10ème **C**ongrès **F**rançais d'**A**coustique

Lyon, 12-16 Avril 2010

Evaluation perceptive des apports de la reproduction sonore par WFS pour une situation de concert

Joseph Sanson[1], Olivier Warusfel[1]

[1] IRCAM, 1 place Igor Stravinsky, 75004 PARIS, joseph.sanson@ircam.fr

Nous présentons dans cet article deux tests d'écoute évaluant la reproduction de scènes sonores en Wave Field Synthesis (WFS). Si la majorité des études spatiales s'intéressent à la localisation, nous nous attachons ici à évaluer d'autres caractéristiques de la reproduction sonore : la synthèse de scènes sonores complexes et la confrontation sur scène entre sources virtuelles et sources réelles.
La première expérience s'attache à la discrimination de sources concurrentes faiblement espacées et prend la forme d'un test AB. Une scène constituée de trois locuteurs espacés de 8° dans le plan horizontal est spatialisée par un système WFS ou par un système stéréo. Entre la scène A et la scène B, les positions de deux des trois sources sont éventuellement interchangées. La tâche des participants est d'indiquer s'ils ont perçu un changement d'organisation spatiale entre les scènes A et B. Les résultats confirment la meilleure discrimination des sources dans le cadre de la WFS.
La deuxième expérience s'attache au réalisme de la reproduction d'une voix parlée par un système de diffusion WFS. Le protocole consiste à alterner de manière aléatoire la déclamation de fragments d'un texte par un acteur et la diffusion de fragments de ce même texte préalablement enregistrés par le même acteur. La tâche des participants est d'indiquer au cours de la lecture quelles sont les parties déclamées par l'acteur ou diffusées par le système de reproduction. Dans ce test, nous portons une attention particulière à la directivité des sources virtuelles en WFS comme facteur susceptible d'influer sur le réalisme. Nous comparons notamment les mérites d'une diffusion avec une directivité omnidirectionnelle, directive ou asservie aux caractéristiques du signal.
Une attention particulière est donnée aux problématiques liées à la mise en place de ces deux tests.

1 Introduction

Les systèmes de spatialisation sonore occupent aujourd'hui une place importante dans beaucoup de manifestations artistiques telles que le théâtre, le cinéma, la musique mixte et électro-acoustique, les installations sonores. Les modes d'utilisation sont alors variés, allant de l'amplification ou le renforcement du son à la création de paysages sonores. Les scènes sonores ainsi créées comprennent souvent plusieurs sources sonores virtuelles interagissant entre elles ou avec des sources acoustiques réelles.

La technique Wave Field Synthesis [1] semble particulièrement bien adaptée à une utilisation incluant une large audience car elle permet une grande zone d'écoute à l'intérieur de laquelle les indices de localisation sont correctement reproduits. Elle permet également un contrôle des caractéristiques de directivité des sources virtuelles [15, 4, 5, 2]. Les études perceptives de la diffusion par WFS ont souvent concerné la précision de localisation des sources sonores [16, 13, 17, 12], [6] dans le cas d'interactions audio-visuelles. Si dans l'ensemble ces études rapportent une acuité de localisation des sources virtuelles comparable au cas d'une source réelle, des effets de coloration ou de différence de site sont également reportés.

La plupart de ces études se sont concentrées sur l'étude d'un seul flux audio, qui est une situation réductrice en regard des applications citées précédemment. Une scène sonore comprend souvent un ensemble de flux audio alternés ou concurrents dont la perception a été étudiée pour des systèmes stéréophoniques [14, 10, 9]. Il est donc intéressant d' évaluer l'influence des caractéristiques de la WFS (reconstruction du front d'ondes, directivité des sources sonores etc) sur la perception de scènes sonores comprenant des flux audio concurrents.

De plus la localisation ne rend pas compte de l'interaction d'une source virtuelle et d'une source acoustique réelle. ici se pose la question du réalisme du système de reproduction qui n'est pas seulement la conjonction des aspects timbraux (réponse fréquentielle, niveaux de distortion...) et spatiaux (taille de la source perçue, sensation d'enveloppement...) mais également l'interaction de ces deux caractéristiques [11].

Plusieurs écoutes informelles ont été conduites à l'IRCAM incluant les cas de figure cités ci-dessus et ont clairement montré un apport de la WFS par rapport à des systèmes de reproduction classiques. En particulier, les sources virtuelles WFS ont montré des similitudes frappantes avec les sources réelles. Il nous a donc semblé intéressant d'objectiver ces impressions par des tests d'écoute et ainsi de tenter de déterminer les caractéristiques déterminantes de la technologie WFS.

L'intention de cet article est de rapporter les résultats de deux tests exploratoires visant à étudier des caractéristiques de l'écoute en WFS. Le premier s'at-

tache à la discrimination de flux audio concurrents en comparant les performances dans le cas de diffusion par un système WFS et un système stéréophonique. Le second évalue le réalisme de la reproduction en WFS en comparant un acteur lisant un texte avec sa reproduction.

2 Discrimination de sources concurrentes en WFS

L'expérience présentée ici évalue la discrimination de flux concurrents spatialisés en WFS et en stéréo. L'utilisation de ces deux systèmes permet d'évaluer l'influence des caractéristiques de la WFS (reconstruction du front d'ondes, manipulation de la directivité des sources virtuelles).

Par extension des précédentes expériences en WFS, le protocole présenté se base sur la localisation des sources, qui n'est cependant pas l'objectif du test. Ainsi, la tâche demandée est de repérer des changements d'organisation spatiale entre les deux scènes d'un test de type AB.

Les scènes sonores sont composées de 3 flux concurrents car les performances de ségrégation semblent se dégrader fortement à partir de ce nombre [14]. Des écoutes informelles ont montré que des scènes composées de 4 sources étaient trop complexes pour être utilisées.

De sorte à rapprocher les résultats de ceux de [9], les sources virtuelles ont été placées à des positions espacées de 8° (approximation de l'angle minimum audible pour des sources concurrentes).

2.1 Méthode et procédure

Matériel Le système WFS considéré ici est constitué de 80 haut-parleurs (Amadeus PMX5) espacés de 16 cm. Le sujet est centré par rapport à l'antenne et placé 5 m devant. Les deux haut-parleurs constituant le système stéréophonique (Amadeus PMX5) sont disposés de sorte à ce que le triangle formé par l'auditeur et les deux haut-parleurs soit équilatéral (voir figure 1). Les deux systèmes ont été égalisés par un filtre appliqué au système stéréo consistant en une différence des réponses fréquentielles des deux systèmes mesurées à la position d'écoute. Le temps de réverbération de la salle dans laquelle le test s'est effectué est de 1,2 s.

La diffusion en WFS a été envisagée sous deux formes : d'une part des sources omnidirectionnelles et d'autre part des sources directives dirigées vers l'auditeur (ouverture à -6 dB de ±25°) . Le test prend la forme d'un test AB. Dans chaque phase, trois voix d'homme sont spatialisées selon 3 directions différentes (0°,8° et -8°) et sont diffusées simultanément. Dans le cas de la WFS, les sources virtuelles sont situées 5 m derrière l'antenne de haut-parleurs de sorte à solliciter un grand nombre de haut-parleur lors de la synthèse du front d'onde associé à chaque source. Dans le cas de la stéréo les sources ont été spatialisées à l'aide du spatialisateur (SPAT) en utilisant des différences d'intensité et de temps d'arrivée.

Entre la scène A et la scène B, les positions de deux des trois sources sont éventuellement interchangées. Selon les conditions, les positions de sources sont soient inchangées, soient subissent un décalage de ±8° (inversion de la source au centre et d'une source extrême) ou de ±16° (inversion des deux sources extrêmes). La tâche du participant est d'indiquer s'il a perçu un changement dans l'organisation spatiale de la scène sonore à l'aide d'un boitier (Cedrus RB-620) muni de deux boutons (oui/non).

Procédure Une session est constituée de 12 épreuves (3 modes de diffusion, 3 interversions de sources, 1 paire sans interversion) pour une durée d'environ 5 minutes. Dans chaque épreuve les scènes A et B sont diffusées pendant 10 secondes avec 2 secondes de pause pour les séparer. Un son de cloche indique au sujet le moment de spécifier son choix. L'épreuve suivante est lancée dès que le sujet a répondu. Une session d'entraînement est proposée aux sujets avant d'effectuer le test. De plus les différentes voix sont présentées individuellement aux sujets de sorte à leur permettre de se familiariser avec les timbres et permettre une meilleure discrimination des flux.

Stimuli Les stimuli sont des voix d'homme parlant français, enregistrés en conditions anéchoïques à l'IRCAM. La voix parlée a été préférée à des stimuli musicaux car elle ne nécessite pas de connaissances préalables. Nous n'utilisons que des voix du même sexe et pas de voix d'enfant afin de limiter les différences entre les stimuli. Les fichiers audio sont quantifiés à 32 bits, échantillonnées à 44100 Hz et égalisées en puissance RMS.

Sujets 16 sujets, agés de 20 à 50 ans ont passé le test. Tous ont indiqué avoir une audition normale et la moitié d'entre eux étaient des auditeurs sans expérience particulière d'écoute. Chaque sujet a passé 6 sessions de test pour une durée totale de 30 minutes environ.

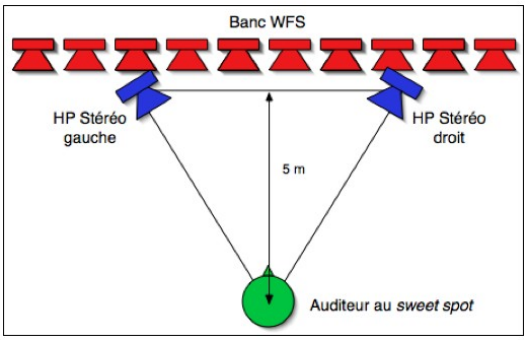

FIG. 1: Position du système WFS, du système Stéréo et de l'auditeur pour le test d'écoute

Analyse Une analyse de la variance (ANOVA) à mesures répétées a été menée avec comme variables indépendantes le système de diffusion (Stéréo, WFS sources omnidirectionnelles, WFS sources directives) et les interversions de sources entre les scènes A et B (centre-gauche, centre-droite et gauche-droite). Le seuil de significativité est défini comme étant $p = 0{,}05$.

2.2 Résultats et discussion

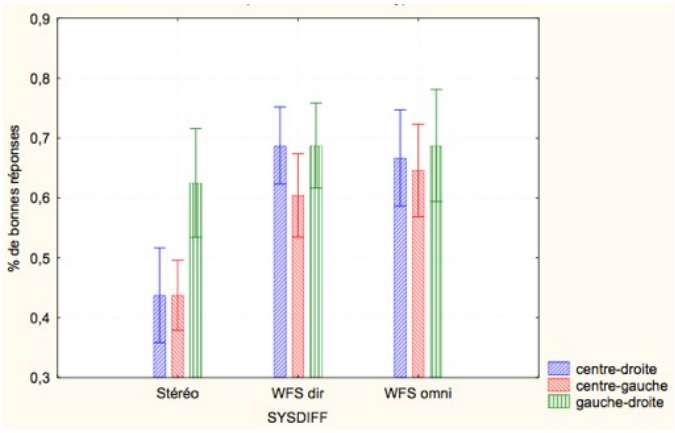

FIG. 2: Pourcentage de bonnes réponses en fonction du système de diffusion et des inversions de position

Résultats Les résultats montrent une meilleure discrimination des changement de position en WFS qu'en Stéréo indiquant ainsi une meilleure lisibilité des scènes sonores, que ce soit avec des sources omnidirectionnelles ou directives. L'interversion des sources n'est pas un paramètre significatif, montrant que la discrimination est aussi bonne pour les interversions de 16° que de 8°. Une analyse croisée des paramètres montre pourtant qu'en Stéréo la discrimination est meilleure pour les inversions de source extrêmes. Il est à noter que les scores de bonnes réponses pour les inversions de sources distantes de 8° (centre-gauche et centre-droit) sont en-dessous du seuil de chance (50 % de bonnes réponses) en Stéréo.

La discrimination n'est pas significativement meilleure en WFS lors d'une diffusion avec des sources directives qu'avec des sources omnidirectionnelles.

Discussion Les résultats montrent des différences significatives pour les taux de réussite entre les deux systèmes de diffusion (WFS et Stéréo). Ceci témoigne de l'influence de la reconstruction du front d'ondes sur la lisibilité des scènes sonores. Ce résultat est à rapprocher de ceux de [17] qui reporte de meilleures performances de localisation en WFS qu'en Stéréo.

[9] reporte dans le cas de sources concurrentes en stéréophonie un angle minimum audible (MAA) de l'ordre de 10°. Nos résultats semblent confirmer cette tendance puisque les scores de réussite pour les interversions de sources proches de 8° (centre-gauche et centre-droite) sont significativement inférieurs aux scores pour les interversions de sources extrêmes. Le fait que la réussite soit inférieure au seuil de chance dans le cas des sources proches de 8° peut s'expliquer par le fait que le choix étant forcé et l'une des épreuves étant une comparaison de scènes identiques, les sujets ont préférentiellement indiqué n'avoir pas perçu de changement en cas de doute. Ce résultat est également à rapprocher de ceux de [14] dans lequel les sujets devaient repérer des mots-clé parmi plusieurs voix concurrentes diffusées au casque. Le taux de réussite était de 63% lors de deux flux simultanés mais chutait à 40% lors de trois flux simultanés. La comparaison est renforcée par le fait que la stratégie de plusieurs sujets était de repérer si certains mots-clé étaient à la même position dans les deux scènes présentées.

Dans le cas de la WFS, les scores de réusssite sont comparables quels que soient les interinversions de sources. Ceci semble indiquer que le MAA en WFS est inférieur à 8° et confirme les résultats de [13] qui reporte un MAA en WFS comparable à celui pour des sources réelles, bien que l'expérience menée par Start n'inclue pas de sources concurrentes.

Le fait que les scores de réussite ne soient pas meilleurs avec des sources de directivité resserrée peut s'expliquer par le choix de directivité que nous avons fait. Les sources sont dirigées vers l'auditeur, ce qui a diminuer l'angle perçu entre les sources

3 Evaluation du réalisme de la reproduction en WFS

Dans le but d'évaluer le réalisme d'un système de reproduction, il semble naturel de comparer un instrument réel à sa reproduction par ce système. Cependant, l'introduction d'un facteur humain dans le protocole expérimental amène un certain nombre de problèmes, dont celui de la reproductibilité du test. Il convient alors de limiter les facteurs pouvant introduire des variations entre les différentes épreuves.

Dans le test présenté ici, plutôt que de procéder à des comparaisons successives de stimuli (test de type ABX par exemple) dans des situations de laboratoire, nous avons préféré rapprocher l'expérience d'une situation réelle, le but étant d'évaluer si un système de reproduction peut *dans une situation réelle* provoquer l'illusion ou du moins créer une ambiguïté et si possible d'identifier les facteurs influençant le réalisme de la reproduction. Ainsi, dans ce test la déclamation d'un texte alterne entre fragments lus par l'acteur et fragments diffusés par des systèmes de reproduction, la tâche des sujets étant d'évaluer en continu s'ils sont en train d'écouter la voix réelle ou sa reproduction.

Le choix d'une voix parlée provient de la familiarité que tout homme possède avec ce type de stimulus. Ainsi, il n'est pas besoin de faire de distinction parmi les sujets. Dans le but de limiter les variations de diction de l'acteur,il est nécessaire de circonscrire les répétitions du test dans un temps limité, une demi-journée nous a semblé être un intervalle raisonnable. Les séquences destinées à être diffusées par les sytèmes de reproduction furent enregistrées également le même jour que le déroulement du test.

Des expériences antérieures ont montré que le facteur principal de réalisme d'une reproduction sonore est le timbre [18]. Les différences de timbre entre les systèmes de reproduction et la voix réelle ont donc été limitées par une procédure d'égalisation de leur réponse fréquentielle de sorte à évaluer d'autres facteurs pouvant influencer le réalisme [11].

Matériel Le test se déroule dans l'espace de projection à l'IRCAM ($20 \times 15 \times 10.5m$). Le temps de réverbération est de 1,2 s. Le système WFS constitué

de 88 haut-parleurs (Amadeus PMX5) espacés de 16 cm est placé en fond de salle et suspendu à une hauteur de 1,80m. La source virtuelle utilisée est une source focalisée située 4.5 m devant l'antenne. L'acteur servant à la récitation se tient debout, la tête à la même hauteur que le système WFS, 4m devant l'antenne. Un haut-parleur est également placé à 4 m devant l'antenne WFS. La disposition est représentée figure 3. La source virtuelle est placée un peu plus en avant de la position de l'acteur et du haut-parleur seul pour limiter les effets de masquage dus à l'acteur.

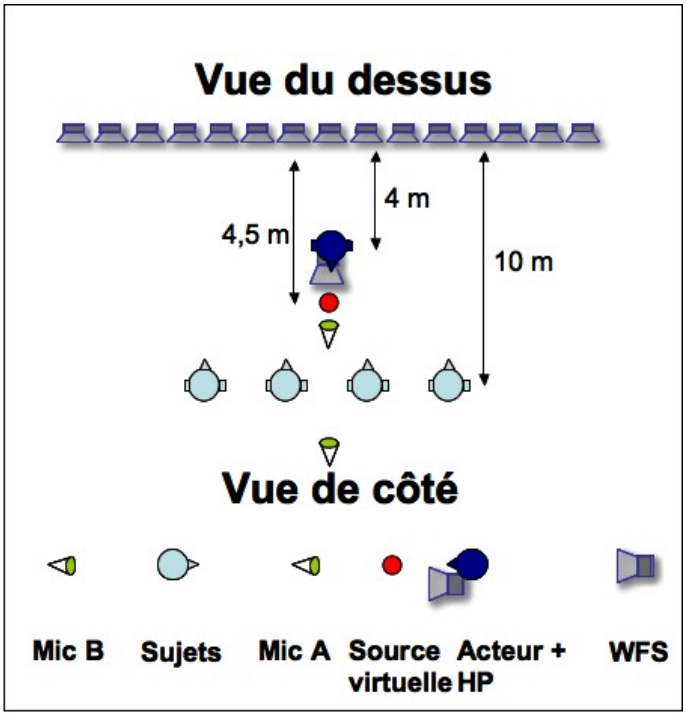

FIG. 3: Disposition adoptée pour le test d'évaluation du réalisme de la reproducton en WFS

Le texte utilisé est un extrait littéraire choisi pour son style fait de phrases courtes et facilement segmentables. Les coupures ont été faites à l'endroit de respirations, la longueur des différents segments est autant que possible semblable.

4 sujets passent l'expérience simultanément. Les sujets sont placés 10 m devant le système WFS avec 1m d'écart entre chaques sujets.

Procédure Préalablement au test, une séance d'enregistrement des segments qui seront diffusés par les systèmes de reproduction est nécessaire. Cette séance a lieu une heure avant le test d'écoute de sorte à vérifier les segments et les réenregistrer si nécessaire.

Lors de la phase préliminaire, le texte est lu par l'acteur et enregistré au moyen d'un microphone Schoeps MK2 (mic A) placé en champ proche (0,5 m de l'acteur). Durant la lecture, le fichier son enregistré est segmenté au fur et à mesure par l'expérimentateur par pression d'une touche indiquant le début et la fin de chaque segment. Cette méthode a été choisie de sorte à obtenir un enchaînement fluide des segments.

Un microphone AKG 414 placé en champ lointain (15m de l'antenne WFS) enregistre également le champ sonore (mic B). A chaque instant, le spectre instantané est calculé et la moyenne des spectres obtenus sur la totalité du texte est gardée en mémoire.

L'enregistrement du microphone A est ensuite rejoué par le système WFS et le spectre moyen en champ lointain est obtenu de la même façon que lors de la lecture par l'acteur . Un filtre d'égalisation fréquentielle du système WFS est alors calculé par division des 2 spectres moyens auquel on associe une phase linéaire. Ce filtre est utilisé pour réduire les différences de timbre entre la voix de l'acteur et la source virtuelle WFS considérée. Cette procéure est répétée pour le haut-parleur seul. Les filtres d'égalisation ainsi calculés sont montrés figure 4.

L'égalisation sur le microphone B (champ lointain) a été préférée à une égalisation en champ proche (microphone A) car elle est moins sensible à l'orientation de l'acteur et à ses mouvements lors de la lecture (respiration, bruits divers).

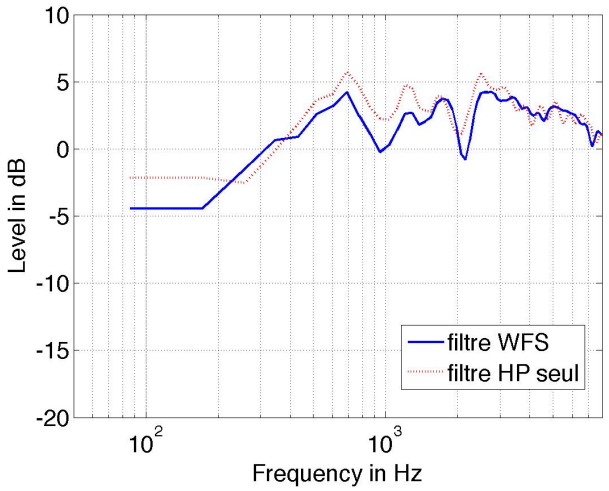

FIG. 4: filtre d'égalisation d'une source virtuelle WFS (resp du haut-parleur) obtenu par différence du spectre moyen de la voix parlée et du spectre moyen de la source WFS (resp du haut-parleur)

Lors de la phase de test, 5 modes de diffusion sont considérés :

- l'acteur lisant la séquence
- le haut-parleur
- la souce virtuelle WFS avec une directivité omnidirectionnelle
- la souce virtuelle WFS avec une directivité resserrée (ouverture à -6dB de $\pm 25^{\circ}$)
- la souce virtuelle WFS avec une directivité variable, asservie à la puissance sonore du signal. Plus la puissance est grande, plus l'ouverture est grande. Ce mode de diffusion n'a pas pour objectif de s'approcher d'une directivité réaliste de la voix, mais d'apporter une variation dans l'interaction de la source avec la salle.

Les différents segments sont déclenchés par l'acteur au moyen d'un télécommande Wiimote pour obtenir un enchaînement fluide des différents segments et diminuer les silences entre chaque segment. Le mode de diffusion des segments est aléatoire. Le nombre d'occurences de chaque mode de diffusion est le même sur la totalité du

test.

Les sujets sont équipés de télécommande Wiimote. Leur tâche est de déterminer les séquences lues par l'acteur. Il leur est donné comme consigne d'appuyer sur le bouton + et de le garder enfoncé tant qu'ils pensent que c'est l'acteur qui déclame et d'appuyer sur le bouton − et de la garder enfoncé dans le cas contraire. Les sujets sont masqués durant tout le test.

Sujets 12 sujets, agés de 18 à 30 ans ont passé le test. Tous ont indiqué avoir une audition normale et la moitié d'entre eux étaient des auditeurs sans expérience particulière d'écoute. Chaque sujet a passé le test 1 fois (3 sessions) pour une durée totale d'environ 15 minutes. Aucune session d'entraînement n'était proposée.

Analyse Un exemple de relevé de réponses comparé aux changements objectifs de diffusion est présenté figure 5.

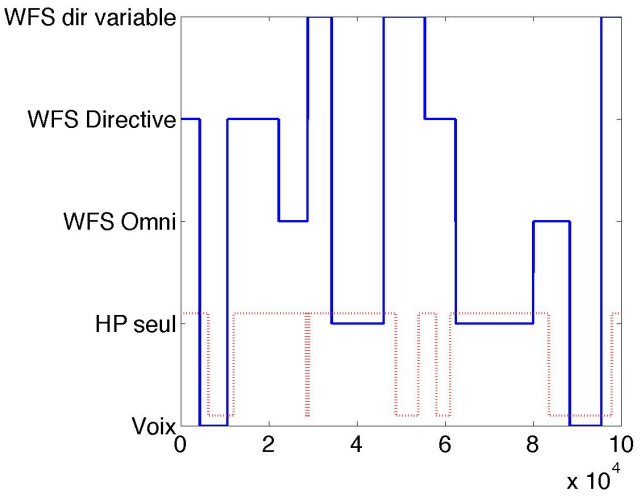

FIG. 5: Extrait de relevé de réponses d'un sujet et de changements objectifs de diffusion. L'axe x correspond au temps (en ms) et l'axe y aux modes de diffusion dans le cas des changements objectifs. Les réponses du sujet ne prennent que 2 valeurs, 0 et 1, 0 correspondant à une diffusion perçue comme étant de la voix et 1 pour un système de reproduction

Dans chaque session, le mode de diffusion de la première séquence n'est jamais l'acteur. Cette séquence a été retirée des résultats servant à l'analyse.

Plusieurs analyses des relevés de réponses des sujets sont possibles. Dans la suite, nous avons choisi de nous intéresser au *score Voix* par sujet et par mode de diffusion qui est le pourcentage de temps passé à indiquer que le mode de diffusion courant est perçu comme étant la voix réelle. La comparaison de la valeur de ce score pour les diffusions par systèmes de reproduction avec la valeur pour la voix réelle nous donne une estimation de l'ambiguïté dans les réponses des sujets.

Le *temps de décision* lors d'un changement de diffusion correspond au temps de réaction moteur plus le temps nécessaire au sujet à décider de la nature du nouveau mode de diffusion. Ce temps peut être une information sur l'hésitation des sujets à apprécier le mode de diffusion lors d'un changement. Le *temps de décision moyen* pour chaque sujet est estimé au moyen d'un calcul de corrélation entre les réponses de chaque sujet et les changements objectifs de diffusion. Le délai correspondant au maximum de corrélation nous donne le temps de décision moyen . Celui-ci est compris entre 1,7 et 2,5 s selon les sujets. Le temps de décision moyen est de 2,0 s. Les relevés de réponses sont ensuite resynchronisés individuellement et les scores Voix sont calculés sur ces relévés resynchronisés. Le resynchronisation permet de réduire les différences inter-sujets tout en conservant l'information des réponses individuelles.

Une analyse de la variance (ANOVA) à mesures répétées a été menée sur les scores Voix avec comme variables indépendantes les différents modes de diffusion (voix, haut-parleur seul, WFS directivité onidirectionnelle, WFS directivité resserrée, WFS directivité variable). Le seuil de significativité est défini comme étant $p = 0{,}05$. Un test post-hoc HSD de Turkey a ensuite été effectué, indiquant la significativité des différences des scores pour les 5 modes de diffusion.

Résultats et discussion Les moyennes des scores Voix sont présentées figure 6. Le score Voix pour le mode de diffusion Voix est significativement différent de tous les autres scores. Aucun des systèmes de diffusion n'a donc créé d'ambiguïté suffisante pour pouvoir se comparer à une voix réelle. Il est cependant intéressant de noter que la voix réelle n'a été perçue comme telle seulement 71 % du temps total.

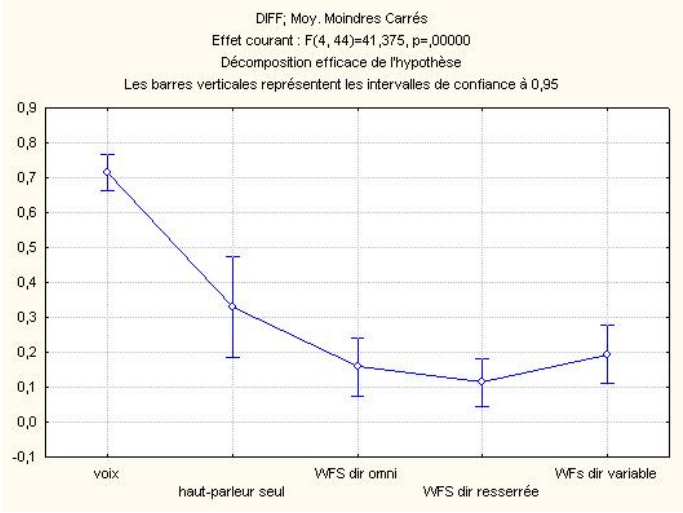

FIG. 6: Moyennes sur les scores Voix pour chaque mode de diffusion

Bien que la WFS avec directivité resserrée soit le seul mode de de diffusion en WFS entraînant des scores significativement différents de ceux pour le haut-parleur seul, les raisons de ces performances sont probablement différentes selon les modes de diffusion. Malgré l'égalisation en fréquence réalisée, une différence de timbre était perceptible entre le haut-parleur seul et la WFS, la présence de l'acteur impliquant un masquage de la source WFS absent dans le cas du haut-parleur seul. Positionner la source virtuelle WFS légèrement devant l'acteur limite ce masquage mais ne l'élimine pas

complètement. Les sujets ont souvent rapporté percevoir des changements de timbre lors d'alternances de systèmes de diffusion les incitant à changer de réponse et ce même lorsqu' aucun changement de réponse n'était requis. 13% des changements de diffusion WFS-WFS ont entraîné un changement de réponse des sujets contre 37% dans le cas de changements WFS - haut-parleur seul ou haut-parleur seul - WFS.

Il convient également de tenir compte de la nature du matériau sonore dont dépend la perception du réalisme au contenu sonore [7]. Des études ont montré des variations minimes de directivité (comparativement à un instrument de musique), en fonction de la voyelle [8], entre des niveaux sonores *normaux* et *forts* [3]. Ainsi il peut sembler normal que des caractéristiques de directivité différentes n'aient que peu d'impact sur le réalisme de la reproduction car la voix n'a que peu de variations de directivité au cours du temps.

4 Conclusion

Nous avons présenté ici 2 tests d'écoute dans lesquels nous avons successivement tenté d'évaluer la perception de scènes sonores complexes en WFS et en stéréo et le réalisme de la reproduction en WFS. Un accent a été mis sur la mise en place des protocoles expérimentaux et les questions qui se posent lors de l'étude de phénomènes perceptifs complexes.

Il a ainsi été montré que les scènes sonores complexes sont plus lisibles en WFS qu'en stéréo (au sens qu'un changement dans l'organisation spatiale de la scène est mieux détecté) et que l'angle minimum audible en WFS pour des sources concurrentes est supérieur à celui en stéréo. L'influence de la directivité n'a pas pu être mis en évidence.

Le deuxième test a montré que dans la situation choisie (déclamation d'un texte dans lequel de segments sont lus par un acteur et d'autres diffusés soit par un système WFS soit par un haut-parler seul), il n'y avait pas d'ambiguïté dans les réponses des sujets entre l'identification de la voix réelle et de ses différentes reproductions. Cependant de multiples interprétations des résultats sont possibles (temps de décision, nombre de changements de réponse par segments, pourcentage de temps). Des analyses complémentaires sont en cours et de futurs travaux s'attacheront à évaluer le réalisme de la reproduction en WFS dans des configurations réduisant les problèmes rencontrés ici (masquage de la source WFS par l'acteur etc)

Références

[1] A.J. Berkhout. A holographic approach to acoustic control. *Journal of the audio engineering society*, 36(12) :977–995, 1988.

[2] T. Caulkins. *Caractérisation et contrôle du rayonnement d'un système de Wave Field Synthesis pour la situation de concert.* PhD thesis, Université de Paris VI, 2007.

[3] W.T Chu and A.C.C Warnock. Detailed directivity of sound fields around human talkers. Technical report, Institute for research in construction, Canada, 2002.

[4] E. Corteel. Objective and subjective comparison of electrodynamic and map loudspeakers for wave field synthesis. In 30^th *conference of the audio engineering society*, March 2007.

[5] J. Daniel, R. Nicol, and J. Moreau. Further investigations of high order ambisonics and wavefield synthesis for holophonic sound imaging. *Proceedings of the 114th convention of the audio engineering society, Amsterdam, the Netherlands*, 2003.

[6] W. de Bruijn. *Application of Wave Field Synthesis in Videoconferencing.* PhD thesis, TU Delft, Netherlands, 2004.

[7] C. Guastavino and B.F.G. Katz. Perceptual evaluation of multi-dimensional spatial audio reproduction. *Journal of the Acoustical Society of America*, 116(2) :1105–1115, August 2004.

[8] B.F.G. Katz and C. d'Alessandro. Directivity measurements of the singing voice. *19th INTERNATIONAL CONGRESS ON ACOUSTICS MADRID*, 2007.

[9] D.R. Perrott. Concurrent minimum audible angle : a re-examination of the concept of auditory spatial acuity. *Journal of the Acoustical Society of America*, 75(4) :1201–1206, April 1984.

[10] G. Rhodes. Auditory attention and the representation of spatial information. *Perception and Psychophysics*, 42(1) :1–14, 1987.

[11] F. Rumsey, S. Zielinski, R. Kassier, and S. Bech. On the relative importance of spatial and timbral fidelities in judgements of degraded multichannel audio quality. *Journal of the Acoustical Society of America*, 118(2) :968 – 976, August 2005.

[12] J. Sanson. Objective and subjective evaluation of localization accuracy in wave field synthesis. *AES 124th convention,*, 2008.

[13] E.W. Start. *Direct sound enhancement by Wave Field Synthesis.* PhD thesis, TU Delft, 1997.

[14] L. J. Stifelman. The cocktail party effect in auditory interfaces : a study of simultaneous presentation. *MIT Media laboratory technical report*, 1994.

[15] E.N.G. Verheijen. *Sound reproduction by Wave Field Synthesis.* PhD thesis, TU Delft, 1997.

[16] P. Vogel. *Application of Wave Field Synthesis in room acoustics.* PhD thesis, TU Delft, 1993.

[17] H. Wittek, F. Rumsey, and G. Theile. On the sound colour properties of wavefield synthesis and stereo. *AES 123rd convention,*, 2007.

[18] S. Zielinski, F. Rumsey, R. Kassier, and S. Bech. Comparison of basic audio quality, timbral and spatial fidelity changes caused by limitation of bandwidth and by down-mix algorithms in 5.1 surround audio systems. *Journal of the Audio Engineering Society*, 2005.