du module de la réponse fréquentielle selon les bandes critiques de l'oreille, dont la largeur croît avec la fréquence. Dans l'étape suivante, on utilise une transformation bilinéaire, afin d'améliorer la modélisation paramétrique en basses fréquences, en se rapprochant de l'échelle des Barks [4]. La modélisation paramétrique consiste à résoudre les équations de Yule-Walker (fonction Yulewalk dans l'environnement Matlab) (fig.2). Après une comparaison auditive avec l'implémentation sous forme de filtres FIR, il a été jugé qu'un filtre IIR d'ordre 20 suffisait pour reproduire correctement les caractéristiques de localisation. Il est envisagé de préciser ce point par une évaluation systématique.

4. IMPLEMENTATION ET ENVIRONNEMENT DE SIMULATION

Pour l'implémentation en temps réel des filtres IIR modélisant les filtres binauraux, on utilise une association en cascade de cellules d'ordre 2, l'appariement des pôles et des zéros et l'ordre de rangement des cellules étant réalisés de façon à minimiser le bruit de calcul global [5]. Une procédure d'interpolation permet de synthétiser n'importe quel azimut à partir de deux mesures effectuées pour des azimuts voisins. Le déplacement continu des sources virtuelles devient alors possible.

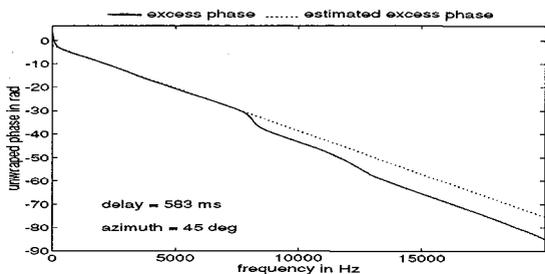


Fig.1 : modélisation de la phase.

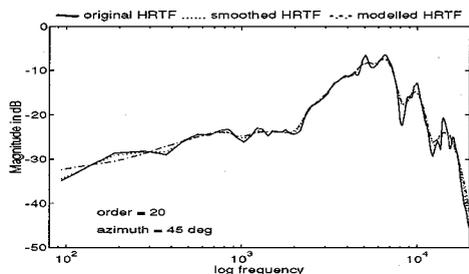


Fig.2 : modélisation de l'amplitude.

L'égalisation du casque est réalisée par la mise en série des filtres binauraux et du filtre inverse du casque, lui aussi réalisé sous la forme de filtres IIR d'ordre 20.

La simulation en temps réel est réalisée sur une station de travail Next Cube munie d'une carte de traitement de signal ISPW (deux micro-processeurs Intel i860) travaillant à la fréquence d'échantillonnage de 44.1 kHz. Les modules de traitement du signal et l'interface de commande sont développés grâce au logiciel Max (environnement graphique orienté objet).

5. VALIDATION DE L'ENVIRONNEMENT DE SIMULATION

Pour obtenir des indications sur les biais éventuels apportés par le "Spatialisateur", différentes étapes de validation sont menées selon une approche structurée consistant à isoler un paramètre, la localisation dans le cas présent, et d'étudier si cette dimension est traduite de manière fidèle par le "Spatialisateur". La validation, décrite ci-dessous (fig.3), consiste à comparer les jugements de localisation donnés en situation d'écoute stéréophonique réelle sur haut-parleurs à ceux obtenus par simulation sur casque [6]. Comme l'indique la fig.3, la prise de son est simulée lors de l'expérience en introduisant des différences de niveau entre les deux canaux (couple XY), auxquelles s'ajoute un décalage temporel dans le cas d'un couple non coïncident (couple AB ORTF). Les signaux de test utilisés sont des EMDIFS [7], répartis de 125 à 8000 Hz, ainsi que deux signaux de parole anéchoïques (une voix masculine et une voix féminine).

Dans un premier temps, l'environnement acoustique choisi est anéchoïque. Dans une étape ultérieure, l'expérience sera reproduite dans un environnement de type salle d'audioconférence. Le "Spatialisateur" permet en effet de synthétiser la distribution spatiale et temporelle d'un effet de salle, en appliquant, en particulier, un traitement binaural simplifié aux premières réflexions et à la réverbération [8].

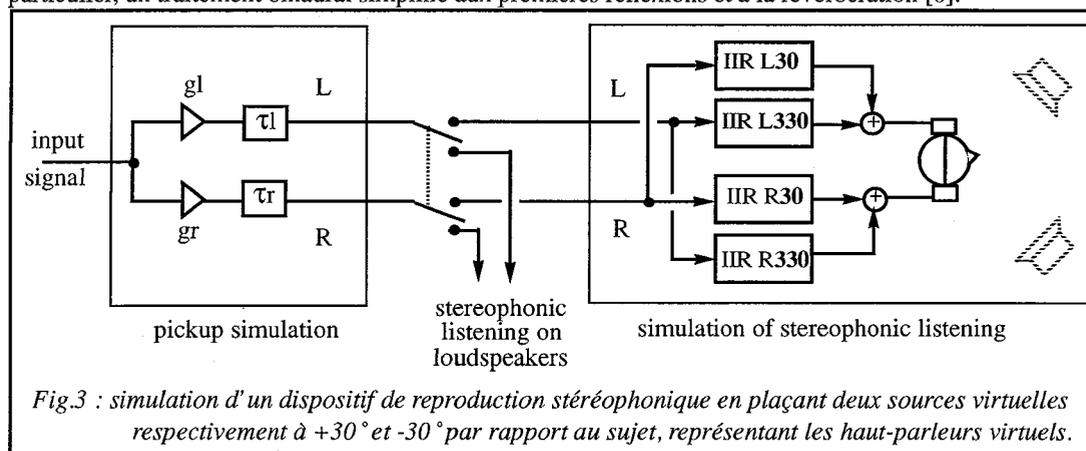


Fig.3 : simulation d'un dispositif de reproduction stéréophonique en plaçant deux sources virtuelles respectivement à $+30^\circ$ et -30° par rapport au sujet, représentant les haut-parleurs virtuels.

6. RESULTATS ET CONCLUSION

La restitution de la latéralisation a été analysée du point de vue des biais éventuels et de la dispersion des résultats entre les sujets, après correction des réponses à l'arrière (par symétrie par rapport à l'axe des oreilles). Pour la latéralisation du couple XY, il apparaît une bonne adéquation entre la simulation et la réalité. Les biais sont faibles et la dispersion comparable (fig.4).

On retrouve les propriétés de latéralisation du couple XY en fonction de la fréquence telles qu'elles sont décrites, par exemple, dans [7]. En ce qui concerne le couple AB ORTF, on constate dans l'expérience sur haut-parleurs une latéralisation plus fidèle à l'angle de prise de son que pour le couple XY. Cette propriété est liée à l'intervention d'indices temporels fournis par les couples non coïncidents. On note cependant une dispersion plus large des réponses des sujets qui peut s'interpréter par le fait que la relation entre indices temporels et indices énergétiques réalisée par le couple AB n'est pas celle de l'écoute naturelle. La concomitance de deux types d'indices rend les performances de localisation plus dépendantes de l'individu et de sa position par rapport aux haut-parleurs. On constate une dispersion comparable lors de l'expérience sur écouteurs (notons que dans ce cas, les différences de positionnement des sujets par rapport aux haut-parleurs se retrouvent dans l'étape de mesure des HRTF).

Pour le couple AB simulé sur écouteurs, l'analyse montre cependant une surestimation de la latéralisation, ainsi qu'une forte dispersion, pour la voix féminine et les signaux centrés sur 1 kHz. Ce phénomène n'apparaît que pour de fortes latéralisations, et peut s'interpréter par l'impossibilité, pour le sujet, de préciser son jugement de localisation par des mouvements de la tête. Dans ce cas, certains sujets ont eu tendance à indiquer une latéralisation maximale ($\pm 90^\circ$). Par ailleurs, comme l'illustre la fig.5, la fréquence des inversions avant-arrière, non négligeable en écoute anéchoïque réelle, est plus élevée encore pour la simulation sur écouteurs. L'analyse montre que cette différence entre simulation binaurale et réalité est plus grande pour la direction frontale et les latéralisations fortes, stimuli susceptibles de provoquer un mouvement réflexe du sujet pour conforter son jugement auditif.

Les conclusions de cette expérience suggèrent donc que les performances du système de simulation, dans la configuration étudiée, seraient rendues satisfaisantes par l'utilisation d'un capteur de position de la tête, permettant un asservissement du traitement binaural relativement à l'orientation du sujet.

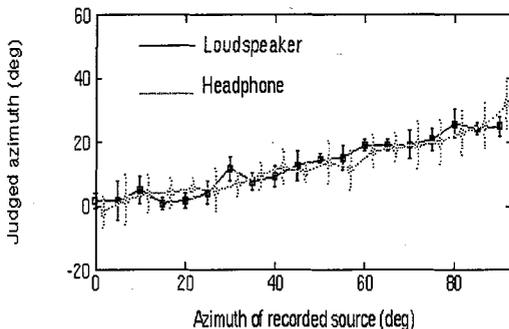


Fig.4 : comparaison des jugements de localisation (voix féminine, couple XY).

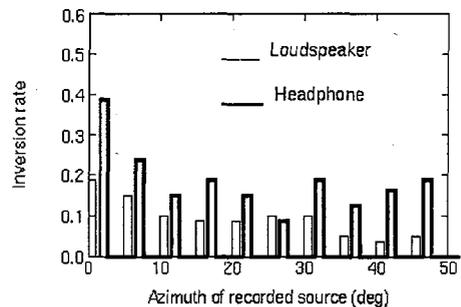


Fig.5 : comparaison du taux d'inversion (couple AB ORTF).

BIBLIOGRAPHIE :

- [1] Blauert J., "Spatial hearing : the psychophysics of human sound localization", Cambridge MIT Press 1983.
- [2] Marin M., "Etude de la localisation en restitution des sons. Application à l'amélioration des systèmes de téléconférence", rapport CNET/LAA/TSS/CMC, 1993.
- [3] Poncet F., "Simulation de localisation de sources sonores dans l'espace", rapport Télécom Paris, Dpt Signal, 1992.
- [4] Smith J., "Techniques for digital filter design and system identification with application to the violin", PhD thesis, CCRMA, Dept of Music, Stanford Univ., June 1983.
- [5] Bellanger M., "Traitement numérique du signal : théorie et pratique" CNET-ENST, Masson, Paris, 2ème édition, 1984.
- [6] Lacoume J.F., "Test d'un environnement de simulation de situations d'écoute binaurale sur des signaux stéréophoniques", rapport CNET/IRCAM, 1993.
- [7] Hugonnet C., "Restitution spatiale des différents systèmes stéréophoniques. Caractérisation objective et subjective", mémoire d'ingénieur CNAM, Paris, 1987.
- [8] Jot J.M., Warusfel O., Kahle E., Mein M., "Binaural concert hall simulation in real time", IEEE digital audio workshop, Mohonk, Oct. 1993.