

Joseph Sanson

*Université de ParisVI*

Master 2 en acoustique, traitement du signal et informatique appliqués à la musique

# Evaluation objective et subjective des erreurs de localisation en Wave Field Synthesis

Sous la direction d'Etienne Corteel et d'Olivier Warusfel

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Théorie et technologie</b>	<b>5</b>
2.1	Origine mathématique . . . . .	5
2.1.1	Intégrale de Kirchhoff-Helmholtz . . . . .	5
2.1.2	Simplifications . . . . .	6
2.2	Haut-parleurs et filtres . . . . .	7
2.2.1	Haut-parleurs . . . . .	7
2.2.2	Egalisation et filtres . . . . .	8
<b>3</b>	<b>Physiologie et modélisation du système auditif</b>	<b>9</b>
3.1	Physiologie . . . . .	10
3.2	Modélisation . . . . .	11
3.2.1	Indice de corrélation interaurale (IACC) . . . . .	12
3.2.2	Différence interaurale de temps d'arrivée . . . . .	13
3.2.3	Différence interaurale de niveau sonore . . . . .	13
3.3	Modèles de localisation . . . . .	13
<b>4</b>	<b>Indices de localisation pour la WFS</b>	<b>14</b>
4.1	SItdVI et SIldVI . . . . .	14
4.2	Droiture . . . . .	14
<b>5</b>	<b>Premières écoutes informelles</b>	<b>16</b>
<b>6</b>	<b>Etude des indices objectifs</b>	<b>17</b>
6.1	Comparaison des méthodes d'extraction d'enveloppe pour le calcul de l'itd . . . . .	17
6.1.1	Comparaisons des fonctions de corrélation interaurales (iacf) . . . . .	17
6.1.2	Comparaison des fréquences d'aliasing . . . . .	21
6.1.3	Comparaison par les indices de droiture . . . . .	21
6.1.4	Conclusion . . . . .	22
6.2	Dépendance des erreurs à la position latérale d'écoute : Mise en évidence d'un phénomène d'accumulation . . . . .	22
6.3	Influence des positions relatives d'écoute et de source . . . . .	23
6.4	Influence des positions relatives d'écoute et de haut-parleurs . . . . .	24
6.5	Influence de l'espacement entre les haut-parleurs . . . . .	27
6.6	Conclusion . . . . .	28
<b>7</b>	<b>Expérience</b>	<b>29</b>
7.1	Buts . . . . .	29
7.2	Protocole . . . . .	29
7.2.1	Résultats et comparaisons avec les résultats objectifs . . . . .	30
7.2.2	Etude du phénomène d'accumulation . . . . .	31
7.2.3	Influence des positions relatives d'écoute et de source . . . . .	32
7.2.4	Effet de l'espacement entre les haut-parleurs . . . . .	32
7.2.5	Effet de la bande passante des sons . . . . .	33
7.3	Discussion . . . . .	33

<b>8 Conclusion</b>	<b>34</b>
<b>A Erreurs sur les indices objectifs pour les positions de source et d'écoute du test psychoacoustique</b>	<b>35</b>
A.1 Pour un espacement de 15 cm entre les haut-parleurs . . . . .	35
A.2 Pour un espacement de 30 cm entre les haut-parleurs . . . . .	39

# 1 Introduction

La Wave Field Synthesis (WFS) [Ber88] ou synthèse de champ d'ondes est un dispositif visant à la reproduction de scènes sonores au même titre que la stéréo, le 5.1, ambisonic, le binaural et le transaural. Elle se différencie de ces techniques par le fait que le champ restitué est valable sur une zone beaucoup plus étendue et n'est pas tributaire de la distance de l'auditeur à chacun des haut-parleurs contrairement à la stéréo, le 5.1 et Ambisonic. Elle permet ainsi aux auditeurs de se déplacer et de se tourner tout en conservant des impressions spatiales cohérentes. Elle permet également de restituer des composantes d'une scène spatiale cohérente à plusieurs auditeurs en même temps sans les isoler, contrairement à la technique binaurale qui nécessitent pour ce faire des casques d'écoute, et rendent donc difficile la communication entre auditeurs.

La WFS s'inspire du principe de Huygens. Ce principe, énoncé en 1678, stipule qu'un front d'ondes est équivalent à un ensemble continu de sources secondaires. Ainsi, une onde qui provient d'une source primaire est strictement équivalente à une onde issue d'une distribution continue de sources secondaires réparties le long du front d'onde primaire. La mise en équation de ce problème a été effectuée par Kirchhoff, Helmholtz et Rayleigh au XIX<sup>e</sup> siècle, en écrivant la synthèse du champ sous forme de problème aux limites de l'acoustique linéaire. L'espace à trois dimensions est divisé en deux sous-espaces complémentaires. Le premier espace (*espace des sources*) contient les sources primaires et le second (*espace de reproduction*) est celui dans lequel la reproduction est assurée.

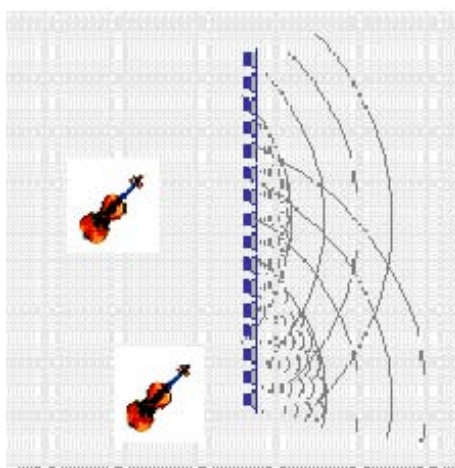


FIG. 1 – Illustration du principe de Huygens

La WFS consiste en un banc linéaire de haut-parleurs faiblement écartés (de 15 à 20 cm). Ce segment peut être considéré comme une ouverture sur un espace contenant les sources sonores. Idéalement, ce medium serait indétectable partout à l'intérieur de la zone d'écoute. Cependant les simplifications par rapport au modèle physique sous-jacent (intégrales de Kirchhoff-Helmholtz et de Rayleigh) et les caractéristiques des haut-parleurs telles que leur directivité induisent des erreurs dans la reproduction qui empêchent une "transparence" parfaite de l'interface de restitution sonore. On peut évaluer cette transparence par le timbre, le positionnement angulaire, la taille de la source et la distance.

L'étude de la reproduction du timbre a été abordée dans [Cor06a], celle de la perception de la distance dans [Ngu06]. Le but de ce stage sera de tenter une approche de la caractérisation des erreurs sur la localisation angulaire. En effet le banc de haut-parleurs est de taille finie et constitué d'un nombre fini de haut-parleurs, ce qui constitue une approximation des principes physiques mis en oeuvre. Il a été opéré

un échantillonnage spatial de la ligne de sources secondaires. La conséquence est la même que celle de l'échantillonnage temporel. Il existe une fréquence de Nyquist au-dessus de laquelle il y a repliement (spatial dans notre cas). Ainsi en hautes fréquences, la transparence du système de reproduction n'est plus assurée.

Ce phénomène de repliement spatial (ou aliasing) a été étudié d'un point de vue psychoacoustique pour en déterminer l'impact sur la perception de la localisation d'une source synthétisée. De Buijn [dB04] dans sa thèse soutient qu'un écartement de haut-parleurs de 15 cm entraîne une fréquence d'aliasing de l'ordre de 1200 Hz, et que cette configuration est suffisante pour assurer la localisation des événements sonores. Les expériences subjectives qu'il a conduites montrent en effet que la localisation est encore assez performante, même pour des sons synthétisés ne contenant pas de basses fréquences. Ces expériences indiquent que la synthèse des sources reste cohérente même en hautes fréquences. Il y a ainsi un décalage entre la théorie sur la synthèse par la WFS et la perception du champ sonore synthétisé.

Les indices de localisation objectifs (différence interaurale de temps d'arrivée, différence interaurale de niveau sonore, fonction de corrélation interaurale) permettent une description du champ sonore arrivant aux oreilles d'un auditeur. Du fait de la présence de la fréquence d'aliasing spatial, ces indices de localisation montrent des erreurs dans la synthèse des sources, qui ne sont pas localisées au même endroit dans les hautes fréquences que dans les basses fréquences. Le but de ce stage est :

- de modéliser au mieux les indices de localisation, notamment en hautes fréquences.
- de quantifier au moyen de ces indices l'importance des erreurs de reproduction, d'identifier les plages dans lesquelles les erreurs sont plus nombreuses.
- d'établir des liens entre les erreurs observées sur les indices objectifs restituées et la localisation des sources au moyen d'expériences psychoacoustiques.

Nous présenterons dans un premier temps le socle physique de la WFS et les simplifications effectuées. Nous verrons les limitations qui découlent de ces simplifications. Nous passerons en revue les technologies employées pour la réalisation du banc. Ces technologies induisent elles aussi des imprécisions sur la synthèse (par exemple la directivité des haut-parleurs). Nous verrons les moyens mis en oeuvre pour minimiser les erreurs de reproduction.

Par la suite, nous décrirons les imprécisions et ambiguïtés que nous avons ressenties lors d'écoutes de la WFS. Ces écoutes furent importantes car elles orientèrent les recherches des erreurs perçues et le choix des indices de localisation utilisés. A la suite de ces écoutes une étude de la littérature existante a été effectuée. Cette étude a porté sur le fonctionnement de l'oreille, notamment en hautes fréquences, les indices de localisation, leur modélisation et les caractéristiques de la WFS. Nous avons ainsi été amenés à comparer dans le cadre de la WFS plusieurs modèles pour l'extraction des différences interaurales de temps. Nous présenterons par la suite plusieurs indices de localisation qui nous ont semblé pertinents ainsi que les indices que nous avons spécialement développé pour l'étude du phénomène d'aliasing spatial en hautes fréquences.

Ces indices nous permettront dans la section 6 de mettre en évidence des différences entre les méthodes d'extraction d'enveloppes dans le but de choisir le modèle le plus adapté à l'étude de la WFS, et ce à l'aide d'une expérience psychoacoustique. Nous quantifierons à l'aide de ces indices les phénomènes observés lors des écoutes qui diminuent la transparence du système. On étudiera en particulier l'influence de la position d'écoute et de la position des sources sur ces indices.

Enfin nous présenterons dans la section 7 les résultats de l'expérience psychoacoustique que nous comparerons aux résultats de l'étude par les indices objectifs.

## 2 Théorie et technologie

### 2.1 Origine mathématique

Le socle mathématique de la WFS est le principe de Huygens et ses formulations mathématiques qui sont l'intégrale de Kirchhoff-Helmholtz et les intégrales de Rayleigh. Ces équations permettent de synthétiser les caractéristiques physiques d'un champ d'onde.

#### 2.1.1 Intégrale de Kirchhoff-Helmholtz

On considère une surface fermée quelconque  $\partial\Omega$  séparant l'espace à trois dimensions  $\Omega$  en deux sous-espaces complémentaires  $\Omega_\psi$  l'espace contenant les sources et  $\Omega_R$  l'espace de reproduction. A partir du problème au limites en  $\partial\Omega$  on en déduit l'intégrale de Kirchhoff-Helmholtz :

$$p_R(\vec{r}_R) = \int_{\partial\Omega} p_\Psi(\vec{r}_S) \vec{\nabla}_S G(\vec{r}_R, \vec{r}_S) - G(\vec{r}_R, \vec{r}_S) \vec{\nabla}_S p_\Psi(\vec{r}_S) \vec{n} dS$$

avec  $p_\Psi(\vec{r}_S)$  est le champ de pression rayonné par l'ensemble des sources primaires sur  $\partial\Omega$ ,  $\vec{n}$  le vecteur normal à  $\partial\Omega$  au point  $r_S, r$  un point de l'espace et G la fonction de Green de l'espace libre en trois dimensions :

$$G(\vec{r}_R, \vec{r}_S) = \frac{e^{-jk\|\vec{r}_R - \vec{r}_S\|}}{4\pi \|\vec{r}_R - \vec{r}_S\|}$$

Le champ sonore est nul dans l'espace des sources  $\Omega_\psi$  et correspond exactement au champ primaire dans l'espace de reproduction  $\Omega_R$ . Cette formulation ne tient aucunement compte ni des caractéristiques des sources primaires ni de leur nombre. Cette formulation indique que la distribution de sources secondaires est un ensemble continu de monopôles alimentées par le gradient de pression du champ rayonné par les sources primaires ainsi qu'un ensemble continu de dipôles alimentés par le signal de pression.

Dans le cas où  $\partial\Omega$  correspond à un plan infini, les contributions des sources monopolaires et dipolaires sont identiques dans  $\Omega_R$ . On peut ne conserver donc qu'un seul type de source. Dans le cas où l'on ne conserve que les sources monopolaires on obtient l'équation de Rayleigh 1 :

$$p_R(\vec{r}_R) = -2 \int_{\partial\Omega} G(\vec{r}_R, \vec{r}_S) \vec{\nabla}_S p_\Psi \vec{r}_S \cdot \vec{n} dS$$

Lorsque l'on ne conserve que les sources dipolaires, on obtient l'équation de Rayleigh 2 :

$$p_R(\vec{r}_R) = 2 \int_{\partial\Omega} p_\Psi \vec{r}_S \vec{\nabla}_S G(\vec{r}_R, \vec{r}_S) \cdot \vec{n} dS$$

La WFS s'appuie sur l'équation de Rayleigh 1.

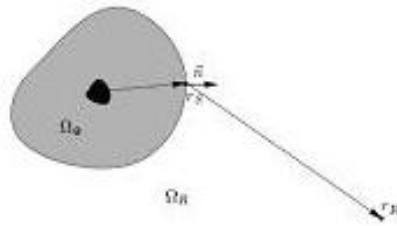


FIG. 2 – Géométrie pour l'intégrale de Kirchhoff-Helmholtz

### 2.1.2 Simplifications

Ces principes ne furent jamais mis en pratique jusqu'à ce que Berkhout de l'Université de Delft propose le concept de la Wave Field Synthesis à la fin des années 1980. Les principes généraux de la WFS sont décrits dans [Ber88] et [Vog93].

**Réduction à une ligne infinie** Dans la pratique, on peut considérer que les sources et les auditeurs sont situés sur la même ligne horizontale, ce qui permet de réduire le plan infini considéré dans l'intégrale de Rayleigh à une ligne infinie. On peut en effet montrer que la contribution principale du champ sonore arrivant aux auditeurs est donnée par les sources secondaires situées dans le plan des oreilles. Cette ligne est l'intersection du plan vertical infini et du plan horizontal comprenant les sources. Chaque point de cette ligne représente alors l'intégrale de toutes les contributions sur la colonne infinie contenant ce point. L'intégrale de Rayleigh résultante se simplifie alors en utilisant l'approximation de la phase stationnaire décrite dans [Cor04]. Cette simplification amène plusieurs erreurs dans la restitution du champ sonore :

- L'utilisation de sources omnidirectionnelles impose une symétrie cylindrique au champ sonore synthétisé.
- L'atténuation du champ sonore n'est parfaitement synthétisée que pour une profondeur d'écoute moyenne.
- Les basses fréquences sont émises en avance d'un huitième de période et avec un niveau plus faible que les hautes fréquences. Ce problème peut être résolu par filtrage [Sta97].
- On a utilisé l'approximation de la phase stationnaire pour déduire les fonctions d'alimentation des sources secondaires, or cette approximation n'est valable qu'en hautes fréquences et en champ lointain. Cependant les erreurs commises ne sont pas significatives au dessus de 100 Hz et au delà de 10 cm du banc.

**Réduction à un segment** Notre plan infini de sources secondaires a maintenant été réduit à une ligne infinie. Il faut réduire cette ligne à un segment. Cela entraîne des problèmes de visibilité, dans le sens que la zone dans laquelle la représentation va être valable est située entre deux lignes reliant la position de la source avec chaque haut-parleur extrême. Pour faire une analogie visuelle, le banc de haut-parleurs peut être assimilé à une fenêtre. On ne peut ainsi voir que les objets se trouvant entre les deux lignes précédemment décrites.

Ce fenêtrage introduit un phénomène de diffraction à l'intérieur de la zone de visibilité qui perturbe le champ synthétisé. Cependant, des études ont montré que cette diffraction peut être limitée en atténuant les contributions des haut-parleurs situés aux extrémités [Sta97], mais cette méthode ne marche que pour

les basses fréquences et pose des problèmes pour les sources proches de l'extrémité du banc de haut-parleurs [Cor04].

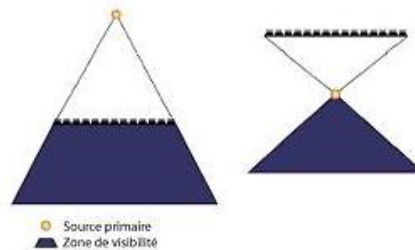


FIG. 3 – Mise en évidence des problèmes de visibilité

**Echantillonnage** La dernière étape permet de passer d'un segment continu de haut-parleurs à un segment discret. On obtient ainsi un segment contenant un nombre fini de sources omnidirectionnelles que l'on a choisi d'espacer régulièrement. On fait donc un échantillonnage spatial. Comme dans le cas de l'échantillonnage temporel, on constate l'existence d'une fréquence au-dessus de laquelle il y a repliement de spectre. On l'appelle "fréquence d'aliasing spatial". En-dessous de cette fréquence, le champ reproduit est identique au champ-cible. Au-dessus, il se crée un repliement des composantes spatiales du champ-cible qui n'assure plus la reproduction exacte. On observe la formation de motifs qui altèrent le champ sonore. Cette altération se traduit par une coloration du son lorsque l'on déplace la tête à une position fixe ou lorsque l'on bouge à l'intérieur de la zone d'écoute, une perception plus floue de la position de la source et peut même conduire à la perception de deux sources distinctes. Les travaux de De Bruijn semblent montrer qu'un écartement des haut-parleurs de 15 à 20 cm est suffisant pour assurer une localisation efficace des sources synthétisées. Cet écartement correspond à une fréquence d'aliasing spatial comprise entre 1500 et 2000 Hz [dB04]. Les travaux plus récents de Corteel ont montré que la fréquence d'aliasing spatial n'est pas seulement liée à l'espacement des haut-parleurs mais est également tributaire de la longueur du banc de haut-parleurs, de la position des sources virtuelles et de la position d'écoute [Cor06b].

## 2.2 Haut-parleurs et filtres

### 2.2.1 Haut-parleurs

Deux types de technologies sont utilisés pour la WFS.

- Un banc constitué de haut-parleurs électrodynamiques (qui seront appelés Elec dans la suite).
- Un banc constitué de panneaux mis en vibration par le biais d'un ou plusieurs transducteurs électrodynamiques appelés excitateurs. Ces excitateurs sont collés à la surface arrière du panneau et sont régulièrement espacés. Les ondes de déformation produites vont rayonner de l'énergie acoustique. Ces panneaux sont appelés MAP (Multi actuator panel). Ils possèdent l'avantage d'avoir une apparence discrète et de pouvoir être intégrés facilement à une salle, suivant ainsi la logique de transparence du système. Cependant leur fonctionnement étant très différent de haut-parleurs classiques, leurs caractéristiques le sont aussi. Leur utilisation en WFS a été étudiée dans [Ngu06], [Cor04], [dB04] et [CNW<sup>+</sup>07].





FIG. 4 – en haut : haut-parleurs MAP, en bas : haut-parleurs Elec

### 2.2.2 Egalisation et filtres

Les *techniques d'égalisation* permettent de prendre en compte les défauts de la Wave Field Synthesis en construisant un ensemble de filtres appliqués au signal d'entrée de chaque haut-parleur du dispositif de restitution. Les filtres d'égalisation sont calculés en référence à un objectif de reproduction à partir d'une caractérisation préalable du champ acoustique synthétisé.

**Egalisation individuelle [Cor04] [Cor06a]** Cette technique considère chaque haut-parleur séparément. A partir d'une description du rayonnement de chaque haut-parleur donné par une fonction de transfert  $c(z)$  on définit le filtre  $h(z)$  que l'on va appliquer pour compenser  $c(z)$ . Typiquement  $h(z)=1/c(z)$ . Cependant la mesure de  $c(z)$  dépend de la position de mesure, on utilise donc soit une moyenne des  $c(z)$  mesurés en différents endroits, soit une position préférentielle pour calculer  $h(z)$ . L'égalisation ne se fait donc qu'en moyenne (spectrale et temporelle) ou à une position donnée. Il est important de noter que le filtrage individuel ne tient aucunement compte de l'interaction qui va s'effectuer entre les différents haut-parleurs.

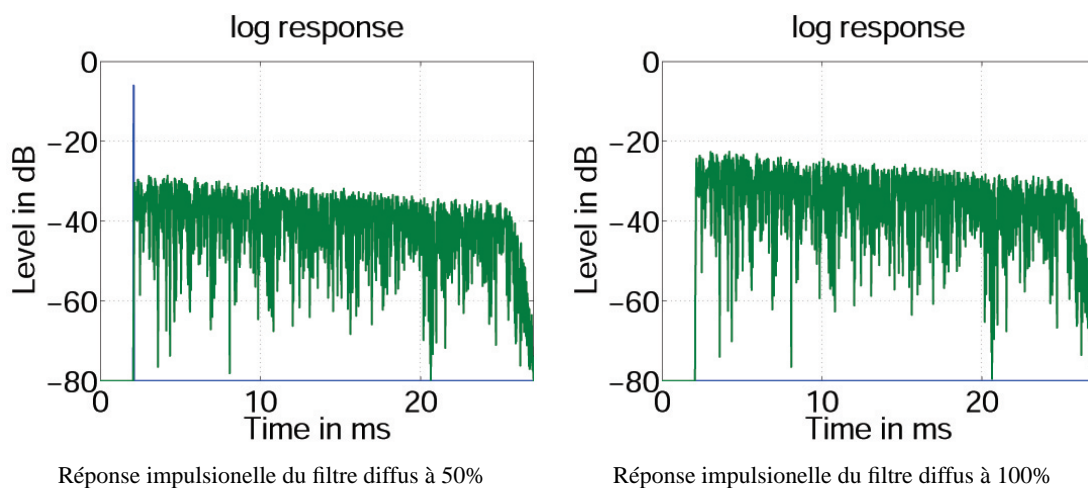
**Egalisation Multicanale [Cor06a]** Ce filtrage a pour ambition de prendre en compte l'ensemble des haut-parleurs lors du calcul du filtre à appliquer à chaque haut-parleur. A partir de la mesure du champ sonore restitué par un banc linéaire de microphones, on essaie de minimiser l'erreur quadratique moyenne entre le champ synthétisé et le champ idéal. Le système peut être assimilé à un système MIMO (Multi Input Multi Output) représenté par une matrice de fonctions de transfert. Le calcul du filtre revient à effectuer une inversion de cette matrice. Les travaux de Corteel ont montré que cette technique assure un contrôle du champ rayonné en zone étendue grâce à l'utilisation d'un banc linéaire de microphones qui assure une description du champ émis par chaque haut-parleur dans une zone étendue. Ce type de traitement n'est réalisé qu'en basses fréquences (en-dessous de la fréquence d'aliasing). Au-dessus le traitement réalisé est une version modifiée de l'égalisation individuelle assurant de plus une égalisation en moyenne (fréquence/espace) du champ synthétisé par l'ensemble des haut-parleurs. Dans toute la suite c'est ce type d'égalisation qui sera utilisé.

**Filtrage diffus à 100% (FD) et à 50% (DD)** Corteel [CNW<sup>+</sup>07] a introduit le filtres diffus à 100% pour limiter l'apparition de motifs spatio-temporels tout en limitant les erreurs de localisation. Ce filtre contient une caractéristique bruitée générée indépendamment pour chaque haut-parleur. Il limite la formation de motifs en hautes fréquences et donc le changement de coloration du son lors de déplacements de l'auditeur à l'intérieur de la zone d'écoute. Il est défini par une longueur de filtre et une pente. Il étale

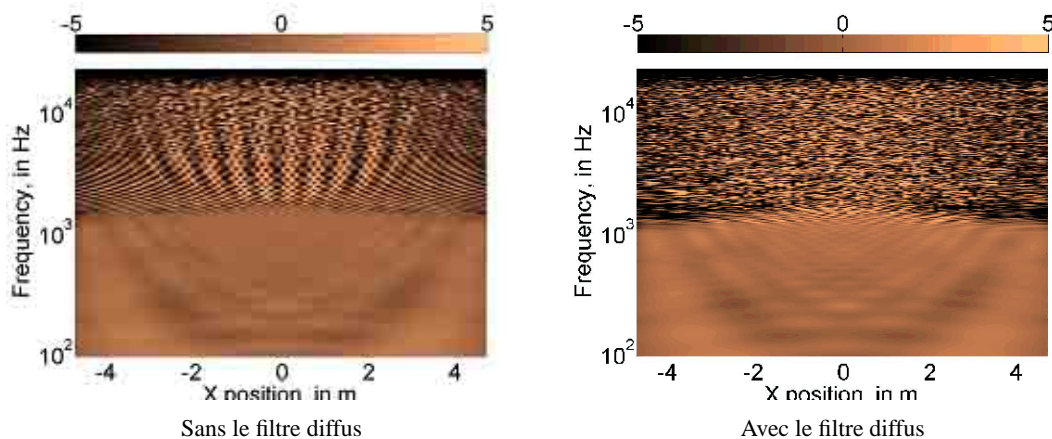
temporellement la réponse impulsionnelle. Son attaque très marquée permet de conserver la localisation spatiale de l'événement. Ce filtre est utilisé seulement au-dessus de la fréquence d'aliasing.

Le filtrage diffus à 50% est une combinaison du filtre individuel et du filtre diffus à 100%. L'énergie est répartie également entre la caractéristique bruitée et la caractéristique non bruitée.

Une étude de ces deux types de filtrage a été effectuée dans [CNW<sup>+</sup>07]. Les résultats d'une expérience subjective ont montré que le filtre diffus à 100% introduit une coloration gênante ce qui n'est pas le cas du filtre diffus à 50%. Dans la suite on ne se servira donc plus que du filtrage diffus à 50%.



TAB. 1 – Réponse impulsionnelle des filtres diffus à 50% et à 100 %



TAB. 2 – Réponse fréquentielle d'un banc de haut-parleurs idéaux. En abscisses la position d'écoute, en ordonnées la fréquence dont le niveau est en dB. On voit l'apparition de motifs au-dessus de la fréquence d'aliasing sans utilisation du filtre diffus

### 3 Physiologie et modélisation du système auditif

L'oreille humaine possède un mode de fonctionnement très complexe. La perception de la provenance d'un son repose sur des indices principalement binauraux, i.e les indices exploitant les différences

entre les signaux arrivant à chaque oreille, [Bla99]. Nous utiliserons par la suite les trois axes du repère sphérique égo-centré ; distance, site et azimut pour décrire les trois dimensions.

La localisation se fait dans les 3 dimensions et les phénomènes mis alors en jeu ne sont pas les mêmes. La perception de la distance, i.e de l'éloignement de la source, est peu performante chez l'homme. Celle-ci exploite principalement des indices monoraux. Le niveau du son direct, la prépondérance du son réfléchi par rapport au son direct, la coloration du signal perçu et l'atténuation des hautes fréquences en fonction de la distance qui est plus forte que celle des basses fréquences sont les quatre indices les plus importants pour la perception de l'éloignement d'une source sonore [dB04].

L'oreille possède peu de moyens pour décider du site d'un son et il semble que le plus important soit le filtrage du son par l'oreille externe. Ce filtrage est en effet dépendant de la provenance du son. Une conséquence de ce phénomène est que l'oreille humaine ne peut efficacement localiser un son selon l'axe vertical que si le spectre contient des hautes fréquences, qui sont ainsi susceptibles d'être filtrées par l'oreille externe [MM02].

La localisation azimutale est de loin la plus performante. Elle repose principalement sur des indices binauraux. Les indices binauraux sont la différence de temps d'arrivée à chaque oreille ou *interaural time difference* (itd) et la différence interaurale d'intensité ou *interaural level difference* (ild). Ces indices n'ont pas la même importance selon les fréquences considérées, ainsi que selon la largeur de bande du son. En basses fréquences, la taille de la tête est négligeable devant la longueur d'onde et il n'y a donc pas de masquage. L'intensité du son arrivant aux deux oreilles est la même. L'itd est donc prépondérant. En hautes fréquences (au-dessus de 1200 Hz), la longueur d'onde est plus petite que la taille de la tête, la différence de phase entre les deux oreilles (ipd) ne peut donc pas être utilisée. Cependant des études ont montré que l'on reste sensible à l'itd. Il a été montré [BT96b] que l'oreille est alors sensible aux différences entre les enveloppes des signaux. L'ild semble néanmoins être l'indice prépondérant en hautes fréquences [MM02]. Si le son possède une bande large contenant des fréquences basses alors l'itd basses fréquences semble déterminer la localisation [WK92].

### 3.1 Physiologie

Dans le but de modéliser au mieux un auditeur, il est nécessaire de s'appuyer sur une description précise du fonctionnement de l'audition. L'oreille humaine est divisée en trois parties dont nous ne détaillerons que les éléments servant à l'audition :

- **L'oreille externe** qui comprend principalement le pavillon et le canal auditif. Le pavillon sert à capter et concentrer les ondes sonores et à amplifier les fréquences autour de 5000 Hz. Le canal auditif conduit au tympan.
- **L'oreille moyenne** est composée en partie du tympan et des osselets. Les vibrations de l'air sont transmises aux osselets par l'intermédiaire du tympan. La fonction principale de l'oreille moyenne est d'adapter les faibles impédances de l'air aux plus grandes impédances de fluides [Hud05]. Sans un système spécial limitant les réflexions à l'interface, il y aurait une perte d'environ 30 dB entre l'onde à l'oreille externe et l'onde arrivant à l'oreille interne.
- **L'oreille interne** comprend en partie la cochlée et les cellules ciliées. Le traitement du son dans l'oreille interne se fait le long de la cochlée qui a la forme d'une spirale. La cochlée est remplie de liquide dans lequel se déplace l'onde. Elle est également séparée en deux chambres par la membrane basilaire sur laquelle on trouve les cellules ciliées qui sont des cellules sensorielles coiffées de structures filamenteuses groupées en touffes et libres de vibrer. Ces cellules sont responsables de la transduction de l'information en impulsions électriques qui seront interprétées par le cerveau. Lorsque l'onde parcourt la membrane basilaire, elle met en vibration les cellules ciliées et ceci en

fonction des fréquences. Les cellules ciliées situées à la base de la cochlée répondent ainsi préférentiellement aux hautes fréquences alors que celles situées à la fin de la cochlée répondent aux basses fréquences. Ce signal nerveux est transmis au nerf auditif puis au cerveau.

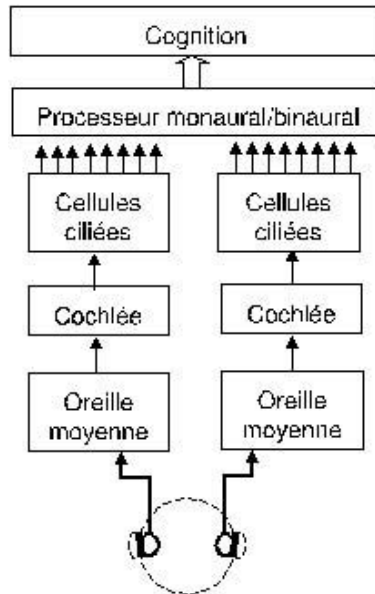


FIG. 5 – schéma du système auditif humain

### 3.2 Modélisation

A partir de cette description de l'oreille, nous avons essayé de prendre en compte les particularités de organes importants pour modéliser de manière réaliste le fonctionnement de l'écoute binaurale. Ainsi, l'utilisation de HRIR (head related impulse responses) garantit une mesure réaliste de réponses impulsionnelles à l'entrée des canaux auditifs et consignent les filtrages naturels dus à la tête et au torse pour une source sonore à une position donnée de l'espace. Selon Hammershoi [HM05], le fait de mesurer les HRIR au début ou à la fin du canal auditif n'introduit pas de différences significatives. On a ainsi capté la contribution de l'oreille externe ainsi que du masquage de la tête et du torse. Il est à noter que les HRIR et leur transformée de Fourier, les HRTF (Head Related Transfer Functions) dépendent de la morphologie de la personne et sont donc individuelles.

Puisque la principale utilité de l'oreille moyenne est d'assurer une bonne transmission de l'onde sonore entre les deux fluides, sa contribution n'a pas été prise en compte [Hud05].

Les indices de localisation sont évalués après analyse tonotopique de la cochlée que l'on peut assimiler à un filtrage par bandes de fréquence. La modélisation de ce filtrage retenue est le formalisme des bandes rectangulaires équivalentes (ERB, Equivalent Rectangular Bandwidth) et leur implémentation sous forme de filtres *gammatone*, dont la largeur de bande augmente avec la fréquence [Pat94].

Les cellules ciliées sont sensibles à la phase des signaux en basses fréquences (en-dessous de 1200 Hz environ) mais perdent cette sensibilité en hautes fréquences. Elles sont dans ce cas sensibles à l'enveloppe des signaux. Leur modélisation en hautes fréquences est sujet à controverse. Nous avons testé trois méthodes d'extraction d'enveloppes qui sont :

- Le modèle de Blauert [BC78]
- Le modèle de Bernstein [BT96a]

– Le modèle de Meddis [Med86]

Nous n'avons pas trouvé dans la littérature de comparaisons de ces modèles d'enveloppes. Il a donc fallu déterminer quelles erreurs et inexactitudes ces modèles introduisaient pour choisir ensuite le meilleur modèle à utiliser. L'étude de la description nous oriente déjà.

### Description des différents modèles pour l'extraction d'enveloppes

**Le modèle de Blauert** est composé d'une rectification demi-onde suivie d'un filtre passe-bas avec une fréquence de coupure à 800 hz.

**Le modèle de Bernstein** a été validé par une expérience psychoacoustique. Il consiste en une rectification demi-onde suivie d'un filtrage passe-bas. Les performances du modèle sont bien établies mais ce modèle n'est pas parfait. La rectification demi-onde se fait au moyen d'une convolution du signal par une fenêtre en  $x^2$ . Ce choix n'est pas du tout motivé dans l'article mais la rectification demi-onde en  $x^2$  suivie d'un filtrage passe-bas renvoie le carré de l'enveloppe et non l'enveloppe elle-même. De plus, l'implémentation de ce modèle dans la toolbox d'Akeroyd<sup>1</sup> (que nous utilisons) semble un peu hasardeuse, une transformée de Hilbert se glissant entre la rectification et le filtrage passe-bas.

**Le modèle de Meddis** est à priori le plus proche de l'oreille humaine car il est une modélisation des cellules cillées et de la transmission de l'information au nerf auditif. Il repose sur l'hypothèse que plus le stimuli est fort, plus le taux de décharge sera important. Une autre hypothèse est qu'une partie des transmetteurs est perdue est qu'une partie retourne dans le réservoir. Cependant, il existe un taux de décharge spontanée qui correspond à une décharge de récepteurs même lorsqu'il n'y a pas de stimulation des cellules cillées. Ce taux dans le modèle de Meddis est très important par rapport aux stimuli considérés. Il a donc fallu, pour rendre ce modèle utilisable, normaliser chaque échantillon par taux de décharge spontanée au repos.

Nous allons maintenant décrire les modèles développés pour calculer ces indices de localisation. Ces modèles proviennent de [Bra05].

#### 3.2.1 Indice de corrélation interaural (IACC)

Cet indice sert à mesurer la cohérence entre le signal qui arrive à chacune des oreilles. Il est très important pour obtenir des informations sur les propriétés spatiales de l'environnement où se déroule l'écoute. Il est obtenu en prenant le maximum de la fonction d'intercorrélation (iacf) calculée comme suit :

$$\varphi_{rl} = \frac{\int_0^{t_0} g_r(t)g_l(t + \tau)dt}{\sqrt{\int_0^{t_0} [g_r(t)]^2 dt \int_0^{t_0} [g_l(t)]^2 dt}}$$

D'après Blauert [Bla99], l'IACC est lié à l'impression d'espace ainsi qu'à la largeur de source apparente (ASW = Auditory source width). Plus l'IACC augmente, plus la corrélation est forte et plus la source apparaît ponctuelle. Une valeur d'IACC de 0.1 équivaut à deux sources séparées. Néanmoins, l'évaluation objective de l'ASW est controversé dans la littérature de l'acoustique des salles.

Nous allons maintenant introduire la notion de seuil de détection (ou encore appelé Just noticeable Difference JND). Le JND d'IACC est généralement donné pour un calcul pleine bande à partir du signal

<sup>1</sup>[http://www.lifesci.sussex.ac.uk/home/Michael\\_Akeroyd/download2.html](http://www.lifesci.sussex.ac.uk/home/Michael_Akeroyd/download2.html)

original (phase) ou de l'extraction de l'enveloppe. D'après [FM04], le JND est de l'ordre de  $\Delta IACC = 0.02$  pour les valeurs d'IACC hautes ( $>0.6$ ) et de  $0.11$  pour les IACC basses ( $<0.5$ ). Le JND est la différence objective la plus faible pour laquelle une différence perceptible est audible. Cependant des valeurs très différentes du JND sont données selon les auteurs. Nous choisirons ici un JND de valeur  $0.01$  pour n'écarter aucune piste lors de l'analyse.

### 3.2.2 Différence interaurale de temps d'arrivée

Un des mécanismes physiologiques mis en jeu pour déterminer la position angulaire d'une source sonore est la différence de temps d'arrivée du signal entre les deux oreilles. On la détermine en considérant le décalage temporel auquel se produit l'IACC. Cette opération s'opère de deux façons en fonction des fréquences contenues dans le signal.

- En basses fréquences (en-dessous d'environ  $1200$  Hz), l'ITD est calculé comme la différence de phase entre les deux signaux.
- En hautes fréquences (au-dessus d'environ  $1200$  Hz), l'ITD est déterminé par analyse de l'enveloppe du signal [Bra05]. Cette sensibilité à l'enveloppe du signal est due au fonctionnement des cellules ciliées.

Une valeur d'ITD en-dessous de  $1$  ms assure la localisation dans le plan azimuthal, une valeur comprise entre  $1$  et  $40$  ms donne l'impression d'une source complètement excentrée. Au-dessus de  $40$  ms, l'oreille perçoit deux sources sonores.

Nous appellerons itd équiprobables les valeurs d'itd pour lesquelles les valeurs d'iacf correspondantes sont inférieures de  $1\%$  à la valeur de l'IACC. Perceptivement, les itd équiprobables ainsi définis ont la même importance que l'itd. Le nombre et la position de ces valeurs équiprobables sont très importantes car elles peuvent provoquer une ambiguïté quant à la localisation de la source sonore. En effet Stern et Trahiotis [ST98] suggèrent que la mesure dans laquelle les maxima de l'iacf sont regroupés influe sur la perception de la provenance de l'événement sonore.

Comme on l'a vu précédemment la synthèse par la WFS est parfaite en basses fréquences, il semble plus intéressant de s'intéresser aux erreurs en hautes fréquences et donc de considérer la différence interaurale d'enveloppes pour le calcul de l'ITD. Pour ce faire, il faut donc extraire l'enveloppe du signal, ce que nous ferons en comparant les trois méthodes d'extraction d'enveloppes présentées plus haut.

### 3.2.3 Différence interaurale de niveau sonore

Ce mécanisme se modélise simplement en regardant dans chaque bande de fréquence la différence d'intensité entre l'oreille droite et l'oreille gauche. L'ILD est donné à un temps  $t$  et pour une fréquence  $k$  par :

$$ILD_k(t) = 20 \times \log_{10} \frac{E_{R,k}(t)}{E_{L,k}(t)} \quad (1)$$

où  $E_{R,k}$  et  $E_{L,k}$  représentent l'énergie des signaux arrivant respectivement aux oreilles droite et gauche dans la bande ERB  $k$ .

## 3.3 Modèles de localisation

La localisation azimutale est basée sur une combinaison des indices d'ITD et d'ILD. Les interactions entre ces indices sont différentes selon les fréquences et largeurs de bande. Il existe plusieurs modèles de combinaison de ces indices mais ils sont peu fiables [Bra05]. Le problème majeur est que notamment pour l'élévation, il est difficile de donner une transformation simple entre les valeurs d'itd et d'ild observées et la localisation. De plus, l'évaluation de la localisation des sources pour des signaux réels sont

sujettes à controverse. La forme et la longueur de la fenêtre d'analyse à partir de laquelle sont extraits les indices de localisation sont sujettes à débat.

Pour contourner ces problèmes nous travaillerons sur les réponses impulsionnelles plutôt que sur les signaux. Nous procéderons à une comparaison des indices idéaux (les indices extraits des HRTF mesurées) et des indices restitués. Les indices restitués proviennent de la simulation de la réponse du banc de haut-parleurs à des positions d'écoute prédéfinies en fonction des caractéristiques de la source synthétisée. Ces réponses sont ensuite binauralisées en les convoluant aux mêmes HRTF que celles utilisées pour les indices idéaux. La différence entre ces deux indices nous donnera ce que nous appellerons *l'erreur sur l'indice*. Nous étudierons chaque indice de cette manière et comparerons les erreurs qui apparaissent sur les indices par bande ERB.

## 4 Indices de localisation pour la WFS

Dans le cadre de la WFS, il est important de modéliser la variation des erreurs sur ces indices de localisation selon les positions d'écoute puisque les auditeurs seront amenés à se déplacer à l'intérieur de la zone d'écoute. Nous étudierons également la variation des erreurs selon les bandes de fréquence pour détecter les possibles incohérences dans la synthèse qui pourraient troubler la localisation.

Pour ce faire nous avons développé trois indices qui rendent compte de ces intégrations. Le SItDVI et le SIIdVI décrivent la variation spatiale des erreurs et la droiture leur variation fréquentielle.

### 4.1 SItDVI et SIIdVI

Dans sa thèse, De Bruijn [dB04] a développé un indice, le SCVI (Spatial Coloration Variation Index), permettant d'évaluer la différence de coloration entre les positions d'écoute. Nous avons produit deux équivalents du SCVI, le SItDVI et le SIIdVI. Le SItDVI (resp le SIIdVI) décrit la variation de l'erreur sur l'itd (resp sur l'ild) qui existe entre plusieurs positions d'écoute. Pour chaque position d'écoute on regarde la variation de l'erreur entre cette position centrale et deux positions situées de part et d'autre. On moyenne ensuite sur le nombre de positions d'écoute de part et d'autre puis sur le nombre de positions centrales. Ces indices sont donnés par la formule :

$$SCVI = \frac{1}{N_x N_{offset}} \sum_{i=N_{offset}+1}^{N_x+N_{offset}} \sum_{j=1}^{N_{offset}} D_{ij}$$

avec

$$D_{ij} = \sqrt{\int_{ERBmin}^{ERBmax} (|S_{x_i}| - |S_{x_{i-j}}|)^2 d(ERB)}$$

où  $D_{ij}$  est la différence entre  $(S_{x_i} - S_{x_{i-j}})$  sur l'itd (resp. l'ILD) dans chaque bande ERB. Le premier itd  $S_{x_i}$  est l'erreur sur l'itd d'un signal reçu à la position  $x_i$  signalée par  $N_{x_i}$ . Le deuxième itd  $S_{x_{i-j}}$  celle d'un signal reçu à la position  $x_i + offset$ .

$N_{offset}$  est le nombre de différences d'écart spatial (l'offset) entre les 2 signaux analysés. On réalise ensuite la moyenne sur les positions d'écoute. Le SItDVI s'exprime en secondes et le SIIdVI en dB.

On peut calculer le SItDVI et le SIIdVI pour chaque bande ERB ou alors moyenner sur les bandes ERB.

### 4.2 Droiture

Le concept de droiture (straightness) a été introduit par Trahiotis et al [TBA01]. Il représente la mesure dans laquelle les IACC sont donnés aux mêmes valeurs d'itd dans toutes les bandes de fréquence.

Selon eux, plus les itd sont "alignés" entre les bandes de fréquence, plus la localisation est aisée. On regarde donc si la direction indiquée par les itd en basses fréquences (en-dessous de 1200 hz) se retrouve dans les itd équiprobables en hautes fréquences.

On calcule la direction de la droite donnée par les basses fréquences. La droiture est ensuite la moyenne de la distance des itd équiprobables à la droite. Plus cet indice est petit, meilleure est la droiture et donc plus la localisation est facile.

Le fait que l'on se réfère aux basses fréquences pour déterminer la direction des itd provient des travaux de Wightman et Kistler [WK92] qui ont montré que lorsqu'un son contient des hautes et des basses fréquences, la direction indiquée par les basses fréquences prime sur la direction indiquée par les hautes fréquences.

$$Droiture = \frac{\sum_{ERB_{1200}}^{ERB_{max}} \sum_k (T_k^{equi} - Dir_{BF})}{N_{ERB} N_{T^{equi}}}$$

$$Dir_{BF} = \frac{\sum_{ERB_{min}}^{ERB_{1200}} T_i}{N_{ERB}}$$

avec  $Dir_{BF}$  est la direction indiquée par les basses fréquences,  $T_i$  la valeur de l'itd sur la bande ERB  $i$  et  $N_{ERB}$  le nombre de bandes ERB.  $T_k^{equi}$  représente le  $k$ -ème itd équiprobable sur la bande ERB considérée et  $N_{T^{equi}}$  le nombre d'itd équiprobables sur cette bande.



## 5 Premières écoutes informelles

Les indices de localisation présentés nous permettront dans la suite de caractériser les erreurs lors de la restitution du front d'ondes, et ce particulièrement en hautes fréquences. Cependant, l'oreille humaine possède un fonctionnement qui n'est pas encore complètement compris, il nous a donc semblé nécessaire de nous rendre compte auditivement des caractéristiques de la synthèse de sources sonores par la WFS. Pour cela nous avons procédé à des écoutes préliminaires qui nous ont permis de mettre en évidence des imprécisions de localisation. Plusieurs configurations ont pu être ainsi testées, ce qui a orienté l'étude par les indices objectifs qui a suivi.

Compte-tenu du temps limité, seulement certaines dimensions parmi celles qui ont une influence sur le repliement ont été étudiées. Les dimensions retenues l'ont été pour des raisons matérielles, de temps ou d'importance supposée de ce paramètre. Nous avons retenu :

- Position de la source en x
- Position de la source en y
- Position d'écoute en x
- Espacement des haut-parleurs
- Traitement (dd ou meq)
- Bande passante des bruits diffusés

Il n'a pas été étudié :

- Directivité des haut-parleurs
- Type de son
- Longueur du banc de haut-parleurs
- Position d'écoute en y

Lors d'écoutes informelles ayant eu lieu dans un studio à l'IRCAM, il a déjà été possible de se rendre compte des caractéristiques de la WFS, notamment des différences de rendu selon les fréquences. Le système installé était composé de 24 ou de 12 haut-parleurs espacés de respectivement 15 et 30 cm. Les haut-parleurs utilisés étaient de marque KEF, modèle HTS2001. Le studio est une pièce de dimensions  $5 \times 8 \times 3$  et possédant une acoustique sèche. Les panneaux MAP n'étant pas disponibles, aucune comparaison n'a été effectuée. Pour une étude de ce sujet, se reporter à [Ngu06] et [CNW<sup>+</sup>07]. Les sons utilisés étaient des bruits blancs filtrés une comparaison du rendu selon les différentes bandes de fréquence. Il était possible de modifier la bande passante des bruits considérés ainsi que la position des sources. Voici les observations générales que nous pouvons faire :

- on observe une instabilité dans la localisation de la source sonore lorsqu'à une position donnée, on bouge la tête de droite et de gauche. Cette instabilité rend imprécise la position de la source qui bouge en même temps que l'on bouge la tête.
- L'utilisation du filtre diffus permet de réduire cette instabilité mais entraîne une légère augmentation de la taille de la source perçue.
- Lorsqu'on augmente l'espacement entre les haut-parleurs, cette instabilité est plus marquée. La source apparaît également plus large.
- Les mêmes phénomènes sont observés pour des écarts de position de tête plus grands, perceptibles lorsque l'on fait un pas sur le côté.
- Plus le son émis contient de hautes fréquences, plus la source apparaît étalée. Ceci peut même dans certains cas conduire à la perception de deux sources distinctes, l'une contenant les basses fréquences et l'autre les hautes fréquences.
- Plus on est proche de l'extrémité du banc de haut-parleurs, plus les erreurs sur la localisation semblent fortes.

- La position de la source est très nette lorsqu'elle est située en face de la position d'écoute. Moins on est en face de la source plus la localisation de la source est vague.

## 6 Etude des indices objectifs

Dans le but d'estimer les erreurs lors de la synthèse à l'aide de la WFS, les indices de localisation restitués ont été comparés à des indices-cible.

Les indices restitués sont les indices calculés suite à la synthèse d'une source par la WFS. Lorsque les caractéristiques de la source virtuelle ont été définies, on calcule le signal envoyé à chaque haut-parleur. La somme des contributions de chaque haut-parleur forme la réponse du banc de haut-parleurs. La réponse du banc arrivant à chaque oreille est ensuite calculée et ce signal est convolué aux HRTF pour former la réponse à une position d'écoute. Les indices restitués sont calculés sur cette réponse.

Les indices-cible sont les HRTF correspondant à la position de la source par rapport à la position d'écoute. Par la suite, lorsque l'on parle d'erreur sur un indice il s'agit de la différence entre l'indice restitué et l'indice-cible.

L'étude s'est faite à l'aide d'un programme Matlab, de HRTF mesurées sur le mannequin KEMAR par Gardner et Martin [GM94]. Les haut-parleurs simulés sont des monopôles omnidirectionnels et le banc composé de 24 haut-parleurs espacés de 15 cm sauf si précisé autrement. Dans la suite, les axes du repère cartésien seront placés de telle sorte que le banc de haut-parleurs est dirigé selon l'axe des abscisses de  $x = -1.8m$  à  $x = 1.8m$ , à une position  $y = 0$ , le banc de microphones virtuels est parallèle au banc de haut-parleurs, de même longueur et situé à une position  $y = -2m$ . Une source sera dite centrée si elle est à la position  $x = 0$ . Toute les distances seront exprimées en mètres et les temps en millisecondes sauf si précisé autrement.

Les fréquences étudiées vont de 50 à 18000 hz divisées en 24 bandes ERB. (cf section 3.2)

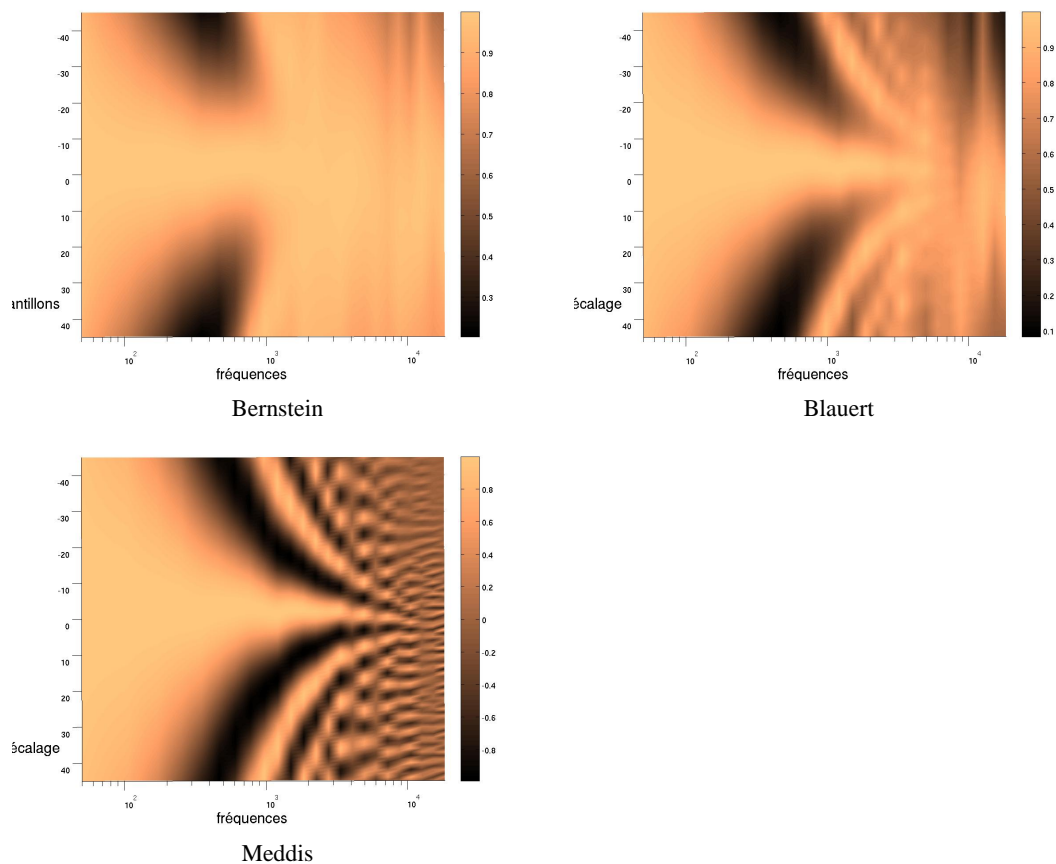
### 6.1 Comparaison des méthodes d'extraction d'enveloppe pour le calcul de l'itd

Comme nous l'avons dit précédemment, l'itd est donné par la différence interaurale de phase en basses fréquences mais elle est plus fortement corrélée à la différence interaurale d'enveloppes en hautes fréquences. Lorsque nous parlerons d'itd dans la suite, cela signifiera la différence interaurale d'enveloppes.

#### 6.1.1 Comparaisons des fonctions de corrélation interaurales (iacf)

Tout d'abord les valeurs de l'iacf par les trois modèles sont différentes, même si elles indiquent globalement les mêmes tendances. En très basses fréquences les valeurs de l'iacf sont proches de 1 pour toutes les valeurs d'itd entre  $-1ms$  et  $1ms$ . Entre environ 200 et 2000 hz, la localisation se fait plus précise et au-dessus la distribution des maxima locaux sur chaque bande de fréquence est très floue. Les modèles de Bernstein et de Blauert continuent d'indiquer une direction préférentielle en hautes fréquences, même si différente de la direction indiquée en basses fréquences, contrairement au modèle de Meddis qui n'indique aucune direction en particulier (cf Tab 4). L'apparition de motifs dus à l'aliasing est très nette pour le modèle de Meddis alors que les deux autres modèles montrent un étalement des valeurs élevées d'iacf.

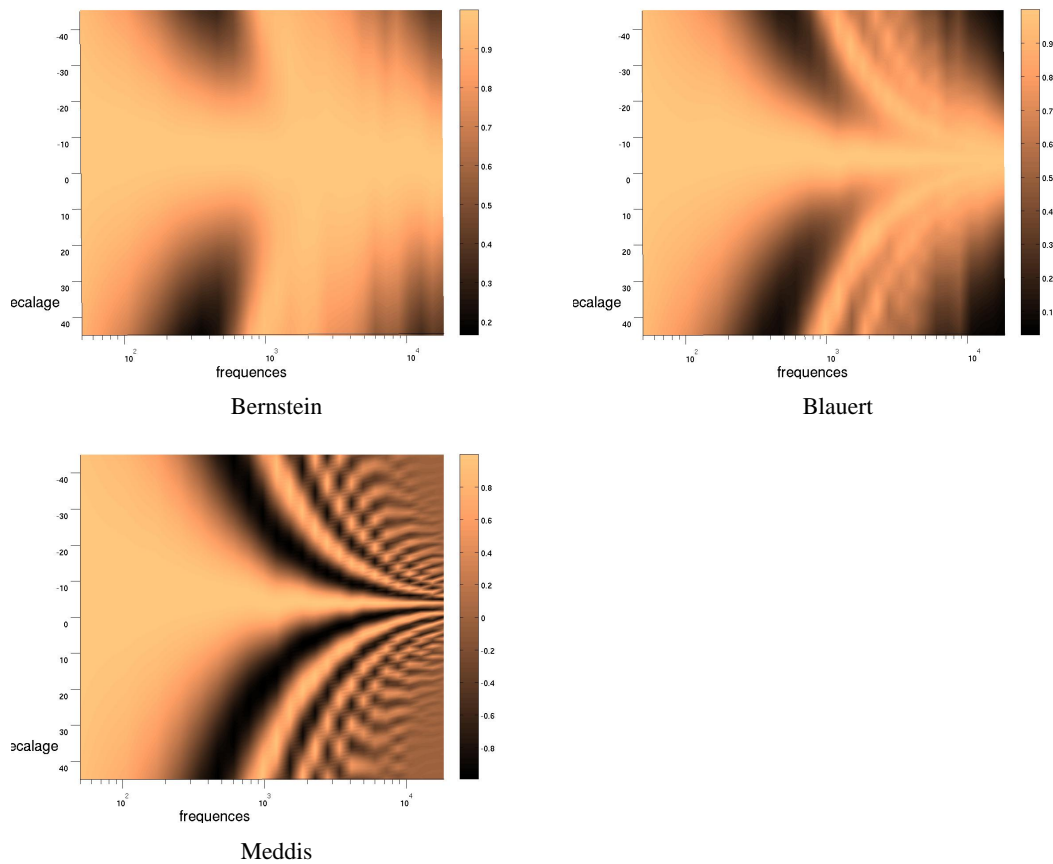
Le modèle de Meddis, indique un IACC très faible en hautes fréquences, ce qui veut dire que les signaux arrivant aux oreilles droite et gauche sont très différents. Selon ce modèle la localisation serait donc très difficile en hautes fréquences (cf fig 6).



TAB. 3 – iacf par Blauert, Meddis et Bernstein pour une source simulée par la WFS, centrée, à 10 m derrière les haut-parleurs et pour une position d’écoute d’abscisse 1. En abscisses les fréquences en Hz, en ordonnées les valeurs d’itd correspondant aux valeurs de l’iacf (en nombre d’échantillons), en cote les valeurs de l’iacf.

On peut voir dans Tab 5 que les trois modèles sont cohérents entre eux lors du calcul de l’IACC en basses fréquences. Les modèles de Blauert et Meddis indiquent la même tendance sur les erreurs même si le modèle de Meddis indique plus d’erreurs. Le modèle de Bernstein indique très peu d’erreurs. Les modèles de Blauert et de Bernstein diffèrent par le nombre et la position des itd équiprobables (cf Tab 5, colonne de gauche). Les itd équiprobables par Meddis sont très peu nombreux et semblent être aléatoirement distribués (cf Tab 5, à droite).

Pour résumer, le modèle de Bernstein indique une position ambiguë de la source. Les itd en hautes fréquences ne fournissent peu d’indications du fait du grand nombre d’itd équiprobables (IACC élevé et itd équiprobables nombreux et regroupés). Ceci pourrait conduire à un étalement voire à un dédoublement de la source. Le modèle de Blauert indique une position de source plus ambiguë encore (IACC élevé et itd équiprobables peu nombreux et dispersés). Ceci montre une plus faible cohérence des hautes fréquences entre elles et pourra dans une plus grande mesure conduire à un dédoublement de la source. Le modèle de Meddis indique que l’itd en hautes fréquences contribuera très peu à la localisation de la source (IACC faible et itd équiprobables très peu nombreux et très dispersés) et n’influenceront pas la localisation permise par les basses fréquences. En effet, il semble que la direction indiquée par les basses fréquences prime sur les hautes fréquences ([WK92]). Un IACC faible en hautes fréquences ne ferait



TAB. 4 – iacf calculés à partir des HRTF mesurées par les modèles de Blauert, Meddis et Bernstein pour une source idéale centrée, à 10 m derrière les haut-parleurs et pour une position d’écoute d’abscisse 1. En abscisses les fréquences en Hz, en ordonnées les valeurs d’itd correspondant aux valeurs de l’iacf (en nombre d’échantillons), en cote les valeurs de l’iacf.

qu’exacerber cette tendance.

Il faut de plus rappeler que l’itd n’est à priori pas l’indice prépondérant en hautes fréquences (cf [MM02]), et que donc ces résultats sur la localisation grâce à l’itd en hautes fréquences ne permettent pas de conclure sur la localisation sans connaissance préalable des valeurs de l’ild.

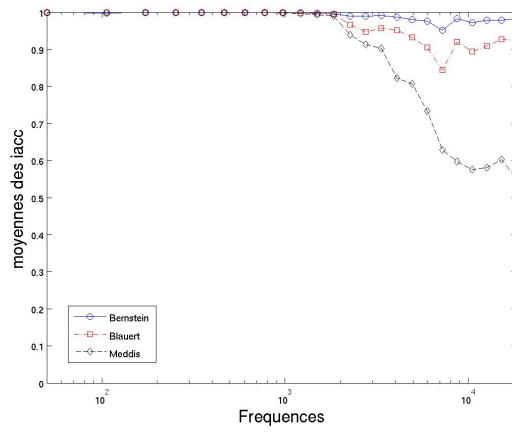
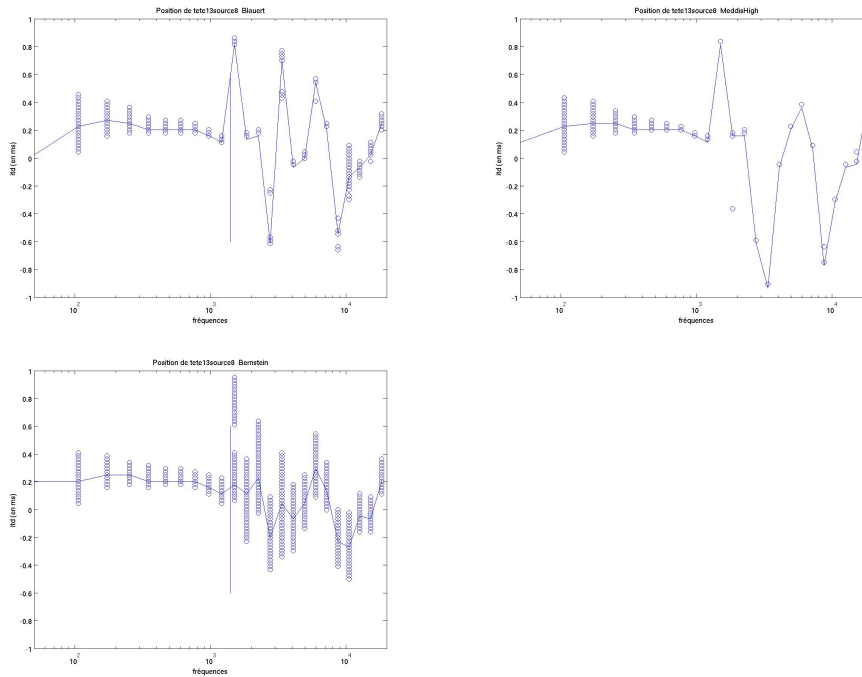
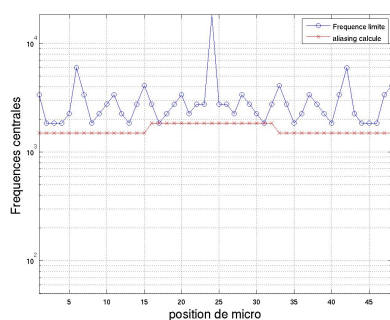


FIG. 6 – Valeurs de l’IACC pour les trois modèles pour la même source et la même position d’écoute que Tab 4. En abscisses les fréquences en Hz, en ordonnées les valeurs de l’IACC

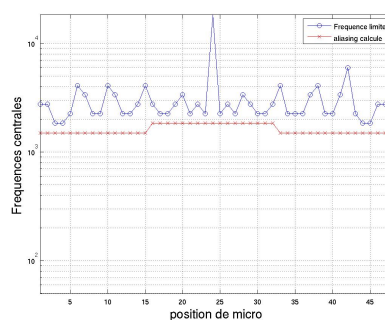


TAB. 5 – itd équiprobables par Blauert, Meddis et Bernstein pour la même source et la même position d’écoute que Tab 4. Les traits pleins représentent les iacc, les ronds les itd équiprobables, la barre verticale la fréquence d’aliasing calculée. En abscisses les fréquences en Hz, en ordonnées les valeurs d’itd en ms

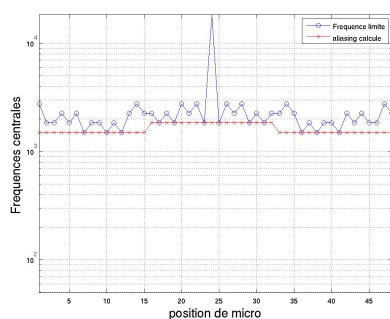
## 6.1.2 Comparaison des fréquences d'aliasing



Bernstein



Blauert



Meddis

TAB. 6 – Position de la fréquence d'aliasing calculées et de la fréquence-limite pour la même source que Tab 4. En abscisses les positions de microphones, en ordonnées les fréquences en Hz

La fréquence-limite est la première bande ERB pour laquelle l'erreur est supérieure à 0.1 ms. La source considérée sur le tableau 6 est une source centrée à 10 m derrière le banc de haut-parleurs.

En hautes fréquences, les erreurs apparaissent, mais plus souvent au-dessus des fréquences d'aliasing calculées selon la méthode proposée par Cordeel [Cor06b] qu'en-dessous (cf Tab 6). La fréquence d'aliasing est plus basse pour le modèle de Meddis, ce qui indique que la plage d'erreurs serait plus grande que celle indiquée par les deux autres indices.

## 6.1.3 Comparaison par les indices de droiture

Nous avons vu que la droiture [TBA01] permet de mesurer la distance moyenne des IACC équiprobables à la direction donnée par les IACC en basses fréquences. Nous appliquons ici cet indice à la comparaison des extracteurs d'enveloppe (cf fig 7).

Le modèles de Bernstein indique le moins d'erreurs et le modèle de Meddis indique le plus d'erreurs. Cela est cohérent avec les résultats obtenus précédemment (Tab 5. On voit ainsi que les IACC équiprobables selon Meddis sont le plus dispersés, indiquant ainsi que la localisation est plus difficile.

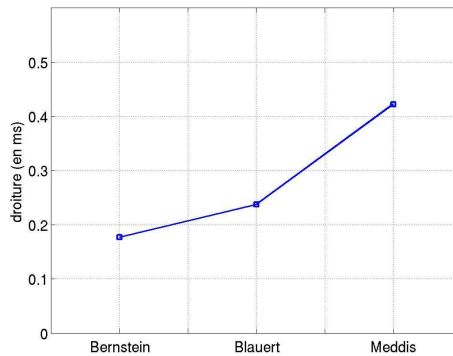


FIG. 7 – Droiture pour les 3 extracteurs d’enveloppe pour la même source et la même position d’écoute que Tab 4. En ordonnées la moyenne de la distance en ms

#### 6.1.4 Conclusion

Suite à l’étude bibliographique présentée dans la partie 3, nous avons étudié ici les différences entre les 3 modèles d’extraction d’enveloppes présentés (Bernstein, Blauert, Meddis) au moyen des indices de localisation pertinents. Nous nous sommes attachés à montrer les caractéristiques de ces modèles qui mettront en valeur des différences perceptibles auditivement. En effet si les erreurs sont très fortes sur la localisation de sons en hautes fréquences seules et sans influence de l’ild, le modèle de Meddis sera alors choisi comme le modèle le plus adapté. Dans le cas contraire, les modèles de Bernstein ou Blauert conviendront mieux.

Il faut cependant noter que le choix d’un modèle d’extraction d’enveloppes sera difficile sinon impossible sans égalisation des ILD qui est l’indice de localisation prépondérant en hautes fréquences [MM02]. Ce point ne sera pas traité plus en détail dans la suite et mériterait une étude à lui seul.

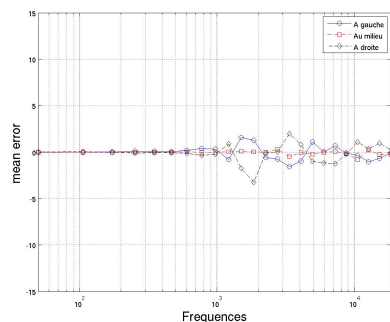
Nous allons maintenant plus nous intéresser aux erreurs selon différentes configurations. Nous regarderons ainsi les erreurs qui apparaissent lorsque l’auditeur bouge à l’intérieur de la zone d’écoute ou lorsqu’il occupe des positions précises. Nous nous intéresserons aux petites variations de position, de l’ordre de 5cm, ce qui correspond au fait de bouger la tête, ainsi qu’aux variations de l’ordre de 50cm, ce qui correspond à faire un pas sur le côté. Il n’est pas nécessaire de s’intéresser aux variations plus grandes car la mémoire auditive humaine est très limitée. Nous considérerons également l’influence de la position d’écoute par rapport à la position de la source et par rapport au banc de haut-parleurs, l’espacement entre les haut-parleurs et l’effet du filtre diffus.

### 6.2 Dépendance des erreurs à la position latérale d’écoute : Mise en évidence d’un phénomène d’accumulation

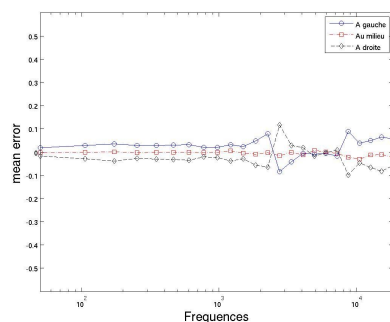
L’accumulation est un phénomène dû à la troncature du banc de haut-parleurs. Il existe en effet un déséquilibre de l’énergie arrivant à chaque oreille lorsqu’on ne se trouve pas au milieu du banc car il y a plus de contributions de haut-parleurs arrivant à une oreille qu’à une autre. Ainsi, plus on s’approche de la limite du banc, plus les erreurs augmentent. Ce phénomène est visible dans l’ild à cause d’un défaut d’énergie à une oreille dû au masquage de la tête, particulièrement en hautes fréquences. Il se voit également dans l’itd car les contributions de chaque haut-parleur arrivant successivement aux oreilles entraînent une enveloppe extraite beaucoup plus longue pour l’oreille du côté où il se trouve le plus de

haut-parleurs. On parle alors d'accumulation temporelle.

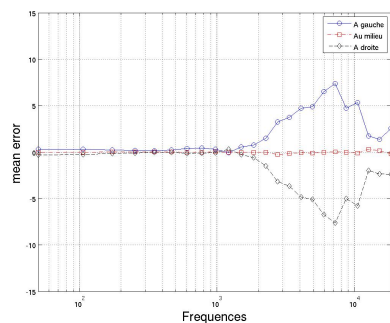
Lorsqu'une source sonore est très proche du banc, l'énergie rayonnée se répartit sur un petit nombre de haut-parleurs. Plus la source est éloignée du banc, plus l'énergie est répartie également entre les haut-parleurs et donc plus le phénomène de troncature du banc est audible. L'accumulation est ainsi d'autant plus importante que la source sonore est éloignée du banc, comme on le voit sur Tab 7.



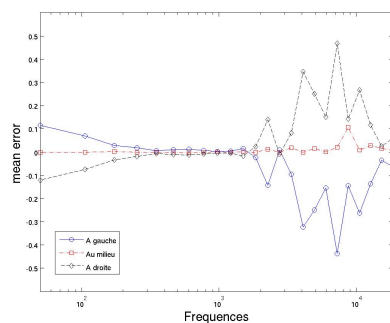
Erreur sur l'ILD pour une source proche du banc



Erreur sur l'ITD pour une source proche du banc



Erreur sur l'ILD pour une source loin du banc



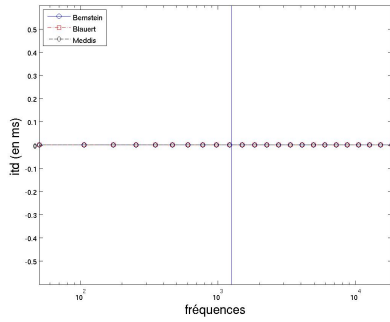
Erreur sur l'ITD pour une source loin du banc

TAB. 7 – Mise en évidence de l'accumulation. En abscisses les fréquences en Hz, en ordonnées les positions de microphones, en cote la valeur de l'erreur sur l'indice considéré

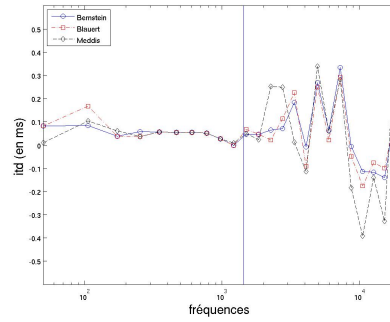
### 6.3 Influence des positions relatives d'écoute et de source

Nous avons ici considéré la différence d'erreur sur l'itd pour les trois extracteurs d'enveloppe entre une position en face de la source et une position non en face. La source considérée sur la figure 8 est une source à 10m derrière le banc de haut-parleurs et à une abscisse de 1,8 m. Comme on le voit sur la figure, les erreurs sont nulles lorsque la position d'écoute est en face de la source alors que l'accumulation est pourtant importante (source éloignée du banc et position d'écoute excentrée). L'erreur est beaucoup plus grande pour une position d'écoute centrée alors que cette position ne subit pas d'accumulation (cf Tab 8, figure de droite).

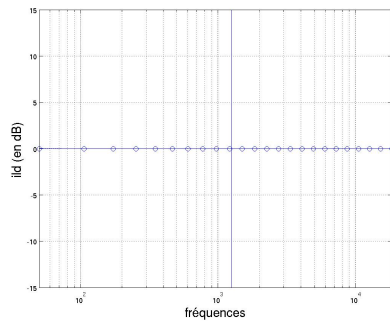




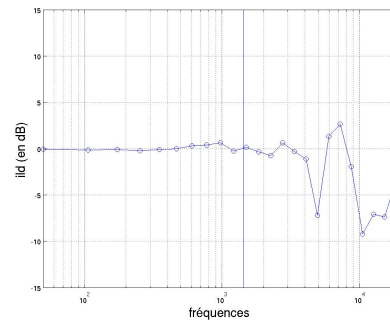
Position d'écoute en face de la source (itd)



Position d'écoute décalée de 1m par rapport à la position en face de la source (itd)



Position d'écoute en face de la source (ild)



Position d'écoute décalée de 1m par rapport à la position en face de la source (ild)

TAB. 8 – Erreur sur l'itd (1<sup>ère</sup> ligne) et sur l'ild (2<sup>ème</sup> ligne) pour une position d'écoute en face de la source (à gauche) ou non en face (à droite). En abscisses les fréquences en Hz, en ordonnées l'erreur sur l'indice

## 6.4 Influence des positions relatives d'écoute et de haut-parleurs

Lors d'écoutes, il fut mis en évidence une coloration du son troublant la localisation lorsque l'on bouge la tête de droite et de gauche. L'analyse objective a montré cette caractéristique du champ restitué par une différence importante des erreurs entre des positions de microphone en face d'un haut-parleur actif et entre deux haut-parleurs actifs. L'étude s'est faite au moyen des SItDVI et SildVI de la manière suivante.

Nous considérons  $N$  microphones. Pour chaque microphone  $k$  tel que  $2 \leq k \leq N-1$ , nous calculons la variation de l'erreur entre ce microphone et les microphones aux positions  $k-1$  et  $k+1$ . Nous moyennons ensuite sur les microphones utilisés (ici  $(N-2) \times 2(N-2)$ ).

Nous avons utilisé plusieurs configurations :

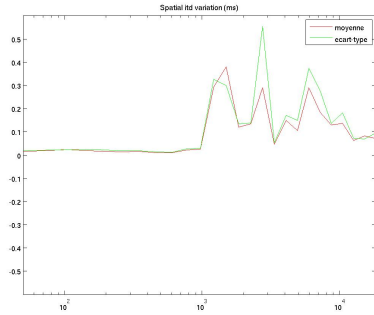
1. Nous avons  $P$  haut-parleurs. Les  $N$  microphones sont situés *en face* des haut-parleurs. Donc  $N = P$ . L'écart entre deux microphones est de 15 cm.
2. Nous avons  $P$  haut-parleurs. Les  $N$  microphones sont situés *entre* deux haut-parleurs. Donc  $N = P - 1$ . L'écart entre deux microphones est de 15 cm.
3. Nous avons  $P$  haut-parleurs. Les  $N$  microphones sont situés tels que *1 microphone sur 2 est en face* d'un haut-parleur. Donc  $N = 2P - 1$ . L'écart entre deux microphones est de 7.5 cm.

La comparaison de ces configurations nous permet de déterminer si l'instabilité dans la localisation qui apparaît lorsque l'on bouge la tête est due au fait qu'il y a plus d'erreurs lorsque l'on est en face ou pas des haut-parleurs ou est due au fait de passer d'une position en face à une position entre les haut-parleurs.

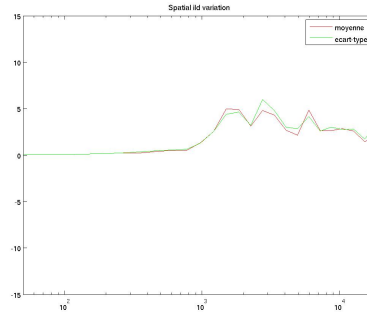
Les résultats de cette analyse sont présentés dans Tab 9.

En basses fréquences la variation de l'erreur sur chaque bande de fréquence est faible, cela veut dire que le champ sonore est cohérent entre les différentes positions d'écoute. Un auditeur se déplaçant parallèlement au banc de haut-parleurs conservera ainsi des impressions de position de la source en cohérence avec ses déplacements. Ces résultats étaient prévisibles, car on sait que la synthèse du champ en basses fréquences est presque parfaite. Cela montre que les erreurs d'atténuation et de diffraction n'influent pas sur la localisation en basses fréquences. En hautes fréquences la variation de l'erreur est élevée, cela veut dire que les erreurs ont un comportement erratique, elles n'évoluent pas de manière continue en fonction de la position en  $x$ . Un auditeur pourra ainsi percevoir des "sauts" de la source d'une position à une autre, d'une façon qui ne correspondra pas à ses déplacements.

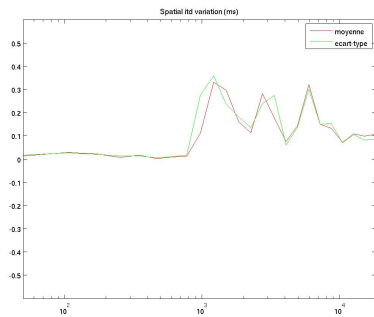
On voit que les variations d'erreurs sont proches dans les deux premières configurations. La variation de l'erreur est plus grande pour la configuration 3, lorsque l'on passe d'en face à entre des haut-parleurs. Cela veut dire que l'instabilité dans la localisation est due *au fait de bouger* plutôt qu'au fait qu'une position statique en face ou entre les haut-parleurs induise plus d'erreurs que l'autre.



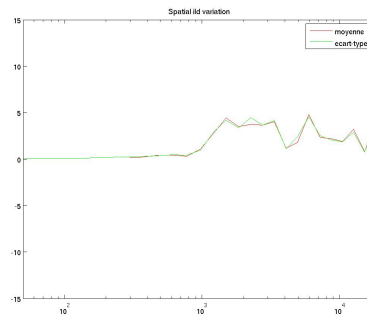
Configuration 1, SItdVI



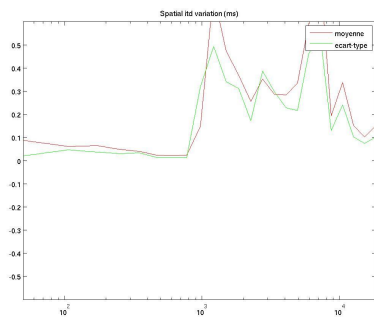
Configuration 1, SIIdVI



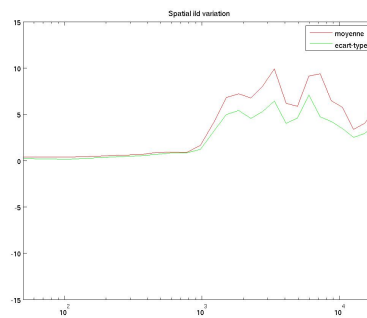
Configuration 2, SItdVI



Configuration 2, SIIdVI



Configuration 3, SItdVI



Configuration 3, SIIdVI

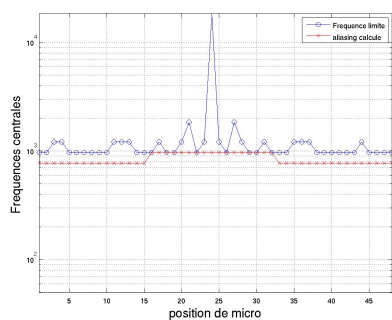
TAB. 9 – Comparaisons des SItdVI et des SIIdVI pour les 3 configurations. En abscisses les fréquences en Hz, en ordonnées la valeur de la variation, en dB pour le SIIdVI, en ms pour le SItdVI

## 6.5 Influence de l'espacement entre les haut-parleurs

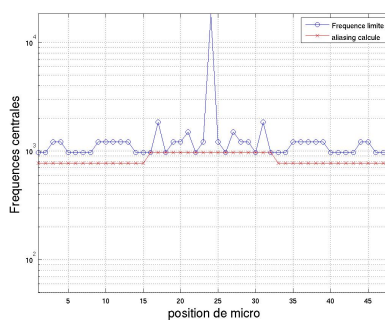
Nous allons ici nous intéresser au cas où le banc compte 12 haut-parleurs espacés de 30 cm, ce qui constitue un banc de la même longueur que ce qui était considéré jusqu'à maintenant mais avec un haut-parleur actif sur deux. En théorie, ce phénomène a pour effet de faire baisser la fréquence d'aliasing. Nous allons ici étudier l'effet de cet espacement sur les indices de localisation par rapport à la configuration avec 24 haut-parleurs.

**Influence sur les erreurs d'itd et d'ild :** Les erreurs sur ces deux indices sont plus importantes dans la configuration 12 haut-parleurs.

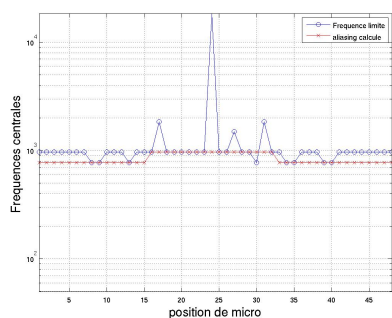
**Influence sur la fréquence d'aliasing :** Comme on le voit sur Tab 10, les fréquences d'aliasing calculées sont plus basses que sur Tab 6, de même pour les fréquences d'aliasing observées. Ceci est conforme aux résultats attendus. Il est cependant important de noter que les fréquences d'aliasing observées dans la configuration 12 haut-parleurs sont souvent plus basses que les fréquences d'aliasing calculées, ce qui est le contraire dans la configuration 24 haut-parleurs.



Bernstein



Blauert

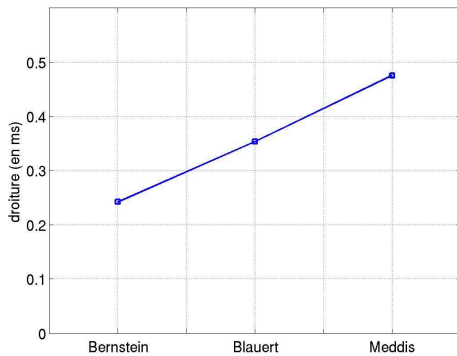


Meddis

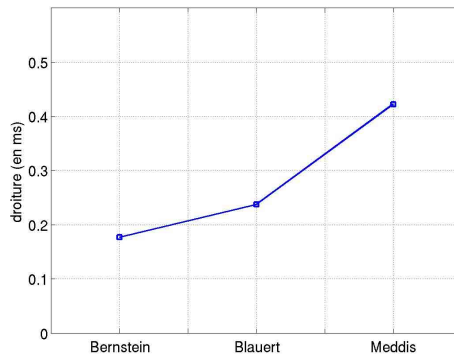
TAB. 10 – Position des fréquences d'aliasing calculées et observées. En abscisses les positions de microphones en mètres, en ordonnées les fréquences en Hz

**Influence sur la droiture :** Les motifs d'IACC sont moins cohérents entre les bandes de fréquence pour la configuration 12 haut-parleurs que dans la configuration 24 haut-parleurs (cf fig 11).

La configuration 12 haut-parleurs montre une baisse de la fréquence d'aliasing qui se répercute sur les



pour un espacement de 15 cm entre les haut-parleurs



pour un espacement de 15 cm entre les haut-parleurs

TAB. 11 – Droiture pour les 3 extracteurs d’enveloppe pour une source simulée par la WFS, centrée, à 10 m derrière les haut-parleurs et pour une position d’écoute d’abscisse 1. En ordonnées la moyenne de la distance en ms

indices de localisation par une hausse des erreurs. Ces résultats sont conformes aux résultats théoriques. Nous devons néanmoins déterminer l’impact de cette différence par une expérience subjective.

## 6.6 Conclusion

Nous avons étudié dans cette partie les particularités de différentes méthodes d’extraction d’enveloppe modélisant le fonctionnement des cellules ciliées dans le traitement des hautes fréquences. Les différences ainsi exhibées permettront de choisir le modèle le plus adéquate pour l’étude de la restitution d’un champ sonore au moyen de la WFS dans le cas des hautes fréquences.

Nous avons également mis en évidence un certain nombre de phénomènes entraînant des erreurs sur la restitution du champ sonore en hautes fréquences. La troncature du banc de haut-parleurs entraîne un déséquilibre de l’énergie arrivant à chaque oreille dans le cas d’une position d’écoute non centrée. Elle entraîne également un étalement de l’enveloppe extraite par l’oreille du côté où il se trouve le plus de haut-parleurs. Nous avons appelé ce phénomène "accumulation".

Le fait de ne pas être en face de la source sonore entraîne une augmentation des erreurs. L’espacement entre les haut-parleurs joue également un rôle dans l’amplitude des erreurs sur les indices de localisation. Enfin le comportement erratique des erreurs en fonction de la position d’écoute en  $x$  entraîne une instabilité sur la localisation de la source lors de déplacements de l’auditeur à l’intérieur de la zone d’écoute.

## 7 Expérience

### 7.1 Buts

Dans le but de valider les observations effectuées à l'aide des indices objectifs, une expérience a été effectuée. Dans cette expérience le sujet va devoir indiquer l'azimut de divers sons dont les caractéristiques diffèrent. Plusieurs dimensions vont ainsi pouvoir être testées. Pour des questions de durée de l'expérience, il a été nécessaire de faire un choix et donc certains paramètres étudiés dans la partie objective n'ont pas été testés. Nous avons choisi comme dimensions :

- l'écartement entre les haut-parleurs.
- la position des sources.
- la position d'écoute.
- les fréquences contenues dans le son.

A partir de ces dimensions il sera possible d'étudier les phénomènes suivants :

- effet de l'accumulation, effet de la distance de la source au banc de hp sur ce phénomène.
- effet d'avoir la source en face ou non.
- effet de l'espacement entre les hp.
- taille apparente de la source.
- Influence des fréquences contenues dans le son sur la localisation.

### 7.2 Protocole

Le test se compose de plusieurs séquences correspondant à des situations différentes et diffusées sur casque d'écoute. Chaque séquence comprend deux stimuli de 200 ms chacun qui alternent. Les HRTF utilisées dans ce test sont individuelles. Elles ont été mesurées dans le plan horizontal, de 0 à 355 degrés par pas de 5 degrés. Ces HRTF ont été ensuite interpolées pour obtenir des réponses tous les 2.5 degrés.

Le premier stimulus, le stimulus-test, est la simulation du signal provenant de la WFS aux oreilles du sujet. Lorsque les caractéristiques de la source virtuelle ont été définies, on calcule le signal envoyé à chaque haut-parleur, considéré comme un monopôle idéal. La somme des contributions de chaque haut-parleur forme la réponse du banc de haut-parleurs. La réponse du banc arrivant à chaque oreille est ensuite calculée et ce signal est convolué aux HRTF du sujet.

Le deuxième stimulus, le stimulus de référence, est simplement la convolution des HRTF du sujet à un son. Le sujet peut agir sur la position de ce stimulus au moyen de 2 touches du clavier, pour changer la HRTF utilisée. Le son utilisé pour ce stimulus est le même bruit blanc filtré de la même manière que le stimulus-test. Pour les différencier, le stimulus de référence est modulé en amplitude à 20 Hz.

Le but de l'expérience est, pour chaque séquence d'aligner le stimulus de référence sur le stimulus-test, de sorte que les deux stimuli semblent provenir de la même direction. Le programme récupère ensuite l'azimut du stimulus de référence tel qu'indiqué par le sujet.

Le test est composé de 48 stimuli-test générés par croisement des conditions suivantes :

- 2 espacements de haut-parleurs : 15 et 30 cm
- 3 sources virtuelles : une onde plane à 0 degré et 2 sources situées 1 mètre derrière le banc de haut-parleurs. L'une est située à l'abscisse 0 et l'autre à l'abscisse 1,5 m.
- 2 positions d'écoute, toutes les deux à 3 mètres devant les haut-parleurs : une à la position centrée ( $x = 0$ ) et l'autre à la position excentrée ( $x = 1,5m$ ).
- 4 bruits blancs filtrés passe-haut à 150, 700, 1500 et 3000 Hz avec 60 dB d'atténuation dans la bande coupée.

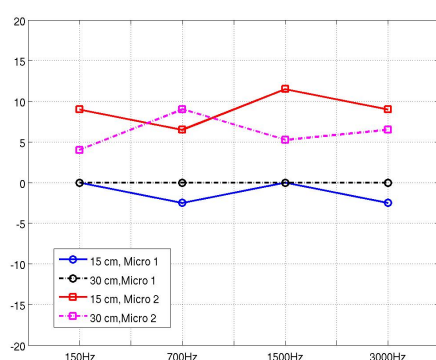
L'ordre de diffusion de ces stimuli est généré aléatoirement et de façon indépendante pour chaque

sujet. Six sujets ont effectué ce test. La durée d'une session variait entre 15 à 20 minutes selon les sujets et chacun d'entre eux a effectué six sessions.

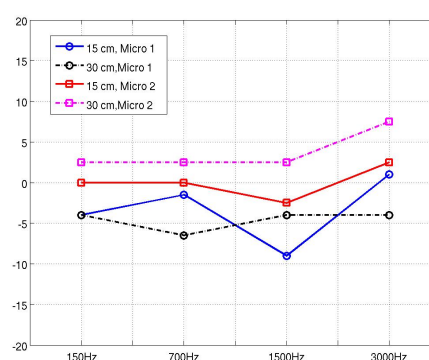
### 7.2.1 Résultats et comparaisons avec les résultats objectifs

L'analyse des résultats se fait en comparant l'azimut indiqué par le sujet dans chaque séquence avec l'azimut réel de la source par rapport à la position du sujet. Ce calcul nous donne ainsi l'erreur commise par les sujet sur l'azimut de la source. Dans les tableaux 13 et 12, "micro 1" signifie la position d'écoute centrée et "micro 2" la position excentrée. "15 cm" et "30 cm" indiquent l'espacement entre les haut-parleurs.

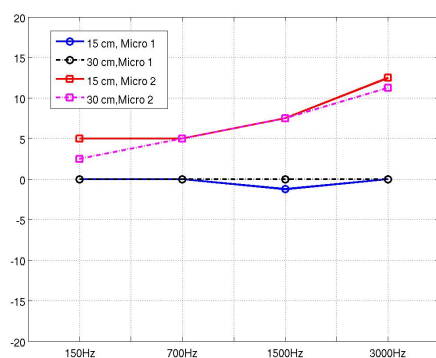
Les valeurs des indices objectifs pour la configuraion de sources et de positions de source du test sont présentées en annexe.



(a) Source centrée

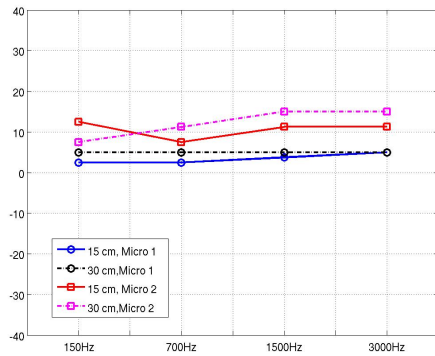


(b) Source excentrée

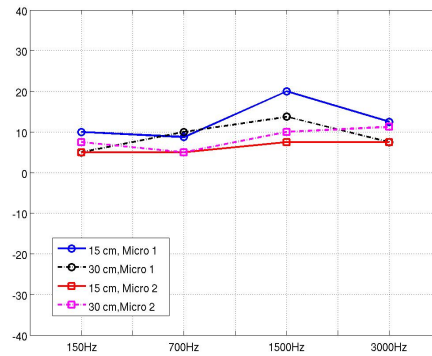


(c) Onde plane à 0°

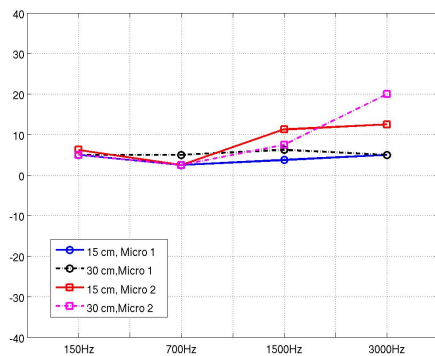
TAB. 12 – Valeur médiane des erreurs de localisation pour les trois sources considérées, les deux espacements de haut-parleurs et les deux positions de source. En abscisses les fréquences de coupure des filtres passe-haut (en Hz), en ordonnées la différence des percentiles à 75% et à 25% (en degrés)



(a) Source centrée



(b) Source excentrée



(c) Onde plane à 0°

TAB. 13 – différence des percentiles d’erreurs à 75% et à 25 % pour les trois sources considérées, les deux espacements de haut-parleurs et les deux positions de source. En abscisses les fréquences de coupure des filtres passe-haut (en Hz), en ordonnées la différence des percentiles à 75% et à 25 % (en degrés)

**La valeur médiane des erreurs** (cf Tab 12) est la valeur pour laquelle 50% des erreurs lui sont supérieures (et donc 50% inférieures). Ceci nous donne une valeur moyenne de l’erreur facilement exploitable et possède l’avantage par rapport à la moyenne de réduire l’influence des valeurs extrêmes.

**différence des percentiles d’erreurs à 75% et à 25 %** (cf Tab 13) nous donne l’intervalle dans lequel 50% des erreurs ont comprises. Ceci équivaut à un écart-type de l’erreur sur la localisation de la source. On peut ainsi relier cette différence à la perception de la largeur de la source. Plus la source paraît étalée, plus l’estimation de la position de cette source va être complexe. Dans le cas d’une source très large, la localisation revient à estimer la position du centre de cette source, ce qui est plus dur que d’estimer la position d’une source ponctuelle.

## 7.2.2 Etude du phénomène d’accumulation

Nous avons vu dans la partie de l’étude objective que le fait d’être situé à une extrémité du banc de haut parleurs entraînait un excès d’énergie ainsi qu’un étalement de l’enveloppe extraite à l’une des oreilles. Ce phénomène augmentait ainsi les erreurs sur les indices de localisation.

L’expérience nous a permis de rendre compte de l’impact perceptif de ce phénomène. Dans le tableau 12, les positions d’écoute excentrées (micro 2) sont les positions sujettes à l’accumulation. Nous avons



ainsi pu comparer les erreurs avec celles commises aux positions centrées (micro 1). Les résultats sont le plus flagrant pour l'onde plane (figure (c)) où ce phénomène n'est pas couplé avec celui de la position relative de l'auditeur et de la source (cf section 7.2.3).

Nous voyons donc sur cette figure que les erreurs sont quasiment nulles pour le micro centré, et ce même avec un espacement de haut-parleurs de 30 cm. Elles augmentent significativement pour une position excentrée et ce quelque soit l'espacement entre les haut-parleurs. L'expérience subjective a ainsi confirmé les résultats de l'analyse objective, montrant ainsi un effet important de la troncature du banc de haut-parleurs. Elle tend de plus à montrer que l'accumulation serait prépondérante par rapport à l'espacement des haut-parleurs par rapport aux signaux test considérés.

Nous avons précédemment indiqué que la distance de la source au banc de haut-parleurs semblait avoir une influence sur l'amplitude des erreurs. En effet dans le cas où une source est située loin derrière les haut-parleurs (le cas extrême étant l'onde plane), l'énergie se répartit sur un grand nombre de haut-parleurs, ce qui exacerbe le défaut d'énergie à une oreille dû à la troncature du banc. La comparaison des figures (b) et (c) des figures 12 et 13 nous donne l'influence de la distance de la source au banc. Nous voyons que l'erreur est légèrement plus haute sur l'onde plane pour les positions d'écoute excentrées en hautes fréquences, ce qui peut être attribué au masquage de la tête (efficace seulement pour les longueurs d'onde plus petites que le diamètre de la tête). Le masquage induirait un défaut d'énergie à une oreille, qui n'est pas compensé du fait de la troncature du banc. Cependant les résultats ne sont pas flagrants et une étude plus poussée est nécessaire pour confirmer ce résultat.

### **7.2.3 Influence des positions relatives d'écoute et de source**

Nous avons comparé dans la section 6.3 les différences sur les erreurs lorsque l'auditeur est situé en face de la source ou pas. Nous avons ainsi pu mettre en évidence de plus grandes erreurs sur les indices de localisation lorsque l'auditeur n'est pas en face de la source.

L'expérience subjective confirme ce résultat. La figure (b) des figures 12 et 13 montre les erreurs pour la source excentrée. Ce graphique montre que la localisation de la source a été plus précise dans le cas où l'auditeur est en face de la source (Micro 2) que lorsqu'il ne l'est pas (Micro 1). La figure (a) des tableaux 12 et 13 montrent le même résultat dans le cas de la source centrée. La localisation a été meilleure pour la position d'écoute en face de la source (Micro 1) que pour la position non en face (Micro 2).

On peut remarquer au passage que ce phénomène est prépondérant par rapport à l'accumulation. Sur la figure (b) du tableau 12, la position en face de la source est excentrée (Micro 2), donc sujette à l'accumulation. Pourtant les erreurs à cette position sont plus faibles que les erreurs pour la position centrée (Micro 1) qui ne subit pas d'accumulation.

### **7.2.4 Effet de l'espacement entre les haut-parleurs**

Le tableau 12 montre pour chaque position de source les valeurs médianes de l'erreur sur la localisation pour un espacement de haut-parleurs de 15 cm et un espacement de 30 cm. L'analyse objective avait montré dans ce cas-ci une augmentation des erreurs des indices de localisation, une moins grande cohérence des erreurs selon les positions d'écoute et selon les fréquences (cf section 6.5) et les écoutes préliminaires semblaient indiquer un étalement de la source.

Comme il a été dit dans la section 7.2.2, l'espacement des haut-parleurs semble avoir une moindre influence sur les erreurs que l'accumulation. De plus le fait que la position d'écoute soit ou non en face de la source induit également des erreurs (cf section 7.2.3). Au vu des figures 13 et 12, l'analyse de l'influence de l'espacement entre les haut-parleurs semble difficile. Les résultats théoriques semblent confirmés dans certains cas (figure (a) Tab 13) ou infirmés dans d'autres (figure (b) Tab 13). Il semble

dès lors impossible de conclure sur l'impact perceptif de cette dimension. Si impact il y a, il est en tous cas plus faible que l'accumulation ou la position relative de la source par rapport à l'auditeur.

Il faut cependant relativiser les résultats, l'expérience est très sommaire concernant l'espacement des haut-parleurs. L'effet de l'espacement des haut-parleurs se voit surtout dans l'instabilité de la position de la source lors de déplacements de l'auditeur à l'intérieur de la source, ce qui n'a pas été envisagé dans le test subjectif.

### **7.2.5 Effet de la bande passante des sons**

Le calcul de la fréquence d'aliasing selon la méthode de Cordeel [Cor06b] a montré que la fréquence d'aliasing varie beaucoup pour les positions d'écoute et de sources considérées ici. Nous devrions en principe voir une différence dans les performances de localisation des sources selon la fréquence.

L'influence de ce facteur dans l'expérience subjective est ténue. On note une inflexion des courbes lors du passage de 700 à 1500 Hz. Cependant cette inflexion ne se remarque pas sur toutes les courbes et les erreurs ne sont pas tout le temps plus fortes qu'en basses fréquences (cf figure (a) Tab 12).

## **7.3 Discussion**

Nous avons ici apporté une contribution à la corrélation de phénomènes perceptifs et d'indices objectifs dans le cadre de la WFS. Le test psychoacoustique réalisé permet d'évaluer l'impact de certains phénomènes dans la reproduction de sources sonores au moyen de la WFS. La corrélation des deux types d'analyse nous permet même d'esquisser un classement des phénomènes induisant des erreurs dans la synthèse pour une position d'écoute fixe. Ainsi par ordre d'importance pour l'amplitude des erreurs, il nous semble avoir :

1. La position relative de la source et de l'auditeur.
2. La position d'écoute par rapport au banc de haut-parleurs (accumulation).
3. Fréquences contenues dans le son et espacement des haut-parleurs

Certaines dimensions telles que les fréquences contenues dans le son ou l'espacement des haut-parleurs nous ont donné des résultats confirmant seulement partiellement ceux de l'analyse objective. Il semble ainsi que les indices de localisation délivrés dans le cas d'un espacement de haut-parleurs de 30 cm et/ou avec des sons contenant uniquement des hautes fréquences sont suffisamment cohérents pour permettre une localisation efficace dans le cas où il n'y a pas de mouvements de l'auditeur. Une expérience supplémentaire pour évaluer l'impact de ces deux dimensions sans influence des deux premiers phénomènes est nécessaire pour produire des résultats plus flagrants.

## 8 Conclusion

Ce stage était destiné à corréler les erreurs sur les indices de localisation avec les résultats d'une expérience subjective. L'objectif était de mettre en évidence un certain nombre de phénomènes induisant des erreurs sur la synthèse de sources sonores au moyen de la WFS et réduisant ainsi la transparence du système.

Il a pour cela fallu procéder à une étude de la littérature sur différents sujets : le fonctionnement de l'organe auditif chez l'homme, les indices de localisation, la modélisation des ces indices et les caractéristiques de la Wave Field Synthesis. Nous avons ainsi été amenés à comparer plusieurs modèles pour l'extraction d'enveloppes en hautes fréquences telle qu'elle est effectuée par les cellules ciliées. Une étude bibliographique suivie d'une analyse par les indices de localisation a été effectuée. Nous avons ainsi mis en évidence des différences entre les 3 extracteurs d'enveloppe étudiés (Bernstein [BT96a], Blauert [Bla99], Meddis [Med86]).

La combinaison des études bibliographiques nous a permis de faire le choix des indices de localisation pertinents pour l'étude des erreurs sur la synthèse. Il a de plus été nécessaire de développer de nouveaux indices pour décrire les erreurs sur la dimension spatiale de la WFS (les erreurs qui apparaissent lorsqu'un auditeur se déplace à l'intérieur de la zone d'écoute) et sur la dimension fréquentielle (la cohérence des indices de localisation selon les fréquences). Suite à l'étude par les indices objectifs, plusieurs phénomènes ont pu être mis en évidence. Nous avons constaté que la position d'écoute par rapport à la position de la source, la position d'écoute par rapport au banc de haut-parleurs et l'espacement entre les haut-parleurs ont une influence sur les erreurs indiquées par les indices de localisation.

Une fois ces phénomènes mis en évidence, nous avons étudié leur impact perceptif lors d'une expérience. L'influence de la position d'écoute par rapport à la position de la source et par rapport au banc de haut-parleurs s'est trouvée confirmée alors que nous avons observé un impact étonnamment faible de l'espacement entre les haut-parleurs et des fréquences contenues dans le son diffusé. Les conclusions que nous avons tirées de ce test ne sont pas définitives. Il reste un grand nombre d'analyses à effectuer pour déterminer la robustesse des conclusions.

L'étude ici effectuée n'est qu'une esquisse de l'étude globale qu'il faudrait entreprendre pour rendre compte des caractéristiques de la WFS ainsi que de leur impact perceptif. Une expérience subjective pour déterminer l'extracteur d'enveloppes le plus adapté à l'étude des erreurs lors de la restitution d'un champ sonore par la WFS serait ainsi nécessaire. a été envisagée pour cela mais elle n'a pas pu être conduite faute de temps.

De nombreuses dimensions pourront être envisagées, telles que le type de haut-parleurs utilisés, la longueur du banc de haut-parleurs, la distance de l'auditeur par rapport à la source, le type de son diffusé, l'impact du filtre diffus sur les déplacements à l'intérieur de la zone d'écoute.

## A Erreurs sur les indices objectifs pour les positions de source et d'écoute du test psychoacoustique

Dans toutes les figures qui suivent, le modèle choisi pour le calcul de l'itd en hautes fréquences. Ce modèle semblait indiquer les mêmes tendances sur les erreurs que les deux autres modèles tout en montrant des amplitudes d'erreurs plus petites que le modèle de Meddis et plus grandes que le modèle de Bernstein.

### A.1 Pour un espacement de 15 cm entre les haut-parleurs

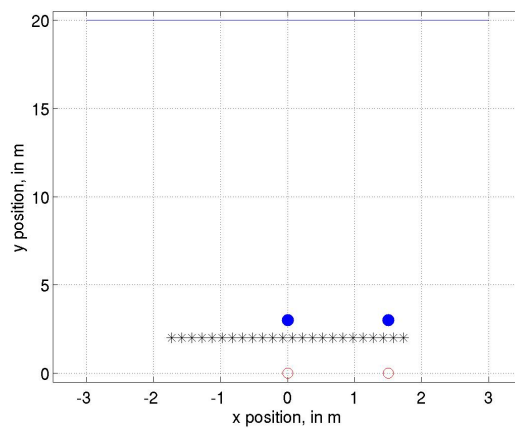
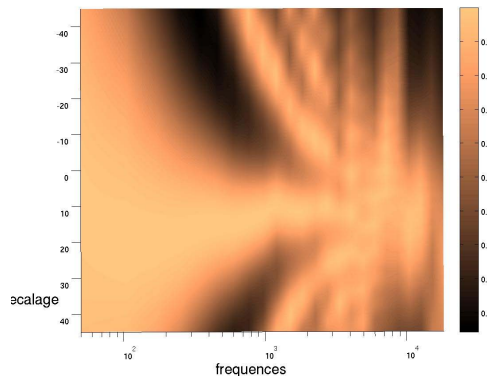
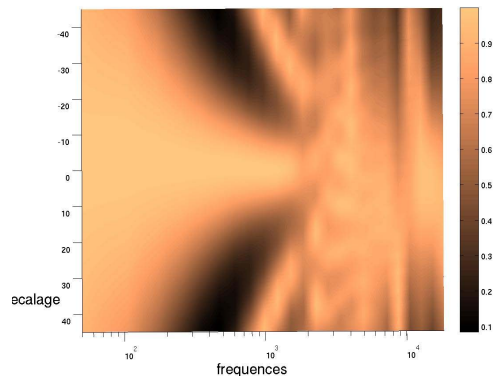


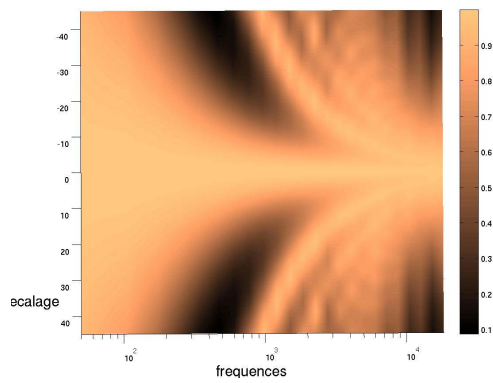
FIG. 8 – Position d'écoute, des haut-parleurs et des sources pour le test avec 24 haut-parleurs. les étoiles noires représentent les haut-parleurs, les rond rouges les positions d'écoute, les ronds bleus les sources centrées et excentrées. La ligne bleue représente l'onde plane



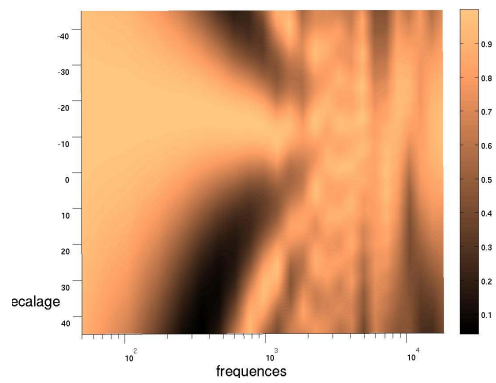
Source excentrée, position d'écoute centrée



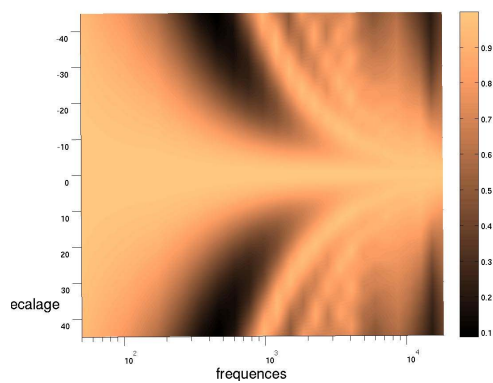
Source excentrée, position d'écoute excentrée



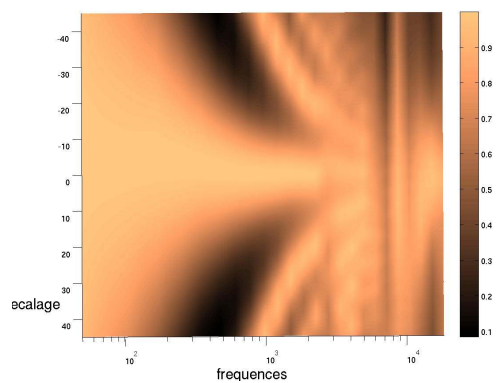
Source centrée, position d'écoute centrée



Source centrée, position d'écoute excentrée

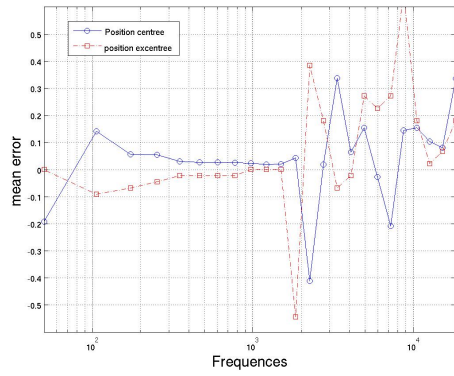


onde plane, position d'écoute centrée

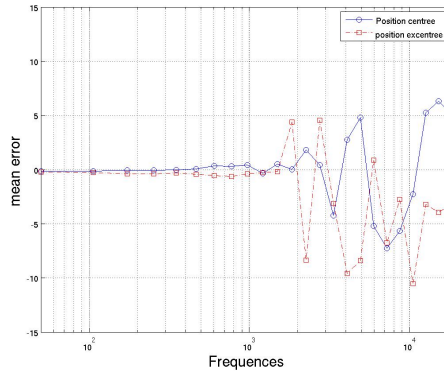


onde plane, position d'écoute excentrée

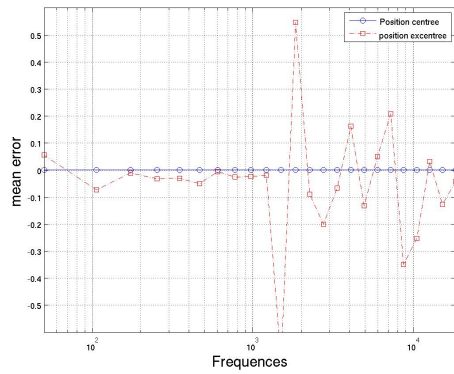
TAB. 14 – iacf par Blauert pour les 3 sources et les 2 positions d'écoute du test. En abscisses les fréquences en Hz, en ordonnées les valeurs d'itd correspondant aux valeurs de l'iacf (en nombre d'échantillons), en cote les valeurs de l'iacf.



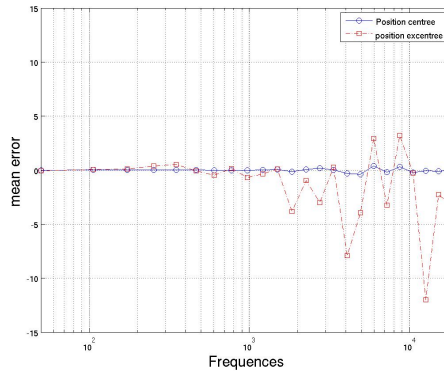
Erreur sur l'itd pour la source excentrée



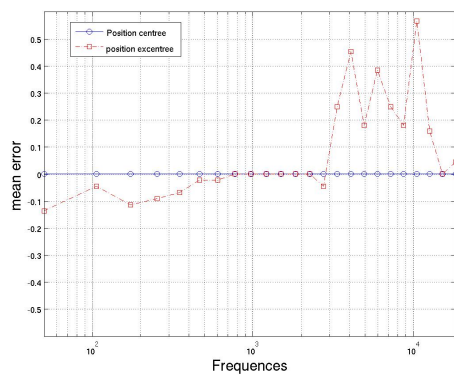
Erreur sur l'ild pour la source excentrée



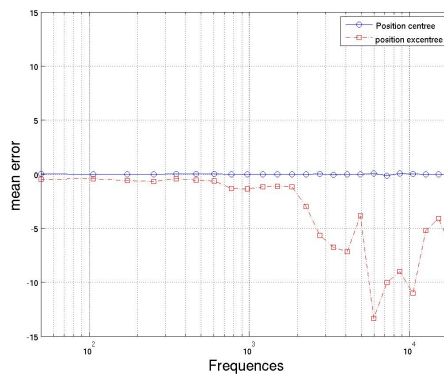
Erreur sur l'itd pour la source centrée



Erreur sur l'ild pour la source centrée

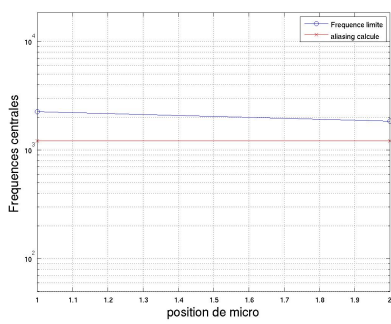


Erreur sur l'itd pour l'onde plane

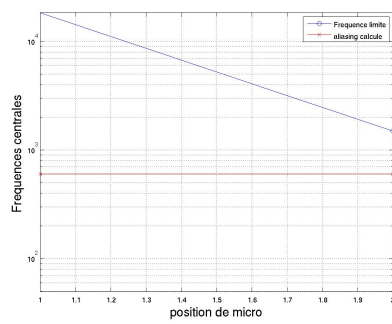


Erreur sur l'ild pour l'onde plane

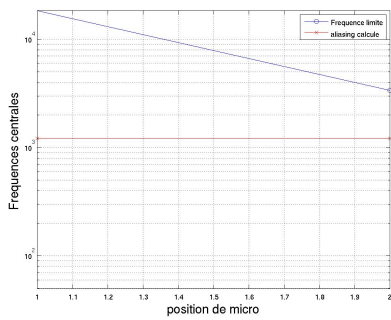
TAB. 15 – Erreurs sur l'itd et l'ild pour les 3 sources et les 2 positions d'écoute du test. En abscisses les fréquences en Hz, en ordonnées les valeurs de l'indice, en dB pour l'ild et en ms pour l'itd



Source excentrée



Source centrée



Onde plane

TAB. 16 – Position de la fréquence d’aliasing calculées et de la fréquence-limite pour les 3 sources du test. En abscisses les positions de microphones, en ordonnées les fréquences en Hz

## A.2 Pour un espacement de 30 cm entre les haut-parleurs

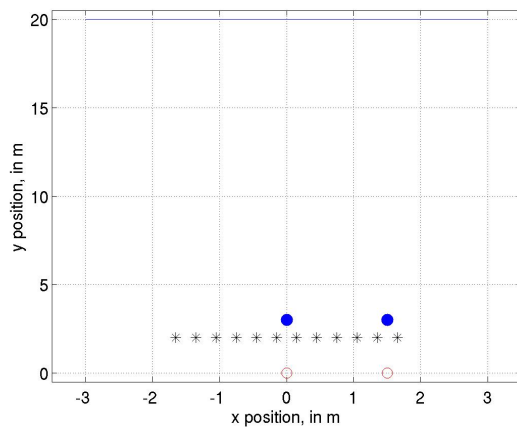
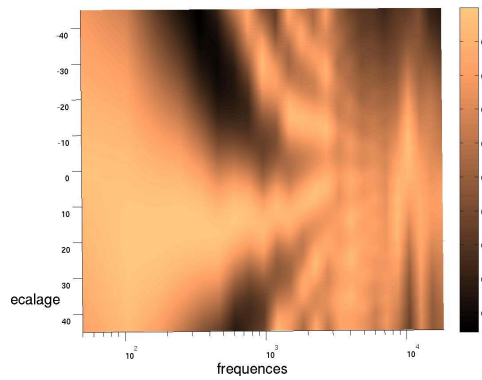
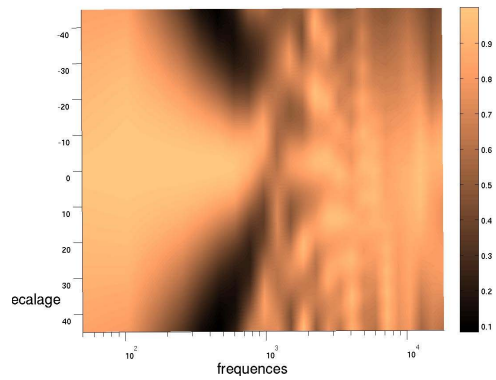


FIG. 9 – Position d'écoute, des haut-parleurs et des sources pour le test avec 12 haut-parleurs. les étoiles noires représentent les haut-parleurs, les rond rouges les positions d'écoute, les ronds bleus les sources centrées et excentrées. La ligne bleue représente l'onde plane

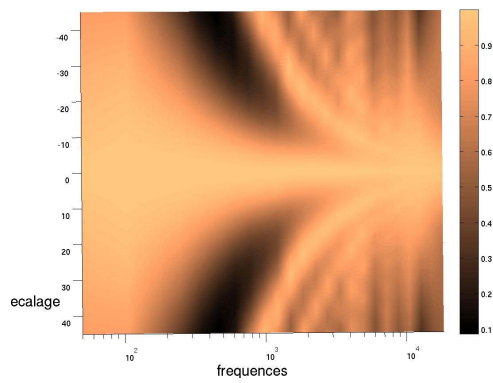




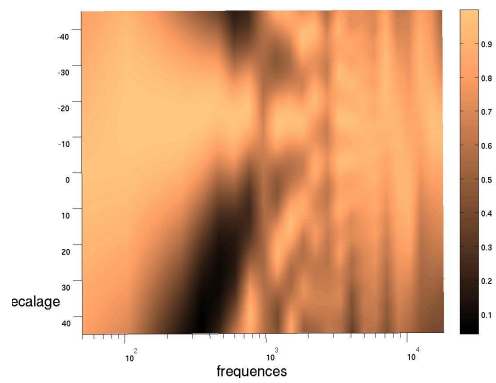
Source excentrée, position d'écoute centrée



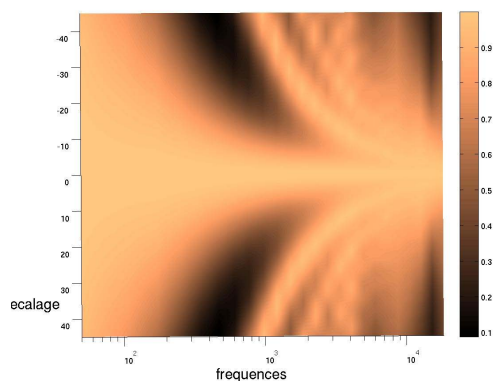
Source excentrée, position d'écoute excentrée



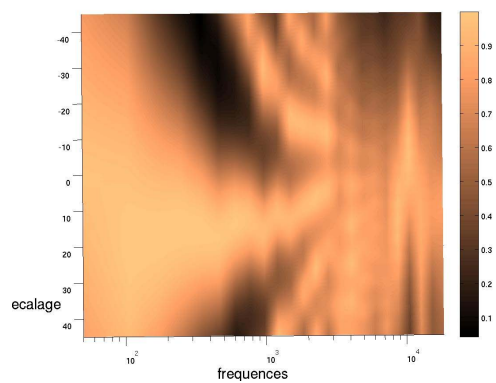
Source centrée, position d'écoute centrée



Source centrée, position d'écoute excentrée

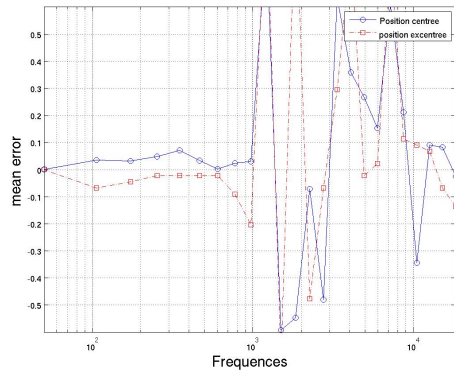


onde plane, position d'écoute centrée

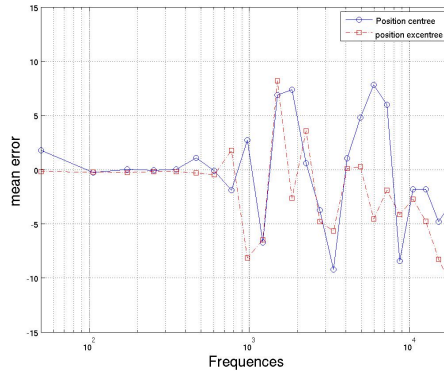


onde plane, position d'écoute excentrée

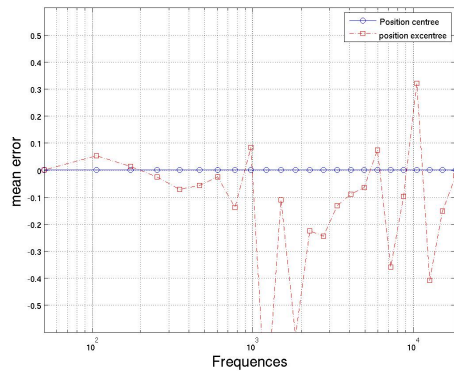
TAB. 17 – iacf par Blauert pour les 3 sources et les 2 positions d'écoute du test. En abscisses les fréquences en Hz, en ordonnées les valeurs d'itd correspondant aux valeurs de l'iacf (en nombre d'échantillons), en cote les valeurs de l'iacf.



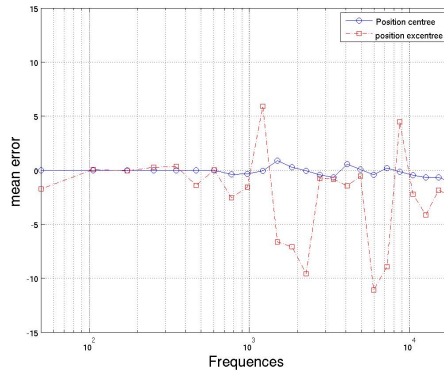
Erreur sur l'itd pour la source excentrée



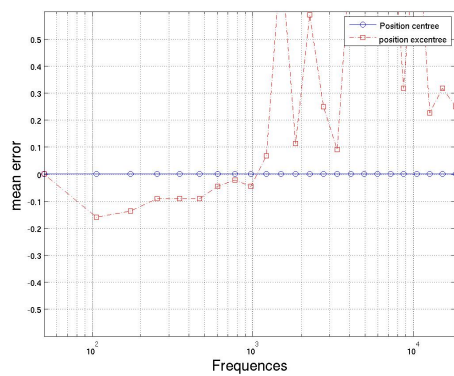
Erreur sur l'ild pour la source excentrée



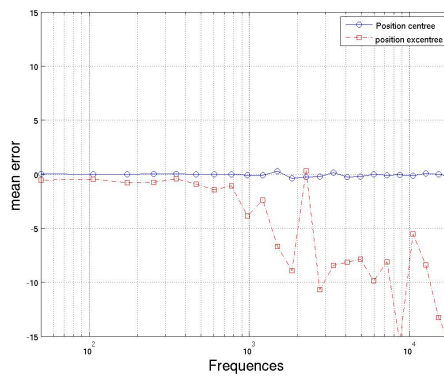
Erreur sur l'itd pour la source centrée



Erreur sur l'ild pour la source centrée

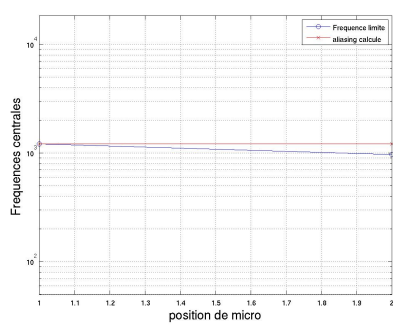


Erreur sur l'itd pour l'onde plane

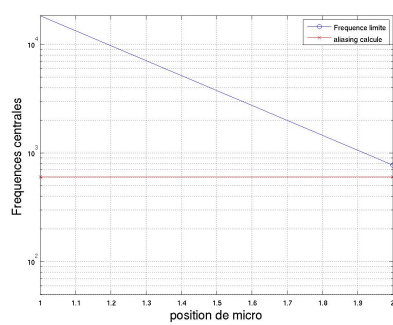


Erreur sur l'ild pour l'onde plane

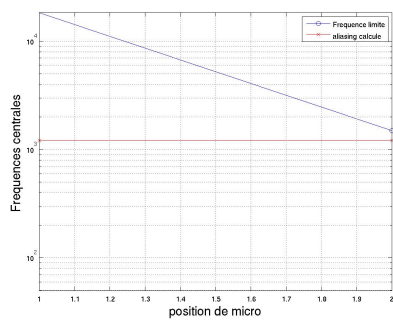
TAB. 18 – Erreurs sur l'itd et l'ild pour les 3 sources et les 2 positions d'écoute du test. En abscisses les fréquences en Hz, en ordonnées les valeurs de l'indice, en dB pour l'ild et en ms pour l'itd



Source centrée



Source excentrée



Onde plane

TAB. 19 – Position de la fréquence d’aliasing calculées et de la fréquence-limite pour les 3 sources du test. En abscisses les positions de microphones, en ordonnées les fréquences en Hz

## Références

- [BC78] J. Blauert and W. Cobben. Some considerations on binaural cross correlation analysis. *Acustica*, 32(2) :96–104, 1978.
- [Ber88] A. J. Berkhout. A holographic approach to acoustic control. *Journal of the Audio Engineering Society*, 36(12) :977–995, Décembre 1988.
- [Bla99] J. Blauert. *Spatial Hearing, The Psychophysics of Human Sound Localization*. MIT Press, 1999.
- [Bra05] J. Braasch. *Communication Acoustics*, chapter Modeling of Binaural Hearing, pages 75–108. Springer, 2005.
- [BT96a] L.R. Bernstein and C. Trahiotis. The normalized correlation : Accounting for binaural detection across center frequency. *Journal of the Acoustical Society of America*, 100(6) :3774–3784, December 1996.
- [BT96b] L.R. Bernstein and C. Trahiotis. On the use of the normalized correlation as an index of interaural envelope correlation. *Journal of the Acoustical Society of America*, 100(6) :3804–3806, September 1996.
- [CNW<sup>+</sup>07] E. Corteel, K-V. NGuyen, O. Warusfel, T. Caulkins, and R. S. Pellegrini. Objective and subjective comparison of electrodynamic and map loudspeakers for wave field synthesis. In *30<sup>th</sup> conference of the Audio Engineering Society*, Saariselkä, Finlande, Mars 2007.
- [Cor04] E. Corteel. *Caractérisation et Extensions de la Wave Field Synthesis en conditions réelles d’écoute*. PhD thesis, Université de Paris VI, Paris, France, 2004.
- [Cor06a] E. Corteel. Equalization in extended area using multichannel inversion and wave field synthesis. *Journal of the Audio Engineering Society*, 54(12), 2006.
- [Cor06b] E. Corteel. On the use of irregularly spaced loudspeaker arrays for wave field synthesis, potential impact on spatial aliasing frequency. In *9th Int. Conference on Digital Audio Effects (DAFx-06)*, Montréal, Canada, Septembre 2006.
- [dB04] W. de Bruijn. *Application of Wave Field Synthesis in Videoconferencing*. PhD thesis, TU Delft, Delft, Pays Bas, 2004.
- [FM04] C. Faller and J. Merimaa. Source localization in complex listening situations : Selection of binaural cues based on interaural coherence. *Journal of the Acoustical Society of America*, 116(5) :3075–3089, Novembre 2004.
- [GM94] B. Gardner and K. D. Martin. Hrtf measurements of a kemar dummy-head microphone. MIT Media Lab Perceptual Computing - Technical Report number 280, 1994.
- [HM05] H. Hammershoi and H. Moller. *Communication Acoustics*, chapter Binaural technique - Basic methods for recording, synthesis and reproduction, pages 224–254. Springer, 2005.
- [Hud05] H. Hudde. *Communication Acoustics*, chapter A functional view o the peripheral human hearing organ, pages 47–74. Springer, 2005.
- [Med86] R. Meddis. Simulation of mechanical to neural transduction in the auditory receptor. *Journal of the Acoustical Society of America*, 79(3), 1986.
- [MM02] E.A. MacPherson and J.C. Middlebrooks. Listener weighing of cues for lateral angle : the duplex theory of sound localization revisited. *Journal of the Acoustical Society of America*, 111(5) :2219–2236, Mai 2002.

- [Ngu06] K.V. Nguyen. Objectivation de la qualité de la restitution par bancs de haut-parleurs. application à la wave field synthesis. Master's thesis, DEA ATIAM, Université Paris 6, Août 2006.
- [Pat94] R. D. Patterson. The sound of a sinusoid : Spectral models. *Journal of the Acoustical Society of America*, 96 :1409–1418, 1994.
- [ST98] R. M. Stern and C. Trahiotis. *Psychophysical and physiological advances in hearing*, chapter Binaural mechanisms that emphasize consistent interaural timing information over frequency, pages 384–395. Whurr, London, 1998.
- [Sta97] E. W. Start. *Direct Sound Enhancement by Wave Field Synthesis*. PhD thesis, TU Delft, Delft, Pays Bas, 1997.
- [TBA01] C. Trahiotis, L.R. Bernstein, and M.A. Akeroyd. Manipulating the "straightness" and "curvature" of patterns of interaural cross correlation affects listeners' sensitivity to changes in interaural delay. *Journal of the Acoustical Society of America*, 109(1) :321–330, 2001.
- [Vog93] P. Vogel. *Application of Wave Field Synthesis in room acoustics*. PhD thesis, TU Delft, Delft, Pays Bas, 1993.
- [WK92] F. L. Wightman and D. J. Kistler. The dominant role of low-frequency interaural time differences in sound localization. *The Journal of the Acoustical Society of America*, 91(3) :1648–1641, Mars 1992.