



**HAL**  
open science

# Holophonie binaurale - Spatialisation sonore sur réseaux de haut-parleurs circumauraux

Raphaël Greff

► **To cite this version:**

Raphaël Greff. Holophonie binaurale - Spatialisation sonore sur réseaux de haut-parleurs circumauraux. Acoustique [physics.class-ph]. Université Pierre et Marie Curie - Paris VI, 2008. Français. NNT: . tel-00619404

**HAL Id: tel-00619404**

**<https://theses.hal.science/tel-00619404>**

Submitted on 6 Sep 2011

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

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

**THESE DE DOCTORAT DE L'UNIVERSITE PARIS VI**

*Spécialité :*  
**ACOUSTIQUE**

*Présentée par*  
**Raphaël GREFF**

*Pour obtenir le grade de*  
**DOCTEUR DE L'UNIVERSITE PARIS VI**

**Holophonie binaurale**  
**Spatialisation sonore sur réseaux**  
**de haut-parleurs circumauraux**

*Soutenue le 25 novembre 2008*

*Jury :*

Directeurs de thèse

Rapporteurs

Examineur

M. Christophe D'ALESSANDRO

M. Brian KATZ

M. Ralph ALGAZI

M. Durand BEGAULT

M. Jean-Dominique POLACK

Les travaux de recherche de thèse présentés dans ce manuscrit ont été encadrés par le laboratoire LIMSI-CNRS<sup>1</sup> et co-financés sous convention CIFRE par l'ANRT et l'entreprise A-Volute<sup>2</sup>.

<sup>1</sup> LIMSI-CNRS  
Université Paris-Sud  
Bâtiment 508  
BP 133  
91403 ORSAY CEDEX

Téléphone : +33 1 69 85 80 80  
Télécopie : +33 1 69 85 80 88  
Courriel : [contact@limsi.fr](mailto:contact@limsi.fr)

<sup>2</sup> A-Volute  
4120 route de Tournai  
59500 DOUAI

Téléphone : +33 3 27 91 65 94  
Télécopie : +33 3 27 91 65 94  
Courriel : [contact@a-volute.fr](mailto:contact@a-volute.fr)

# Résumé

Titre : Holophonie binaurale – spatialisation sonore sur réseaux de haut-parleurs circumauraux

Les systèmes de réalité virtuelle ont pour objectif de plonger les utilisateurs dans un espace suggéré artificiellement par des informations multi-sensorielles, auditive entre autres. Le positionnement des sources sonores dans l'espace constitue un facteur important d'immersion et contribue grandement au réalisme d'une scène virtuelle. La synthèse binaurale est la technique de spatialisation sonore qui s'approche le plus de l'écoute naturelle. Elle permet de spatialiser un événement sonore monophonique à l'aide de seulement deux filtres, les HRTF (*Head-Related Transfer Functions*), caractérisant la propagation acoustique de la source vers les oreilles d'un auditeur. Les différences de temps d'arrivée et de niveau de pression au niveau des tympans sont des indices de localisation primordiaux pour estimer de la latéralité d'une source, mais sont insuffisants pour la situer précisément en élévation. Le contenu spectral des HRTF contient les indices permettant cette discrimination, notamment les pics et creux en hautes fréquences induits par les effets de diffraction au sein des pavillons d'oreille. L'inconvénient majeur de la technique binaurale repose donc sur le fait que les HRTF sont propres à chaque utilisateur, car liés aux déformations subies par l'onde sonore lorsqu'elle rencontre les différentes parties du corps humain. Une écoute avec des filtres non individuels comporte des artéfacts perceptibles. Il est donc nécessaire d'acquérir des HRTF individuelles. La mesure ou le calcul numérique sont des méthodes couramment utilisées pour l'estimation de ces dernières, mais sont trop coûteuse en temps et en moyen pour être employées dans des applications grand public. D'autres procédés de personnalisation des HRTF tentent de modéliser le lien entre la morphologie d'un auditeur et les indices de localisation qui lui sont associés, en se basant sur des mesures anthropométriques. Cependant, l'estimation correcte des paramètres morphologiques ne peut être garantie que via l'utilisation d'outils de mesure robuste et souvent onéreux, notamment pour les dimensions liées aux pavillons d'oreille.

Cette thèse propose une nouvelle approche d'individualisation des HRTF, fondée sur une séparation des indices de localisation liés au torse et à la tête de ceux liés aux pavillons. La composante tête/torse est modélisée numériquement alors que la composante pavillonnaire est recrée à l'aide de réseaux de haut-parleurs entourant chacun des pavillons d'oreille de l'auditeur. L'idée de cette méthode est d'essayer d'exciter les véritables indices acoustiques induits par la diffraction du front d'onde reconstruit sur les pavillons. Ce nouveau concept, que nous appellerons « holophonie binaurale », car dérivé de techniques holophonique et binaurale, est développé dans ce manuscrit : différentes simulations ont été menées dans des cas de figure idéaux, et différentes optimisations sont proposées pour améliorer le fonctionnement du procédé pour un nombre réaliste de transducteurs. Ces études ont conduit à la conception et à l'évaluation d'un prototype de casque multi-haut-parleur dédié.

Mots-clef : spatialisation sonore, individualisation des HRTF, synthèse binaurale, holophonie, réseaux de haut-parleurs circumauraux, calcul BEM.

# Abstract

Title: Binaural holophony – Sound spatialization over circumaural transducer arrays

Virtual reality systems aim to create an artificial space suggested by multiple sensorial informations, including an auditive one. Sound source positioning is an important factor of immersion that improves the realism of a virtual scene. Binaural synthesis is a spatialization technique derived directly from normal hearing. It offers a simple way for positioning a monophonic sound event with only two filters, the HRTFs (Head-Related Transfer Functions), which characterize the acoustic paths from the source location to the ears of a listener. Arrival time and level differences at both tympanic membranes are primordial cues for estimating the lateralization of a sound source, but are not sufficient for achieving the exact localization in elevation. Spectral content of the HRTFs provides cues that permit this discrimination, notably high frequency peaks and notches created by diffraction effects within the pinnae. Major disadvantage of binaural synthesis is that the HRTFs are representative of a corresponding morphology. When using non-personal filters, perceptible artefacts may occur. For high quality binaural synthesis, the HRTFs need to be individualized, matching the morphology of the listener. Typical means for this are to measure or calculate the HRTFs of the listener, but these lengthy and costly methods are not feasible for general public applications. Other ways for individualizing the HRTFs consist of modelling the link between the morphology of an auditor and the associated localization cues, based on anthropometric measurements. However, an accurate estimation of the morphological parameters can only be guaranteed by using accurate and often costly measurement tools, notably for the estimation of pinna-related parameters.

This thesis explores a new approach for HRTF individualization, based on a physical separation of the localization cues linked to the head and torso from them related to the pinnae. The head/torso component is numerically modeled while the pinna components are created using a multiple transducer array placed around each pinna. The philosophy of this method consists in trying to excite the correct localization cues provided by the diffraction of the reconstructed wave front on the listener's own pinnae. This new concept, that we will call "binaural holophony" – application of holophonic techniques to binaural synthesis, is developed in this manuscript. Different simulations were conducted assuming ideal conditions, and different optimizations are proposed to improve the reconstruction of the HRTFs when using a realistic number of transducers. This study has conducted to the conception and the evaluation of a spatialization system dedicated to binaural holophony.

Keywords: sound spatialization, HRTF individualization, binaural synthesis, holophony, circumaural transducer arrays, BEM calculation.

# Executive summary

Title: Binaural holophony – Sound spatialization over circumaural transducer arrays

## INTRODUCTION AND BACKGROUND

This thesis is a contribution to the investigation of the problem of the individualization of the HRTF for general public applications. To provide accurate localization cues, binaural synthesis requires individualized HRTF, *i.e.* matching the morphology of the listener. Measuring or calculating the HRTF of a listener are common but lengthy and costly methods that are not feasible for general public applications. In order to simplify the estimation procedure of individualized HRTF, a great deal of research efforts aim to model the relation between morphology and the HRTF and to provide fast individualization procedures based on anthropometric measurements (see I.3.2.3). However, an accurate estimation of morphological parameters may only be guaranteed using accurate measurement tools (which are often expensive), essentially for estimating pinnae dimensions.

In this manuscript, a novel approach for HRTF individualization is proposed, by separating the head and torso effect from that of the pinnae using a multiple transducer array placed around each pinna. The goal is to reduce the individualization of the HRTF to the modeling of the head/torso component, and to try to excite the correct localization cues provided by the diffraction of the reconstructed wave front on listener's own pinnae. For achieving this, we define the concept of binaural holophony which consists in using holophonic techniques to reconstruct the sound fields in each pinna area. In practice, this should be done by using circumaural transducer arrays mounted on a headset and by feeding them correctly in order to reconstruct the desired binaural signal at listener's eardrums.

## BINAURAL HOLOPHONY CONCEPT

The basic principle of holophony is based on the discretization of the Kirchhoff-Helmholtz integral which gives a straightforward way of reproducing a sound field. At the recording stage, the listening area is surrounded by a microphone array composed by pressure and velocity microphones which record the primary sound field due to external sources. For the reproduction stage, loudspeakers are substituted for the microphones by replacing pressure microphones by dipole sources and velocity microphones by monopole sources. Each loudspeaker is fed with the signal recorded by its associated microphones (assuming the use of some weighting factors due to the spatial distribution of the secondary sources).

It has been remarked that the two elementary transducers of the secondary source are highly redundant, so that only one of them is necessary. In practice, only monopole loudspeakers are used and fed by figure-of-eight or even cardioid microphones. Figure-of-eight solution is an exact solution if one is able to record the contribution of each primary source separately. The recorded contribution of a primary source is taken into account only if this source is viewed by the positive pole of the microphone. For planar array, this selection is done by construction, but for complex array geometries, the logical weighting can only be performed if the exact location and nature of each primary

source are known. For this reason cardioid microphones are often preferred when recording real complex sound fields using complex array geometry. Nevertheless, the location of a sound source relative to an arbitrary field point is given by the active intensity at this point. Therefore, we proposed to generalize the logical weighting - when using figure-of-eight microphones - to any arbitrary primary sound field, *i.e.* created by unknown sources. Three kinds of holophonic recording solutions have been defined and employed for investigating the concept of binaural holophony: CARD (cardioid recording), FIG-8 S-LOG (figure-of-eight recording, considering the contribution of each primary source separately) and FIG-8 T-LOG (figure-of-eight recording, considering the total sound field).

The array designed for binaural holophony application has been chosen as having a hemispherical geometry. Based on Listen and CIPIC anthropometric databases, we evaluated that a radius of 50mm is sufficient to surround all of the pinnae, and 90% of them when centring the array on ear canal entrance. Only centred positioning was investigated. The primary array was composed of 640 equally distributed secondary sources to avoid spatial aliasing in the considered frequency range (3.5 to 16kHz).

## **SIMULATIONS USING RAW HOLOPHONIC TECHNIQUES**

The goal was to compare the original HRTF to a reconstructed HRTF using raw holophonic techniques for different array sizes. By introducing two kind of filters, the transducer driving filters (CAHF for Circum-Aural Holophonic Filters) and the transfer functions representing the path from a considered transducer to the entrance of the ear canal (PRTF for near-field Pinnae-Related Transfer Functions), the reconstructed HRTF can be expressed as in equation (III-1) in III.2.1. For estimating the quality of reconstruction, two metrics have been introduced, correlation and distance, both calculated on a logarithmic scale and based on decibel spectral values of the HRTF.

Simulations have been performed first on 6 different baffled pinnae meshes. In this case, the three kinds of holophonic recording solutions previously mentioned could be used, because all sources are known (the primary source and its image). CAHF have been calculated analytically due to the simplicity of the problem. Original HRTF and PRTF have been estimated via BEM techniques, respecting the sampling rule of 6 nodes per wavelength. When using the array of 640 secondary sources, results indicate a good match between original HRTF and reconstructed HRTF using CARD and FIG-8 S-LOG solutions, this for all pinnae models. FIG-8 T-LOG is dependant on lateral angle: quality of reconstruction decreases with lateralization of the primary source, because contributions of the primary source and those of the image source tend to oppose. If the number of secondary sources is too small, spatial aliasing occurs which degrades the quality of reconstruction. By decreasing the number of transducers from 640 to 8 (keeping a homogeneous distribution), the correlation decreases from more than 0.9 to less than 0.2 and spectral distances increases from approximately 1dB to more than 5dB (FIG-8 S-LOG method).

Then, simulations have been done on a model of the Neumann KU-100 dummy head, obtained via a MRI scan. In this case, only CARD and FIG-8 T-LOG solutions have been calculated, because of the complexity of the problem. Moreover, all transfer functions have been estimated via BEM calculations. Because of the size of the model, a sampling of 6 nodes per wavelength has been only respected up to 10kHz, but

simulations have been still conducted up to 16kHz. When using the array of 640 secondary sources, results of the ipsilateral HRTF are equivalent to those of the baffled pinnae simulations. For the contralateral HRTF, quality of reconstruction decreases with respect to lateralization for both recording methods, but CARD solution seems to be more robust. Decreasing the number of secondary sources degrades the reconstruction for both recording method, wherever the primary source is located. FIG-8 T-LOG method offers better results than CARD for median plane HRTF, but can not reconstruct the contralateral HRTF properly. A hybrid recording method (HYBRID) has been defined using FIG-8 T-LOG for median plane primary sources and CARD for lateralized primary sources.

## **SIMPLIFICATION OF THE HEAD MODEL**

BEM calculations dedicated to CAHF estimation are very time-consuming (the current calculation required more than three weeks on a cluster of three dual core processors with a total of 9GB RAM). So it was interesting to evaluate the degradation of HRTF reconstruction using an equivalent spherical head model which permits an analytical resolution of the problem. A similar approach from that Algazi proposed [AAD01b] has been used to find the radius of the iso-ITD sphere relative to the KU-100 dummy head. CAHF have then been calculated for this sphere. No relevant difference has been noticed compared to KU-100 results. The spherical head approximation seems to offer an easy way to design a binaural holophonic system.

## **OPTIMIZATIONS FOR SMALL ARRAY SIZES**

It was desired, for practical reason, to arrive at a solution for a binaural holophonic system using a maximum of 16 channels, *i.e.* 8 transducers per pinna. Precedent simulations have shown that raw applications of holophony principles provide bad results when spatial aliasing is occurring. Transducer driving filters and/or transducer positioning need to be optimized.

### **CAHF enhancement for a given array topology**

The idea is to use the information recorded by the array of  $N = 640$  secondary sources - corresponding to an accurate discrete representation of the hemispherical surface in the considered frequency band - to feed an array of a reduced number  $r$  of secondary sources (for example  $r = 8$ ). Holophony principles involve that one transducer, considered as punctual, need to be representative for a part of the surface, defined by the array. When the number of transducers is reduced, the area associated to each of them becomes bigger. If the number of transducer is too small, the information measured at the location of the transducer may not be representative for what happens on its associated area.

A simple optimization (AZ method) consists in averaging the information on each transducer area, instead of using only the information measured at the locations of each transducer.

For a considered array topology, the CAHF can also be enhanced according to a particular pinna morphology (or a set of pinnae). For  $K$  transducers and  $M$  frequency points, the PRTF associated to a pinna can be viewed as a  $K \times M$  matrix  $PRTF^K$ , of  $K$  filters or  $M$  directivities. For a considered pinna, it is possible to approximate the HRTF

reconstruction offered by  $N = 640$  transducers with only  $r = 8$  transducers, projecting  $PRTF^N$  on  $PRTF^r$ . This produces a  $N \times r$  matrix  $C$  of  $r$  spatial functions (of length  $N$ ), defined by equation (IV-8) in IV.3.2.3. The spatial function  $n^{\circ j}$  is used to weight the signals measured at the  $N = 640$  points on the surface and to feed the transducer  $n^{\circ j}$  of the reduced array configuration (PROJ method).

### Array topology optimization

To optimize the locations of the transducers in an array, we investigated only pinna-related methods, by decomposing the  $N \times M$  matrix  $PRTF^N$  of PRTF into a  $N \times r$  matrix  $C$  of spatial functions and a  $r \times M$  matrix  $L$  of filters using statistical analysis. Here both  $C$  and  $L$  are unknown. PCA (Principal Component Analysis) and ICA (Independent Component Analysis) permit to find these matrices simultaneously. The maxima of the  $r$  spatial functions, obtained with PCA or ICA, may be used to define optimized locations of the  $r$  transducers.

Another method, called SUBSET for “subset selection”, and based on a PCA decomposition of the  $N \times M$  matrix of PRTF and a QR factorization of the resulting spatial functions, permits to order the  $N$  transducer locations in a decreasing order and in a least square sense. The  $r$  first re-ordered locations provide an optimized array topology.

As these techniques are related to one or more pinnae morphologies, it has been decided to optimize the locations of the secondary sources in a heuristic way. As pinna-related localization cues are predominant for primary sound sources located in the median plane, and as the corresponding primary sound fields are essentially reconstructed by secondary sources of grazing incidences, a circular array topology has been designed (secondary sources equally distributed at the basis of the hemispherical surface).

## EVALUATION OF PROPOSED OPTIMIZATIONS

### Objective evaluation via simulations

Simulations have been re-conducted using all combinations of driving filter calculations (raw holophonic captation, AZ method and PROJ method) and array topologies (8 secondary sources located according an equi-reparted distribution, PCA, ICA, SUBSET optimized topologies and circular distribution). Results indicated that CAHF calculation optimization have a greater impact than topology optimization, except the circular distribution. Configurations “SUBSET PROJ” and “circular AZ” show the most interesting improvements.

### Subjective evaluation via listening tests

A modular hand-made headset prototype has been constructed to perform subjective evaluations of different array configurations. Localization tests have been conducted on 4 subjects in the median plane, and have conducted to design a modified array topology derived from the circular distribution by moving one transducer in the most lateralized position (top of the hemispherical surface). Localization tests have then been conducted on 4 other subjects to compare the performance of this binaural holophonic configuration with individualized HRTF. Results show that both spatialization systems are comparable.

## **CONCLUSION AND PERSPECTIVES**

Simulations and subjective evaluations have shown that it is possible to excite pinnae cues using multiple transducers well-placed around listener's pinnae, and using transducer driving filters related only to a simple spherical head model. Comparison of the proposed system to individualized HRTF don't permit to conclude directly that the proposed binaural holophonic system offers localization performances as accurate as individualized HRTF system do, because statistics on 4 subjects may not be representative of the entire population. However, it justifies the use of binaural holophony for general public application as a promising alternative to common generic HRTF-based binaural systems.

Moreover, the holophonic approach offers a way to decompose the HRTF into head/torso component - CAHF - and pinnae component - PRTF (exact physical decomposition in comparison to Algazi's heuristic signal approach). This allows us to study separately the link between pinnae morphology and associated spectral modifications. Pinnae-related optimizations presented above offer also another point of view for characterizing this relation and open the way of fast HRTF composing.



# Tables des matières

<b>Résumé.....</b>	<b>3</b>
<b>Abstract .....</b>	<b>4</b>
<b>Executive summary .....</b>	<b>5</b>
<b>Tables des matières.....</b>	<b>11</b>
<b>Liste des figures .....</b>	<b>15</b>
<b>Liste des tableaux .....</b>	<b>19</b>
<b>Introduction .....</b>	<b>21</b>
<b>I Perception auditive et synthèse binaurale .....</b>	<b>23</b>
1 INTRODUCTION .....	24
2 LA LOCALISATION AUDITIVE .....	25
2.1 Le référentiel auditeur et les systèmes de coordonnées utilisés.....	25
2.2 Définition des HRTF : approche système de la propagation acoustique .....	26
2.3 Indices acoustiques de la localisation auditive .....	29
2.3.1 <i>Théorie duplexe : définition des indices interauraux classiques</i> .....	29
2.3.2 <i>Ambiguïtés de localisation liées à la théorie duplexe</i> .....	30
2.3.3 <i>Pertinence d'indices de type spectraux</i> .....	30
2.3.4 <i>Indices spectraux monauraux</i> .....	31
2.3.5 <i>Indices spectraux interauraux</i> .....	31
2.3.6 <i>Indices spectraux et identification du timbre</i> .....	32
2.3.7 <i>Natures des indices spectraux</i> .....	32
2.4 Localisation auditive et autres modalités sensorielles.....	33
2.4.1 <i>Indices de localisation dynamiques</i> .....	33
2.4.2 <i>Influence de la vision</i> .....	33
2.4.3 <i>Modalités proprioceptives</i> .....	34
2.5 Performances de localisation du système auditif.....	34
2.5.1 <i>Précision de localisation en azimuth</i> .....	34
2.5.2 <i>Précision de localisation en élévation</i> .....	35
2.5.3 <i>Jugements confus liés à la tâche de localisation auditive statique</i> .....	35
3 HRTF ET TECHNIQUE BINAURALE.....	38
3.1 Extraction des indices perceptifs à partir des HRTF .....	39
3.1.1 <i>Décomposition des HRTF</i> .....	39
3.1.2 <i>Estimation de l'ITD</i> .....	40
3.1.3 <i>Egalisation de la composante spectrale des HRTF</i> .....	41
3.2 Individualisation des HRTF .....	42
3.2.1 <i>Mesure des HRTF</i> .....	42
3.2.2 <i>Acquisition des HRTF par des méthodes numériques</i> .....	43
3.2.3 <i>Individualisation rapide des HRTF</i> .....	45
3.2.4 <i>Vers une nouvelle approche de l'individualisation des HRTF</i> .....	48
4 CONCLUSION.....	49
<b>II Concept d'holophonie binaurale .....</b>	<b>51</b>
1 INTRODUCTION .....	52
2 HOLOPHONIE.....	52
2.1 Principe de Huygens.....	52
2.1.1 <i>Concept de source secondaire</i> .....	52
2.1.2 <i>Problèmes aux limites</i> .....	53
2.2 L'intégrale de Kirchhoff.....	54
2.3 Les intégrales de Rayleigh .....	55

2.4	Application à la reproduction d'un champ sonore.....	56
2.4.1	<i>Principe et limitations</i> .....	56
2.4.2	<i>Discrétisation de la distribution de sources secondaires</i> .....	56
2.4.3	<i>Généralisation des intégrales de Rayleigh</i> .....	57
2.4.4	<i>Pondération logique à partir de l'intensité active</i> .....	58
3	HOLOPHONIE BINAURALE.....	59
3.1	Présentation du concept.....	59
3.2	Solutions d'enregistrement holophonique retenues.....	59
3.3	Calcul des signaux alimentant le réseau.....	60
3.3.1	<i>Cas d'une captation FIG-8</i> .....	61
3.3.2	<i>Cas d'une captation CARD</i> .....	61
3.4	Exemple d'application : réseau sphérique.....	61
3.5	Quelle géométrie de réseau ?.....	64
3.6	Définition des différentes fonctions de transfert .....	65
4	CONCLUSION.....	66
<b>III</b>	<b>Reconstruction holophonique des HRTF .....</b>	<b>67</b>
1	INTRODUCTION .....	68
2	EVALUATION DE LA QUALITE DE RECONSTRUCTION.....	68
2.1	Reconstruction des HRTF .....	68
2.2	Définition des outils de mesure objective .....	68
2.2.1	<i>Distance spectrale</i> .....	69
2.2.2	<i>Corrélation spectrale</i> .....	69
2.3	Prédominance des résonances ou des anti-résonances ?.....	70
2.3.1	<i>Expérimentation perceptive</i> .....	70
2.3.2	<i>Evaluation de la qualité de reconstruction des anti-résonances</i> .....	73
3	CALCUL NUMERIQUE DES DIFFERENTES FONCTIONS DE TRANSFERT .....	73
3.1	Calcul BEM en théorie.....	74
3.1.1	<i>Enoncé du problème</i> .....	74
3.1.2	<i>Conditions aux limites</i> .....	74
3.1.3	<i>Fonction de Green</i> .....	75
3.1.4	<i>Formulation intégrale</i> .....	75
3.1.5	<i>Formulation isoparamétrique</i> .....	76
3.2	Calcul BEM en pratique.....	77
3.2.1	<i>Règle de discrétisation générale</i> .....	77
3.2.2	<i>Modèle symétrique ou modèle bafflé par un plan infini</i> .....	77
3.2.3	<i>Principe de réciprocité</i> .....	78
4	SIMULATION DE RECONSTRUCTION : CAS DE PAVILLONS BAFFLES.....	79
4.1	Modélisation numérique du problème.....	79
4.1.1	<i>Modèles de pavillons utilisés</i> .....	79
4.1.2	<i>Réseau de sources secondaires considéré</i> .....	79
4.1.3	<i>Positions de sources primaires considérées</i> .....	80
4.1.4	<i>Estimation des HRTF de référence et des PRTF</i> .....	81
4.1.5	<i>Estimation des CAHF</i> .....	81
4.2	Analyse des résultats .....	82
4.2.1	<i>Estimation de l'atténuation induite par le réseau</i> .....	82
4.2.2	<i>Sans aliasing spatial</i> .....	82
4.2.3	<i>Réduction de la taille du réseau</i> .....	85
5	SIMULATION DE RECONSTRUCTION : CAS D'UN MODELE DE TETE .....	89
5.1	Modélisation numérique du problème.....	89
5.1.1	<i>Modèle de tête utilisé</i> .....	89
5.1.2	<i>Estimation des HRTF de référence et des PRTF</i> .....	90
5.1.3	<i>Estimation des CAHF</i> .....	90
5.2	Analyse des résultats .....	91
5.2.1	<i>Sans aliasing spatial</i> .....	91
5.2.2	<i>Réduction de la taille du réseau</i> .....	93
5.3	Vers une captation hybride.....	96
6	CONCLUSION.....	98

	COMPLEMENT DE FIGURES.....	99
<b>IV</b>	<b>Simplifications et optimisations.....</b>	<b>105</b>
1	INTRODUCTION .....	106
2	MODELE DE TETE SPHERIQUE EQUIVALENTE .....	106
2.1	Sphère iso-ITD.....	106
2.2	Détermination de la sphère équivalente .....	107
2.2.1	<i>Estimation de l'ITD</i> .....	107
2.2.2	<i>Obtention du rayon équivalent</i> .....	107
2.3	Reconstruction des HRTF à partir du modèle sphérique.....	108
2.3.1	<i>Sans aliasing spatial</i> .....	108
2.3.2	<i>Réduction de la taille de réseau</i> .....	110
3	OPTIMISATION DES CAHF POUR UNE DISPOSITION DONNEE.....	113
3.1	Méthode de moyennage par zone.....	113
3.1.1	<i>Illustration de la méthode</i> .....	113
3.1.2	<i>Obtention des CAHF optimisés</i> .....	114
3.2	Méthode par projection .....	115
3.2.1	<i>Décomposition linéaire des PRTF</i> .....	115
3.2.2	<i>Projection des PRTF sur un ensemble réduit</i> .....	116
3.2.3	<i>Calcul des CAHF optimisés</i> .....	116
4	OPTIMISATION STATISTIQUE DE LA DISPOSITION DU RESEAU.....	117
4.1	Description des méthodes de décomposition.....	117
4.1.1	<i>Méthodes PCA et ICA</i> .....	117
4.1.2	<i>Méthode SUBSET</i> .....	121
4.2	Application à la base de données calculée de PRTF .....	121
5	OPTIMISATION HEURISTIQUE DE LA DISPOSITION DU RESEAU.....	126
5.1	De l'importance du plan médian .....	126
5.2	Vers une distribution circulaire .....	127
6	EVALUATION DES DIFFERENTES OPTIMISATIONS.....	127
7	CONCLUSION.....	130
	COMPLEMENT DE FIGURES.....	131
<b>V</b>	<b>Mise en œuvre et évaluation perceptive.....</b>	<b>143</b>
1	INTRODUCTION .....	144
2	SYSTEME DE RESTITUTION DEDIE A L'HOLOPHONIE BINAURALE .....	144
2.1	Modèle sphérique de rayon moyen.....	144
2.2	Compensation du phénomène d'atténuation.....	145
2.3	Casque multi-transducteur modulaire.....	146
2.4	Application de test.....	147
2.5	Autres matériels .....	148
3	EVALUATION PERCEPTIVE .....	148
3.1	Evaluation subjective de différentes configurations dans le plan médian .....	148
3.1.1	<i>Configurations testées</i> .....	148
3.1.2	<i>Evaluation qualitative dans le plan médian</i> .....	149
3.1.3	<i>Test de localisation dans le plan médian</i> .....	150
3.2	Evaluation subjective de la configuration C dans le plan horizontal.....	153
3.2.1	<i>Evaluation qualitative dans le plan horizontal</i> .....	153
3.2.2	<i>Test de localisation dans le plan horizontal</i> .....	154
3.3	Modification de la configuration C .....	155
3.4	Evaluation subjective de la configuration D.....	155
3.4.1	<i>Ecoutes informelles</i> .....	155
3.4.2	<i>Tests de localisation</i> .....	156
3.5	Holophonie binaurale vs. synthèse binaurale individuelle .....	157
4	CONCLUSION.....	160
<b>VI</b>	<b>Implémentation de la spatialisation .....</b>	<b>163</b>
1	INTRODUCTION .....	164

2	TRANSPOSITION DE L'APPROCHE BICANALE .....	164
2.1	Synthèse binaurale bicanale .....	164
2.2	Application à la synthèse holophonique binaurale .....	165
3	TRANSPOSITION DE L'APPROCHE MULTICANALE.....	165
3.1	Synthèse binaurale multicanale .....	165
3.2	Application à la synthèse holophonique binaurale .....	166
3.2.1	<i>Approche K-multicanale</i> .....	167
3.2.2	<i>Approche multicanale-K : paradigme des haut-parleurs virtuels</i> .....	167
4	IMPLEMENTATION SIMPLIFIEE : DECOMPOSITION DES CAHF .....	167
4.1	Décomposition des CAHF associés à la configuration <i>D</i> .....	168
4.2	Définition de l'implémentation P-BIN .....	169
4.3	Evaluation rapide de l'implémentation P-BIN .....	169
5	CONCLUSION.....	170
	<b>Conclusion et perspectives .....</b>	<b>173</b>
	<b>Références bibliographiques.....</b>	<b>179</b>
	<b>Publications .....</b>	<b>185</b>
	ARTICLE A – PRESENTE AU 19 <sup>EME</sup> CONGRES ICA .....	187
	ARTICLE B – PRESENTE A LA 123 <sup>EME</sup> CONVENTION AES .....	195
	ARTICLE C – PRESENTE A LA CONFERENCE ACOUSTICS'08.....	203

# Liste des figures

<i>Fig. I-1 : Représentation schématique de l'auditeur et repère associé</i> .....	25
<i>Fig. I-2 : Systèmes de coordonnées sphériques utilisés en spatialisation sonore</i> .....	26
<i>Fig. I-3 : Définition des HRTF pour une position de source et un auditeur donnés</i> .....	27
<i>Fig. I-4 : Exemple d'une paire de HRTF et de HRIR associées</i> .....	28
<i>Fig. I-5 : Représentation d'un cône de confusion au sens de la théorie duplexe</i> .....	30
<i>Fig. I-6 : Précision de localisation en azimuth dans le plan horizontal d'après Blauert [Bla97]</i> .....	36
<i>Fig. I-7 : Précision de localisation en azimuth d'après Oldfield et Parker [OP84a]</i> .....	36
<i>Fig. I-8 : Précision de localisation en élévation dans le plan médian d'après Blauert [Bla97]</i> .....	37
<i>Fig. I-9 : Précision de localisation en élévation d'après Oldfield et Parker [OP84a]</i> .....	37
<i>Fig. I-10 : Exemple de têtes binaurales utilisée pour la technique binaurale</i> .....	38
<i>Fig. I-11 : Illustration des méthodes d'estimation de l'ITD (ms)</i> .....	41
<i>Fig. II-1 : Illustration du principe de Huygens</i> .....	53
<i>Fig. II-2 : Géométrie du problème de Kirchoff</i> .....	54
<i>Fig. II-3 : Géométrie du problème selon l'approche de Rayleigh</i> .....	56
<i>Fig. II-4 : Illustration de la pondération logique à partir de l'intensité active</i> .....	58
<i>Fig. II-5 : Réseau sphérique de 5cm de rayon à 1280 sources secondaires</i> .....	62
<i>Fig. II-6 : Reconstruction holophonique par un réseau sphérique - 12kHz</i> .....	62
<i>Fig. II-7 : Erreur de reconstruction en fonction de la méthode de captation - réseau sphérique</i> .....	63
<i>Fig. II-8 : Erreur de reconstruction en fonction de la taille du réseau sphérique - méthode FIG-8</i> .....	63
<i>Fig. II-9 : Erreur de reconstruction en fonction de la taille du réseau sphérique - méthode CARD</i> .....	64
<i>Fig. II-10 : Paramètres anthropométriques du pavillon d'oreille (CIPIC et LISTEN)</i> .....	65
<i>Fig. III-1 : Illustration de l'extraction des motifs spectraux</i> .....	71
<i>Fig. III-2 : Erreur de localisation moyenne (localisation absolue et relative)</i> .....	72
<i>Fig. III-3 : Normalisation des motifs spectraux liés aux anti-résonances</i> .....	73
<i>Fig. III-4 : Modèles de pavillons d'oreille utilisés</i> .....	79
<i>Fig. III-5 : Réseau hémisphérique de 5cm de rayon à 640 sources secondaires</i> .....	80
<i>Fig. III-6 : Positions de la source primaire</i> .....	80
<i>Fig. III-7 : Estimation de l'atténuation en basse fréquence induite par le réseau hémisphérique</i> .....	82
<i>Fig. III-8 : Reconstruction des HRTF via 640 sources secondaires - pavillon Cortex</i> .....	84
<i>Fig. III-9 : Reconstruction des N-P via 640 sources secondaires - pavillon Cortex</i> .....	84
<i>Fig. III-10 : Corrélations et distance entre HRTF en fonction de l'incidence – tous pavillons confondus</i> .85	
<i>Fig. III-11 : Fréquence d'aliasing spatial en fonction de la taille du réseau</i> .....	86
<i>Fig. III-12 : HRTF reconstruites en fonction de la taille du réseau - pavillon Cortex</i> .....	87
<i>Fig. III-13 : N-P reconstruits en fonction de la taille du réseau - pavillon Cortex</i> .....	87
<i>Fig. III-14 : Corrélations et distance entre HRTF en fonction de la taille de réseau</i> .....	88
<i>Fig. III-15 : Corrélations et distance entre HRTF en fonction de la taille de réseau - plan médian</i> .....	88

Fig. III-16 : Corrélation et distance entre N-P en fonction de la taille de réseau - plan médian .....	88
Fig. III-17 : Tête acoustique Neumann KU-100 et modèle numérique dérivé.....	89
Fig. III-18 : Illustration du problème de diffraction induit par des objets métalliques.....	90
Fig. III-19 : Reconstruction des HRTF ipsilatérales via 640 sources secondaires - tête KU100.....	92
Fig. III-20 : Reconstruction des HRTF contralatérales via 640 sources secondaires - tête KU100.....	92
Fig. III-21 : Corrélation et distance en fonction de l'incidence - côté ipsilatéral.....	93
Fig. III-22 : Corrélation et distance en fonction de l'incidence - côté contralatéral.....	93
Fig. III-23 : HRTF ipsilatérales reconstruites en fonction de la taille du réseau - tête KU-100 .....	94
Fig. III-24 : HRTF contralatérales reconstruites en fonction de la taille du réseau - tête KU-100.....	95
Fig. III-25 : Corrélation et distance en fonction de la taille de réseau - côté ipsilatéral.....	96
Fig. III-26 : Corrélation et distance en fonction de la taille de réseau - côté contralatéral.....	96
Fig. III-27 : Corrélation et distance en fonction de la taille de réseau - plan médian.....	97
Fig. III-28 : Corrélation et distance en fonction de la taille de réseau - côté ipsilatéral.....	97
Fig. III-29 : Corrélation et distance en fonction de la taille de réseau - côté contralatéral.....	97
Fig. III-30 : Reconstruction des HRTF via 640 sources secondaires - pavillon Yuvi .....	100
Fig. III-31 : Reconstruction des N-P via 640 sources secondaires - pavillon Yuvi.....	100
Fig. III-32 : Reconstruction des HRTF via 640 sources secondaires - pavillon DB60.....	101
Fig. III-33 : Reconstruction des N-P via 640 sources secondaires - pavillon DB60.....	101
Fig. III-34 : Reconstruction des HRTF via 640 sources secondaires - pavillon DB65 .....	102
Fig. III-35 : Reconstruction des N-P via 640 sources secondaires - pavillon DB65.....	102
Fig. III-36 : Reconstruction des HRTF via 640 sources secondaires - pavillon DB90.....	103
Fig. III-37 : Reconstruction des N-P via 640 sources secondaires - pavillon DB90.....	103
Fig. III-38 : Reconstruction des HRTF via 640 sources secondaires - pavillon DB95 .....	104
Fig. III-39 : Reconstruction des N-P via 640 sources secondaires - pavillon DB95.....	104
Fig. IV-1 : Illustration de l'estimation de l'ITD.....	107
Fig. IV-2 : ITD estimées et modélisées de la tête KU-100 (en ms) – plan horizontal .....	108
Fig. IV-3 : Evaluation de la modélisation sphérique iso-ITD - HRTF ipsilatérales .....	109
Fig. IV-4 : Evaluation de la modélisation sphérique iso-ITD - HRTF contralatérales.....	109
Fig. IV-5 : Corrélation et distance en fonction de l'incidence - côté ipsilatéral.....	110
Fig. IV-6 : Corrélation et distance en fonction de l'incidence - côté contralatéral.....	110
Fig. IV-7 : Evaluation du modèle sphérique en fonction de la taille du réseau - HRTF ipsilatérales .....	111
Fig. IV-8 : Evaluation du modèle sphérique en fonction de la taille du réseau - HRTF contralatéral...	112
Fig. IV-9 : Corrélation et distance en fonction de la taille de réseau - côté ipsilatéral.....	113
Fig. IV-10 : Corrélation et distance en fonction de la taille de réseau - côté contralatéral .....	113
Fig. IV-11 : Illustration de la répartition de la surface en zones de proximité.....	114
Fig. IV-12 : Valeurs singulières issues de la SVD et pourcentage de variance associé.....	122
Fig. IV-13 : Dispositions optimisées pour un réseau réduit à 8 sources secondaires.....	123
Fig. IV-14 : Repère de lecture des figures liées aux fonctions spatiales et de pondération.....	123

Fig. IV-15 : Fonctions spatiales issues de la PCA .....	124
Fig. IV-16 : Fonctions spatiales issues de l'ICA.....	125
Fig. IV-17 : Activité du réseau pour des sources primaires situées dans le plan médian - 8kHz.....	127
Fig. IV-18 : Récapitulatif des dispositions évaluées.....	128
Fig. IV-19 : Exemple de fonctions de pondération issues de l'optimisation PROJ.....	129
Fig. IV-20 : Evaluations des différentes configurations optimisées .....	130
Fig. IV-21 : Evaluations des différentes configurations optimisées dans le plan médian.....	130
Fig. IV-22 : Fonctions de pondération - configuration EQUI - optimisation AZ.....	132
Fig. IV-23 : Fonctions de pondération - configuration PCA - optimisation AZ.....	133
Fig. IV-24 : Fonctions de pondération - configuration ICA - optimisation AZ.....	134
Fig. IV-25 : Fonctions de pondération - configuration SUBSET - optimisation AZ .....	135
Fig. IV-26 : Fonctions de pondération - configuration CIRC - optimisation AZ .....	136
Fig. IV-27 : Fonctions de pondération - configuration EQUI - optimisation PROJ.....	137
Fig. IV-28 : Fonctions de pondération - configuration PCA - optimisation PROJ.....	138
Fig. IV-29 : Fonctions de pondération - configuration ICA - optimisation PROJ.....	139
Fig. IV-30 : Fonctions de pondération - configuration SUBSET - optimisation PROJ .....	140
Fig. IV-31 : Fonctions de pondération - configuration CIRC - optimisation PROJ .....	141
Fig. V-1 : Estimation de l'atténuation basse-fréquence .....	145
Fig. V-2 : Prototype modulaire de casque multi-transducteurs .....	146
Fig. V-3 : Directivité théorique d'un piston vibrant (en dB).....	147
Fig. V-4 : Interface de l'application de test de localisation (plan médian).....	148
Fig. V-5 : Distributions relatives aux configurations testées .....	149
Fig. V-6 : Report des jugements après correction des confusions – configuration A – plan médian.....	152
Fig. V-7 : Report des jugements après correction des confusions – configuration B – plan médian.....	152
Fig. V-8 : Report des jugements après correction des confusions – configuration C – plan médian.....	153
Fig. V-9 : Report des jugements après correction des confusions – configuration C – plan horizontal..	154
Fig. V-10 : Distribution relative à la configuration D .....	155
Fig. V-11 : Report des jugements après correction des confusions – configuration D – plan médian ....	156
Fig. V-12 : Report des jugements après correction des confusions – configuration D – plan horizontal	157
Fig. V-13 : Report des jugements – holophonie binaurale – plan médian .....	158
Fig. V-14 : Report des jugements – HRTF individuelles – plan médian .....	159
Fig. V-15 : Report des jugements – holophonie binaurale – plan horizontal .....	159
Fig. V-16 : Report des jugements – HRTF individuelles – plan horizontal.....	160
Fig. VI-1 : Implémentation bicanale de la synthèse binaurale.....	164
Fig. VI-2 : Implémentation K-bicanale de la synthèse holophonique binaurale.....	165
Fig. VI-3 : Implémentation multicanale de la synthèse binaurale.....	166
Fig. VI-4 : Implémentation simplifiée de la synthèse holophonique binaurale .....	168
Fig. VI-5 : Filtrés et directivités issus de la décomposition .....	169

*Fig. VI-6 : Filtres et directivités simplifiés..... 170*  
*Fig. VI-7 : Corrélation et distance en fonction de l'implémentation..... 170*

# Liste des tableaux

<i>Tab. III-1 : Index des positions de la source primaire – coordonnées polaire interaurales .....</i>	<i>81</i>
<i>Tab. III-2 : Index des positions de la source primaire – coordonnées polaire interaurales .....</i>	<i>91</i>
<i>Tab. V-1 : Résultats moyens par individu et par configuration – plan médian .....</i>	<i>151</i>
<i>Tab. V-2 : Résultats par individu pour la configuration C – plan horizontal .....</i>	<i>154</i>
<i>Tab. V-3 : Résultats par individu pour la configuration D – plans médian et horizontal .....</i>	<i>156</i>
<i>Tab. V-4 : Comparaison des résultats par individu – holophonie binaurale vs. HRTF individuelles.....</i>	<i>158</i>



# Introduction

Les techniques de spatialisation des sons permettent de suggérer la présence d'une source sonore en tout lieu de l'espace, et constituent à ce titre un facteur important d'immersion dans les applications de réalité virtuelle. L'une de ces techniques, la synthèse binaurale, a pour objectif de restituer aux tympans de l'auditeur le champ sonore qui aurait été engendré en situation d'écoute réelle. Sa simplicité de mise en œuvre, matérielle (carte son deux canaux et casque d'écoute stéréophonique) et logicielle (coût de calcul raisonnable), en font un procédé de simulation sonore idéal. Toutefois les mécanismes perceptifs mis en jeu dépendent fortement de l'auditeur. En effet, le positionnement d'une source virtuelle s'effectue en ajoutant au signal sonore des indices de localisation qui traduisent des modifications subies par l'onde acoustique de la source aux tympans. Une synthèse binaurale haute-fidélité nécessite par conséquent l'emploi d'indices personnalisés.

La mesure ou le calcul numérique sont des méthodes couramment utilisées pour l'estimation de ces derniers, mais sont trop coûteuse en temps et en moyen pour être employées dans des applications grand public. Une grande part des travaux de recherche dans le domaine tente de modéliser le lien entre la morphologie d'un auditeur et les indices de localisation qui lui sont associés, pour ainsi offrir des procédés d'individualisation basés sur des mesures anthropométriques. Cependant, l'estimation correcte des paramètres morphologiques ne peut être garantie que via l'utilisation d'outils de mesure robuste, notamment pour les dimensions liées aux pavillons d'oreille. De plus, il est établi que les modifications spectrales subies par l'onde et engendrées par les phénomènes de diffraction et de réflexion sur les pavillons fournissent des indices prépondérants pour la localisation d'événement sonore en élévation ou encore la discrimination avant-arrière.

Les travaux présentés dans cette thèse découlent de ce constat, et proposent d'aborder l'individualisation de la synthèse binaurale sous un angle nouveau. L'idée est d'utiliser les pavillons d'un auditeur en reconstruisant l'onde sonore arrivant sur ces derniers. L'objectif est donc d'essayer d'exciter les indices de localisation propre à l'auditeur, fournis par la diffraction de l'onde synthétisée sur les pavillons. En pratique, la méthode envisagée consiste à utiliser des techniques holophoniques pour reconstruire les champs sonores adéquats dans chaque zone pavillonnaire, en utilisant des réseaux de haut-parleurs circumauraux montés sur un casque. Par nature, cette méthode de spatialisation n'est plus à proprement parlée un procédé de synthèse binaurale au sens où on l'entend couramment, car la caractérisation des signaux ne s'effectue pas ponctuellement au niveau des tympans, mais en amont, sur une surface entourant les pavillons. On appellera cette technique l'« holophonie binaurale », faisant ainsi explicitement référence aux procédés dont elle dérive.

Ce manuscrit synthétise les recherches que nous avons menées sur ce sujet, de la théorisation du concept d'holophonie binaurale jusqu'au développement d'un système de restitution dédié. Il est composé de six chapitres :

- Le premier chapitre synthétise l'état des connaissances sur la localisation auditive et les techniques binaurales, et aborde différents travaux dédiés à

l'individualisation de la synthèse binaurale, en justifiant de ce fait l'étude menée lors de cette thèse.

- Le deuxième chapitre s'intéresse aux principes physiques sur lesquels se fonde l'holophonie, et présente le nouveau concept d'holophonie binaurale.
- Les troisième et quatrième chapitres constituent le cœur de cette thèse ; ils permettent, par le biais de simulations numériques, de valider le fonctionnement théorique de l'holophonie binaurale, et proposent différentes simplifications et optimisations destinées à prototyper un système de restitution dédié.
- Le cinquième chapitre rapporte les résultats issus de tests perceptifs menés sur différentes configurations de casque multi-transducteur, et offre une première comparaison d'une synthèse holophonique binaurale générique par rapport à une synthèse binaurale conventionnelle individualisée.
- Le sixième et dernier chapitre aborde le problème de la spatialisation du son sur un système d'holophonie binaurale et propose des solutions pour alléger le processus de calcul.

# I

## Perception auditive et synthèse binaurale

## 1 INTRODUCTION

La problématique générale des systèmes de reproduction sonore est de donner l'illusion d'une scène sonore réelle. Dans une telle scène, notre système auditif capte principalement le message sonore, mais il est également capable d'identifier le positionnement des sources et la nature de leur environnement acoustique. Les systèmes de réalité virtuelle ont toujours été conçus en mettant l'accent sur le réalisme et la richesse du rendu visuel, mais depuis quelques années, le son est devenu une composante tout aussi importante que l'image, et des efforts importants sont consacrés à l'amélioration du rendu sonore de ces systèmes, notamment en termes de spatialisation des sons. La spatialisation sonore peut être présentée comme la reproduction, que ce soit de manière physique ou perceptive, des propriétés spatiales d'une scène sonore. En général, on souhaite à partir de signaux monophoniques, être capable de positionner des événements sonores par rapport à l'auditeur dans un environnement virtuel. Le positionnement des sources sonores et l'« effet de salle », lié à l'espace acoustique dans lequel sont situées les sources et l'auditeur, ne sont pas des phénomènes complètement indépendants, mais sont souvent étudiés séparément. Nous ne nous intéresserons ici qu'à l'aspect de positionnement des sources sonores et ferons donc abstraction de l'environnement dans lequel celles-ci émettent.

La conception d'un système de spatialisation sonore doit prendre en compte les performances du système auditif humain. Il est important d'identifier les mécanismes mis en jeu lors d'une écoute naturelle, afin de les simuler efficacement. Deux approches de la perception auditive permettent d'étudier les questions de perception spatiale et de localisation d'événements sonores : l'approche physique, qui s'arrête à l'étude de la propagation des ondes acoustiques jusqu'à l'oreille externe, et l'approche physiologique, qui intègre également l'étude de l'oreille interne et des traitements neuronaux utilisés lors de la tâche de localisation. L'idée principale de la première approche est que la pression acoustique au niveau des tympanes varie en fonction de la position d'une source dans l'espace, mais également en fonction de la morphologie du corps tout entier. La synthèse binaurale est un procédé de spatialisation basé sur ce concept. Le principe est d'ajouter des indices acoustiques de localisation à un message sonore et de diffuser le résultat au niveau des tympanes de l'auditeur. Ces indices sont représentés sous forme de filtres, appelés en anglais *Head-Related Transfer Functions* (HRTF), littéralement les fonctions de transfert liées à la tête, spécifiques à chaque incidence, et constituant l'empreinte des transformations subies par le son lorsqu'il rencontre les obstacles formés par le torse, la tête et les oreilles. Or, d'importantes variations anatomiques peuvent être remarquées d'un individu à l'autre, particulièrement au niveau de la taille et de la forme des pavillons d'oreilles, et celles-ci ont un impact direct sur la forme des HRTF, et par conséquent sur les indices acoustiques de localisation. Chaque individu possède ses propres HRTF, relatives à sa propre anatomie. Pour approcher une situation d'écoute réelle, il est donc nécessaire que les filtres utilisés lors de la spatialisation soit en accord avec la morphologie de l'auditeur. Ceci constitue l'un des problèmes majeurs de la synthèse binaurale.

Dans ce premier chapitre, nous ferons le point sur l'état des connaissances dans le domaine de la localisation auditive. Nous présenterons les différents mécanismes mis en

jeu lors de la tâche de localisation et les performances moyennes du système auditif humain. Puis nous aborderons les techniques binaurales et les axes de recherche suivis par la communauté scientifique afin, notamment, d'améliorer les performances perceptives de la synthèse binaurale. Cet état de l'art nous permettra d'introduire et de justifier les travaux de recherche entrepris au cours de cette thèse.

## 2 LA LOCALISATION AUDITIVE

La localisation sonore désigne la capacité du système auditif à déterminer la position relative d'une source sonore. Cette capacité est à différencier de la perception spatiale qui englobe la localisation sonore et la perception de l'espace dans lequel se situent les sources et l'auditeur. Dans la suite de la section, la description des mécanismes de localisation sera restreinte au cas d'une seule source placée en champ libre. De plus, la distance ne sera pas prise en compte et seule l'incidence de l'onde émise par la source sera considérée.

### 2.1 Le référentiel auditeur et les systèmes de coordonnées utilisés

Pour spécifier la position d'une source sonore relativement à l'auditeur, il apparaît naturel d'utiliser un système de coordonnées dont le centre est confondu avec celui de la tête. Si l'on fixe les axes du repère comme indiqué sur la figure I-1, on peut définir trois plans standards :

- le plan horizontal ( $xz$ ), définissant une séparation haut/bas,
- le plan frontal ( $yz$ ), marquant une distinction haut/bas,
- le plan médian ( $xy$ ), séparant le demi-espace gauche du demi-espace droit.

Le centre du repère est généralement situé au milieu de l'axe interaural, ici représenté par l'axe ( $z$ ), et définit comme la droite passant par le centre des deux oreilles.

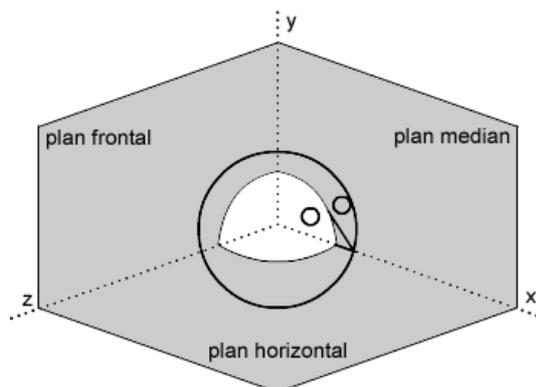


Fig. I-1 : Représentation schématique de l'auditeur et repère associé

Vu que seule l'incidence de la source sonore nous intéresse, il est pratique d'utiliser un système de coordonnées sphériques, car le positionnement de la source se restreint alors

à deux coordonnées d'espace (pour une distance fixe). Cependant, il existe plusieurs façons de définir ces coordonnées (voir Fig. I-2).

- Le système polaire vertical (partie gauche de la figure I-2) est le plus usité et place l'axe vertical ( $y$ ) comme axe principale. Dans ce système, l'azimut est mesuré comme l'angle  $\theta$  entre le plan médian et un autre plan vertical défini par l'axe vertical et le lieu de la source, l'élévation, quant à elle, est mesurée comme l'angle  $\varphi$  du plan horizontal à la source. Ainsi, les surfaces d'azimut constant forment des plans passant par l'axe vertical, et les surfaces d'élévation constante sont des cônes concentriques autour de l'axe vertical.
- Le système polaire interaural (partie droite de la figure I-2) est une alternative au précédent système et place l'axe interaural ( $z$ ) comme axe principal. Dans ce cas, on mesure l'azimut comme l'angle  $\theta$  entre le plan médian et la source, et on mesure l'élévation comme l'angle  $\varphi$  entre le plan horizontal et un plan défini par l'axe interaural et le lieu de la source. Les surfaces d'azimut constant sont alors des cônes concentriques autour de l'axe interaural, et les surfaces d'élévation constante sont des plans passant par l'axe interaural.

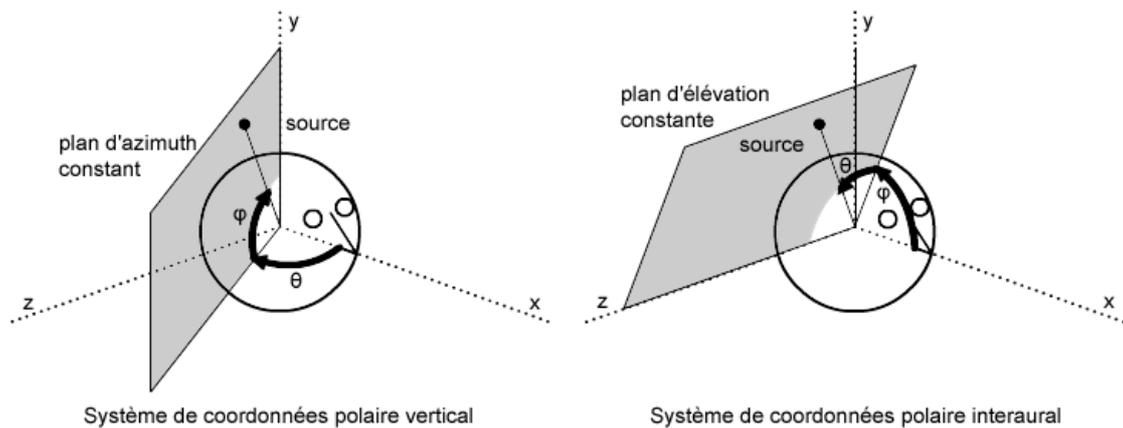


Fig. I-2 : Systèmes de coordonnées sphériques utilisés en spatialisation sonore

Pour distinguer les angles définis par l'une ou l'autre des conventions, on gardera l'appellation azimut et élévation pour les angles définis par la convention polaire verticale, et on nommera respectivement angle latéral et angle polaire l'azimut et l'élévation définis par la convention polaire interaurale.

## 2.2 Définition des HRTF : approche système de la propagation acoustique

Notre capacité à localiser les sons est grandement liée à notre morphologie. En effet, les modifications que subit une onde sonore, générée par une source sonore placée en une position donnée de l'espace, et se propageant jusqu'à l'entrée de nos conduits auditif, constituent un codage du positionnement de cette source.

Considérons une source sonore émettant un signal acoustique  $x$ . On peut définir les signaux acoustiques parvenant aux tympanes de l'auditeur par  $x_R$  pour l'oreille droite et  $x_L$  pour l'oreille gauche. Dans le cas de figure le plus simple (voir Fig. I-3), où la source

sonore est considérée comme fixe par rapport à l'auditeur, le phénomène de propagation acoustique entre la source et l'entrée de nos conduits auditifs peut être modélisé par un système linéaire et invariant dans le temps. On peut donc le caractériser par deux réponses impulsionnelles  $h_L$  et  $h_R$  définies par les relations de convolution suivantes :

$$\begin{cases} x_L(t) = x(t) \otimes h_L(t) \\ x_R(t) = x(t) \otimes h_R(t) \end{cases} \quad (\text{I-1})$$

Les réponses impulsionnelles  $h_L$  et  $h_R$  caractérisent la propagation acoustique entre la source et nos oreilles gauche et droite. L'expression I-1 peut donc également s'écrire sous sa forme fréquentielle où  $H_L$  et  $H_R$  sont respectivement les transformées de Fourier de  $h_L$  et  $h_R$  :

$$\begin{cases} X_L(f) = X(f) H_L(f) \\ X_R(f) = X(f) H_R(f) \end{cases} \quad (\text{I-2})$$

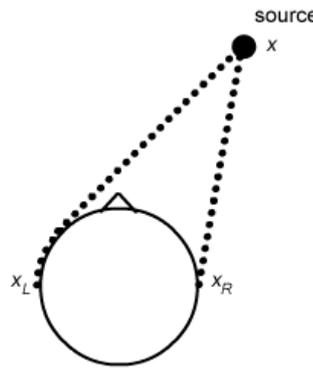


Fig. I-3 : Définition des HRTF pour une position de source et un auditeur donnés

Les HRTF sont souvent assimilées aux filtres  $H_L$  et  $H_R$ . Cependant la définition exacte des HRTF implique que ces dernières ne soient représentatives que des phénomènes de diffraction, réflexion et résonance induits par la présence de l'auditeur dans le champ créé par la source, et qu'elles n'intègrent aucunement les atténuations induites par la propagation de l'onde dans le milieu considéré. En notant  $x_O$  le signal acoustique présent au centre de la tête de l'auditeur, en l'absence de celui-ci, et donc généré par la source en champ libre, et en notant  $X_O$  sa transformée de Fourier, on définit les HRTF par le rapport suivant :

$$\begin{cases} HRTF_L(f) = \frac{X_L(f)}{X_O(f)} \\ HRTF_R(f) = \frac{X_R(f)}{X_O(f)} \end{cases} \quad (\text{I-3})$$

L'expression temporelle d'une HRTF est appelée HRIR pour *Head-Related Impulse Response* (réponse impulsionnelle liée à la tête). L'un des moyens pour obtenir les HRTF d'un individu est de les mesurer. Nous aborderons plus précisément l'acquisition des HRTF au paragraphe I.3.2. La figure I-4 présente un exemple de spectres de HRTF et les réponses impulsionnelles associées.

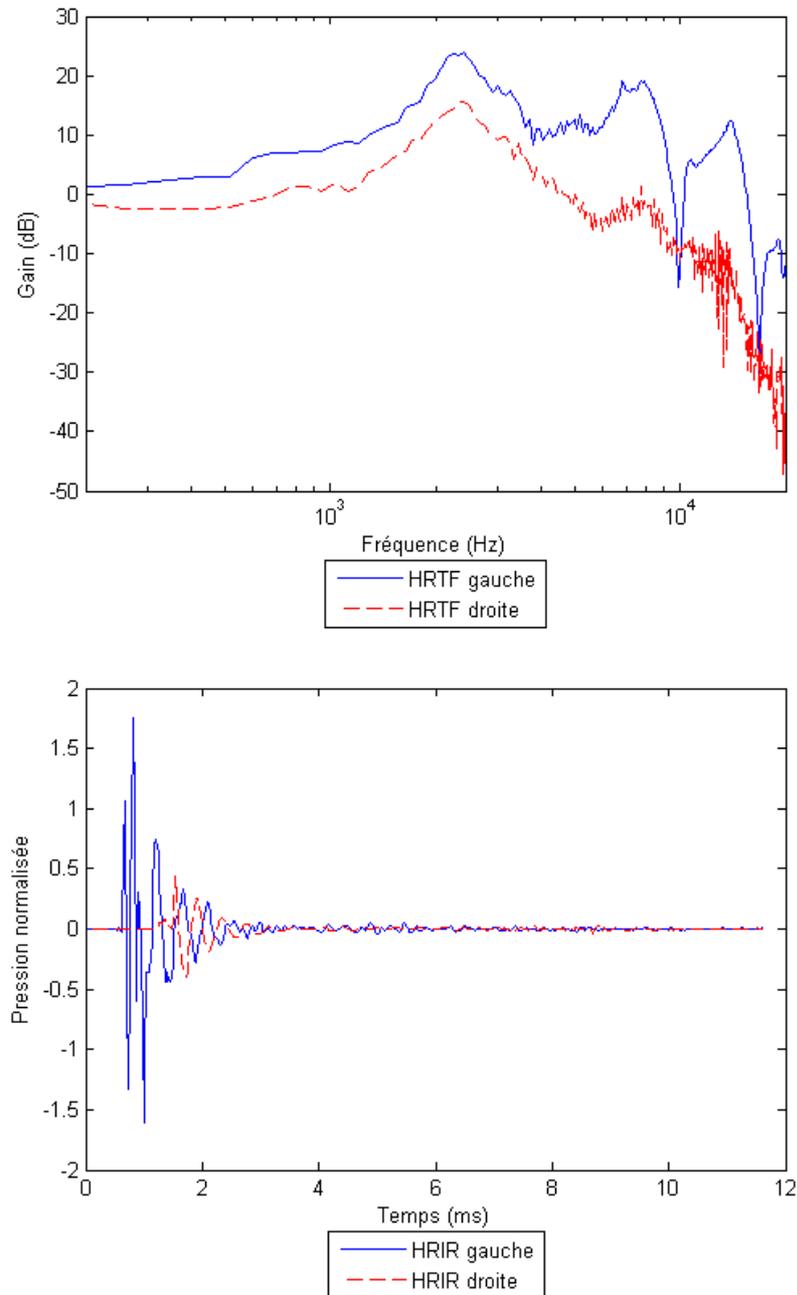


Fig. I-4 : Exemple d'une paire de HRTF et de HRIR associées

Si l'on souhaite concevoir un système de spatialisation sonore, il est intéressant de comprendre à quelles grandeurs mesurables, contenues dans les HRTF, notre système auditif est sensible en termes de localisation. Ces grandeurs, calculables à partir des

HRTF, sont appelées « indices acoustiques de localisation », dans la mesure où se sont les grandeurs acoustiques exploitées par notre système auditif pour estimer la position d'une source sonore. On pourra également les appeler « indices perceptif de localisation ».

Le paragraphe suivant présente les indices connus ou susceptibles de jouer un rôle lors de la tâche de localisation, et discute de leur contribution et de leur importance. D'autres phénomènes, comme l'influence d'autres modalités perceptives, seront également abordés.

## 2.3 Indices acoustiques de la localisation auditive

### 2.3.1 Théorie duplexe : définition des indices interauraux classiques

La localisation d'une source sonore est principalement effectuée par notre système auditif à partir de l'information acoustique présente aux deux oreilles. En 1907, Rayleigh publie la *Duplex Theory* qui décrit de manière simple les mécanismes de perception par rapport à l'angle latéral. Il y définit deux indices acoustiques que sont la différence interaurale de temps ou *Interaural Time Difference* (ITD), et la différence interaurale de niveaux ou *Interaural Level Difference* (ILD) [Bla97].

- La différence de temps d'arrivée de l'onde sonore entre les deux oreilles est l'indice de localisation prédominant. Plus cette différence tend vers sa valeur maximale, plus la source est latéralisée. La contribution majeure de l'ITD se situe en basses fréquences. En effet, au-delà de 1500Hz, fréquence à laquelle la longueur d'onde des signaux est du même ordre de grandeur que la distance interaurale, la différence de phase entre les deux oreilles devient une information ambiguë, car celle-ci n'est plus comprise dans un seul cycle. En hautes fréquences, il semblerait que l'ITD ne soit plus donnée par la phase mais estimée à partir des modulations de l'enveloppe d'amplitude des signaux de pression présents aux tympans
- La différence de niveaux entre les signaux reçus par chacune des oreilles, l'ILD, est l'indice de localisation complémentaire de l'ITD, et contrairement à celui-ci, devient une information importante en hautes fréquences, au-delà de cette même limite approximative de 1500Hz. En effet, le phénomène de diffraction dû à la présence de la tête devient important lorsque la longueur d'onde des sons incidents devient faible par rapport aux dimensions de la tête. Cette dernière joue alors le rôle de masque et des différences de niveaux entre les deux oreilles apparaissent.

Malgré la simplicité de cette théorie, elle constitue une bonne approximation de la perception de latéralisation. Par exemple, l'onde sonore émise par une source située dans l'hémisphère droit atteindra d'abord l'oreille droite, puis l'oreille gauche ; de même le niveau sonore sera plus fort à droite qu'à gauche. Cependant, l'ITD et l'ILD fournissent des informations ambiguës. En effet, pour un individu et pour une bande de fréquence donnée, un même couple ITD/ILD peut être représentatif de plusieurs incidences. Ces positions forment ce que l'on appelle couramment des « cônes de confusion » et ne peuvent être distinguées les unes des autres par la seule théorie duplexe.

### 2.3.2 Ambiguïtés de localisation liées à la théorie duplexe

Un cône de confusion (voir Fig. I-5) est défini comme le lieu des points pour lesquels l'ITD et l'ILD sont constants [Cha96]. L'ITD et l'ILD dépendant d'un individu à l'autre, un même couple ITD/ILD (pour une bande de fréquence donnée) décrira deux cônes différents pour deux individus différents. Par souci de normalisation, on désignera par « cône de confusion » les lignes iso-ITD du modèle de tête sphérique, c'est-à-dire des trajectoires à angle latéral et rayon constants en coordonnées polaires interaurales. Les principaux artéfacts liés aux cônes de confusions sont des erreurs de localisation en élévation ou des problèmes de discernement entre les hémisphères avant et arrière.

La théorie duplexe nécessite donc d'être complétée par la considération d'autres indices de localisation, permettant d'expliquer la faculté que possède le système auditif de distinguer les positions situées sur un même cône de confusion.

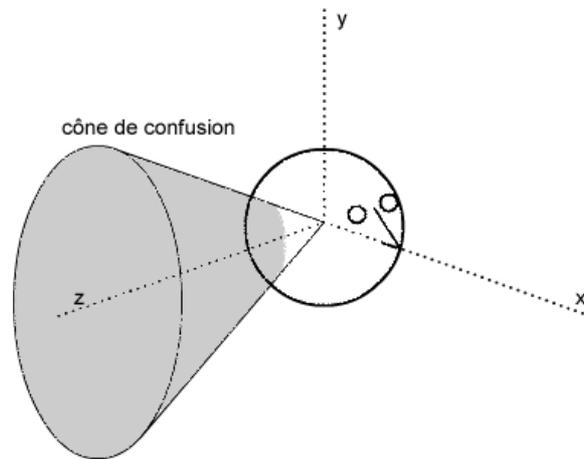


Fig. I-5 : Représentation d'un cône de confusion au sens de la théorie duplexe

### 2.3.3 Pertinence d'indices de type spectraux

De nombreuses preuves indiquent que l'information utilisée pour résoudre les ambiguïtés liées aux seuls indices ITD/ILD est de type spectral [ASS90]. D'abord, la perception de l'élévation d'une source à bande étroite peut être influencée par la fréquence centrale du son considéré : en effet, ce type de stimulus peut être perçu à une élévation autre que son élévation réelle [Bla97, Mid92]. De plus, la précision avec laquelle l'élévation d'un stimulus large bande est perçue peut être dégradée lorsqu'une fine bande est enlevée [HK74, But86] ou lorsque le spectre est brouillé [WK97]. Enfin lorsqu'on atténue le relief des pavillons d'oreille d'un auditeur en remplissant leur cavité, celui-ci perd tout ou partie de ses capacités à déterminer l'élévation d'une source ou à distinguer l'avant de l'arrière [Mor01, OP84b]. Or, les pavillons diffractent l'onde sonore générée par une source et jouent le rôle de filtres acoustiques. Ils ont un impact très marqué sur le spectre des HRTF [Sha82, Kah00]. En modifiant leur relief, l'effet de filtrage est modifié.

Les HRTF traduisent les effets de filtrage réalisé par le corps tout entier. Elles permettent de définir deux classes d'indices spectraux : des indices spectraux monauraux, qui s'appuie sur le fait que le système auditif pourrait analyser

l'information spectrale présente à chaque oreille indépendamment, ainsi que les indices spectraux interauraux, qui serait liés à une analyse conjointe des spectres monauraux ou à partir de la différence interaurale de spectre (ISD pour *Interaural Spectral Difference*). La mesure dans laquelle le système auditif utilise l'information spectrale monaurale et interaurale est une question importante et très débattue dans la littérature.

#### 2.3.4 Indices spectraux monauraux

Les travaux qui ont examiné la capacité d'un auditeur à localiser les sons en utilisant l'information présente à une seule oreille montrent à la fois la suffisance des indices spectraux monauraux et la nécessité des indices spectraux monauraux. Ces études ont généralement consisté en des tests de localisation menés sur des sujets sourd d'une oreille (que nous qualifierons de « déficients ») ou des sujets sains dont une oreille avait été rendue sourde artificiellement (que nous qualifierons de « dégradés »). Elles montrent que la localisation monaurale, à l'aide d'une seule oreille, est moins précise que la localisation binaurale, à l'aide de deux oreilles [BHM90, OP86], mais qu'il est possible que des sujets déficients estime l'élévation d'une source aussi bien que des auditeurs binauraux normaux [WO04] et que des auditeurs dégradés peuvent apprendre à discerner l'élévation aussi précisément que des auditeurs binauraux normaux non entraînés [HK74b].

La localisation monaurale semble être plus précise pour une source ipsilatérale, *i.e.* positionnée du côté de l'oreille intacte par rapport au plan médian, que pour une source contralatérale, *i.e.* placée du côté opposé [SM94]. Ceci soulève la question de la contribution relative de l'information spectrale ipsi- et contralatérale. En remplissant les cavités des deux pavillons, la capacité de localisation en élévation et la distinction avant/arrière sont abolies. En remplissant les cavités de l'oreille ipsilatérale, la précision de localisation en élévation et la discrimination entre les hémisphères avant et arrière sont fortement dégradées. Cependant, lorsque les reliefs d'une seule oreille sont estompés et que la source est positionnée du côté contralatéral par rapport à celle-ci, la dégradation de la localisation en élévation et de la distinction avant/arrière est moins marquée, et l'est d'autant moins que la source s'éloigne du plan médian. Il semblerait donc, en situation d'écoute binaurale normale, que plus une source est latéralisée, plus le système auditif se sert de l'information spectrale monaurale de l'oreille ipsilatérale.

#### 2.3.5 Indices spectraux interauraux

Une écoute binaurale offre de meilleures performances de localisation en élévation ainsi qu'une meilleure discrimination avant/arrière qu'une écoute monaurale, même dans le plan médian. Ceci semble indiquer que le système auditif analyse conjointement les informations spectrales monaurales ou qu'il utilise un indice tel que l'ISD.

Le fait que l'on puisse localiser des sons dont on ne connaît pas le timbre peut laisser supposer que l'ISD est un indice de localisation. Cependant, la plupart des études appuyant la thèse de l'exploitation de l'ISD sont sujettes à discussion. Les autres semblent plutôt indiquer que l'ISD n'est pas un indice acoustique de localisation. En effet, le brouillage spectrale d'une source dégrade aussi bien l'écoute binaurale que l'écoute monaurale, et ce d'une commune mesure [HW74b]. Une étude plus récente montre que l'ISD seul ne suffit pas à maintenir une bonne localisation en élévation et à distinguer correctement l'avant de l'arrière [JCCS04].

### 2.3.6 Indices spectraux et identification du timbre

L'ISD n'étant a priori pas exploité, la localisation auditive à partir des indices spectraux laisse supposer que le système auditif effectue la déconvolution du spectre de la source et des indices spectraux liés à la morphologie et à la position de la source [Mid92]. Les indices spectraux seraient donc stockés en mémoire. Cette hypothèse est cohérente avec le modèle d'association de Theile [The80] qui propose une identification conjointe de la position d'une source sonore et de son timbre. A l'instar de la précédente hypothèse, cette interprétation suppose que nous possédons une base de données personnelle d'indices de localisation, acquise et mémorisée au cours de notre vie. Toutefois, ce modèle suggère que nous possédons également une banque de timbres stockée en mémoire. Les observations de Blauert sont concordantes. Il a mis en évidence que de meilleures performances de localisation sont obtenues lorsque le stimulus est une voix familière plutôt qu'une voix inconnue [Bla97].

### 2.3.7 Natures des indices spectraux

Il apparaît donc que la relation entre la position perçue et les indices spectraux est complexe, et que le processus de localisation est certainement associé au processus d'identification du timbre. Cependant, on ne sait pas exactement si c'est la forme spectrale toute entière qui code la position d'une source, ou si seules certaines caractéristiques spectrales sont suffisantes. Ce qui est sûr, c'est que le pavillon d'oreille joue un rôle prédominant dans le codage des indices spectraux.

L'effet de filtrage induit par les pavillons engendre de fortes variations au niveau des spectres présents aux oreilles, à savoir des pics et des creux de fortes amplitudes. On peut expliquer ses variations dans le spectre des HRTF de deux manières différentes : soit on voit le pavillon d'oreille comme un résonateur, les pics et les creux sont alors respectivement les résonances et anti-résonances du pavillon ; soit on voit le pavillon comme un réflecteur, les pics et les creux étant alors induits respectivement par des réflexions constructives et destructives sur le pavillon.

Des études ont montré que la localisation auditive serait sensible principalement à la position des résonances des indices spectraux [Mid92, Bla97], et d'autres que ce serait plutôt les anti-résonances qui gouverneraient la localisation [Han94]. Dans une étude sur l'importance relative de certaines caractéristiques spectrales dans le codage de certaines positions de l'espace, Hebrank et Wright [HK74a] concluent que :

- la position avant est indicée par un creux d'une octave, situé entre 4 et 10kHz, et une augmentation de l'énergie au-delà de 13kHz,
- la position zénithale est codée par un pic d'un quart d'octave situé entre 7 et 9kHz et dont la fréquence de coupure haute est de l'ordre de 10kHz,
- l'indice de localisation vers l'arrière est un pic de faible amplitude compris entre 10 et 12kHz,
- la sensation d'élévation des positions frontale est liée à la fréquence de coupure basse du premier creux d'une octave.

D'après Langendijk et Bronkhorst [LB02], ce seraient les deux premières résonances et la première anti-résonance, qui coderaient la position des sources en élévation et suivant

l'axe avant/arrière, la première résonance étant due à un effet « résonateur de Helmholtz » de la conque [Sha82, Kah01] et la première anti-résonance étant induite par une réflexion destructive sur la partie postérieure de la conque [HW74a, RDY04, RR05].

Quelque soit la nature exacte des indices spectraux, des zones spectrales ont pu être associées à différentes dimensions spatiales et différentes parties morphologiques.

- Les informations spectrales perceptivement prédominantes pour localiser des sources dans le plan médian sont situées en hautes fréquences. Suivant l'axe haut/bas, elles se situeraient dans la bande de 6kHz à 12kHz, suivant l'axe devant/derrière dans la bande 4kHz à 16kHz [LB02]. Ces indices sont a priori liés aux modifications spectrales engendrées par la diffraction de l'onde sonore sur les pavillons d'oreille.
- Les informations spectrales perceptivement prédominantes pour localiser des sources dans le plan frontal sont situées en basses fréquences, en dessous de 3kHz, et associés au phénomène de diffraction par le corps et la tête [AAD01a].

## **2.4 Localisation auditive et autres modalités sensorielles**

### **2.4.1 Indices de localisation dynamiques**

Les indices acoustiques statiques expliquent pour une part importante les mécanismes de localisation auditive. Les informations de latéralisation sont portées par l'ITD et l'ILD, l'indice le plus important étant la différence interaurale de phase pour des signaux contenant des basses fréquences [WK92]. Ces informations seules ne permettent pas de localiser précisément la position d'une source sur un cône de confusion. Les indices spectraux permettent alors de lever les indéterminations. Cependant, pour une source placée dans le plan médian, il semblerait que seule l'information haute fréquence permette de la localiser. Si la source ne possède de l'information qu'en basses fréquences, la discrimination est impossible.

Perret et Noble [PN97] présentent une étude assez complète sur le rôle des mouvements de la tête lors de la tâche de localisation dans le plan médian. Leurs travaux montrent que l'indice dynamique servant à déterminer l'élévation fonctionne pour des signaux n'ayant de l'énergie qu'en dessous de 2kHz. Une étude de Wightman et Kistler [WK99] montre de plus que seules les variations des indices de localisation dynamiques sont exploités et non leurs variations corrélées aux mouvements de la tête. De plus, leur étude indique que le nombre de jugements confus, notamment les confusions avant/arrière, diminue de manière significative lorsque les indices dynamiques sont exploités.

### **2.4.2 Influence de la vision**

La vision jouerait un rôle de calibration des indices de localisation, notamment lors du plus jeune âge, lors de la phase d'apprentissage de localisation des sons. Lors de la tâche de localisation proprement dite, la vision n'aurait cependant aucune influence significative. La plupart des études réalisées sur des sujets aveugles, ou des sujets voyants dont les yeux ont été bandés, n'ont montré aucune dégradation des performances de localisation.

Cependant, un phénomène montre la prédominance de la vision sur l'ouïe si l'on fournit au cerveau des informations visuelles et auditives conflictuelles sur la position de la source. Il s'agit de l'« effet ventriloque ». Un écart angulaire allant jusqu'à 40° entre les positions réelles des sources visuelle et auditive peut être toléré pour que les deux événements semblent être spatialement coïncidents [Rad94].

### 2.4.3 Modalités proprioceptives

La proprioception désigne la perception du corps dans son environnement. Les informations proprioceptives, provenant de multiples capteurs situés par exemple dans les muscles, les tendons, les ligaments articulaires ou aussi la peau, codent très précisément la position relative des différents segments corporels. Les informations provenant du système vestibulaire entrent également dans cette catégorie. En effet, celui-ci est le système sensoriel principal de la perception du mouvement et de l'orientation par rapport à la verticale, et est à la base du sens de l'équilibre.

Les mouvements de tête sont en partie codés par le système vestibulaire. Or, nous avons vu précédemment que ceux-ci n'intervenaient pas directement lors de la tâche de localisation. On peut donc supposer que le retour d'information proprioceptive, tout comme le retour d'information visuelle, joue essentiellement un rôle de calibration des indices de localisation.

## 2.5 Performances de localisation du système auditif

La localisation auditive est un processus complexe d'intégration d'informations auditives, visuelles et cognitives, que le cerveau améliore au cours du plus jeune âge, et qui continue tout au long de la vie. De nombreux travaux psychoacoustiques ont étudié la capacité du système auditif humain à localiser une source sonore réelle, et ont permis d'établir des performances moyennes de localisation en situation d'écoute réelle. Ces études, généralement effectuées en milieu anéchoïque, exposent distinctement les performances de localisation en azimuth et en élévation. Cette dichotomie ne présente qu'un intérêt pratique, et ne présume en rien d'un traitement cérébral équivalent. De plus, elles explorent la précision de localisation dite « statique » du système auditif, c'est-à-dire pour des événements sonores et des auditeurs immobiles, où du moins considérés comme tel.

### 2.5.1 Précision de localisation en azimuth

Blauert [Bla97] fait un rapport exhaustif des études sur la localisation auditive. La précision de localisation est estimée ici comme l'amplitude de déplacement de la source sonore engendrant la perception d'un changement de position de celle-ci. Cette valeur est couramment appelée *localization blur* (flou de localisation en français). La figure I-6 expose de manière synthétique les performances de localisation en azimuth dans le plan horizontal estimées en termes de flou de localisation. La capacité de discrimination des sources apparaît bien plus marquée devant et derrière que sur les côtés, et meilleure devant que derrière. La précision est de l'ordre de  $\pm 4^\circ$  devant,  $\pm 6^\circ$  derrière,  $\pm 10^\circ$  pour les positions latérales.

Une autre étude importante menée par Oldfield et Parker [OP84a] estime la précision de localisation en azimuth comme l'erreur entre la position réelle d'une source et la position pointée par les différents sujets testés. La précision de localisation est donc indiquée en termes d'erreur de localisation, et non de flou (voir Fig. I-7, image du haut). On remarque que pour l'avant et l'arrière, les résultats sont cohérents avec ceux de Blauert. On note cependant que l'erreur estimée ici est plus importante pour les positions latérales arrière pouvant atteindre des valeurs de l'ordre de  $15^\circ$ . L'erreur absolue moyenne en azimuth est également évaluée en fonction de l'élévation (voir Fig. I-7, image du bas). On remarque que celle-ci évolue peu et reste inférieure à  $10^\circ$ .

### **2.5.2 Précision de localisation en élévation**

La figure I-8 présente les performances de localisation en élévation dans le plan médian rapportées par Blauert, et indique que les positions frontales sont les mieux discriminées, et que, pour les positions zénithales, le *localization blur* atteint des valeurs de l'ordre de  $\pm 20^\circ$ . Oldfield et Parker reporte quand à eux une précision de localisation en élévation dans le plan médian quasi-constante de l'ordre de  $8^\circ$ , et une erreur moyenne de localisation en élévation en fonction de l'azimut quasi-constante également de l'ordre de  $10^\circ$  (voir Fig. I-9).

### **2.5.3 Jugements confus liés à la tâche de localisation auditive statique**

Nous appelons jugement confus une erreur de localisation de forte amplitude. Les jugements confus reportés dans la littérature sont généralement des confusions avant/arrière pour lesquels les sons sont perçus derrière alors que la source est positionnée devant, confusions arrière/avant pour lesquels les sons sont perçus devant alors que la source est positionnée derrière.

Oldfield et Parker reporte un taux de confusion moyen selon la direction avant/arrière de l'ordre de 4%. Wightman et Kistler [WK89] reporte eux une moyenne de l'ordre de 6%. On peut donc considérer qu'en milieu anéchoïque et lors d'une tâche de localisation « statique », une probabilité de confusion avant/arrière de l'ordre de 5% est représentative des performances du système auditif humain.

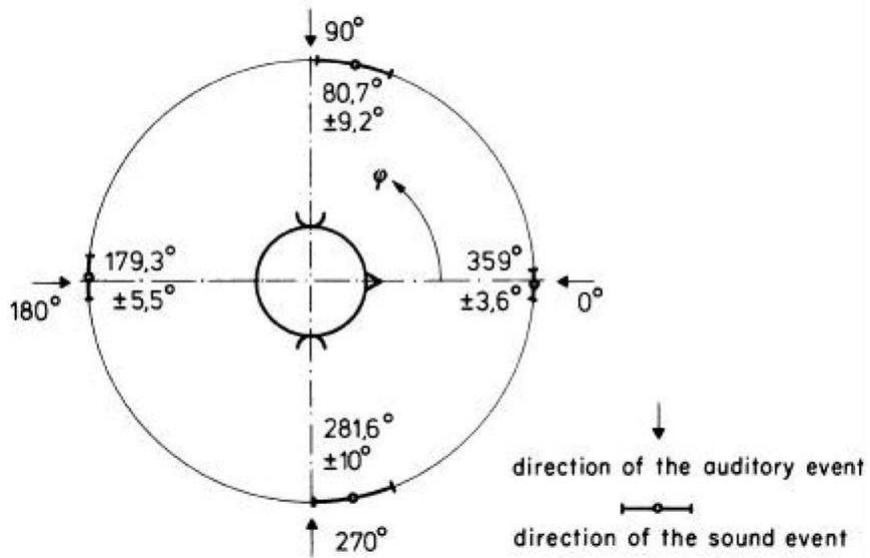


Fig. I-6 : Précision de localisation en azimut dans le plan horizontal d'après Blauert [Bla97]

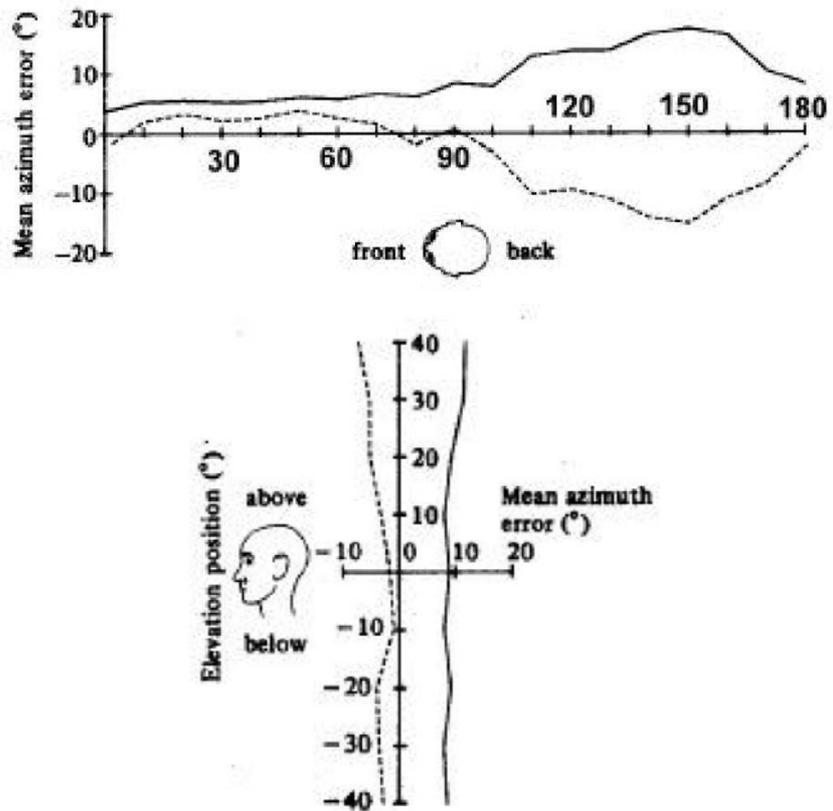


Fig. I-7 : Précision de localisation en azimut d'après Oldfield et Parker [OP84a]

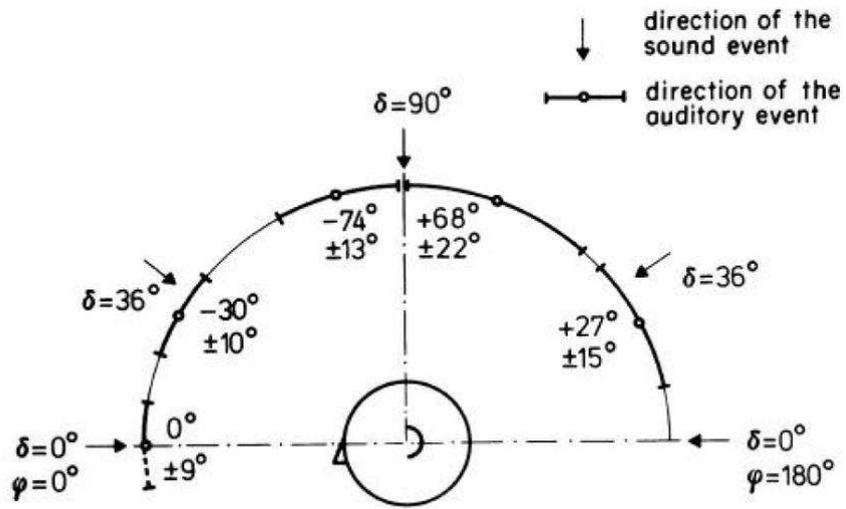


Fig. I-8 : Précision de localisation en élévation dans le plan médian d'après Blauert [Bla97]

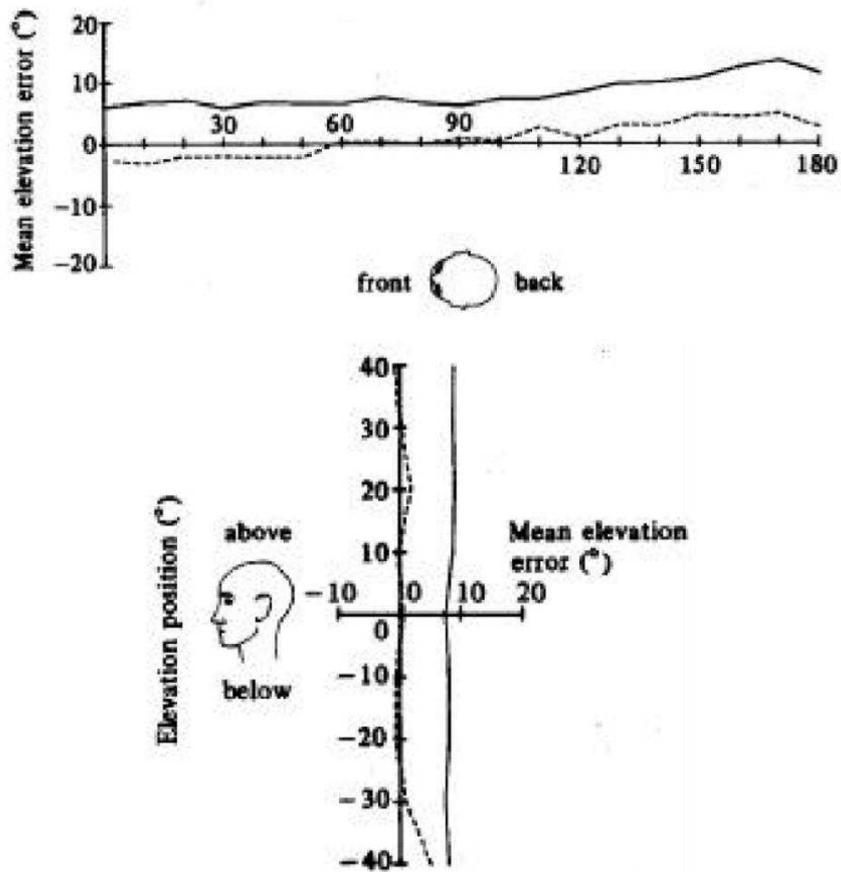


Fig. I-9 : Précision de localisation en élévation d'après Oldfield et Parker [OP84a]

### 3 HRTF ET TECHNIQUE BINAURALE

La technique binaurale repose sur une reproduction au niveau du conduit auditif de l'auditeur des informations acoustiques nécessaires à la construction par le système auditif d'une image sonore spatiale. Cette technique peut être exploitée de deux manières, différentes par leur implémentation, mais néanmoins très proches, puisque toutes deux basées sur l'écoute binaurale naturelle.

- L'enregistrement binaural consiste à placer des microphones dans les oreilles d'une tête, humaine ou artificielle, et à enregistrer une scène sonore. La diffusion au casque de l'enregistrement se fait alors sans matriçage des signaux.
- La synthèse binaurale consiste à simuler un enregistrement binaural à l'aide des HRTF. Il suffit d'effectuer la convolution entre un son monophonique, représentant une source sonore, et les HRTF associées à l'auditeur et à la position simulée, et de diffuser le résultat au casque.



Fig. I-10 : Exemple de têtes binaurales utilisée pour la technique binaurale

Le principe des enregistrements binauraux a été développé dans les années 1960 et a surtout contribué à la définition de têtes artificielles (voir Fig. I-10), sensées être représentatives d'une morphologie d'un auditeur moyen. Les indices perceptifs présents dans un enregistrement binaural effectué avec une tête artificielle sont donc liés à la morphologie de celle-ci. La synthèse binaurale a connu un véritable essor dans les années 1990, les progrès de l'informatique rendant alors abordable la conception et l'étude de systèmes de spatialisation basés sur ce principe. Différentes dénominations (le plus souvent en langue anglaise) sont utilisées dans la littérature pour faire référence à ce genre de systèmes : *3D-audio* [Beg94], VAD pour *Virtual Auditory Display* [WAKW93], VAS pour *Virtual Auditory Space* [Car96], ou encore VAI pour *Virtual Acoustic Imaging* [NKTH97]. A l'instar d'un enregistrement binaural, une scène sonore

synthétisée contient les indices perceptifs liés aux filtres utilisés pour effectuer cette synthèse, et sont donc liés à la morphologie associée à ces filtres. Cependant, dans ce dernier cas, il est possible d'intervenir sur les filtres et donc de les choisir en fonction de l'auditeur. Dans cette section, nous nous intéresserons uniquement à la synthèse binaurale et au processus de création de filtres de spatialisation.

### 3.1 Extraction des indices perceptifs à partir des HRTF

#### 3.1.1 Décomposition des HRTF

Les filtres définis en I.2.2 par l'expression (I-3) sont appelés HRTF à phase mixte ( $HRTF_{mixte}$ ). Il est alors possible de les décomposer en une composante à phase minimale ( $HRTF_{min}$ ) et une composante à excès de phase ( $HRTF_{excess}$ ). La composante à excès de phase est un filtre passe-tout qui contient les principales informations de phase. Elle peut elle-même être décomposée en un retard pur (un filtre passe-tout à phase linéaire  $\varphi_{lin}$ ) et une composante résiduelle contenant l'information de phase résiduelle. La composante à phase minimale a un module égal à celle de la composante à phase mixte. Sa phase  $\varphi_{min}$  se déduit du module de  $HRTF_{mixte}$  par une transformée de Hilbert.

$$HRTF_{mixte} = H \cdot e^{\varphi} \quad (I-4)$$

$$\begin{cases} |HRTF_{min}| = H \\ \varphi_{min} = \mathfrak{I}[\text{Hilbert}(-\log(H))] \end{cases} \quad (I-5)$$

$$\begin{cases} |HRTF_{excess}| = 1 \\ \varphi_{excess} = \varphi - \varphi_{min} \end{cases} \quad (I-6)$$

$$\varphi_{lin} = 2\pi f \tau \quad (I-7)$$

$$\varphi_{res} = \varphi_{excess} - \varphi_{lin} \quad (I-8)$$

$$\begin{cases} HRTF_{mixte} = H_{min} \cdot H_{excess} \\ HRTF_{mixte} = H_{min} \cdot H_{lin} \cdot H_{res} \end{cases} \quad (I-9)$$

Plogsties *et al.* considèrent la décomposition (I-9) [PMOCM00]. Ils étudient par un test perceptif si la phase résiduelle est audible. Pour la plupart des positions mesurées, la phase résiduelle est inaudible et la composante à excès de phase peut être considérée comme linéaire et peut donc être remplacée par un retard pur sans artéfact audible. Cependant, il existe des positions principalement situées près de l'axe interaural pour lesquelles cette information doit être reproduite. Ils montrent que la phase résiduelle est audible si le retard interaural de groupe évalué à 0Hz, noté IGD0 (*Interaural Group Delay at 0Hz*) est supérieur à 30 $\mu$ s. Dans le cas de figure où l'IGD0 est supérieur à cette valeur plafond, c'est-à-dire lorsque la phase résiduelle devient audible, leur étude montre également que la composante résiduelle peut être remplacée par un deuxième retard pur calculé comme le retard de groupe de la composante résiduelle évalué à 0Hz.

On en conclut qu'un couple de HRTF à phase mixte gauche et droite peut simplement être modélisé par une ITD correspondant au retard interaural de groupe évalué à 0Hz de l'excès de phase et une fonction de transfert à phase minimale.

Cette modélisation est puissante dans le sens où elle sépare l'ITD et le spectre. L'implémentation de la synthèse binaurale est alors simplifiée, d'autant plus que la composante fréquentielle est à phase minimale, c'est-à-dire que sa compacité temporelle est maximale. Avec l'évolution des capacités de calcul, cette modélisation perd de son utilité en tant que processus de synthèse binaurale. Cependant, elle effectue une séparation des indices, et montre qu'un couple de HRTF peut être caractérisé par une ITD (constante pour chaque fréquence) et deux spectres d'amplitude. La méthode d'estimation de l'ITD proposée par Plogsties *et al.* n'est pas la seule (et est loin d'être la plus usitée), et le choix de la méthode influe grandement sur les valeurs estimées, d'autant plus que l'estimation s'effectue sur des HRTF mesurées. La composante spectrale est obtenue plus directement et correspond simplement au module des HRTF à phase mixte. Cependant, il est utile d'effectuer certains traitement sur cette composante afin, par exemple, de réduire les HRTF à l'information dépendante de la direction.

Dans la suite, la caractérisation des HRTF se fera uniquement par l'ITD et les modules des spectres gauche et droit, c'est pourquoi nous nous intéresserons à l'évaluation de l'ITD et aux méthodes d'égalisation des spectres d'amplitude.

### 3.1.2 Estimation de l'ITD

Il existe différentes méthodes d'estimation de l'ITD, chacune fondée sur des principes différents, et soumises à divers problèmes de biais, de robustesse et de confiance. Nous présentons ici les trois principaux groupes d'estimateurs présentés dans la littérature. La figure I-11 illustre l'estimation de l'ITD suivant une variante de chaque méthode.

#### a) Méthode par détection de seuil

La méthode décrite par Moller *et al.* [MSHJ95] consiste à détecter le temps d'arrivée du son sur les HRIR gauche et droite. Ils définissent le temps d'arrivée comme le moment où l'amplitude atteint un certain pourcentage de la valeur maximale de la réponse impulsionnelle (10% dans leur cas). L'ITD est alors donné par la différence des temps d'arrivée gauche et droit.

Cette méthode dépend fortement de la valeur du seuil choisie. Comme le rapport signal sur bruit diminue pour les positions contralatérales, une valeur faible du seuil peut correspondre à une estimation au niveau du bruit, et donc entraîner une sous-estimation de l'ITD. De plus et toujours pour les positions contralatérales, l'apparition d'un pic secondaire de même niveau que le pic principale couplée à une valeur de seuil élevée peut entraîner par exemple une sur-estimation de l'ITD.

#### b) Méthode par approximation linéaire de la phase

Cette première méthode prolonge la réflexion sur la linéarité de l'excès de phase. La phase de la composante à excès de phase peut être approximée par une droite dont la pente représente le retard monaural. Larcher [Lar01, JLW95] propose d'effectuer une régression linéaire directement sur la différence des excès de phase gauche et droite plutôt que d'estimer séparément les retards monauraux gauche et droit pour ensuite en déduire l'ITD.

Pour garantir la robustesse de la méthode, elle préconise d'isoler les intervalles fréquentiels sur lesquels l'excès de phase est quasi-linéaire (c'est-à-dire sans rupture de pente). Une solution proposée consiste à détecter les pics de la dérivée de l'excès de phase, et à retenir comme intervalle de régression le plus grand intervalle sans pic. Cette solution fait appel à un seuil permettant de détecter les pics. Or, un mauvais réglage de ce seuil peut engendrer une mauvaise estimation de l'ITD.

### c) Méthode par corrélation

Initialement proposée par Wightmann et Kistler [WK92], la méthode souvent appelée MaxIACC (*Maximum of the Inter-Aural Cross Correlation function*) effectue le calcul de la fonction de corrélation interaurale, c'est-à-dire entre les HRIR gauche et droite. Cette méthode est basée sur l'hypothèse que le système auditif effectue la corrélation entre les signaux gauche et droit pour réaliser des tâches d'identification et de localisation de sources sonores. L'ITD est alors le temps pour lequel la fonction d'intercorrélation atteint son maximum.

Afin d'améliorer la prédiction, l'intercorrélation peut être effectuée sur l'enveloppe des HRIR ou encore l'enveloppe énergétique des HRIR. D'après Busson [Bus05], cette méthode semble être la plus robuste des méthodes d'estimation. Sa fiabilité pour les positions proches de l'axe interaural reste cependant discutable, car l'apparition d'un pic secondaire de même niveau que le pic principal sur la réponse contralatérale peut entraîner une sur-estimation de l'ITD.

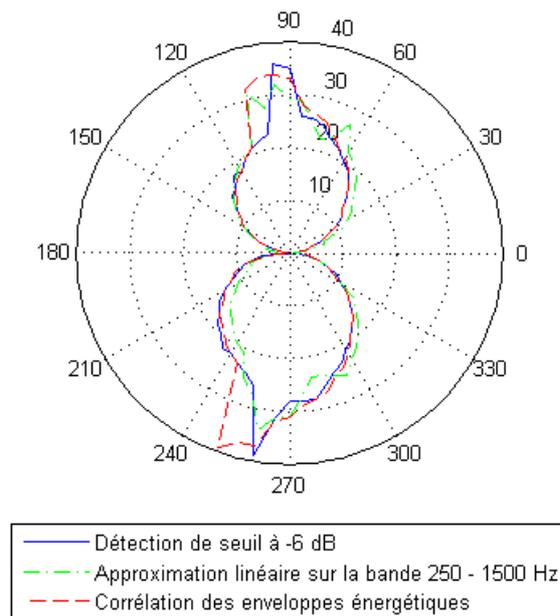


Fig. I-11 : Illustration des méthodes d'estimation de l'ITD (ms)

### 3.1.3 Egalisation de la composante spectrale des HRTF

Différents traitements peuvent être appliqués aux HRTF. Outre le lissage des spectres ou l'application de pondérations perceptives, on peut présenter les différents types

d'égalisation envisageables pour améliorer la fidélité de la diffusion de contenus binauraux [JLW95].

Deux principaux champs de référence ont été préconisés par les constructeurs de casques et de têtes artificielles :

- le champ libre, constitué d'une onde plane provenant d'une incidence donnée, le plus souvent l'incidence frontale,
- le champ diffus, constitué d'ondes planes décorrélées provenant d'incidences uniformément distribués autour du récepteur.

Une égalisation champ libre favorisera la conservation du timbre d'une source provenant de la direction choisie pour effectuer l'égalisation, mais altérera le timbre d'une source provenant d'une autre incidence. Une égalisation champ diffus préservera le timbre du champ sonore diffus, mais modifiera celui d'une onde provenant d'une seule incidence.

Dans la littérature [The86, Bla97], l'égalisation champ diffus est préférée pour garantir une bonne compatibilité entre les techniques d'enregistrement binaural et les techniques de prise de son stéréophoniques traditionnelles. De plus, elle permet de réduire nettement les différences entre sessions de mesures ou les différences entre têtes.

## **3.2 Individualisation des HRTF**

### **3.2.1 Mesure des HRTF**

Une reproductibilité fidèle de l'écoute naturelle grâce à la synthèse binaurale est obtenue avec les HRTF individuelles de chaque auditeur : les mesures sont alors le mètre étalon, et permettent d'obtenir les fonctions transfert les plus proches des HRTF d'un individu. Cependant la mesure des HRTF est un problème majeur de la synthèse binaurale. C'est une tâche délicate, soumise à de nombreuses sources d'erreurs, longue et nécessitant un matériel coûteux (salle anéchoïque, système mécanique pour la mise en position de la source, etc.).

#### ***a) Principe***

Une HRTF est une fonction de transfert complexe, calculée comme le rapport entre la pression  $P$  délivrée par une source au niveau du tympan, par la pression  $P_0$  résultante en champ libre, c'est-à-dire au lieu du centre de la tête de l'auditeur, en l'absence de celui-ci. Le principe de la mesure est donc de placer une source sonore à une position de l'espace (généralement un haut-parleur diffusant un signal balayant l'ensemble de la bande de fréquence audible), de mesurer  $P$  grâce à un microphone situé au niveau des tympans et de mesurer  $P_0$  au lieu du centre de la tête, auditeur absent.

Pour des raisons physiologiques évidentes, il est très difficile et peu pratique de mesurer un signal acoustique au niveau du tympan. Cependant, il a été reporté dans la littérature que les modifications introduites par le conduit auditif sont indépendantes de la position de la source. Le microphone est donc généralement positionné à l'entrée du canal auditif. De plus, les variabilités inter-individuelles étant atténuées en conduit bloqué, cette méthode de mesure est souvent préférée aux méthodes en conduit ouvert.

### **b) Sources d'erreur**

Plusieurs types d'erreur sont référencés dans la littérature ; en voici les principales sources.

- Réflexions sur le système de mesure : le système de mesure est souvent complexe et peut comporter de nombreux éléments mécaniques, qui sont susceptibles de diffracter et réfléchir l'onde sonore, voire même d'entrer en résonance, et d'ainsi perturber les signaux mesurés.
- Positionnement du microphone : le positionnement du microphone à l'entrée du canal auditif est difficile à garantir. Ce problème peut influencer sur la reproductibilité des mesures.
- Positionnement du sujet : il est difficile de positionner deux sujets exactement de la même manière. De plus, il est pratiquement impossible d'immobiliser complètement un sujet, et de petits mouvements de tête peuvent perturber les signaux mesurés.
- Erreurs basses fréquences : les sources utilisées étant généralement des haut-parleurs, ceux-ci ont une bande passante limitée, notamment en basse fréquence. De plus, le caractère anéchoïque du lieu de mesure est difficilement garanti en basse fréquence. La composante basse fréquence des HRTF mesurées est souvent négligée et d'autres techniques d'acquisition ou des méthodes de modélisation doivent être utilisées pour la recréer.

Le problème majeur de l'acquisition des HRTF par mesure semble donc être la reproductibilité : deux mesures effectuées sur un même individu et un même dispositif peuvent conduire à des résultats différents ; ce problème est d'autant plus marqué si l'on utilise différents dispositifs de mesure [KB07].

#### **3.2.2 Acquisition des HRTF par des méthodes numériques**

Un des moyens permettant de s'affranchir de la lourde tâche de mesure des HRTF, est la modélisation numérique du problème acoustique correspondant. Plusieurs méthodes numériques sont disponibles au sein desquelles plusieurs formulations existent. A la différence des mesures acoustiques, ces méthodes garantissent la reproductibilité des résultats. De plus ces techniques permettent le placement des microphones et des sources avec une très grande précision, sans souci d'encombrement. Il est également possible de contrôler l'impédance des surfaces et les caractéristiques du fluide entourant l'objet d'analyse. Ces conditions idéales permettent un travail systématique sur l'influence des paramètres de modélisation de l'objet d'analyse.

Le calcul BEM (*Boundary Element Method* ou méthode par élément de frontière), est une technique de calcul numérique très employée pour la résolution de problèmes de diffraction. La méthode permet l'obtention du champ de pression dans un volume à partir du calcul de la pression à la surface limitant le volume. Les solutions sont obtenues par la résolution de la formulation intégrale de l'équation d'onde. Ce type de calcul est tout à fait approprié pour l'estimation des HRTF, le problème physique équivalent étant la diffraction d'une onde par un volume représentant la tête de l'auditeur.

### a) Historique du calcul BEM appliqué à la prédiction des HRTF

Weinrich [Wei84] fut le premier à montrer la validité de la méthode BEM pour l'acquisition de HRTF. Il a calculé la réponse en champ proche d'un modèle simplifié de tête jusqu'à une fréquence de 1700Hz. Malgré l'utilisation d'un modèle grossier sans pavillon, et donc une simulation de HRTF bien éloignée d'une HRTF de tête humaine, ce travail a montré la validité de la méthode, en obtenant une bonne concordance avec les HRTF mesurées sur le même modèle.

Poursuivant la même approche, les travaux de Katz [Kat98, Kat01a] ont montré que la BEM pouvait travailler avec des maillages aux formes complexes comme le maillage de la tête d'un auditeur. Ne pouvant atteindre des fréquences supérieures à 5400Hz pour cause de ressources informatiques limitées, il a mis en évidence par la BEM l'influence du pavillon externe et l'influence de l'impédance des cheveux et de la peau sur les HRTF [Kat01b].

Les travaux de Kahana [Kah00, KNPC99] ont permis une meilleure compréhension des techniques d'éléments finis et de frontière pour la prédiction des HRTF. Il a effectué des comparaisons entre différentes méthodes de calcul numérique et montré que la technique BEM était la plus appropriée et la plus rapide. Il a également travaillé sur des formulations permettant la réduction des coûts de calculs (principe de réciprocité, symétrie). Grâce à ces méthodes optimisées et aux progrès de l'informatique, il a calculé la réponse d'un modèle plus complet d'une tête artificielle comprenant les épaules et le torse. Limitée par le stockage mémoire, la résolution du problème sur ce modèle n'a pu être réalisée que jusqu'à 2500Hz. Kahana a aussi calculé des réponses de pavillons bafflés et ce pour des modèles très précis (jusqu'à 16000 éléments), et à vérifier les modes de résonances des pavillons établis jadis par Shaw [KN00, Sha82].

Motivés par la faisabilité de prédiction de HRTF à partir de modélisation simple de la morphologie de l'auditeur, Algazi *et al.* [AD02] ont mis au point un modèle baptisé *Snowman*, où la tête est modélisée par une sphère et le torse soit par une sphère, soit par une ellipsoïde (le cou n'étant pas pris en compte). Les résultats des simulations BEM, effectuées pour des fréquences comprises entre 500 et 5000Hz, montrent de bonnes correspondances avec les mesures sur une réplique physique du modèle, ceci donnant une autre preuve de la validité de la BEM pour l'acquisition des HRTF. Ces simulations ont également permis l'identification des contributions de la tête et du torse par l'observation de figures de diffraction. Les auteurs concluent quant à la validité de modèles géométriques simples pour la prédiction des HRTF en basses fréquences. Le modèle *Snowman* étant individualisable, cette étude montre aussi que l'on peut adapter des modèles géométriques simples à la morphologie d'un auditeur.

Fels *et al.* [FBV04] ont adapté numériquement le modèle de tête paramétrique mis au point par Genuit lors de sa thèse [Gen84] et ont conduit des simulations BEM sur celui-ci. En se servant de la formulation réciproque pour abaisser la charge de travail et en définissant l'entrée du canal auditif comme un piston vibrant, ils ont mené les calculs jusqu'à des fréquences de l'ordre de 6 à 8kHz. Ils ont validé les résultats issus des calculs BEM à l'aide d'une réplique physique du modèle de Genuit.

L'étude, débutée par Pernaux [Per03] et continuée par Busson [Bus05] au cours de leurs travaux de thèse, se situe dans la droite lignée de ceux d'Algazi. Ils ont calculées les HRTF de modèles géométriques simples, partant de la sphère avec oreilles centrées à un

modèle complet, comprenant une tête sphérique un cou cylindrique et un demi torse ellipsoïdal, jusqu'à des fréquences de l'ordre de 3 à 4kHz. Le modèle complet, baptisé modèle *Teletuby*, permet, à partir de photographies d'un sujet, d'être adapté à une morphologie donnée et ainsi d'être utilisé pour générer des HRTF individualisées. De plus, en analysant les figures de diffraction, ils ont pu mettre en évidence les contributions des différentes parties du corps modélisées par le modèle complet, notamment celle du torse dans les fréquences basses (inférieures à 2kHz).

### **b) Limitations**

Le calcul BEM est une méthode extrêmement coûteuse en ressources mémoires et en temps. L'utilisation de cette technique pour acquérir les HRTF propres à un sujet apparaît tout aussi contraignante que la mesure de celles-ci : il faut générer un modèle numérique de l'auditeur, et donc avoir accès à des systèmes de scan 3D, puis être en mesure d'effectuer les calculs sur une large bande de fréquence et donc de posséder un système de calcul dédié. Toutefois, cette méthode peut servir à retrouver l'information basse-fréquence souvent manquante ou inexploitable dans les HRTF mesurées.

L'intérêt des méthodes de calcul BEM semble donc résider plus dans l'étude des HRTF que dans l'acquisition de filtres individuels. Les conditions idéales offertes par la technique permettent un travail systématique, qui a d'ores et déjà conduit, à partir de modélisations géométriques simples, à la mise en évidence des contributions spectrales de certaines parties du corps.

### **3.2.3 Individualisation rapide des HRTF**

L'amélioration des performances de localisation induites par l'utilisation de HRTF individuelles apparaît évidente, mais dépend du type de stimulus employé et de la position de la source sonore. En effet, une expérience de Begault *et al.* a montré que, pour des stimuli vocaux et positionnés dans le plan horizontal, l'emploi d'une synthèse binaurale individualisée n'apportait pas d'amélioration significative en termes de performance de localisation par rapport à une synthèse binaurale générique [BWA01]. L'auteur mentionne toutefois que la largeur de bande de la voix ne permet pas d'exploiter pleinement les indices spectraux, et que la localisation est essentiellement induite par l'ITD et l'ILD. En utilisant des stimuli à bande large, permettant donc d'exciter la totalité des indices spectraux, notamment les indices liés aux pavillons d'oreille, Wenzel *et al.* ont mis en évidence l'importance de l'individualisation, notamment pour la localisation précise d'une source virtuelle sur un cône confusion [WAKW93]. Le taux de confusion avant/arrière semble également moindre moyennant l'utilisation de HRTF individuelles. En comparant les résultats issus des tests de localisation menés par Wenzel *et al.*, basés sur une synthèse binaurale générique, à ceux de Wightman et Kistler [WK89], basés sur une synthèse binaurale individuelle, le taux de confusion avant/arrière est divisé d'un facteur 3, passant d'environ 30 à 10%.

Si un utilisateur verra donc ses performances accrues en utilisant ses propres HRTF, nous avons vu que l'obtention même de celles-ci est un processus coûteux en temps et en moyen qui n'est réalisable que pour une poignée de privilégiés. En effet, si l'on désire mesurer ou calculer les HRTF d'un individu, les moyens à mettre en œuvre sont considérables. Il est donc utile de trouver une alternative pour améliorer les performances.

L'utilisation d'indices de localisation dynamiques est un des moyens pour atteindre en partie cet objectif. En effet, Begault *et al.* ont montré que l'utilisation d'un système de tracking de tête couplé à une synthèse binaurale générique permettait de faire chuter considérablement le taux de confusion avant/arrière [BWA00]. Cependant nous nous intéressons ici à la tâche de localisation dite statique, et à la seule exploitation des indices binauraux classiques, à savoir l'ITD, et l'ILD, et des indices spectraux.

On peut distinguer deux grands axes de recherche ayant pour but d'améliorer les performances de localisation de la synthèse binaurale, sans avoir à mesurer ou calculer les HRTF d'un individu, ou sans imposer de contrainte autre, qui rendrait leur application délicate pour le grand public (comme l'emploi d'un système de tracking de tête par exemple). La première approche consiste à utiliser la plasticité du système auditif humain en apprenant à un utilisateur à « écouter avec les oreilles d'un autre ». En effet, des travaux ont montré que les biais de localisation et les jugements confus diminuent après une certaine phase d'apprentissage [BKW04]. Cette étape d'apprentissage peut être vue comme une phase de recalibration des indices perceptifs. La deuxième approche consiste à fournir rapidement à l'utilisateur des filtres s'approchant au mieux de ses propres HRTF, autrement dit à « individualiser » les filtres de spatialisation. Ces méthodes ne consistent pas à mesurer ou calculer les HRTF d'un individu ou d'une représentation simplifiée de celui-ci, mais s'appuient généralement sur de grandes bases de données de HRTF mesurées ou calculées pour modéliser, chacune à leur manière, les liens entre les indices acoustiques de localisation et un certain nombre de paramètres morphologiques. Nous développerons ici les stratégies d'individualisation rapide qui nous semblent être les plus intéressantes.

Les techniques rapportées ci-dessous sont généralement utilisées pour individualiser la composante spectrale des HRTF. Elles sont donc habituellement associées à une méthode d'individualisation de l'ITD. Nous aborderons séparément les techniques d'individualisations de l'ITD.

#### **a) Sélection de HRTF en fonction d'un critère de proximité morphologique**

Cette méthode, proposée entre autres par Zotkin *et al.* [ZDDRM02, ZDD02], se base sur l'utilisation d'une importante base de données de HRTF, la base de données du laboratoire CIPIC [ADTA01, Copic]. Les HRTF d'environ 150 sujets ont été mesurées, ainsi qu'un nombre réduit de grandeurs anthropométriques. L'idée consiste à trouver le jeu de HRTF dont les paramètres morphologiques associés sont les plus proches de ceux de l'utilisateur.

Les auteurs reportent une amélioration sensible des résultats en terme de localisation avec une diminution de 30% du taux de confusion avant/arrière pour certains sujets. Cependant, les résultats de l'individualisation dépendent fortement de la proximité morphologique de l'utilisateur avec un des sujets de la base.

#### **b) Modélisation basée sur une décomposition linéaire des HRTF**

L'objectif de cette approche est de reconstituer un jeu de HRTF « adaptées » à un utilisateur à partir des mesures d'un nombre réduit de paramètres anthropométriques. La décomposition linéaire des HRTF ne constitue pas en elle-même un moyen de synthétiser des HRTF propres à l'auditeur, mais va permettre, par exemple, d'isoler certaines composantes présentant ou non des variations inter-individuelles, pour certaines positions de l'espace ou pour certaines bandes de fréquences.

Afin de pouvoir être utilisées à des fins d'individualisation, les données issues de l'analyse statistique doivent être reliées à des paramètres caractéristiques du sujet. En procédant à un certain nombre de mesure morphologique sur un panel de sujets, puis en effectuant une PCA (*Principal Component Analysis* ou analyse en composantes principales) sur ces données, Jin *et al.* [JLLCC00] ont constaté que seul un petit nombre des coefficients permettent de décrire les variations des grandeurs anthropométriques en fonction des sujets. En effectuant une seconde PCA sur les HRTF de ces sujets et une régression non linéaire entre les coefficients issus de chaque analyse, les auteurs ont pu définir une première correspondance entre paramètres morphologiques et HRTF.

Des études perceptives préliminaires ont fourni des résultats encourageants quant aux performances envisageable d'un tel système.

#### **c) Modélisation par warping fréquentiel**

Cette approche, développée par Middelbrooks [Mid99a, Mid99b] et poursuivie par Larcher [Lar01], consiste, à partir d'un jeu de HRTF, à synthétiser les HRTF d'un auditeur en effectuant des opérations de dilatation ou compression sur l'échelle des fréquences.

La modélisation est basée sur l'observation du spectre des HRTF où l'on constate que les variations anthropométriques inter-individuelles, notamment au niveau du pavillon d'oreille, se traduisent au niveau signal par des translations et des dilatations ou compressions des formes spectrales caractéristiques (sur l'échelle fréquentielle).

A notre connaissance, aucune évaluation perceptive n'a encore été rendue publique sur ce type de modélisation.

#### **d) Modélisation dite « structurelle »**

L'approche structurelle proposée par Genuit [Gen84] consiste à représenter les HRTF comme une combinaison de filtres, chacun traduisant des effets d'une partie du corps. Cette décomposition des HRTF est une approximation heuristique, car elle ne rend pas compte des interactions des ondes diffractées par les différentes structures anatomiques.

En suivant cette même idée, Algazi *et al.* [ADMT01] ont proposé une décomposition des HRTF à l'aide de deux filtres en cascade, le premier correspondant aux effets combinés de la tête et du torse, le deuxième traduisant les contributions du pavillon. En mesurant les HRTF d'un mannequin acoustique, puis séparément les filtres du même mannequin sans pavillon et les filtres des pavillons bafflés, ils ont montré une bonne concordance entre la mesure directe et le modèle composite. Cependant, aucune évaluation perceptive n'a pour l'instant été menée sur cette approche.

#### **e) Individualisation de l'ITD**

Les techniques d'individualisation de l'ITD consistent généralement à proposer un modèle plus ou moins simple, souvent obtenu par des considérations géométriques, et à établir le lien entre les paramètres du modèle et certaines grandeurs anthropométriques. La formulation de Woodworth, la plus simple et aussi la plus utilisée, donnant l'ITD par rapport à l'angle latéral, est obtenue en considérant la tête de l'auditeur comme une sphère rigide et en positionnant les oreilles diamétralement opposées. En notant  $a$  le rayon,  $c$  la célérité du son et  $\theta$  l'azimut de la source dans le repère polaire interaural, la formule de Woodworth s'écrit :

$$ITD(\theta) = \frac{a}{c}(\sin(\theta) + \theta) \quad (I-9)$$

En minimisant l'erreur entre les ITD estimés d'une grande base de données de HRTF et les ITD modélisés correspondants, puis en effectuant une régression non linéaire entre certains paramètres morphologiques et les rayons de tête sphériques optimaux, Algazi *et al.* [AAD01b] établissent la relation suivante, où  $d_1$ ,  $d_2$  et  $d_3$  sont respectivement la demi-largeur, la demi-profondeur et la demi-hauteur de la tête :

$$a = 0,51d_1 + 0,019d_2 + 0,18d_3 + 3,2 \quad (I-10)$$

Le modèle de tête sphérique ne permet cependant pas de reproduire certaines variations présentes sur les cônes de confusion ou les fluctuations de l'ITD en élévation, dans un système de coordonnées polaires-interaures, peuvent atteindre 18% de la valeur maximale de l'ITD sur un plan d'azimut constant. Les sources identifiées de ces erreurs sont la non sphéricité de la tête et la position des oreilles par rapport au centre de la tête. Le modèle sphérique permet une prédiction robuste et assez fidèle de l'ITD mais qui ne varie pas avec l'élévation.

D'autres modélisations ont été mises au point pour traduire les variations de l'ITD sur les cônes de confusion, comme par exemple le modèle ellipsoïdale proposé par Duda *et al.* [DAA99] ou le modèle sphérique prenant en compte le décalage des oreilles proposé par Busson [Bus05, BNW04]. Mais, à notre connaissance, les relations entre ces modèles et les paramètres anthropométriques n'ont pas encore été formulées.

### 3.2.4 Vers une nouvelle approche de l'individualisation des HRTF

Quelque soit la stratégie d'individualisation envisagée, l'obtention de filtres HRTF adaptés à un auditeur demande une démarche rigoureuse de la part de ce dernier. Si la méthode choisie est basée sur des paramètres anthropométriques, le problème est de fournir les outils permettant de garantir une précision et une répétabilité des mesures morphologiques, notamment pour les grandeurs liées aux pavillons d'oreilles. Si la méthode choisie nécessite une phase de réglage empirique, par exemple par test d'écoute, celle-ci risque d'être longue et fastidieuse. Pour des applications professionnelles, ce type de contraintes peut être envisagé. Pour des applications grand public, il se peut que les auditeurs fassent abstraction de la phase d'individualisation si celle-ci est trop complexe ou trop longue. Le risque d'obtenir un jeu de HRTF inadaptées à l'utilisateur final est alors grand.

Une nouvelle idée, initié par Katz en 1997, est d'essayer d'exploiter les pavillons de l'auditeur lors de la phase de restitution, espérant ainsi limiter l'individualisation à quelques paramètres morphologiques, non liés aux pavillons et donc plus facilement mesurables. L'approche consiste à utiliser deux réseaux de transducteurs, un autour de chaque pavillon d'oreille, pour reconstruire les fronts d'onde incidents. Les indices de localisation liés aux pavillons sont alors naturellement présents dans les signaux captés par les tympanes, de part la modification de l'onde engendrée par les phénomènes de

diffraction et réflexion sur ces derniers. Nous n'avons connaissances que de deux études basées sur cette idée. Chacunes ont abouties au prototypage d'un casque multi-transducteurs et à une première évaluation perceptive du dispositif. Celles-ci restent succinctes quant à la théorisation du concept, et tentent essentiellement d'apporter une réponse rapide quant aux performances envisageables d'un tel système de spatialisation. La première étude [Sas97] propose un réseau de cinq transducteurs positionnés en demi-cercle dans le plan horizontal (pour chaque pavillon). Les filtres associés à chaque transducteur sont synthétisés à partir d'un calcul par élément de frontière (BEM) pour une tête donnée, et ce jusqu'à 10kHz. La projection de l'intensité active (calculée au lieu) du transducteur sur la normale du transducteur est utilisée pour pondérer l'information de pression (calculée au même lieu), et ainsi alimentée le transducteur. Une évaluation perceptive a montré une certaine équivalence des performances de localisation entre ce système et une synthèse binaurale générique dans le plan horizontal. Cependant, les stimuli utilisés étaient des voix filtrés à 4 ou 10kHz, et l'évaluation ne concernait que le plan horizontal, soit une exploitation relativement pauvre des indices spectraux pavillonnaires.

La deuxième étude a été menée par les fondateurs de la société A-Volute. Le réseau possède ici six transducteurs positionnés sur une coque hémisphérique. Les filtres associés à chaque transducteur ont été mesurés directement en chambre anéchoïque en positionnant des microphones omnidirectionnels aux lieux des transducteurs. Une première évaluation a également conclu à une équivalence des performances de localisation entre ce système et une synthèse binaurale générique, voire à une légèrement amélioration des confusions avant/arrière. L'évaluation comparative concernait cette fois tout l'espace, mais a été effectuée moyennant une synthèse dynamique (pour les deux systèmes). Or nous avons vu précédemment que l'ajout d'indices de localisation dynamiques permettait de lever certaines ambiguïtés de localisation, comme les confusions avant/arrière.

Le principe, sur lequel se fonde le concept de casque multi-transducteurs dédié à la reconstruction de l'onde sonore arrivant sur les pavillons d'oreille, est le principe de Huygens. Cependant, les deux études citées précédemment l'appliquent de façon approximative, en interprétant chacune à sa manière l'énoncé de Huygens, plutôt qu'en se basant sur la théorie physique qui en découle : l'holophonie.

#### **4 CONCLUSION**

Dans ce premier chapitre, nous avons abordé l'audition spatiale et les différents mécanismes mis en jeu lors de la tâche de localisation. L'ITD et l'ILD sont des indices perceptifs importants qui permettent d'estimer le degré de latéralisation d'une source sonore. Cependant ils ne suffisent pas pour localiser précisément un son, et définissent ce que l'on appelle des « cônes de confusions ». Les indices nécessaires pour percevoir l'élévation d'une source et pouvoir distinguer l'hémisphère avant de l'hémisphère arrière sont contenus dans le spectre des HRTF. Bien que toute la lumière n'est pas encore été faite sur la nature exacte des indices spectraux, certaines zones de l'espace ont pu être associées à certaines zones fréquentielles, et la capacité de localisation dans ces directions particulières a ainsi pu être liée à certaines parties du corps humain. Le pavillon d'oreille, notamment, a un impact marqué sur la forme spectrale des HRTF

pour des fréquences supérieures à 4kHz, et semble avoir une responsabilité prédominante quant à la localisation des sons dans le plan médian, et plus généralement suivant les axes haut/bas et avant/arrière.

Nous nous sommes également intéressés aux techniques binaurales et aux différents travaux dont elles font l'objet à l'heure actuelle. Nous avons pu dégager l'idée principale des différentes recherches en la matière, à savoir rendre la synthèse binaurale configurable à partir de paramètres morphologiques, pour essayer de recréer ou d'adapter des HRTF à un auditeur donné. Cependant, ces techniques se heurtent essentiellement au problème que constitue le pavillon d'oreille, et le lien entre les grandeurs caractéristiques de ce dernier et les HRTF est loin d'être facile à établir.

Ce constat nous a permis d'introduire l'idée que nous allons développer dans la suite de la thèse : utiliser les propres pavillons d'un individu pendant la phase d'écoute en tentant de reconstruire l'onde sonore arrivant sur ces derniers.

# II

## Concept d'holophonie binaurale

## 1 INTRODUCTION

Le concept que nous allons développer est celui de l'holophonie binaurale. Le principe consiste à utiliser des techniques holophoniques pour reconstruire les champs sonores à proximité des pavillons d'oreille d'un auditeur. En pratique, il s'agit de concevoir des réseaux de haut-parleurs entourant chacun des pavillons et de les alimenter de telle sorte à ce que les signaux de pression aux tympan s'approchent de ceux qui auraient été induits par une scène sonore réelle et la diffraction des ondes acoustiques sur le corps de l'utilisateur.

Avant de caractériser ces réseaux ainsi que les filtres qui leurs seront associés, il est nécessaire de comprendre les principes généraux de l'holophonie, et de voir comment ceux-ci sont employés pour synthétiser des champs sonores. C'est pourquoi nous commencerons ce chapitre par une présentation générale de l'holophonie avant d'aborder son application à la synthèse binaurale.

## 2 HOLOPHONIE

Pour comprendre l'holophonie et ses principes, nous nous sommes essentiellement basés sur les travaux de thèse de Nicol [Nic99].

### 2.1 Principe de Huygens

L'holophonie est une technique de reproduction sonore qui consiste à recréer un champ acoustique dans un volume à partir de la connaissance du champ à la surface délimitant ce volume. Elle est basée sur le principe de Huygens. Dans la littérature, l'holophonie est également nommée WFS pour *Wave Field Synthesis* (reconstruction de champ acoustique) [BVV93].

#### 2.1.1 Concept de source secondaire

Au XVII<sup>ème</sup> siècle, Huygens a observé que, lorsqu'une onde se propage, chaque particule du milieu de propagation se comporte, une fois qu'elle a été mise en vibration, comme une source qui rayonne à son tour une ondelette. La superposition des différentes ondelettes émises par l'ensemble des particules excitées à un instant donné reforme l'onde primaire aux instants ultérieurs. Le principe de Huygens peut donc être formulé de la manière suivante : le front d'onde rayonné par une source se comporte comme une distribution de sources dites secondaires, par opposition à la source primaire ayant engendrée le front d'onde.

Deux types d'applications utilisent ce principe : des applications d'annulation de champ acoustique ou des applications de reproduction de champ acoustique dont l'holophonie fait partie.

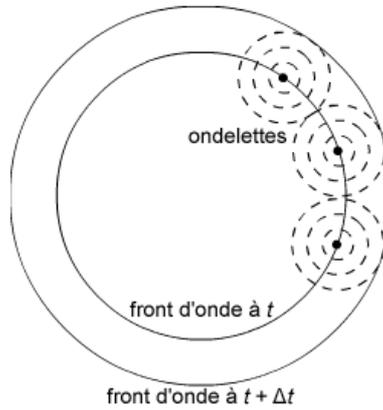


Fig. II-1 : Illustration du principe de Huygens

### 2.1.2 Problèmes aux limites

Huygens n'énoncera son principe que de manière qualitative, apportant la notion de source secondaire. Il faudra attendre les travaux de Kirchhoff et Helmholtz qui, au cours des XIX<sup>ème</sup> et XX<sup>ème</sup> siècles, établiront plusieurs énoncés quantitatifs du principe de Huygens.

Si l'on désigne l'espace dans lequel rayonnent les sources primaires par  $\Omega$ , on cherche à le décomposer en deux sous-espaces :

- $\Omega_1$  constituant le domaine des sources et contenant donc les sources primaires,
- $\Omega_2$  constituant le domaine de restitution et ne contenant aucune source primaire.

L'objectif est d'exprimer le champ à l'intérieur de  $\Omega_2$  en considérant uniquement une distribution de sources secondaires réparties sur la frontière  $\partial\Omega$  entre les domaines  $\Omega_1$  et  $\Omega_2$  (voir Fig. II-2). Il s'agit d'un problème aux limites pouvant être résolu par la méthode des fonctions de Green. Nous n'aborderons pas les développements mathématiques liés à la résolution de ce problème, mais nous intéresserons uniquement à l'expression de la solution : le signal de pression  $p$  présent à l'instant  $t$  et au point  $\vec{r}$  du domaine  $\Omega_2$  est donné par l'équation intégrale suivante :

$$p(\vec{r}, t) = \iint_{\partial\Omega} \left\{ \int_{t_1}^{t_2} [g(\vec{r} - \vec{r}_0, t - t_0) \vec{\nabla} p_0(\vec{r}_0, t_0) - p_0(\vec{r}_0, t_0) \vec{\nabla} g(\vec{r} - \vec{r}_0, t - t_0)] dt_0 \right\} \cdot \vec{n}_0(\vec{r}_0) dS_0 \quad (\text{II-1})$$

Le vecteur  $\vec{r}_0$  désigne la variable d'intégration suivant la double intégrale et le vecteur  $\vec{n}_0$  la normale unitaire à la surface  $\partial\Omega$  au point  $\vec{r}_0$ . La variable d'intégration temporelle est  $t_0$  et les bornes  $t_1$  et  $t_2$  correspondent aux instants initial et final d'observation du champ. La fonction  $p$  représente la grandeur caractéristique du champ acoustique, à savoir ici la pression acoustique. La fonction  $g$  désigne la fonction de Green associée au problème.

Cette expression correspond à la formulation la plus générale du principe de Huygens. Les deux termes  $g(\vec{r} - \vec{r}_0, t - t_0) \vec{\nabla} p_0(\vec{r}_0, t_0)$  et  $p_0(\vec{r}_0, t_0) \vec{\nabla} g(\vec{r} - \vec{r}_0, t - t_0)$  s'interprètent

comme la pression induite au point  $\vec{r}$  par deux sources situées en un point  $\vec{r}_0$  de la surface  $\partial\Omega$ . Les fonctions  $g(\vec{r} - \vec{r}_0, t - t_0)$  et  $\vec{\nabla}g(\vec{r} - \vec{r}_0, t - t_0)$  caractérisent le rayonnement de ces sources, tandis que les fonctions  $\vec{\nabla}p_0(\vec{r}_0, t_0)$  et  $p_0(\vec{r}_0, t_0)$  correspondent aux signaux qui les alimentent.

Différents énoncés sont obtenus suivant la fonction de Green choisie pour résoudre ce problème. Le choix de la fonction de Green est a priori arbitraire, mais il peut être orienté par la géométrie de la surface  $\partial\Omega$ . De plus, la forme de la fonction de Green influe directement sur la nature du rayonnement des sources secondaires. La représentation la plus utilisée est l'intégrale de Kirchhoff, qui impose une condition sur la géométrie du problème et fixe l'expression de la fonction de Green associée.

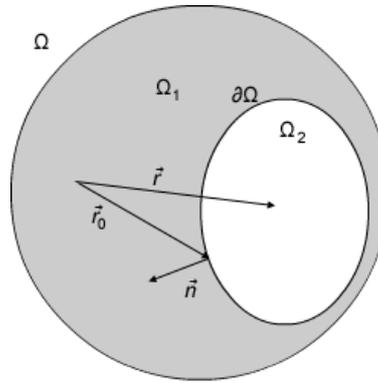


Fig. II-2 : Géométrie du problème de Kirchhoff

## 2.2 L'intégrale de Kirchhoff

Deux conditions spécifiques sont imposées par l'approche de Kirchhoff quant à la formulation du principe de Huygens, et concernent comme nous venons de le mentionner :

- la géométrie du problème,
- la fonction de Green associée.

En effet, le domaine de restitution  $\Omega_2$  ne doit contenir aucune source primaire et est entouré par une surface  $\partial\Omega$  fermée. De plus, la fonction de Green choisie correspond à son expression en espace libre notée  $g_\infty$  :

$$g_\infty(\vec{r} - \vec{r}_0, t - t_0) = \frac{\delta\left(\frac{R}{c} - \tau\right)}{4\pi R} \quad (\text{II-2})$$

avec  $R = |\vec{R}| = |\vec{r} - \vec{r}_0|$  et  $\tau = t - t_0$ . Si l'on se place dans les conditions décrites ci-dessus, l'expression, sous sa forme temporelle, de la pression  $p$  induite à l'intérieur du volume  $\Omega_2$  s'écrit  $\forall \vec{r} \in \Omega_2$  :

$$p(\vec{r}, t) = \frac{1}{4\pi} \iint_{\partial\Omega} \left\{ \frac{\vec{\nabla} p_0(\vec{r}_0, t_0) \cdot \vec{n}_0}{R} - \frac{\vec{R}}{R} \cdot \vec{n}_0 \left[ \frac{p_0(\vec{r}_0, t_0)}{R^2} + \frac{1}{Rc} \frac{\partial p_0(\vec{r}_0, t_0)}{\partial t_0} \right] \right\}_{t_0=t-Rc} dS_0 \quad (\text{II-3})$$

L'expression fréquentielle de l'intégrale de Kirchhoff conduit à l'intégrale de Kirchhoff-Helmholtz. La pression est alors exprimée en fonction de la pulsation  $\omega$  et la nouvelle formulation s'obtient par application de la transformée de Fourier. On a alors  $\forall \vec{r} \in \Omega_2$  :

$$p(\vec{r}, \omega) = \iint_{\partial\Omega} \left[ \vec{\nabla} p_0(\vec{r}_0, \omega) \cdot \vec{n}_0 - \frac{\vec{R}}{R} \cdot \vec{n}_0 (1 + jkR) \frac{p_0(\vec{r}_0, \omega)}{R} \right] \frac{e^{-jkR}}{4\pi R} dS_0 \quad (\text{II-4})$$

La nature du rayonnement des sources secondaires est ici bien identifiée, et les deux types de sources secondaires sont des monopôles, dont l'intensité est proportionnelle au gradient de pression  $\vec{\nabla} p_0 \cdot \vec{n}$ , et des dipôles, dont l'intensité est proportionnelle à la pression  $p_0$ . En des termes plus simples, chaque source secondaire est composée de deux sources élémentaires :

- une source monopolaire alimentée par un microphone à gradient de pression,
- une source dipolaire alimentée par un microphone de pression.

### 2.3 Les intégrales de Rayleigh

En choisissant convenablement l'expression de la fonction de Green associée au problème dans l'équation (II-1), il est possible de restreindre la composition des sources secondaires à une seule source élémentaire. Par exemple, si l'on choisie l'expression de la fonction de Green en espace semi-infini et que l'on impose à son gradient de s'annuler sur la surface  $\partial\Omega$ , on obtient la première intégrale de Rayleigh : chaque source secondaire est composé d'une seule source monopolaire alimentée par un microphone à gradient de pression. L'expression fréquentielle de cette intégrale est la suivante  $\forall \vec{r} \in \Omega_2$  :

$$p(\vec{r}, \omega) = \iint_{\partial\Omega} \vec{\nabla} p_0(\vec{r}_0, \omega) \cdot \vec{n}_0 \frac{e^{-jkR}}{2\pi R} dS_0 \quad (\text{II-5})$$

De façon similaire, si l'on impose à la fonction de Green elle-même de s'annuler sur la surface  $\partial\Omega$ , on obtient la deuxième intégrale de Rayleigh : chaque source secondaire est composée d'une seule source dipolaire alimentée par un microphone de pression. L'expression fréquentielle de cette intégrale est la suivante  $\forall \vec{r} \in \Omega_2$  :

$$p(\vec{r}, \omega) = - \iint_{\partial\Omega} \frac{\vec{R}}{R} \cdot \vec{n}_0 (1 + jkR) \frac{p_0(\vec{r}_0, \omega)}{R} \frac{e^{-jkR}}{2\pi R} dS_0 \quad (\text{II-6})$$

L'expression de la fonction de Green en espace semi-infini est applicable, par exemple, lorsqu'au lieu d'effectuer la partition de l'espace  $\Omega$  en entourant l'un des deux sous-espaces  $\Omega_1$  ou  $\Omega_2$  par une surface fermée, on sépare ces deux sous-espaces par un plan infini. Il faut donc comprendre que, dans le cas d'une distribution planaire de sources secondaires, le champ induit par le réseau monopolaire est identique au champ induit par le réseau dipolaire (dans le sous-espace d'écoute  $\Omega_2$ ), et que par conséquent il est possible de se limiter à l'utilisation de l'un ou l'autre des réseaux.

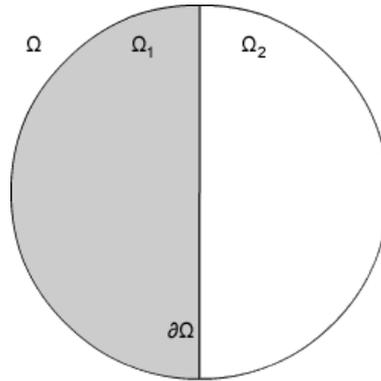


Fig. II-3 : Géométrie du problème selon l'approche de Rayleigh

## 2.4 Application à la reproduction d'un champ sonore

### 2.4.1 Principe et limitations

A partir de l'intégrale de Kirchhoff, on peut définir une méthode de reproduction sonore basée sur l'utilisation d'un réseau de transducteurs entourant la zone de reconstruction du champ acoustique, couplé à un réseau de microphones captant le champ primaire généré par des sources placées en dehors de la zone de reconstruction. Ces deux réseaux doivent être coïncidents : un microphone à gradient de pression doit être positionné au même lieu qu'un transducteur monopolaire qu'il va alimenter ; de la même manière, un microphone à pression doit être situé au même endroit qu'un transducteur bipolaire qu'il va également alimenter. On peut donc distinguer deux étapes que sont la prise de son et la restitution.

Une première limitation pratique apparaît évidente : la théorie suppose une distribution continue de sources secondaires, ce qui est impossible en réalité. On est obligatoirement amené à considérer un réseau discret de transducteurs, ce qui introduit le problème d'échantillonnage spatial. De plus, il est intéressant de pouvoir simplifier le réseau de sources secondaire en n'utilisant par exemple que des transducteurs monopolaires, et ce quelque soit la géométrie du réseau. Nous verrons comment est effectuée cette simplification en pratique.

### 2.4.2 Discrétisation de la distribution de sources secondaires

En reprenant les notations utilisées précédemment, l'intégrale de Kirchhoff peut s'écrire sous la forme d'un produit de convolution :

$$p(\vec{r}) = \iint_{\partial\Omega} [g_\infty(\vec{r} - \vec{r}_0) \vec{\nabla} p_0(\vec{r}_0) \cdot \vec{n}_0 - p_0(\vec{r}_0) \vec{\nabla} g_\infty(\vec{r} - \vec{r}_0) \cdot \vec{n}_0] dS_0 \quad (\text{II-7})$$

La discrétisation du réseau de sources secondaires s'obtient en découpant la surface  $\partial\Omega$  en un nombre donné  $N$  de surfaces élémentaires  $\Delta S_0^k$  ( $k = 1 \dots N$ ) et en disposant une source secondaire au centre de chacune de ces surfaces, repéré par le vecteur  $\vec{r}_0^k$ . La discrétisation du réseau équivaut à substituer aux intégrales continues des sommes discrètes. Le signal de pression  $\hat{p}$  restitué par le réseau discret s'exprime alors :

$$\hat{p}(\vec{r}) = \sum_{k=1}^N [g_\infty(\vec{r} - \vec{r}_0^k) \vec{\nabla} p_0(\vec{r}_0^k) \cdot \vec{n}_0^k - p_0(\vec{r}_0^k) \vec{\nabla} g_\infty(\vec{r} - \vec{r}_0^k) \cdot \vec{n}_0^k] \Delta S_0^k \quad (\text{II-8})$$

L'échantillonnage d'un signal à support spatial est équivalent à celui d'un signal temporel. Un champ sonore évalué sur une surface varie en fonction de coordonnées d'espace. Dans le domaine dual, le spectre correspondant est alors fonction de fréquences spatiales. La question essentielle de tout échantillonnage réside dans le choix du pas d'échantillonnage : plus le champ varie rapidement sur la surface  $\partial\Omega$ , plus le pas d'échantillonnage spatial  $\Delta$  doit être faible pour assurer une reconstruction acceptable du champ.

Cette idée se traduit par la condition de Shannon. Sans développer plus précisément la notion de spectre spatial qui permet d'effectuer la transposition du théorème de Shannon d'un signal discret à support temporel à un signal discret à support spatial, nous donnons ici la condition la plus générale et la plus restrictive sur le pas d'échantillonnage spatial. Si l'on nomme  $f_{max}$  la fréquence maximale pour laquelle nous désirons reconstruire correctement le champ acoustique (ce qui équivaut à une longueur d'onde minimale  $\lambda_{min}$ ) et que l'on note  $c$  la vitesse du son dans le milieu considéré, la condition de Shannon peut s'exprimer ainsi :

$$\Delta < \frac{\lambda_{min}}{2} = \frac{c}{2f_{max}} \quad (\text{II-9})$$

### 2.4.3 Généralisation des intégrales de Rayleigh

La redondance des champs induits par un réseau monopolaire et un réseau bipolaire associé conduit à la généralisation des intégrales de Rayleigh pour une géométrie quelconque. En effet, il est possible d'approximer la reconstruction d'un champ généré par différentes sources primaires en n'utilisant qu'un seul type de source élémentaire. Cependant, l'application directe des intégrales de Rayleigh n'est pas correcte si le réseau n'est pas planaire : un signal parasite apparaît, celui-ci n'étant plus annulé par le deuxième réseau de sources élémentaires.

Deux solutions permettent d'atténuer voire d'annuler l'influence de ce signal d'erreur. La première consiste à activer ou non un transducteur en fonction de l'emplacement relatif de la source primaire par rapport à ce transducteur. La deuxième consiste à

modifier la directivité des microphones utilisés pour approximer le fonctionnement de la première solution : le signal émis par un transducteur lié à la reconstruction du champ généré par une source primaire est naturellement pondéré à l'enregistrement en fonction de l'emplacement relatif de la source primaire par rapport à ce transducteur.

Dans le cas de l'utilisation exclusive d'un réseau quelconque de transducteurs monopolaires, la première technique consiste à activer un transducteur si l'onde émise par la source primaire est captée par le pôle positif du microphone à gradient de pression associé. De la même manière, si l'onde est captée par le pôle négatif, le transducteur est désactivé et n'intervient donc pas dans la reconstruction du champ acoustique. Pour des réseaux planaires vérifiant la géométrie du problème de Rayleigh, cette pondération logique est induite par construction, car toute source primaire ne peut être vue que par les pôles positifs du réseau. Cette solution est cependant difficile à mettre en œuvre pour l'enregistrement par un réseau quelconque d'une scène complexe, car il est nécessaire de connaître l'emplacement et la nature de chaque source primaire. En théorie, cette intervention sur l'activité des transducteurs revient à simuler l'utilisation de microphones dont la directivité correspond à une demie « figure en huit », or ce type de microphone n'existe pas en réalité. C'est là que la deuxième solution intervient, en proposant l'utilisation de microphones unidirectionnels, dont la directivité s'approche d'une demie « figure en huit », par exemple des microphones cardioïdes. Cette deuxième technique est généralement utilisée pour l'enregistrement d'une scène sonore complexe, car il n'est plus nécessaire de connaître l'emplacement et la nature de chaque source primaire.

#### 2.4.4 Pondération logique à partir de l'intensité active

Dans le cas d'une scène simple composée d'une seule source primaire en champ libre, la position relative de celle-ci par rapport à un microphone du réseau est donnée par le vecteur d'intensité active au lieu du microphone. La pondération logique de la captation peut donc être effectuée par le biais de l'intensité active (voir Fig. II-4).

Dans le cas d'une scène complexe, même si la nature et le positionnement de chaque source primaire sont inconnus, il est tout à fait possible d'évaluer l'intensité active du champ global à la position de chaque source secondaire. Il est donc envisageable d'appliquer une pondération logique commandée par l'intensité active, et ce quelque soit le champ acoustique primaire.

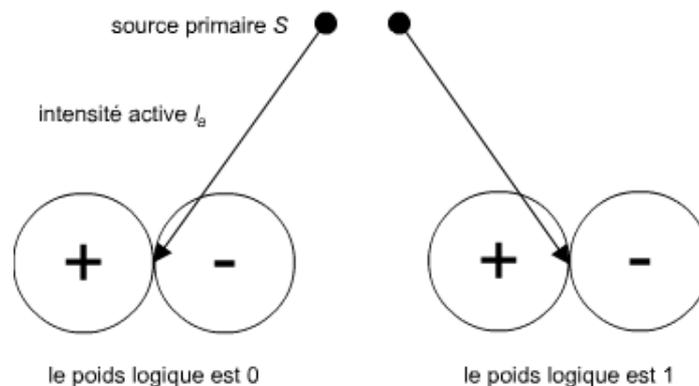


Fig. II-4 : Illustration de la pondération logique à partir de l'intensité active

### **3 HOLOPHONIE BINAURALE**

#### **3.1 Présentation du concept**

Les indices de localisation codés dans les HRTF sont liés aux modifications que subit une onde sonore lors de son trajet de la source aux tympans d'un auditeur. Ces transformations sont dues aux phénomènes de réflexion et de diffraction engendrés par les différentes parties du corps, à savoir le torse, les épaules, la tête et les pavillons. Les différentes techniques d'individualisation des HRTF que nous avons abordées au chapitre précédent se heurtent à un problème complexe : comment faire le lien entre la géométrie des différentes parties du corps et les modifications spectrales qu'elles engendrent ? Même si des formes géométriques simples peuvent constituer de bons substituts pour le corps (torse et épaules) et la tête, la forme complexe des pavillons est difficile à modéliser, ainsi que la relation entre les différentes grandeurs anthropométriques des pavillons et le codage spectral qu'elles impliquent.

L'holophonie binaurale a pour but d'utiliser les pavillons d'un auditeur en laissant ces derniers codés naturellement les indices de localisation qui leur sont liés et qui sont propre à l'auditeur. Il s'agit donc d'appliquer des techniques holophoniques, pour reconstruire les ondes arrivant sur chacun des deux pavillons d'oreille à l'aide de deux réseaux de sources secondaires distincts, englobant chacun des pavillons. L'approche holophonique offre un découplage des indices de localisation liés aux pavillons de ceux liés au reste du corps : les deux réseaux diffusent des signaux qui contiennent les indices associés à la tête et au corps, entre autre les indices interauraux classiques (ITD, ILD), et l'onde recrée par chacun des réseaux est modifiée par chacune des deux oreilles qui codent donc naturellement les indices spectraux pavillonnaires propre à l'auditeur.

Le problème d'individualisation des HRTF est simplifié, dans le sens où la diffusion ne s'effectue plus au niveau des conduits auditifs, mais en amont des pavillons : on s'affranchit ainsi du problème complexe de modélisation des modifications spectrales dues aux pavillons. Cependant, la mise en œuvre de l'holophonie binaurale est plus complexe que celle de techniques binaurales classiques, car on ne s'efforce plus de caractériser le champ simplement en deux points (les tympans ou l'entrée des conduits), mais sur deux surfaces, à la position de chaque source secondaire.

#### **3.2 Solutions d'enregistrement holophonique retenues**

A partir des éléments que nous avons présentés en II.2.4.3 et II.2.4.4, nous proposons d'utiliser deux types de captation holophonique la deuxième pouvant être appliquée de deux manières différentes suivant que l'on connaisse ou non les sources qui composent le champ primaire.

- **CARD** : utilisation de microphones cardioïdes. Application à l'enregistrement de tout type de champ primaire.
- **FIG-8** : utilisation de microphone à gradient de pression et pondération logique commandée par l'intensité active. Lorsque cette méthode est appliquée séparément à l'enregistrement de chaque champ induit par chaque source primaire, on parlera de solution FIG-8 S-LOG. Lorsque la scène sonore est trop

complexe, cette technique ne peut être appliquée que sur le champ total. On parlera dans ce cas de solution FIG-8 T-LOG.

Il est attendu que la méthode FIG-8 S-LOG permette la meilleure reconstruction du champ, dans le sens où elle consiste en une généralisation exacte de l'approche de Rayleigh, mais elle n'est applicable à un champ complexe que si l'on peut enregistrer séparément les contributions de chaque source qui le compose. La solution CARD est la solution généralement employée lorsque la solution FIG-8 S-LOG ne peut être utilisée (typiquement pour l'enregistrement d'une scène sonore réelle par un réseau non planaire). La robustesse de la méthode FIG-8 T-LOG, quant à elle, devrait être dépendante de la scène sonore et du réseau considéré : par exemple, si deux sources ont des incidences opposées par rapport à un microphone du réseau, c'est-à-dire que l'une est vue par le pôle positif et l'autre par le pôle négatif, le champ total d'intensité active induit par les deux sources ne sera représentatif d'aucun des deux champs induits séparément par chacune des deux sources. En d'autres termes plus, les sources seront dispersées autour du réseau holophonique, moins la captation FIG-8 T-LOG devrait être efficace ; a contrario, plus les sources seront rassemblées dans une même direction (par rapport au réseau), plus la méthode FIG-8 T-LOG se comportera comme la méthode FIG-8 S-LOG. On pourrait hâtivement décider ne pas considérer la variante T-LOG de la solution FIG-8. Cependant, elle pourrait constituer une alternative à la technique CARD, quand la scène à enregistrer est complexe et que la géométrie du réseau utilisé est quelconque. Si l'on se place par exemple dans le cas d'une seule source en présence d'un objet diffractant, la méthode FIG-8 S-LOG ne peut être appliquée (à moins de trouver un ensemble de sources induisant un champ équivalent au champ diffracté). Toutefois, la variante T-LOG pourrait être intéressante si, vue du réseau, la contribution du champ direct (émis par la seule source en champ libre) est prépondérante par rapport au champ diffracté (réfléchi par l'objet diffractant).

### 3.3 Calcul des signaux alimentant le réseau

Pour tout réseau discret composé de sources monopolaires, la première intégrale de Rayleigh, décrite par l'équation (II-5), et l'expression discrète de l'intégrale de Kirchhoff (II-8) nous permettent d'écrire :

$$\hat{p}(\vec{r}, \omega) = \sum_{i=1}^N s(\vec{r}_i, \omega) \frac{e^{-jk|\vec{r}-\vec{r}_i|}}{2\pi |\vec{r}-\vec{r}_i|} \Delta S_i \quad (\text{II-10})$$

où  $\vec{r}$  définit un point de la zone d'écoute,  $\vec{r}_i$  et  $\Delta S_i$  indiquent respectivement le lieu de la  $i^{\text{ème}}$  source secondaire et la surface associée, et  $s$  le signal mesuré en  $\vec{r}_i$ . Chaque source secondaire monopolaire du réseau discret est donc alimentée par le signal  $s(\vec{r}_i, \omega)\Delta S_i$ . En fonction de la méthode de captation considérée,  $s$  pourra s'exprimer de différentes manières.

### 3.3.1 Cas d'une captation FIG-8

Pour la méthode FIG-8,  $s_i$  est calculée ainsi :

$$s(\vec{r}_i, \omega) = K(\vec{r}_i, \omega) \vec{\nabla} p_i(\vec{r}_i, \omega) \cdot \vec{n}_i(\vec{r}_i) \quad (\text{II-11})$$

où  $K$  correspond au poids logique et est défini par les relations suivantes :

$$\begin{cases} K(\vec{r}_i, \omega) = 1 \text{ si } \Re(\vec{I}(\vec{r}_i, \omega)) \cdot \vec{n}_i(\vec{r}_i) > 0 \\ K(\vec{r}_i, \omega) = 0 \text{ sinon} \end{cases} \quad (\text{II-12})$$

$\vec{I}$  représente l'intensité, et  $\Re(\vec{I})$  correspond à l'intensité active.  $\vec{I}$  peut s'exprimer à partir du gradient de pression et de la pression :

$$\vec{I}(\vec{r}_i, \omega) = p^*(\vec{r}_i, \omega) \frac{\vec{\nabla} p(\vec{r}_i, \omega) \cdot \vec{n}_i(\vec{r}_i)}{j\rho\omega} \quad (\text{II-13})$$

où  $\rho$  est la masse volumique du milieu considéré et l'opérateur \* désigne la conjugaison complexe.

Le signal  $s$  est donc calculable à partir des seules informations de pression et de gradient de pression suivant la normale à la surface.

### 3.3.2 Cas d'une captation CARD

La captation CARD implique l'utilisation de microphones cardioïdes. On peut calculer  $s$  de la manière suivante :

$$s(\vec{r}_i, \omega) = \vec{\nabla} p_i(\vec{r}_i, \omega) \cdot \vec{n}_i(\vec{r}_i) + \frac{j\omega}{c} p_i(\vec{r}_i, \omega) \quad (\text{II-14})$$

Tout comme pour la captation FIG-8,  $s$  est calculable à partir des seules informations de pression et de gradient de pression suivant la normale à la surface.

## 3.4 Exemple d'application : réseau sphérique

Dans l'exemple suivant, nous considérons une scène sonore composée d'une seule source monopolaire. Les variantes S-LOG et T-LOG de la captation FIG-8 sont alors identiques. Le but n'est pas d'évaluer la robustesse des différentes solutions (ceci sera effectué dans le prochain chapitre), mais uniquement de valider les principes holophoniques appliqués à un cas simple de réseau non planaire.

Considérons un réseau sphérique de 5cm de rayon composé de 1280 sources secondaires (voir Fig. II-5). La condition de Shannon est vérifiée jusqu'à une fréquence de 38kHz. La source est positionnée à 2m du centre du réseau. La figure II-6 illustre la reconstruction temporelle d'une onde harmonique de fréquence 12kHz. Les deux méthodes fournissent de bons résultats. Cependant l'onde recrée par la méthode CARD est sujette à quelques artéfacts. La captation CARD étant moins directive que la captation FIG-8, certaines sources secondaires peuvent émettre en s'opposant au front d'onde synthétisé. La reconstruction du champ peut donc être dégradée à proximité de telles sources secondaires.

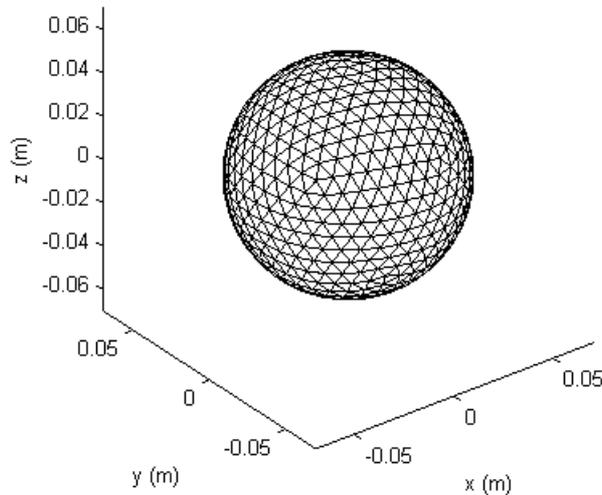


Fig. II-5 : Réseau sphérique de 5cm de rayon à 1280 sources secondaires

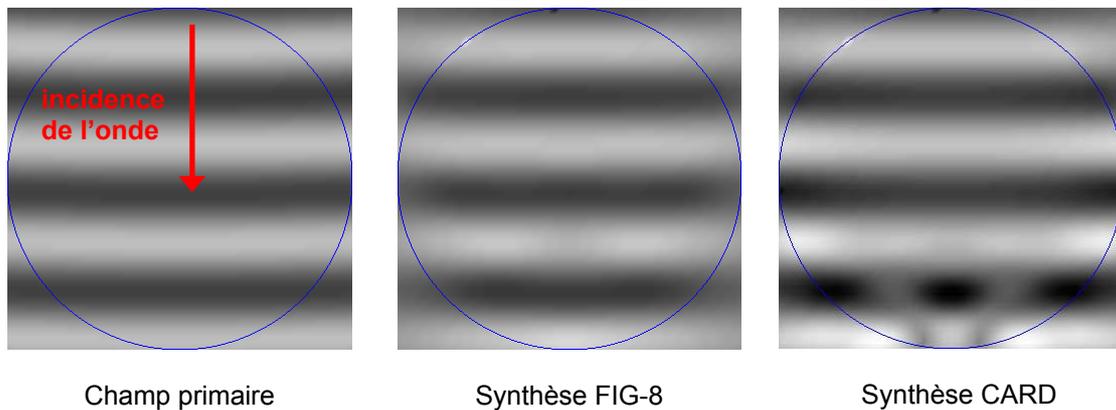


Fig. II-6 : Reconstruction holophonique par un réseau sphérique - 12kHz

La figure II-7 montre l'erreur de reconstruction en fonction de la fréquence, estimée au centre du réseau sphérique pour les deux méthodes. On s'aperçoit que plus on descend en fréquence plus le réseau a du mal à synthétiser le champ. Cette atténuation est liée dans une moindre mesure au type de captation employé, et dans une plus grande mesure à la taille du réseau (voir Fig. II-8 et II-9). En dessous d'une certaine fréquence, le réseau devient en quelque sorte « aveugle ». La géométrie du réseau agit comme un

filtre passe-haut et la taille du réseau définit la fréquence de coupure : plus le réseau sera petit, plus la fréquence de coupure sera élevée.

Il conviendra par la suite de prendre en considération ce phénomène, soit pour définir une bande de fréquence d'étude non sujette à cette atténuation, soit pour définir un filtre correctif.

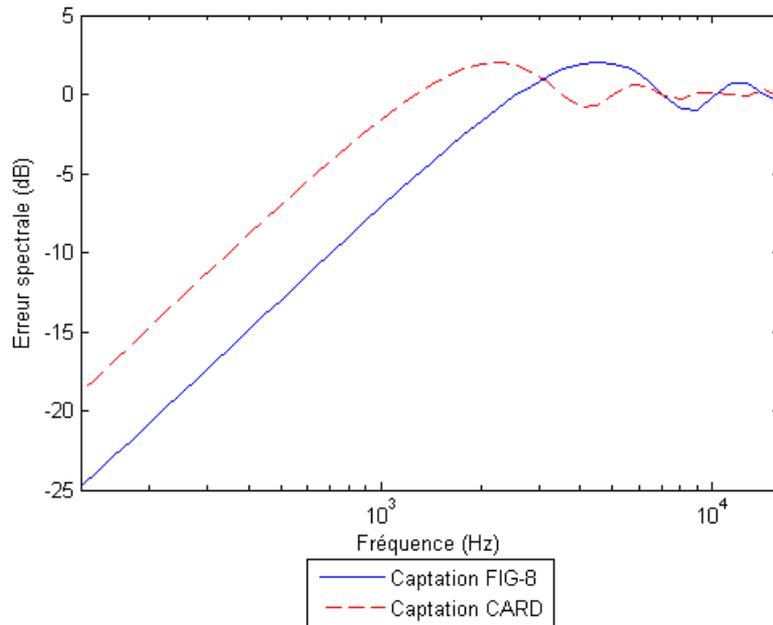


Fig. II-7 : Erreur de reconstruction en fonction de la méthode de captation - réseau sphérique

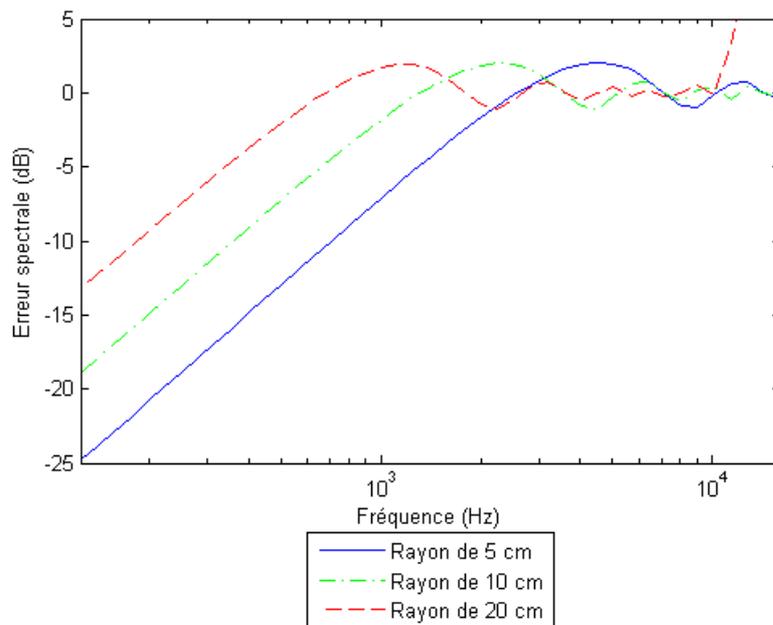


Fig. II-8 : Erreur de reconstruction en fonction de la taille du réseau sphérique - méthode FIG-8

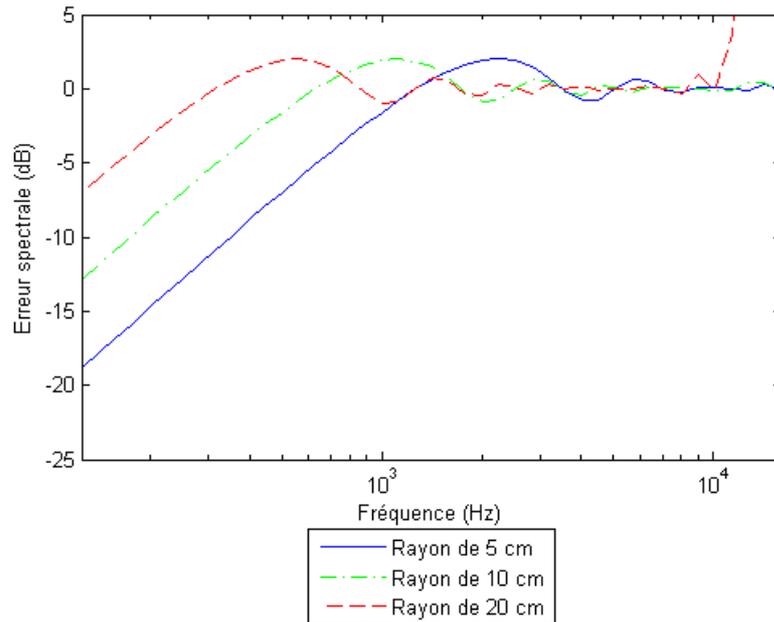


Fig. II-9 : Erreur de reconstruction en fonction de la taille du réseau sphérique - méthode CARD

### 3.5 Quelle géométrie de réseau ?

Les réseaux utilisés devront entourer les pavillons d'oreilles, leurs dimensions minimales sont donc principalement fonction des dimensions de l'oreille. De plus, l'objectif final étant la matérialisation de ces réseaux sous la forme d'un casque à plusieurs transducteurs, les dimensions maximales peuvent être déduites de considérations pratiques voire esthétiques.

Pour la géométrie, nous choisissons une surface hémisphérique dont le centre correspond à l'entrée du canal auditif. Cette géométrie permet d'entourer complètement un pavillon. De plus, chaque point d'un hémisphère est positionné à égale distance de son centre, et chaque normale à sa surface est dirigée vers son centre. Cette homogénéité des rapports source secondaire - entrée du canal, nous est parue intéressante.

Les données anthropométriques accompagnant les bases de données HRTF du laboratoire CIPIC [ADTA01, Cipic] ou du projet LISTEN [Listen] nous permettent d'effectuer quelques statistiques sur les dimensions des pavillons, notamment leur hauteur définie sur la figure II-5 comme la longueur  $d_5$ .

Pour la base de données CIPIC, la hauteur moyenne est de 64mm avec un écart-type de 5mm ; de plus les hauteurs minimale et maximale sont respectivement de 55mm et 76mm. En ce qui concerne la base de données LISTEN, la moyenne est de l'ordre de 65mm, l'écart-type de 7mm, la valeur minimale de 51mm et la valeur maximale de 81mm. Cependant, aucune mesure ne nous permet de situer l'entrée du canal auditif sur cette hauteur. En observant les pavillons d'un certain nombre d'individus, nous avons estimé que l'entrée du canal se situe environ au tiers de la hauteur. La distance limitante est alors en moyenne de 43mm. En choisissant un rayon de 50mm, on estime que plus

de 90% des pavillons des bases CIPIC et LISTEN peuvent être entourés sans collision par le réseau hémisphérique centré sur l'entrée du canal auditif, et que 100% peuvent être englobés si la condition de centrage du réseau est levée.

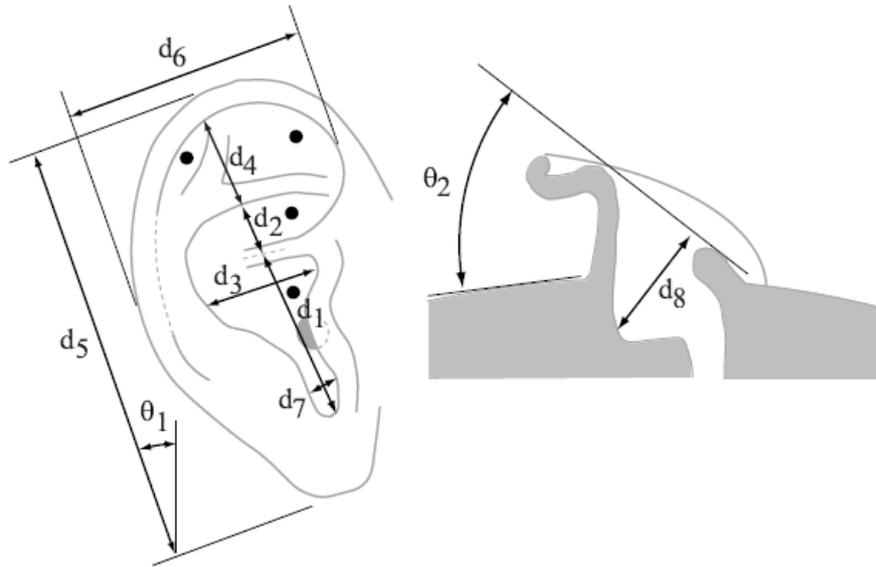


Fig. II-10 : Paramètres anthropométriques du pavillon d'oreille (CIPIC et LISTEN)

### 3.6 Définition des différentes fonctions de transfert

Pour clarifier le discours, nous allons définir ici les différents filtres ou signaux que nous allons être amenés à manipuler.

- $HRTF^o$  : HRTF liées à une morphologie donnée. L'exposant o signifie que ce sont les filtres originaux de référence, donc engendrés par la modification de l'onde lors de son trajet de la source primaire jusqu'au tympan. Les variables sont la position de la source primaire et la fréquence.
- $HRTF^r$  : l'indice r signifie que ce sont les HRTF reconstruits par application de méthodes holophoniques, par opposition avec les  $HRTF^o$  dits « originaux ». Les variables sont la position de la source primaire et la fréquence.
- CAHF : ce sont les filtres holophoniques qui vont permettre de recréer l'onde sonore arrivant sur les pavillons, et sont donc liés à une morphologie (hors pavillons) donnée. CAHF signifie en anglais *Circumaural Array Holophonic Filters* (littéralement filtres holophoniques liés au réseau circumaural). Les variables sont la position de la source secondaire dans le réseau holophonique, la fréquence, ainsi que la position de la source primaire.
- PRTF : ce sont les filtres traduisant les modifications d'une onde lors de son trajet de chaque source secondaire jusqu'au tympan. Ce sont en quelque sorte des HRTF uniquement liées à une morphologie de pavillons donnée. Les variables sont la position de la source secondaire dans le réseau holophonique et la fréquence.

Les données que nous manipulerons seront des données numériques discrètes. On pourra donc représenter ces différents filtres par des matrices de dimension égale au nombre de variables qui leur sont rattachés. Par exemple, les HRTF seront généralement vu comme une matrice de dimension 2, les CAHF une matrice de dimension 3, etc.

#### **4 CONCLUSION**

Ce chapitre a permis de présenter le concept d'holophonie binaurale, et de définir les différents éléments qui vont nous permettre de concevoir un système de restitution dédié à la mise en application de ce concept. Le point a été fait sur les techniques holophoniques et des méthodes de captation ont été établies. Celles-ci serviront à caractériser les filtres alimentant le réseau de sources secondaires. Une géométrie et une dimension de réseau ont été arrêtées. Celles-ci seront utilisées dans les prochains chapitres pour simuler et optimiser le fonctionnement d'un système de restitution dédié à l'holophonie binaurale.

# III

## Reconstruction holophonique des HRTF

## 1 INTRODUCTION

Ce chapitre fait rapport de différentes simulations que nous avons effectuées pour évaluer la qualité de reconstruction des HRTF par le biais des techniques holophoniques. Différents modèles de pavillon ainsi qu'un modèle de tête ont été utilisées pour conduire les différents calculs nécessaires (calcul des HRTF de référence, et calcul des CAHF et des PRTF). Outre la comparaison des trois méthodes de captation employées, nous essayerons de quantifier l'impact d'une réduction du nombre de sources secondaires : l'objectif final étant la conception d'un casque à plusieurs transducteurs, il est important d'évaluer le fonctionnement d'un système constitué d'un nombre de sources secondaires réaliste.

## 2 EVALUATION DE LA QUALITE DE RECONSTRUCTION

### 2.1 Reconstruction des HRTF

En utilisant les notations proposées au paragraphe II.3.3, on peut simuler la reconstruction des HRTF à partir de l'expression suivante :

$$HRTF^r(\theta, \varphi, f) = \sum_{i=1}^N CAHF_i(\theta, \varphi, f) \cdot PRTF_i(\theta, \varphi, f) \quad (\text{III-1})$$

où  $N$  est le nombre de sources secondaires du réseau considéré,  $CAHF_i$  est le filtre holophonique correspondant à la  $i^{\text{ème}}$  source secondaire et lié à la géométrie de tête considérée, et  $PRTF_i$  la fonction de transfert liant la  $i^{\text{ème}}$  source secondaire et le pavillon considéré.

La qualité de reconstruction est évaluée en comparant les HRTF reconstruites ( $HRTF^r$ ) et les HRTF originales ( $HRTF^o$ ) correspondant à la même géométrie de problème (même tête et même pavillon). Pour effectuer cette comparaison, nous utiliserons deux indices de mesure objective décrites ci-après.

### 2.2 Définition des outils de mesure objective

L'évaluation de la qualité de reconstruction des HRTF se fera par comparaison des HRTF reconstruites avec les HRTF de référence. Pour une position de source primaire donnée, il s'agit alors simplement de comparer deux variables observées un certain nombre de fois (chaque fréquence correspondant à une observation). Nous utiliserons donc des outils statistiques classiques, en définissant une mesure de distance et une mesure de corrélation.

Il est cependant utile de choisir un repère adéquat de représentation de ces vecteurs, pour garantir une cohérence perceptive. Pour se faire, nous choisissons une base de

représentation approximant la sensibilité du système auditif humain : nous comparerons donc les spectres des HRTF en décibel, évalués sur une échelle logarithmique en fréquence.

### 2.2.1 Distance spectrale

Soit  $A$  et  $B$  deux séries de  $N$  mesures, notées  $a_i$  et  $b_i$  avec  $i = 1, 2, \dots, N$ . Nous définissons la distance  $E$  de  $B$  par rapport à  $A$  comme l'erreur absolue moyenne entre  $A$  et  $B$  :

$$E = \frac{1}{N} \sum_{i=1}^N |b_i - a_i| \quad (\text{III-2})$$

Si l'on note  $A = 20 \log_{10} |HRTF^o|$  et  $B = 20 \log_{10} |HRTF^r|$ , où  $HRTF^o$  et  $HRTF^r$  sont des vecteurs contenant respectivement les spectres des HRTF originales et des HRTF reconstruites, calculés pour une position de source fixe et en  $N$  points fréquentiels répartis sur une échelle logarithmique, nous définissons la distance  $E$  entre le filtre reconstruit et le filtre de référence par :

$$E = \frac{1}{N} \sum_{i=1}^N 20 \log_{10} \frac{|hrtf_i^r|}{|hrtf_i^o|} \quad (\text{III-3})$$

où  $hrtf_i^o$  et  $hrtf_i^r$  sont respectivement les  $i^{\text{ème}}$  éléments des vecteurs  $HRTF^o$  et  $HRTF^r$ .

### 2.2.2 Corrélation spectrale

#### a) Coefficient de corrélation de Pearson

La formulation la plus usitée pour l'étude de la corrélation entre deux variables observées un nombre discret de fois  $N$ , correspond à celle du coefficient de corrélation selon Pearson. Si l'on considère  $A$  et  $B$  définies précédemment et que l'on note de plus  $m_A$  et  $m_B$  leurs moyennes respectives, ainsi que  $s_A$  et  $s_B$  leurs écart-types respectifs, le coefficient de corrélation  $R$  entre ces variables s'écrit :

$$R = \frac{\sum_{i=1}^N a_i b_i - N m_A m_B}{(N-1) s_A s_B} \quad (\text{III-3})$$

A l'instar de la définition que nous avons faite de la distance, le coefficient de corrélation spectrale entre une HRTF reconstruite et la HRTF de référence correspondante est obtenu en considérant les spectres en décibels estimés sur une échelle logarithmique.

### ***b) Interprétation géométrique de la corrélation***

Le coefficient de corrélation peut être vu comme le cosinus de l'angle entre deux vecteurs qui correspondent aux deux séries de mesures préalablement centrées (c'est-à-dire auxquelles on aura retranché leurs moyennes respectives).

Lorsqu'on désire effectuer une moyenne sur différents coefficients de corrélation, il est commun et pratique de calculer cette moyenne directement, voire sur les carrés des coefficients. Cependant, l'approche géométrique nous invite plutôt à effectuer ce calcul sur les angles associés.

Nous utiliserons donc les relations suivantes pour passer du coefficient de corrélation  $R$  à l'angle de corrélation  $\theta$  correspondant, et inversement :

$$\begin{cases} R = \cos \theta \\ \theta = \cos^{-1} R \end{cases} \quad (\text{III-4})$$

## **2.3 Prédominance des résonances ou des anti-résonances ?**

Les résonances et anti-résonances, présentes dans les HRTF et dues aux phénomènes de diffraction sur les pavillons d'oreilles, sont des indices importants de localisation en élévation. Nous avons vu que le rôle de ces indices n'est pas encore bien défini. Loin de prétendre déterminer le rôle de chaque résonance et chaque anti-résonance, nous avons voulu savoir simplement, si l'un ou l'autre de ces groupes d'indices a une influence prédominante en termes de localisation en élévation.

### ***2.3.1 Expérimentation perceptive***

Pour tenter de répondre à cette question, nous avons effectué un test de localisation dans le plan médian sur trois type de jeux de HRTF comprenant un jeu de référence, choisi pour chaque sujet participant à l'expérience parmi les HRTF mesurées de la base de données LISTEN [Listen] à l'aide d'une écoute informelle préalable, ainsi que deux jeux dérivés du premier, l'un conservant uniquement le motif spectral lié aux anti-résonances, l'autre conservant uniquement le motif spectral lié aux résonances. La distinction entre les zones spectrales appartenant à chacun des motifs a été réalisée à partir d'un jeu de spectres dit « de frontière » : pour chaque HRTF, le motif des résonances est considéré comme la partie du spectre située au dessus de la frontière ; de même, le motif des anti-résonances est choisi comme la partie située en dessous du spectre de frontière. Une question importante était de déterminer le jeu de spectres de frontière. Le lissage des HRTF nous est apparu être une approche intéressante, dans le sens où, par nature, elle permet d'atténuer les variations, et par conséquent de réduire l'amplitude des résonances et des anti-résonances. Différents algorithmes de lissage existent, cependant la méthode RLOESS utilisant une fenêtre de largeur égale à la taille du spectre d'amplitude échantillonné nous est apparue être la meilleure méthode, fournissant des spectres de frontière approchant les HRTF d'une tête sphérique équivalente (voir IV.2 pour plus d'information sur la modélisation des HRTF par une

sphère iso-ITD). La figure III-1 illustre le traitement d'extraction des motifs spectraux liés aux résonances et anti-résonances.

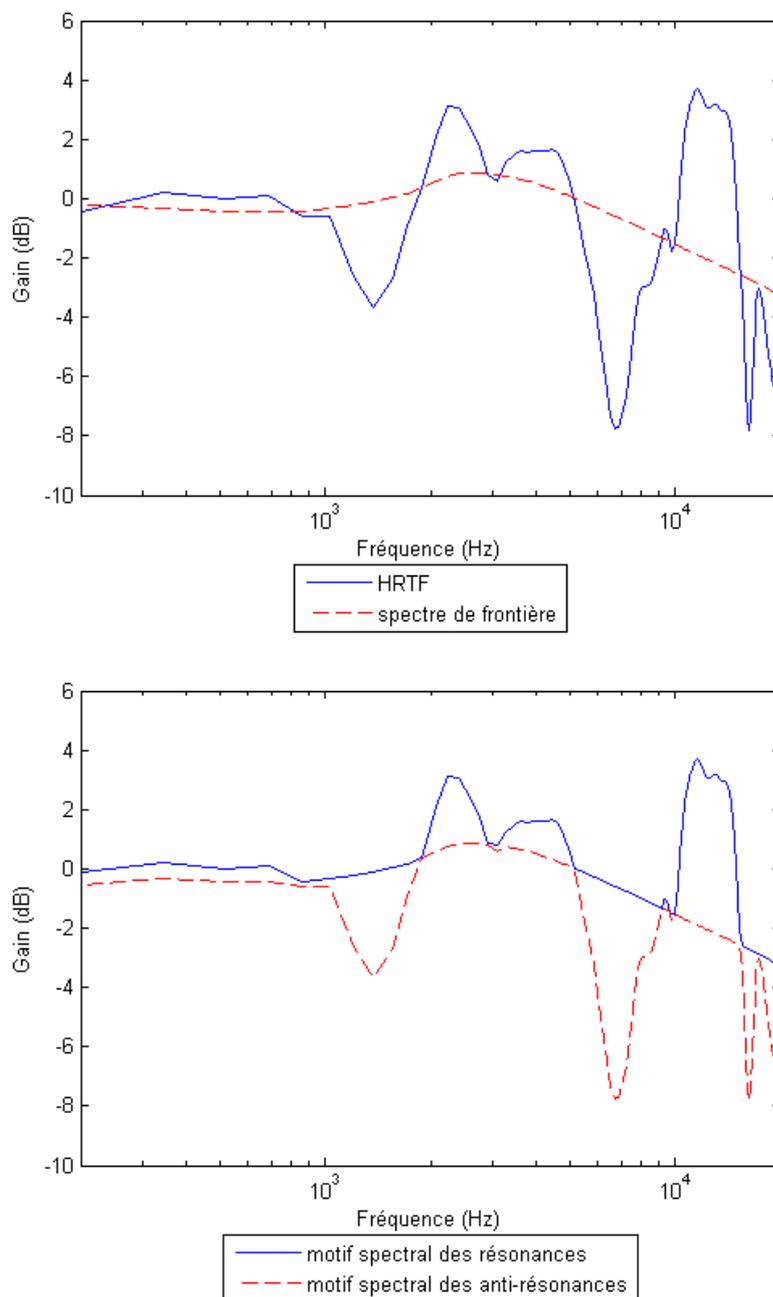


Fig. III-1 : Illustration de l'extraction des motifs spectraux

Huit individus ont participé au test de localisation, mais seuls les résultats de quatre d'entre eux ont pu être considéré comme pertinents (les quatre autres n'étaient pas capable de localiser la source virtualisée à partir des HRTF non modifiées). L'analyse des résultats exploitable semblent indiquer que les résonances seules ne peuvent gouverner la localisation en élévation dans le plan médian, et que les anti-résonances ont une influence prédominante, comme l'indique les courbes d'erreur de localisation

absolue, calculée comme l'écart entre la position reportée par un individu et la position cible pour chaque jeu de filtres, et d'erreur relative, calculée comme l'écart entre la position reportée par un individu lorsque le son est spatialisé à partir des HRTF modifiées (résonances seules ou anti-résonances seules) et la position reportée lorsque le son est spatialisé à partir des HRTF non modifiées (voir Fig. III-2). Cette expérience a conduit à la publication d'un article. Pour de plus amples informations quant à la mise en œuvre et aux résultats de ce test, le lecteur est prié de se reporter à cet article, présenté en annexe (voir Publications, Article A [GK07a]).

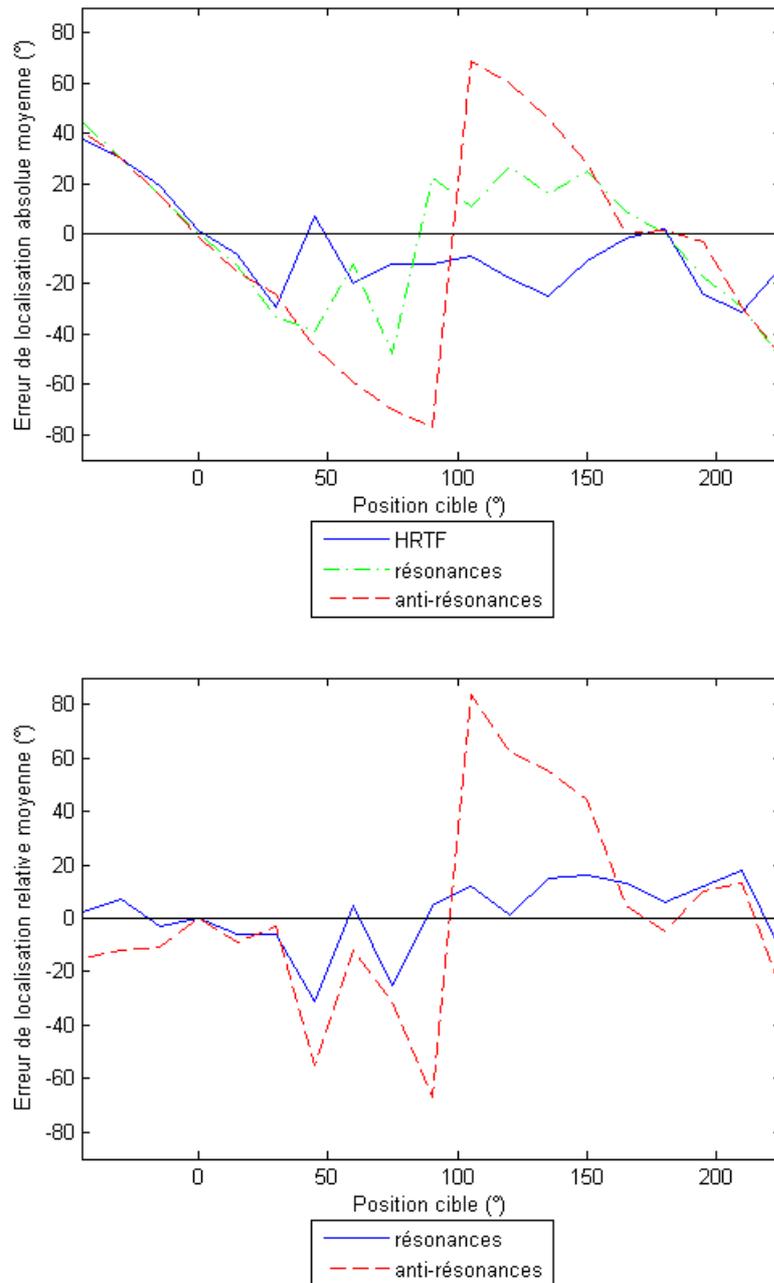


Fig. III-2 : Erreur de localisation moyenne (localisation absolue et relative)

### 2.3.2 Evaluation de la qualité de reconstruction des anti-résonances

Les anti-résonances jouant apparemment un rôle prédominant par rapport aux résonances dans la tâche de localisation en élévation, il pourra être intéressant d'évaluer leur bonne reconstruction dans le plan médian, comme on le fera pour le spectre des HRTF. On pourra donc appliquer les outils de corrélation et de distance sur les spectres d'anti-résonances des HRTF originales et reconstruites.

Pour ce faire, nous appliquerons la méthode décrite au paragraphe précédent, mais au lieu d'exploiter directement les spectres d'anti-résonances ainsi obtenu, nous retrancherons le spectre de frontière pour en quelque sorte obtenir des schémas spectraux normalisés (voir Fig. III-3).

Pour simplifier notre discours, nous désignerons par la suite ces spectres d'anti-résonances normalisés par l'acronyme N-P pour *normalized Notch-Patterns*.

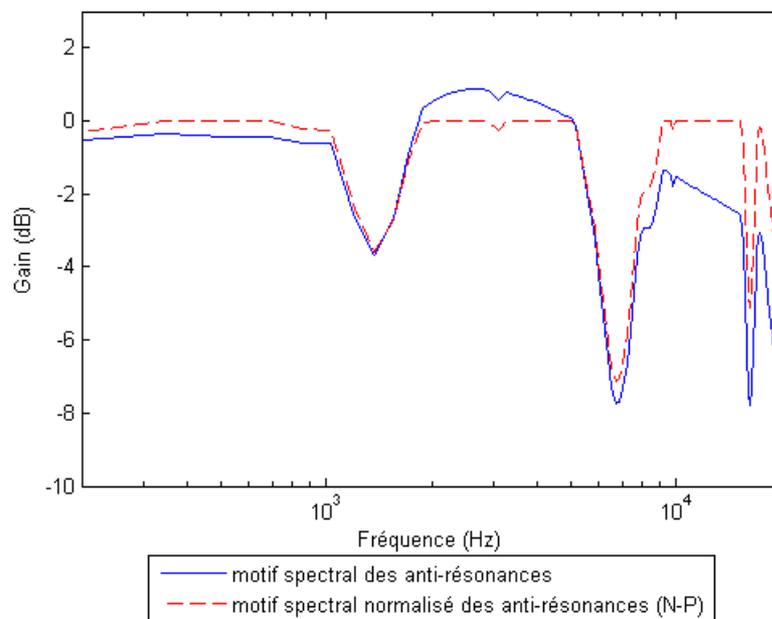


Fig. III-3 : Normalisation des motifs spectraux liés aux anti-résonances

## 3 CALCUL NUMERIQUE DES DIFFERENTES FONCTIONS DE TRANSFERT

L'estimation des différentes fonctions de transfert, nécessaires pour simuler la reconstruction des HRTF et évaluer la qualité de celle-ci, sera effectuée par la méthode des éléments finis de frontière ou BEM (sauf cas particulier où une résolution analytique pourra être entreprise de manière simple). Dans cette section, nous expliquerons la théorie régissant le calcul BEM, en nous appuyant sur le manuel théorique [Fra03] accompagnant le logiciel BEM que nous avons utilisé, ainsi que sur les travaux de thèse de Kahana [Kah00], et nous aborderons les questions pratiques liées à la mise en œuvre des méthodes BEM pour l'estimation de HRTF, par exemple.

### 3.1 Calcul BEM en théorie

La BEM est une méthode de calcul numérique très employée pour la résolution de problèmes de diffraction. Les solutions sont estimées à partir de la discrétisation de la formulation intégrale de l'équation d'onde. Pour le calcul de HRTF et de CAHF, le problème physique équivalent est la diffraction d'une onde par un volume représentant tout ou partie de l'auditeur.

#### 3.1.1 *Enoncé du problème*

Le principe de la BEM est de remplacer un problème différentiel aux limites du domaine d'étude par un problème intégral sur les bords du domaine. Dans un domaine  $\Omega$  contenant un fluide homogène et incompressible, le champ de pression dû à la radiation d'un objet ou à la diffraction par celui-ci est donné par l'équation d'onde :

$$\Delta p(\vec{r}, t) - \frac{1}{c^2} \frac{\partial^2 p(\vec{r}, t)}{\partial t^2} = 0 \quad (\text{III-5})$$

où  $p$  est la pression instantanée en un point  $\vec{r}$  de  $\Omega$  au temps  $t$ ,  $c$  la vitesse du son dans le fluide et  $\Delta$  l'opérateur de Laplace. En exprimant l'équation d'onde dans le domaine fréquentiel, en régime harmonique, on définit l'équation de Helmholtz :

$$\Delta p(\vec{r}, \omega) - k^2 p(\vec{r}, \omega) = 0 \quad (\text{III-6})$$

où  $k = \omega / c$  est le nombre d'onde et  $\omega$  est la pulsation définie à partir de la fréquence par  $\omega = 2\pi f$ .

#### 3.1.2 *Conditions aux limites*

Trois types de conditions aux limites sont associés à l'équation d'Helmholtz : pression invariante, vitesse normale ou impédance acoustique à la surface de l'objet. Ces conditions sont appelées respectivement les conditions aux limites de Dirichlet, Neumann et Robin.

La vitesse particulière normale  $u_0$  à la surface  $\partial\Omega$  d'un objet vibrant est donné par :

$$u_0(\vec{r}, \omega) = \frac{j}{\omega\rho_0} \vec{\nabla} p(\vec{r}, \omega) \cdot \vec{n}_0(\vec{r}) \quad (\text{III-7})$$

où  $\rho_0$  désigne la masse volumique du fluide et  $\vec{n}_0$  la normale à la surface.

Une condition supplémentaire est nécessaire pour la résolution du problème lorsque le domaine  $\Omega$  est considéré comme infini. La pression à l'infini doit satisfaire la condition aux limites de Sommerfield, qui permet de s'assurer que l'énergie s'évanouit à l'infini :

$$\lim_{r \rightarrow \infty} [r \left( \frac{\partial p(\vec{r}, \omega)}{\partial r} + jk p(\vec{r}, \omega) \right)] = 0 \quad (\text{III-8})$$

### 3.1.3 Fonction de Green

La résolution de l'équation d'onde est simplifiée en introduisant la fonction de Green en espace infini, définissant une source ponctuelle monopolaire au point  $\vec{r}_0$  de  $\Omega$  :

$$g(\vec{r} - \vec{r}_0, \omega) = \frac{e^{-jk|\vec{r} - \vec{r}_0|}}{4\pi |\vec{r} - \vec{r}_0|} \quad (\text{III-9})$$

La fonction de Green satisfait l'équation d'Helmholtz inhomogène suivante :

$$\Delta g(\vec{r} - \vec{r}_0, \omega) + k^2 g(\vec{r} - \vec{r}_0, \omega) = -\delta(\vec{r} - \vec{r}_0, \omega) \quad (\text{III-10})$$

où  $\delta$  est la fonction de Dirac. La fonction de Green satisfait également la condition de Sommerfield exprimée alors par :

$$\lim_{r \rightarrow \infty} r \left( \frac{\partial g(\vec{r} - \vec{r}_0, \omega)}{\partial r} + jk g(\vec{r} - \vec{r}_0, \omega) \right) = 0 \quad (\text{III-11})$$

### 3.1.4 Formulation intégrale

Dans le cas d'un objet diffractant, défini par sa surface périphérique  $\partial\Omega$ , et placé dans un domaine infini  $\Omega$ , le champ total  $p_t$  correspond à la superposition des champs incident  $p_i$  et diffracté  $p_s$  :

$$p_t(\vec{r}, \omega) = p_i(\vec{r}, \omega) + p_s(\vec{r}, \omega) \quad (\text{III-12})$$

La formulation intégrale du problème permet d'exprimer la pression acoustique diffractée  $p_s$  en tout point de  $\Omega$  (incluant  $\partial\Omega$ ) à partir de la pression et de la vitesse normale sur la surface fermée  $\partial\Omega$  délimitant le domaine acoustique  $\Omega$  :

$$C(\vec{r}) \cdot p_s(\vec{r}, \omega) = \iint_{\partial\Omega} [p_{s0}(\vec{r}_0, \omega) \vec{\nabla} g(\vec{r} - \vec{r}_0, \omega) \cdot \vec{n}_0(\vec{r}_0) - g(\vec{r} - \vec{r}_0, \omega) \vec{\nabla} p_{s0}(\vec{r}_0, \omega) \cdot \vec{n}_0(\vec{r}_0)] dS_0 \quad (\text{III-13})$$

où le coefficient  $C$  représente l'angle solide exprimé comme une fraction de  $4\pi$ . Dans le cas qui nous intéresse, la normale  $\vec{n}_0$  à la surface  $\partial\Omega$  est dirigée vers l'extérieur. De plus, pour des positions dans  $\Omega$ , le coefficient  $C$  vaut 1, et pour des positions sur la

surface  $\partial\Omega$ , le coefficient  $C$  vaut  $\frac{1}{2}$ . On voit que la solution intégrale s'exprime alors comme la superposition de champs générés par des sources monopolaires et dipolaires réparties à la surface  $\partial\Omega$ .

### 3.1.5 Formulation isoparamétrique

Pour résoudre numériquement l'équation (III-13),  $\partial\Omega$  est généralement discrétisée en un nombre donné  $n_s$  de surfaces élémentaires triangulaires  $\Delta S_0^k$  ( $k = 1 \dots n_s$ ), chacune étant donc définie par trois points, appelés également noeuds. Pour chaque facette  $\Delta S_0^k$ , la pression et son gradient peuvent être approximés à partir des valeurs estimées au lieu de chaque sommet :

$$\begin{cases} p_k(\vec{\sigma}_k, \omega) \approx \sum_{m=1}^3 S_{km}(\vec{\sigma}_k) \cdot p_k(\vec{\sigma}_{km}, \omega) \\ \frac{\partial p_k(\vec{\sigma}_k, \omega)}{\partial n} \approx \sum_{m=1}^3 S_{km}(\vec{\sigma}_k) \cdot \frac{\partial p_k(\vec{\sigma}_{km}, \omega)}{\partial n} \end{cases} \quad (\text{III-14})$$

où  $\vec{\sigma}_k$  désigne les coordonnées locales d'un point sur la  $k^{\text{ème}}$  surface élémentaire,  $\vec{\sigma}_{km}$  les coordonnées locales du sommet indicé  $m$  de cette surface indicée  $k$ , et  $S_{km}$  la fonction de pondération liée au sommet  $m$  de la surface  $k$ . L'équation (III-13) peut donc être reformulée à partir de la pression et de son gradient, exprimés localement sur chacune des surfaces composant  $\partial\Omega$  :

$$C(\vec{r}) \cdot p_t(\vec{r}, \omega) = \sum_{k=1}^{n_s} \left[ \iint_{\Delta S_0^k} \left( \frac{\partial g(\vec{\sigma}_k, \omega)}{\partial n_0} p_k(\vec{\sigma}_k, \omega) - g(\vec{\sigma}_k, \omega) \frac{\partial g(\vec{\sigma}_k, \omega)}{\partial n_0} \right) dS_0 \right] \quad (\text{III-15})$$

En développant l'équation (III-15) à partir des expressions données en (III-14), on peut obtenir une formulation simplifiée du problème. Pour tout point  $\vec{r}$  de  $\Omega$ , on a :

$$p(\vec{r}, \omega) = \sum_{i=1}^{n_p} a_i(\vec{r}, \omega) p_i(\omega) - b_i(\vec{r}, \omega) v_i(\omega) \quad (\text{III-16})$$

où  $p_i$  et  $v_i$  sont respectivement les valeurs de la pression et la vitesse particulière en chaque nœud  $i$  du maillage ( $i = 1 \dots n_p$ , où  $n_p$  est le nombre total de nœuds dans le maillage), et que l'on cherche à estimer, et où  $a_i$  et  $b_i$  sont connus et définis à partir de la fonction de Green et des fonctions de pondérations, traduisant ainsi des contributions monopolaires et dipolaires de chaque nœud du maillage au point  $\vec{r}$ . Nous n'explicitons pas les expressions de  $a_i$  et  $b_i$ .

En exprimant l'équation (III-16) en chaque point du maillage, on obtient la formulation matricielle suivante :

$$P = AP - BV \quad (\text{III-17})$$

où  $P$  et  $V$  sont des vecteurs de taille  $n_s$  composés respectivement des valeurs  $p_i$  et  $v_i$ , et où  $A$  et  $B$  sont des matrices  $n_p \times n_p$  composés respectivement des éléments  $a_{ij}$  et  $b_{ij}$  définis par :

$$\begin{cases} a_{ij}(\omega) = a_i(\vec{r}_j, \omega) \\ b_{ij}(\omega) = b_i(\vec{r}_j, \omega) \end{cases} \quad (\text{III-18})$$

où  $\vec{r}_j$  désigne le lieu du  $j^{\text{ème}}$  nœud du maillage. Le travail principal des codes de calcul BEM consiste à calculer et inverser les matrices  $A$  et  $B$ , pour ensuite déterminer les vecteurs  $P$  et  $V$ , et enfin pouvoir estimer  $p$  en chaque point de  $\Omega$ .

### 3.2 Calcul BEM en pratique

L'élaboration d'un code de calcul BEM n'étant pas l'objet de l'étude, nous avons utilisé un logiciel commercial, VNoise 3.0, développé par la société STS. Toutefois, la tâche de calcul ne peut être réalisée qu'après avoir préalablement discrétiser le modèle numérique de l'objet diffractant considéré en un maillage exploitable, et ceci constitue une étape non négligeable, nécessitant le respect de certaines règles.

Le calcul BEM est un outil puissant, mais la charge de calcul moyenne est souvent conséquente, notamment si l'on s'efforce d'estimer des HRTF sur une large bande de fréquence. Nous allons voir que différentes considérations pratiques permettent d'accélérer et d'alléger les calculs.

#### 3.2.1 Règle de discrétisation générale

La fréquence maximale pouvant être calculée est déterminée par la finesse du maillage (et réciproquement). La pratique recommande une résolution de 4 à 10 nœuds par longueur d'onde [Kat98, Kah00], cette résolution dépendant des fonctions d'approximation du champ de pression à la surface. Une résolution de 6 nœuds par longueur d'onde est une discrétisation couramment employée.

Si l'on note  $n$  la résolution, à savoir le nombre de nœuds par longueur d'onde, et  $d_{max}$  la distance internodale maximale du maillage considéré, la fréquence maximale pouvant être calculée est exprimée par :

$$f_{max} = \frac{c}{d_{max} \cdot n} \quad (\text{III-19})$$

#### 3.2.2 Modèle symétrique ou modèle bafflé par un plan infini

Si le modèle possède une symétrie planaire (au niveau de la géométrie et des matériaux) ou est bafflé par un plan infini, il est possible de ne considérer que la moitié du maillage en répercutant la symétrie au niveau de l'écriture des fonctions de Green et donc des matrices  $A$  et  $B$ , décrite par les expressions (III-17) et (III-18). Le nombre de surfaces et le nombre de nœuds considérés sont alors approximativement deux fois plus petits.

Si le logiciel de calcul utilisé permet d'utiliser ce principe, en offrant la possibilité de définir des plans infinis d'admittance uniforme quelconque (c'est le cas de VNoise 3.0), la procédure pour résoudre un problème de diffraction sur un maillage symétrique, en n'utilisant donc qu'une des moitiés du maillage, est composée de trois étapes.

- Le problème est d'abord résolu en définissant un plan de symétrie d'admittance nulle. Chaque source possède une image par rapport au plan, de même amplitude et de même phase.
- Le problème est ensuite conduit en définissant un plan d'antisymétrie, c'est-à-dire d'admittance infinie. Dans ce cas, une source image possède une amplitude identique mais une phase opposée.
- La solution du problème complet est obtenue en moyennant les solutions des problèmes symétrique et antisymétrique.

Par exemple, si l'on note  $p'$  le champ de pression calculé à partir du problème symétrique et  $p''$  le champ de pression calculé à partir du problème antisymétrique, le champ de pression  $p$  du modèle complet s'exprime :

$$p = \frac{p' + p''}{2} \quad (\text{III-20})$$

Bien que cette procédure nécessite de conduire deux calculs sur un demi maillage au lieu d'un calcul sur le maillage entier, elle permet environ de réduire la charge mémoire d'un facteur 4 (lors de chaque phase d'inversion) et de diminuer le temps de calcul global de moitié, car charge et temps de calcul sont approximativement proportionnels au carré du nombre de nœuds du maillage (soit le nombre d'éléments dans chaque matrice  $A$  et  $B$ ).

### 3.2.3 Principe de réciprocité

Que ce soit par la mesure ou le calcul, une base de données HRTF est généralement construite en évaluant ces dernières successivement pour chaque position de source considérées. Une méthode alternative plus rapide est suggérée par l'utilisation du principe de réciprocité [KFCS82] : la pression acoustique  $p_1$ , générée en un point  $P_1$  d'un fluide par une source placée en un point  $P_2$  et de vitesse particulière  $v$ , est égale à la pression acoustique  $p_2$ , générée au point  $P_2$  par cette même source placée au point  $P_1$  et de même vitesse particulière  $v$ . Il devient donc possible d'évaluer simultanément les HRTF pour toutes les positions de sources considérées, en positionnant des points de mesure (microphones omnidirectionnels virtuels) à la place des sources et en positionnant les sources (monopolaires) à l'entrée des conduits auditifs.

Les travaux de Kahana [Kah00] lui ont permis de dégager une règle importante nécessaire à la convergence du calcul réciproque : la distance internodale des éléments de la surface à proximité de la source doit être de l'ordre de grandeur de la distance entre la source et la surface.

## 4 SIMULATION DE RECONSTRUCTION : CAS DE PAVILLONS BAFFLES

Cette partie a pour but d'examiner la qualité de reconstruction des HRTF dans le cas simple de pavillons bafflés. Dans ce cas précis, il est courant d'utiliser l'acronyme PRTF pour *Pinna-Related Transfer functions* (littéralement les filtres liés au pavillon), mais pour éviter toute confusion avec les PRTF que nous avons définies, nous garderons l'appellation HRTF.

### 4.1 Modélisation numérique du problème

#### 4.1.1 Modèles de pavillons utilisés

La modélisation des pavillons présentés à la figure III-4 a été effectuée par Kahana au cours de sa thèse [Kah00]. Il s'agit de six modèles d'oreille droite différents : cinq correspondent à des pavillons utilisés sur des mannequins acoustiques (Cortex, DB60, DB65, DB90, DB95), un correspond au pavillon d'oreille numérisé de Kahana (Yuvi).

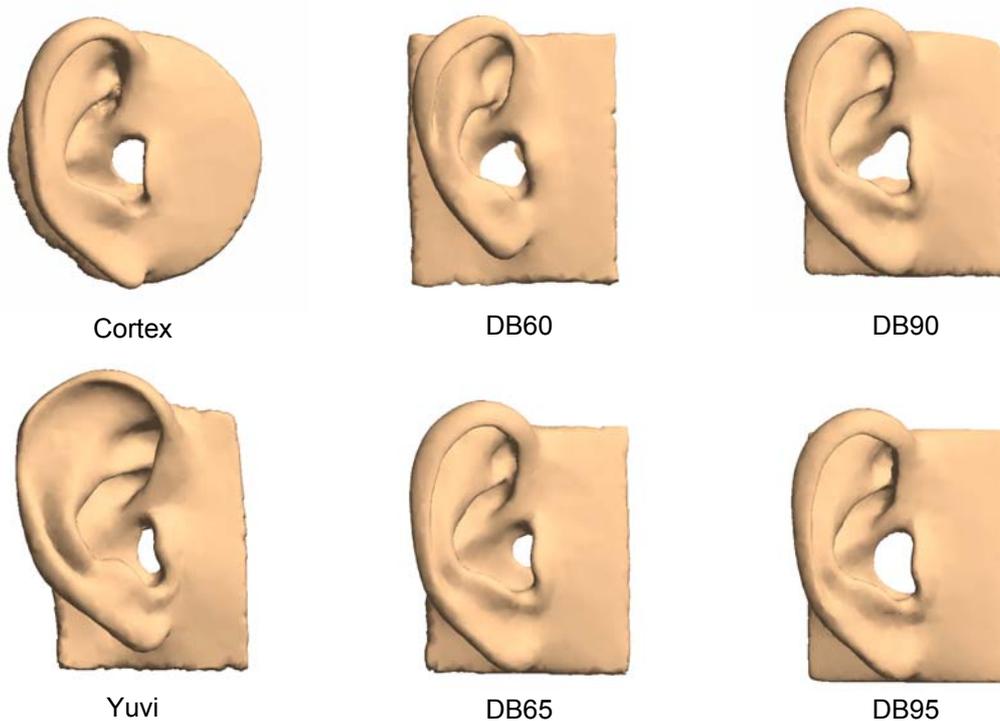


Fig. III-4 : Modèles de pavillons d'oreille utilisés

#### 4.1.2 Réseau de sources secondaires considéré

Le réseau possède une géométrie hémisphérique de rayon 5cm. Il est composé de 640 sources secondaires. La condition de Shannon est vérifiée jusqu'à une fréquence de 36kHz. Le réseau est positionné de manière à ce que son centre coïncide avec le centre du repère et avec un point de la zone d'entrée des canaux auditifs de chaque pavillon, et que sa base soit située dans le plan infini bafflant les pavillons (voir Fig. III-5).

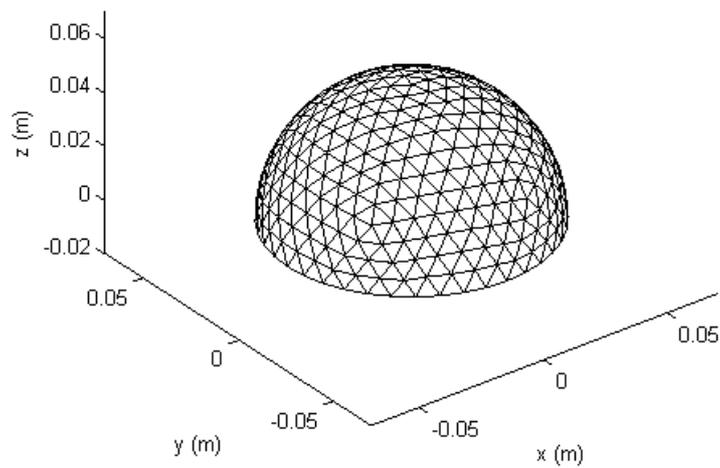


Fig. III-5 : Réseau hémisphérique de 5cm de rayon à 640 sources secondaires

#### 4.1.3 Positions de sources primaires considérées

Les différentes positions de sources considérées sont données par la figure III-6 et le tableau III-1. Il sera utile de se référer à l'index de positionnement pour lire les différents graphiques présentés par la suite.

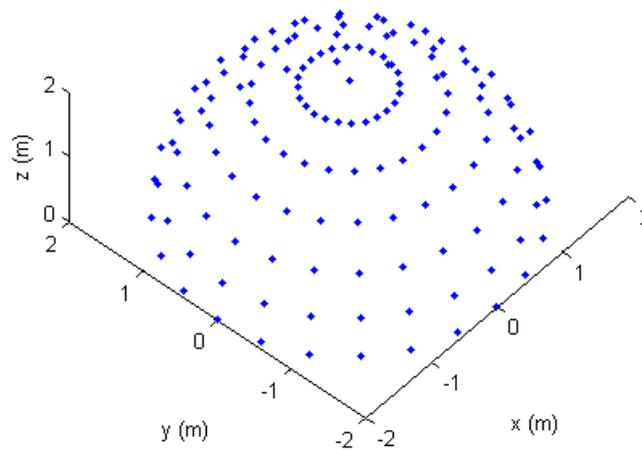


Fig. III-6 : Positions de la source primaire

<i>i</i>	$\theta$	$\varphi$																		
1	0	0	25	15	0	49	30	0	73	45	0	97	60	0	121	75	0	145	90	0
2	0	15	26	15	15	50	30	15	74	45	15	98	60	15	122	75	15			
3	0	30	27	15	30	51	30	30	75	45	30	99	60	30	123	75	30			
4	0	45	28	15	45	52	30	45	76	45	45	100	60	45	124	75	45			
5	0	60	29	15	60	53	30	60	77	45	60	101	60	60	125	75	60			
6	0	75	30	15	75	54	30	75	78	45	75	102	60	75	126	75	75			
7	0	90	31	15	90	55	30	90	79	45	90	103	60	90	127	75	90			
8	0	105	32	15	105	56	30	105	80	45	105	104	60	105	128	75	105			
9	0	120	33	15	120	57	30	120	81	45	120	105	60	120	129	75	120			
10	0	135	34	15	135	58	30	135	82	45	135	106	60	135	130	75	135			
11	0	150	35	15	150	59	30	150	83	45	150	107	60	150	131	75	150			
12	0	165	36	15	165	60	30	165	84	45	165	108	60	165	132	75	165			
13	0	180	37	15	180	61	30	180	85	45	180	109	60	180	133	75	180			
14	0	195	38	15	195	62	30	195	86	45	195	110	60	195	134	75	195			
15	0	210	39	15	210	63	30	210	87	45	210	111	60	210	135	75	210			
16	0	225	40	15	225	64	30	225	88	45	225	112	60	225	136	75	225			
17	0	240	41	15	240	65	30	240	89	45	240	113	60	240	137	75	240			
18	0	255	42	15	255	66	30	255	90	45	255	114	60	255	138	75	255			
19	0	270	43	15	270	67	30	270	91	45	270	115	60	270	139	75	270			
20	0	285	44	15	285	68	30	285	92	45	285	116	60	285	140	75	285			
21	0	300	45	15	300	69	30	300	93	45	300	117	60	300	141	75	300			
22	0	315	46	15	315	70	30	315	94	45	315	118	60	315	142	75	315			
23	0	330	47	15	330	71	30	330	95	45	330	119	60	330	143	75	330			
24	0	345	48	15	345	72	30	345	96	45	345	120	60	345	144	75	345			

Tab. III-1 : Index des positions de la source primaire – coordonnées polaire interaurales

#### 4.1.4 Estimation des HRTF de référence et des PRTF

Le calcul des HRTF<sup>0</sup> et des PRTF a été réalisé par des méthodes de calcul BEM. Nous avons utilisé le logiciel BEM VNoise 3.0 en cluster sur trois ordinateurs possédant chacun un processeur double-cœur à 3GHz et 4Go de mémoire vive. Ce logiciel a été utilisé entre autre par Pernaux [Per03] qui a démontré sa robustesse en résolvant le problème de diffraction par une sphère rigide. Nous avons effectué le même type de vérification.

Les maillages des modèles de pavillon ont été discrétisés de manière à vérifier une distance inter-nodale inférieure au 6<sup>ème</sup> de la longueur d'onde pour une fréquence maximale de 16kHz. Grâce à l'utilisation du principe de réciprocité, l'estimation des HRTF<sup>0</sup> et des PRTF pour une fréquence donnée a pu être réalisé en un seul calcul. L'évaluation des fonctions de transferts a été effectuée pour 85 fréquences, de 125Hz à 16kHz par 12<sup>ème</sup> d'octave. La durée d'un calcul complet pour un pavillon n'a jamais excédé quatre heures.

#### 4.1.5 Estimation des CAHF

L'estimation des filtres holophoniques a pu être effectuée de manière analytique. En effet, le problème est ici simple, car le plan infini introduit une symétrie que l'on peut

simuler en considérant une deuxième source primaire, symétrique de la première par rapport à ce même plan.

## 4.2 Analyse des résultats

### 4.2.1 Estimation de l'atténuation induite par le réseau

Comme nous l'avons remarqué en II.3.4, la géométrie du réseau, essentiellement, et la méthode de captation dans une moindre mesure, engendre un effet passe-haut que l'on va retrouver au niveau des HRTF reconstruite. Pour estimer l'atténuation et trouver une bande de fréquence épargnée par ce phénomène, nous avons dans un premier temps calculé une erreur spectrale moyennée sur toutes les positions calculées et tous les pavillons, définie comme la différence des spectres d'amplitude mesurés en décibel des HRTF reconstruites et originales.

La figure III-7 présente cette erreur évaluée pour chaque méthode de captation. On voit que le phénomène d'atténuation en basse fréquence n'est plus effectif après une fréquence voisine de 3kHz. Comme dans la littérature il est courant de considérer que les pavillons ont une influence perceptive prédominante pour des fréquences comprises entre 4kHz et 16kHz (voir I.2.3.7), on pourra se contenter d'analyser les résultats des simulations à partir d'une fréquence de 3 ou 4kHz.

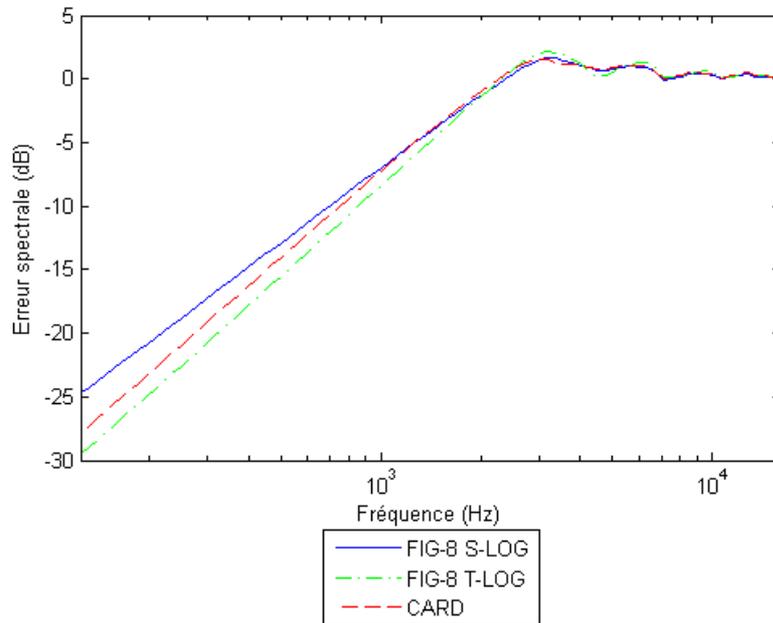


Fig. III-7 : Estimation de l'atténuation en basse fréquence induite par le réseau hémisphérique

### 4.2.2 Sans aliasing spatial

Les résultats présentés dans cette section concernent un réseau hémisphérique à 640 sources secondaires. La condition de Shannon est vérifiée jusqu'à une fréquence de 36kHz, soit plus du double de la fréquence maximale calculée. Nous choisissons

comme zone fréquentielle d'analyse des résultats une bande comprise entre 3,4kHz et 16kHz, exempte de tout phénomène d'atténuation.

Dans un premier temps, il est judicieux d'évaluer visuellement les résultats des différentes méthodes holophoniques employées. Les HRTF de référence et les HRTF reconstruites sont présentées sous la forme d'un graphique à deux dimensions : l'axe des abscisses correspond à l'axe des fréquences réparties suivant une échelle logarithmique, l'axe des ordonnées indique la position de la source primaire (le tableau III-1 permet d'obtenir la position de la source par rapport à son index), l'échelle des couleurs correspond au niveau en décibel. Ce type de visualisation est couramment appelé figure de diffraction. Les résultats étant similaires d'un pavillon sur l'autre, nous utiliserons uniquement les figures de diffraction liées au pavillon Cortex pour illustrer nos commentaires. Les figures liées aux autres pavillons sont présentées dans le complément de figures annexé à ce chapitre.

La figure III-8 montre une très bonne correspondance entre les HRTF de référence et les HRTF reconstruites obtenues par le biais des méthodes CARD et FIG-8 S-LOG. Les résonances et anti-résonances sont correctement reconstituées. La méthode FIG-8 T-LOG, quant à elle, fournit de moins bons résultats : en effet, on remarque que plus l'angle latéral augmente, plus la qualité de reconstruction se dégrade. De part la géométrie du problème, on peut considérer que le réseau est éclairé par deux sources en champ libre. Or, nous avons supposé dans le chapitre précédent que, dans le cas d'une scène sonore complexe (au-delà d'une source), plus les contributions de chaque source tendraient à s'opposer, moins la méthode FIG-8 T-LOG serait fiable. L'observation des HRTF reconstruites conforte ces suppositions.

La figure III-9 représente les « *notch patterns* » (N-P) dans le plan médian. On remarque bien que les méthodes FIG-8 S-LOG et T-LOG ont un comportement strictement identique dans le plan médian, de part la géométrie du problème. Les N-P sont globalement bien reconstruits. Dans le cas du pavillon Cortex, la méthode CARD semble offrir une meilleure reconstruction des N-P. Cependant, ceci ne se vérifie pas forcément pour les autres pavillons (voir le complément de figures en fin de chapitre).

Les indices de corrélation ou de distance entre les spectres des HRTF de référence et les HRTF reconstruites montrent sans équivoque la dégradation de la qualité de reconstruction des HRTF via l'utilisation de la méthode FIG-8 T-LOG liée à la latéralisation de la source primaire. La figure III-10 présente le coefficient de corrélation ainsi que la distance en fonction de l'index de la source primaire, tous pavillons confondus. On voit bien que, pour la captation FIG-8 T-LOG, plus l'angle latéral augmente, plus les indices sont mauvais. Ce que montre également cette figure, c'est une homogénéité des résultats par rapport aux différents modèles de pavillon utilisés, indiquée par une équivalence des indices de corrélation et de distance.

Les résultats montrent donc que les techniques holophoniques permettent de simuler l'onde sonore primaire arrivant sur un pavillon d'oreille dans le cas idéal où le réseau vérifie la condition de Shannon sur toute la bande considérée et moyennant l'utilisation d'une méthode de captation pertinente, à savoir la méthode FIG-8 S-LOG ou la méthode CARD (coefficient de corrélation supérieur à 0,75 et distance inférieure à 3dB).

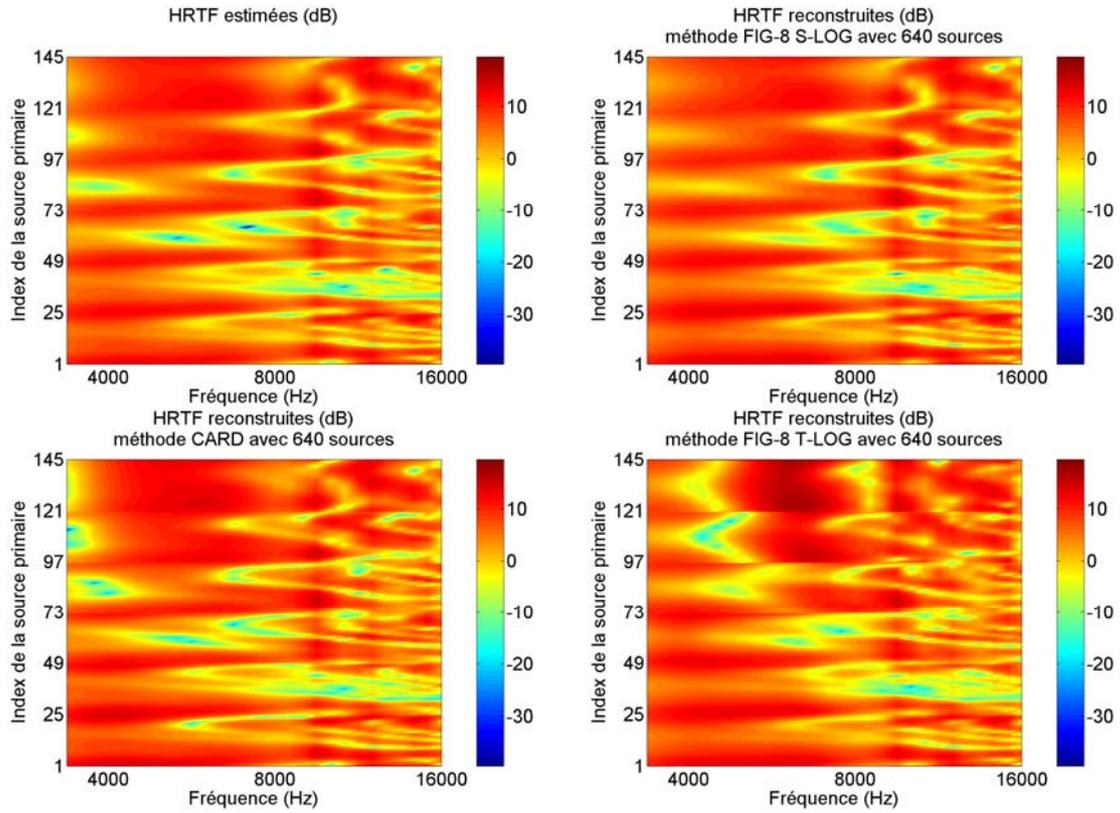


Fig. III-8 : Reconstruction des HRTF via 640 sources secondaires - pavillon Cortex

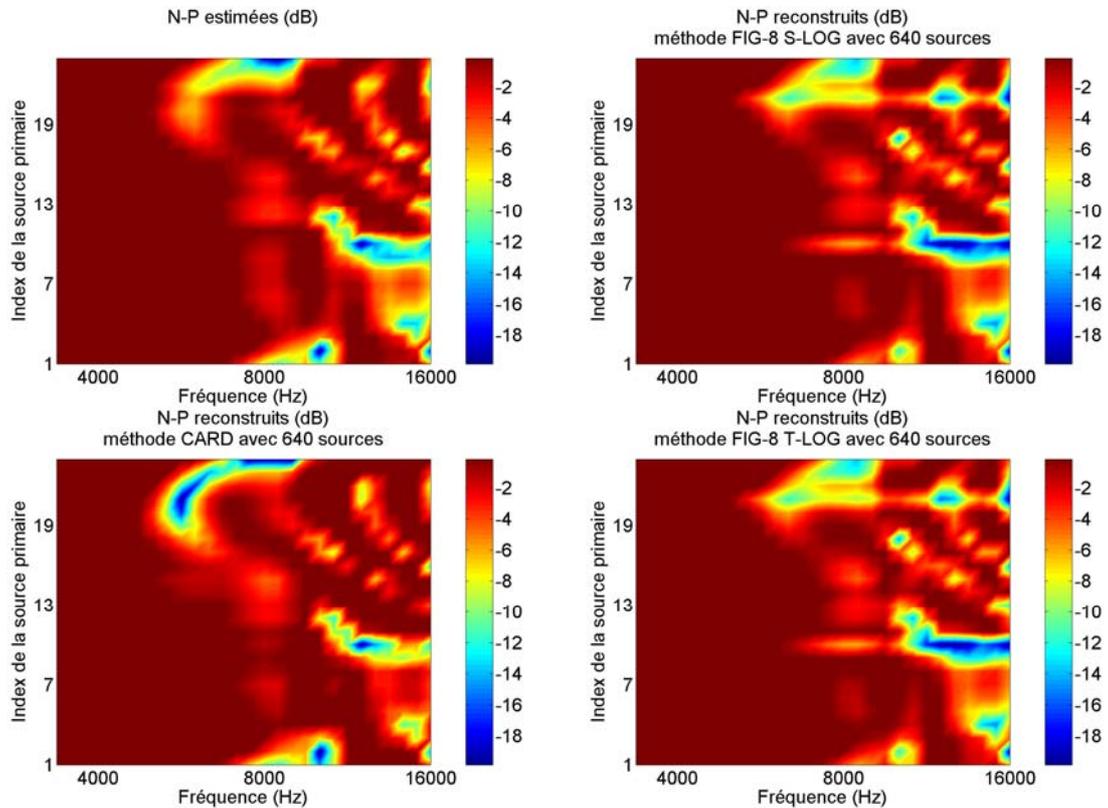


Fig. III-9 : Reconstruction des N-P via 640 sources secondaires - pavillon Cortex

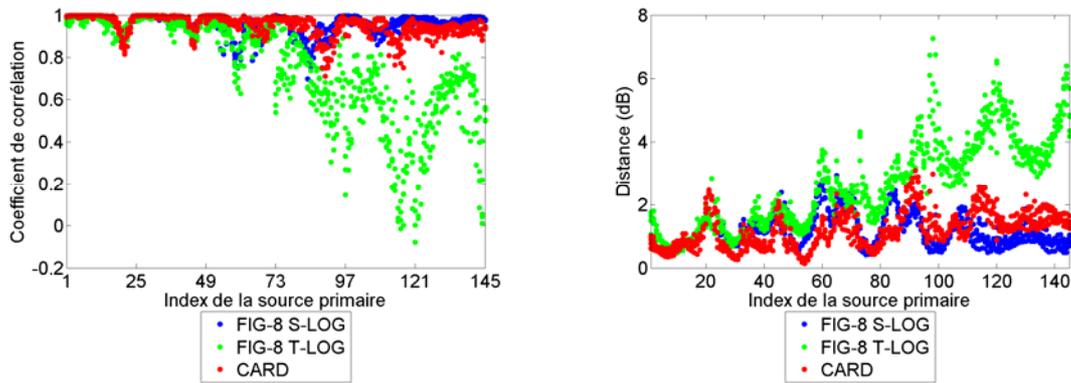


Fig. III-10 : Corrélation et distance entre HRTF en fonction de l'incidence – tous pavillons confondus

Le nombre de sources secondaires utilisées lors de ces premières simulations est extravagant si on garde à l'esprit que l'objectif final est le prototypage d'un système réel. En effet, si l'on considère les puissances de calcul disponibles sur les ordinateurs du commerce et que l'on reste réaliste sur le nombre de canaux à traiter, il est raisonnable de se limiter à un réseau composé de 8 sources secondaires (soit un système de restitution à 16 canaux). C'est pourquoi, la prochaine section sera consacrée à l'étude de la dégradation de la qualité de reconstruction des HRTF liée à la réduction du nombre de sources secondaires.

#### 4.2.3 Réduction de la taille du réseau

Différentes tailles de réseaux sont ici considérées, autrement dit différentes fréquences d'aliasing spatial. Les sources secondaires sont sélectionnées parmi les 640 sources du réseau initial, en s'appuyant sur des méthodes de création de surfaces hémisphériques à partir de tétraèdres ou d'icosaèdres, de telle sorte à obtenir la répartition la plus uniforme possible. On notera que la fréquence d'aliasing est proportionnelle à la racine carrée du nombre de sources secondaires (voir Fig. III-11).

La figure III-12 témoigne de la dégradation de la qualité de reconstruction des HRTF du pavillon Cortex en fonction du nombre de sources secondaires utilisées dans le cas d'une utilisation de la méthode FIG-8 S-LOG. On constate cela pour chacun des six pavillons considérés et chacune des trois techniques de captation. Plus le réseau est petit, plus l'aliasing spatial est effectif. On s'aperçoit d'ores et déjà qu'une application brute des techniques holophoniques risque de ne plus permettre l'activation des indices de localisation pavillonnaire d'un individu. Les N-P suivent bien entendu la même tendance (voir Fig. III-13).

Les indices moyens de corrélation et de distance entre HRTF mettent en évidence ce problème (voir Fig. III-14). Ils montrent également que, pour un nombre très réduit de sources secondaires, la méthode FIG-8 T-LOG devient meilleure que la méthode CARD (ceci se vérifie pour tous les pavillons). Pour ces tailles de réseau, les indices étant suffisamment mauvais, leur comparaison peut ne pas être pertinente. Cependant, ils nous invitent à ne pas écarter la méthode FIG-8 T-LOG pour les études suivantes.

Les indices de corrélations et de distance liés aux N-P suivant globalement la même tendance que ceux liés aux HRTF (voir Fig. III-15 et III-16), il apparaît donc redondant

d'évaluer à la fois la qualité de reconstruction des HRTF et des N-P. C'est pourquoi nous nous focaliserons à présent uniquement sur l'analyse des HRTF dans leur globalité.

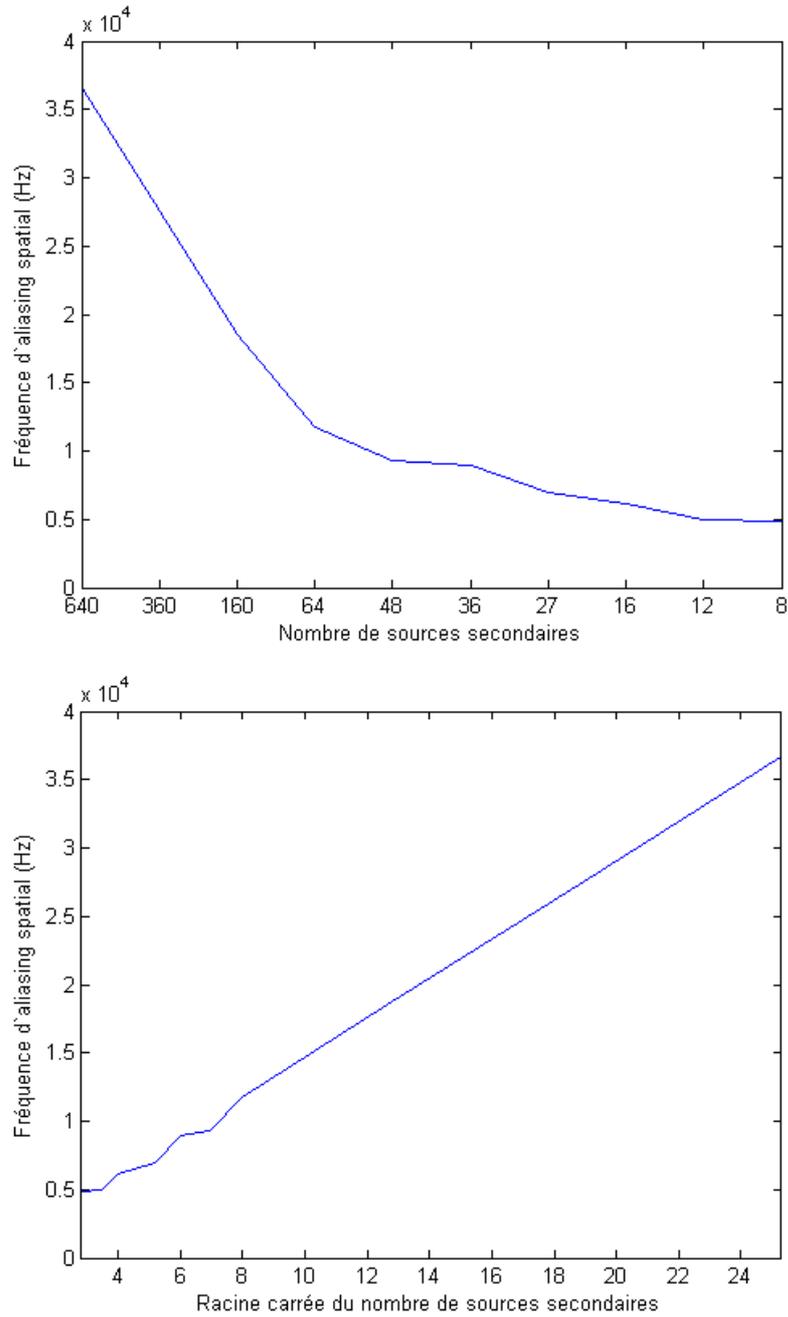


Fig. III-11 : Fréquence d'aliasing spatial en fonction de la taille du réseau

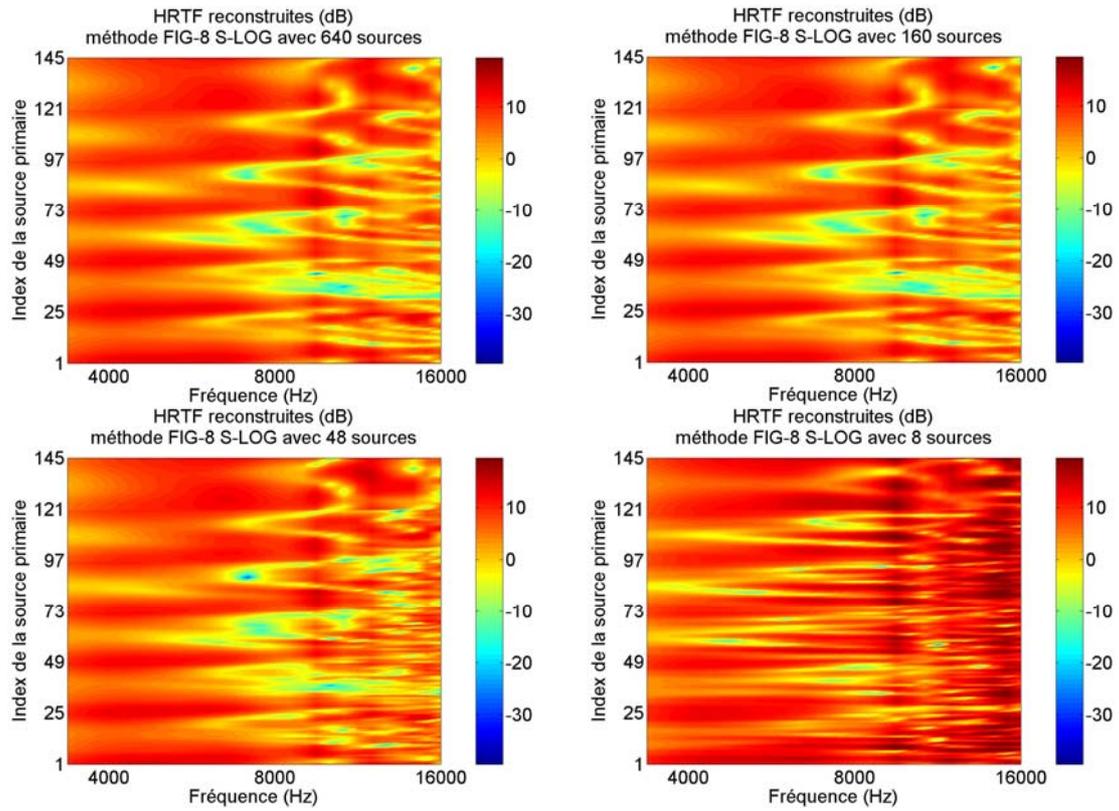


Fig. III-12 : HRTF reconstruites en fonction de la taille du réseau - pavillon Cortex

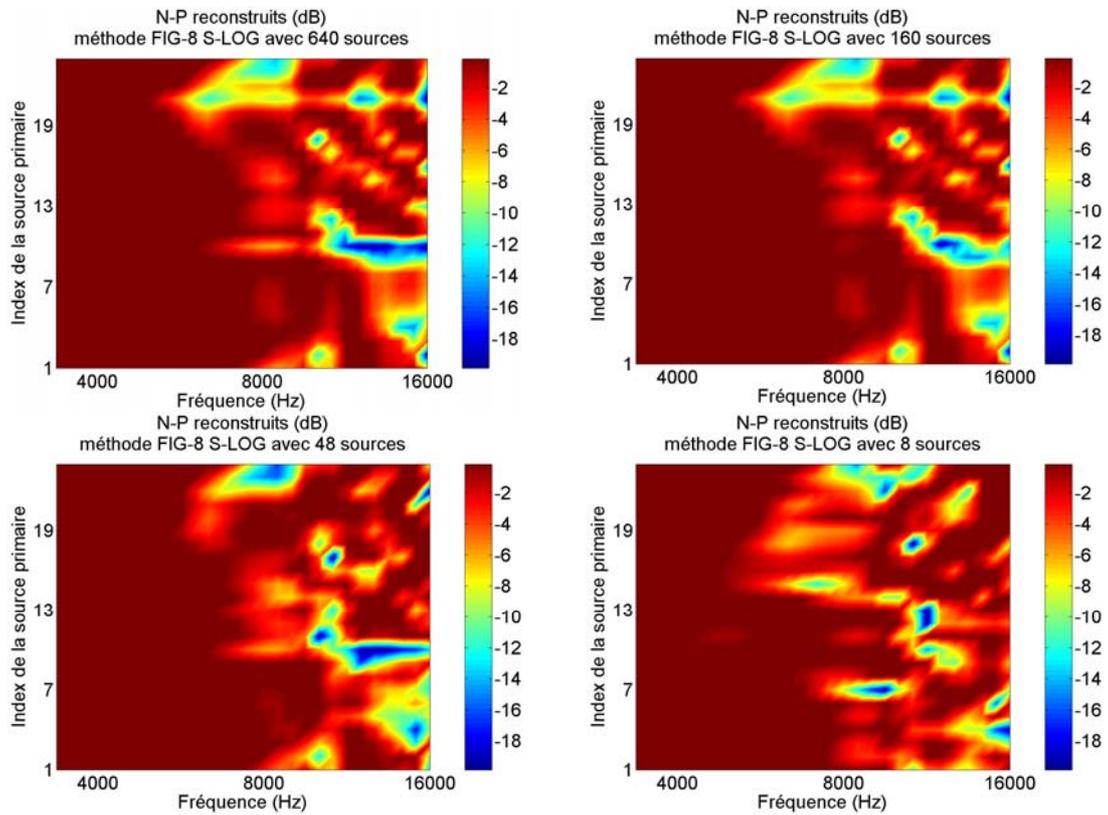


Fig. III-13 : N-P reconstruits en fonction de la taille du réseau - pavillon Cortex

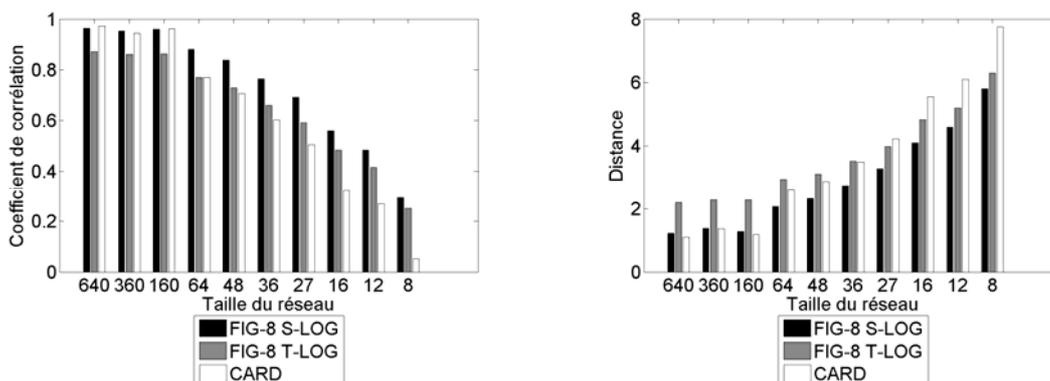


Fig. III-14 : Corrélation et distance entre HRTF en fonction de la taille de réseau

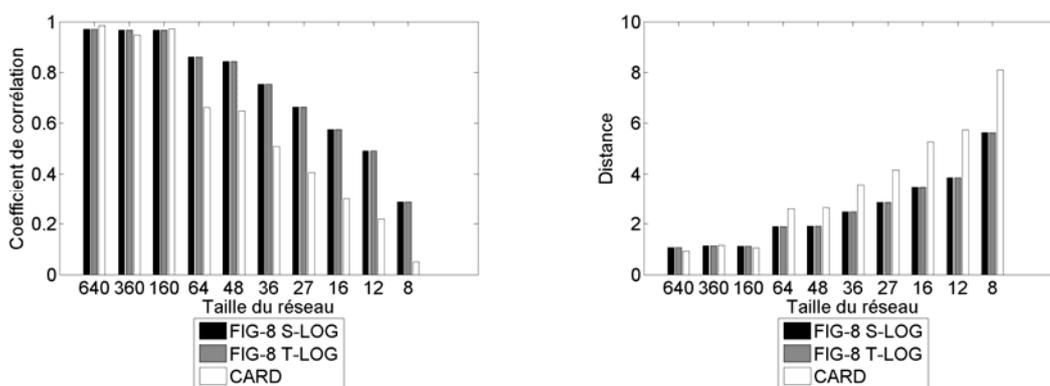


Fig. III-15 : Corrélation et distance entre HRTF en fonction de la taille de réseau - plan médian

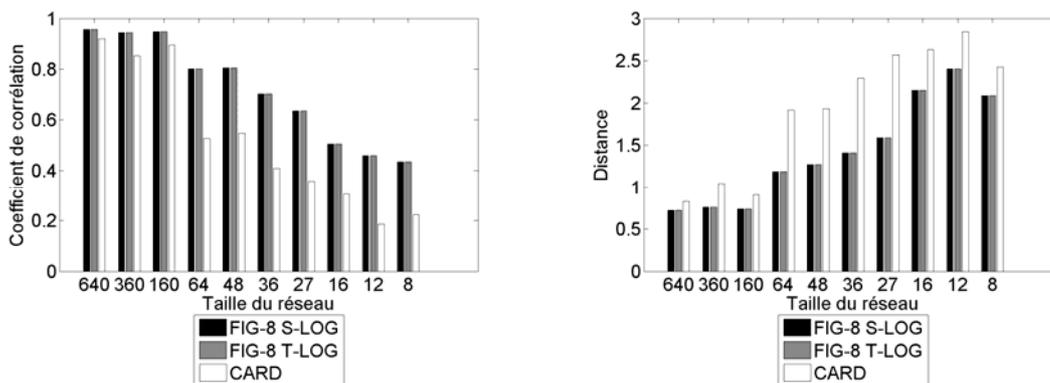


Fig. III-16 : Corrélation et distance entre N-P en fonction de la taille de réseau - plan médian

A la vue de ces résultats, il apparaît nécessaire de trouver un moyen pour améliorer la qualité de reconstruction des HRTF dans le cas d'un réseau composé d'un petit nombre de sources secondaires. Avant d'aborder les différentes pistes d'optimisation que nous avons étudiées et qui seront présentées dans le prochain chapitre, nous allons effectuer les mêmes analyses sur un problème plus complexe, le cas d'un modèle de tête. Ce problème permet entre autres d'évaluer la reconstruction de HRTF contralatérales,

c'est-à-dire pour des positions de source situées de l'autre côté de la tête par rapport au pavillon considéré.

## 5 SIMULATION DE RECONSTRUCTION : CAS D'UN MODELE DE TETE

### 5.1 Modélisation numérique du problème

#### 5.1.1 Modèle de tête utilisé

Le modèle numérique utilisé correspond à la tête acoustique Neumann KU-100 (voir Fig. III-17). Nous avons effectué la modélisation en deux étapes : une étape d'acquisition et une étape de traitement. La tête a tout d'abord été scannée en laboratoire médical par une méthode MRI (*Magnetic Resonance Imagery* ou imagerie par résonance magnétique), ceci produisant une multitude d'images numériques, chaque image correspondant à une tranche volumique d'épaisseur 0,8mm, et la résolution d'un pixel étant de 0,63mm. Toutes les parties métalliques de la tête ont dû être retirées pour obtenir des images exploitables (voir Fig. III-18), perdant ainsi l'emplacement exact de la membrane du microphone. Ensuite, l'ensemble des images ont été traitées manuellement de manière à assurer une extraction de contours fermés. Puis ces contours ont été utilisés pour obtenir le maillage de la surface externe de la tête artificielle.

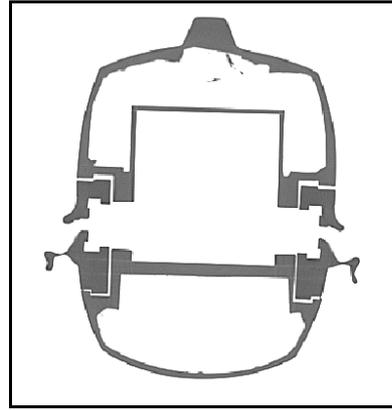
Une étude a été réalisée pour valider l'utilisation de techniques par éléments finis de frontière pour le calcul de HRTF, en comparant les résultats issus de simulations BEM à partir de cette modélisation et des mesures effectuées sur la vraie tête acoustique KU-100 [GK07b] (voir Publications, Article B). Cependant, même si dans le cas précis des simulations conduites dans ce chapitre, la validité du calcul par rapport aux mesures ne nous intéresse pas, celle-ci nous conforte sur la robustesse des résultats que nous allons exploiter. En effet, contrairement aux simulations faites sur les pavillons bafflés, la règle des 6 nœuds par longueur d'onde n'a pu être vérifiée que jusqu'à 10kHz, ceci pour garder une taille de maillage raisonnable et pouvoir effectuer les simulations dans des délais acceptables.



Fig. III-17 : Tête acoustique Neumann KU-100 et modèle numérique dérivé



Composantes métalliques présentes



Composantes métalliques absentes

Fig. III-18 : Illustration du problème de diffraction induit par des objets métalliques

### 5.1.2 Estimation des HRTF de référence et des PRTF

Le calcul des  $HRTF^o$  et des  $PRTF$  a été réalisé par des méthodes de calcul BEM. Le maillage du modèle a été construit de manière à vérifier une distance internodale inférieure au 6<sup>ème</sup> de la longueur d'onde pour une fréquence maximale de 10kHz. L'évaluation des fonctions de transferts a été effectuée pour 43 fréquences, de 125Hz à 16kHz par 6<sup>ème</sup> d'octave.

Comme pour les simulations conduites sur les pavillons, le principe de réciprocité a été employé. Nous avons de plus considéré le modèle comme symétrique par rapport au plan médian en n'utilisant que la partie droite du modèle. Grâce à la procédure décrite en III.3.2.2, le temps de calcul a été divisé par deux, soit une durée totale d'environ une journée.

### 5.1.3 Estimation des CAHF

La complexité de la géométrie du modèle ne permettait pas une résolution analytique du problème. Nous avons donc utilisé des techniques BEM pour estimer les CAHF. Pour ce calcul, l'utilisation du principe de réciprocité n'était pas intéressante, car il aurait fallu effectuer la résolution en activant une source secondaire du réseau à la fois, leur nombre étant plus important que le nombre de positions de la source primaire initialement envisagé. Toutefois, pour gagner quelques semaines de temps de calcul, le nombre de positions de source ainsi que le nombre de points fréquentiels ont été réduits par rapport aux simulations conduites sur les pavillons bafflés, passant de 145 à 89 positions (comme indiqué par le tableau III-2) et de 85 à 43 fréquences. L'estimation complète des CAHF a duré approximativement quatre semaines.

Le problème étant complexe à première vue, la méthode de captation holophonique FIG-8 S-LOG n'a pas été mise en œuvre. Les résultats présentés ci-dessous ne concerneront donc que les méthodes CARD et FIG-8 T-LOG. Toutefois, si nous avions eu accès aux matrices  $A$  et  $B$  (décrites en III.3.1.5), nous aurions pu envisager d'évaluer la méthode FIG-8 S-LOG, car le problème de diffraction aurait pu être restreint à un problème de radiation d'une distribution équivalente de sources monopolaires et dipolaires, de position et de nature connues.

<i>i</i>	$\theta$	$\varphi$																																	
1	0	0	25	15	0	49	30	0	61	45	0	73	60	0	81	75	0	89	90	0															
2	0	15	26	15	15																62	45	30	74	60	45	82	75	45						
3	0	30	27	15	30																51	30	60	63	45	60	75	60	45	83	75	90			
4	0	45	28	15	45																52	30	90	64	45	90	76	60	135	84	75	135			
5	0	60	29	15	60																53	30	120	65	45	120	77	60	180	85	75	180			
6	0	75	30	15	75																54	30	150	66	45	150	78	60	225	86	75	225			
7	0	90	31	15	90																55	30	180	67	45	180	79	60	270	87	75	270			
8	0	105	32	15	105																56	30	210	68	45	210	80	60	315	88	75	315			
9	0	120	33	15	120																57	30	240	69	45	240									
10	0	135	34	15	135																58	30	270	70	45	270									
11	0	150	35	15	150																59	30	300	71	45	300									
12	0	165	36	15	165																60	30	330	72	45	330									
13	0	180	37	15	180																														
14	0	195	38	15	195																														
15	0	210	39	15	210																														
16	0	225	40	15	225																														
17	0	240	41	15	240																														
18	0	255	42	15	255																														
19	0	270	43	15	270																														
20	0	285	44	15	285																														
21	0	300	45	15	300																														
22	0	315	46	15	315																														
23	0	330	47	15	330																														
24	0	345	48	15	345																														

Tab. III-2 : Index des positions de la source primaire – coordonnées polaire interaurales

## 5.2 Analyse des résultats

### 5.2.1 Sans aliasing spatial

En ce qui concerne les HRTF ipsilatérales, la méthode CARD fournit de meilleurs résultats que la méthode FIG-8 T-LOG. Cependant, la dépendance de cette dernière par rapport à l'incidence de la source primaire est moins marquée que pour les pavillons bafflés (voir Fig. III-19 et III-21). En effet, le champ direct, assimilable à la contribution seule de la source primaire en champ libre, semble prédominant par rapport au champ diffracté par la tête, même pour des angles latéraux élevés. De ce fait, une captation FIG-8 T-LOG constitue une alternative pertinente à une captation CARD.

La reconstruction des HRTF contralatérales est également de meilleure qualité si l'on utilise la méthode CARD. Toutefois, les deux méthodes sont sujettes à une dépendance par rapport à l'angle latéral. Néanmoins, cette dépendance est moins marquée pour la méthode CARD. Plus encore, pour une incidence latérale maximale, une captation FIG-8 S-LOG ne permet quasiment plus de reconstruire les HRTF contralatérales (voir Fig. III-20) : les signaux captés sont trop faibles et la distance spectrale atteint des valeurs supérieures à 15dB, comme l'indique la figure III-22.

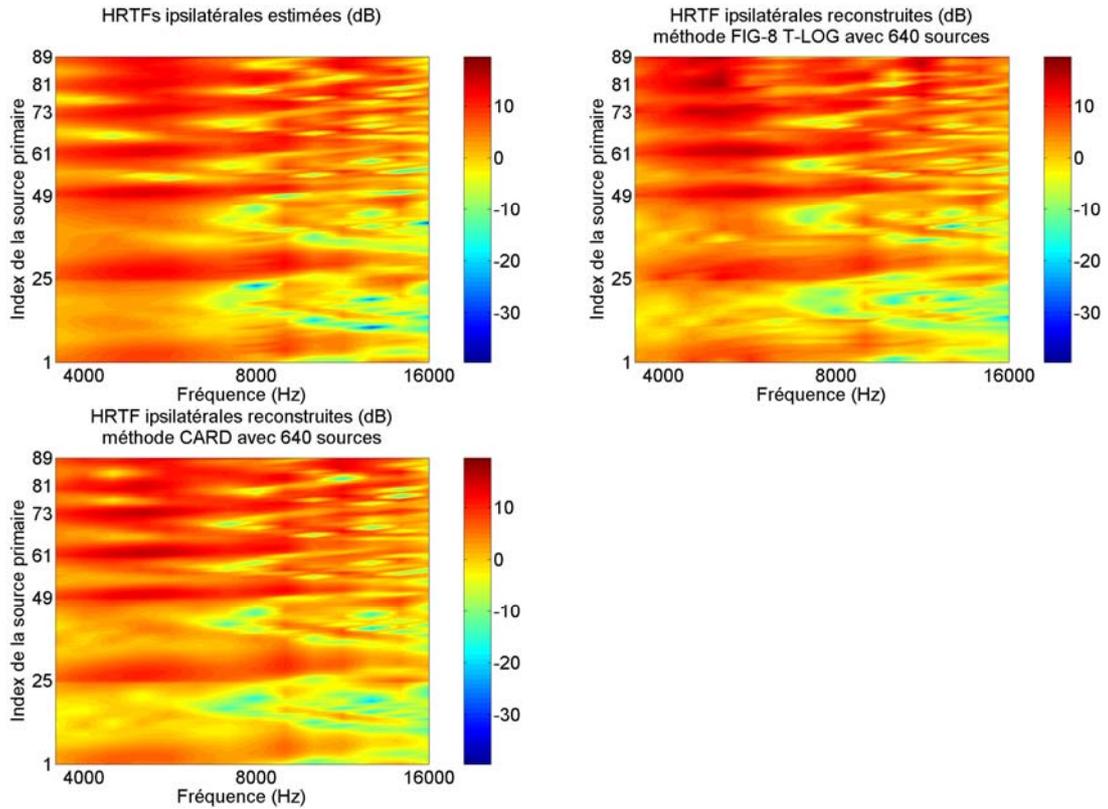


Fig. III-19 : Reconstruction des HRTF ipsilatérales via 640 sources secondaires - tête KU100

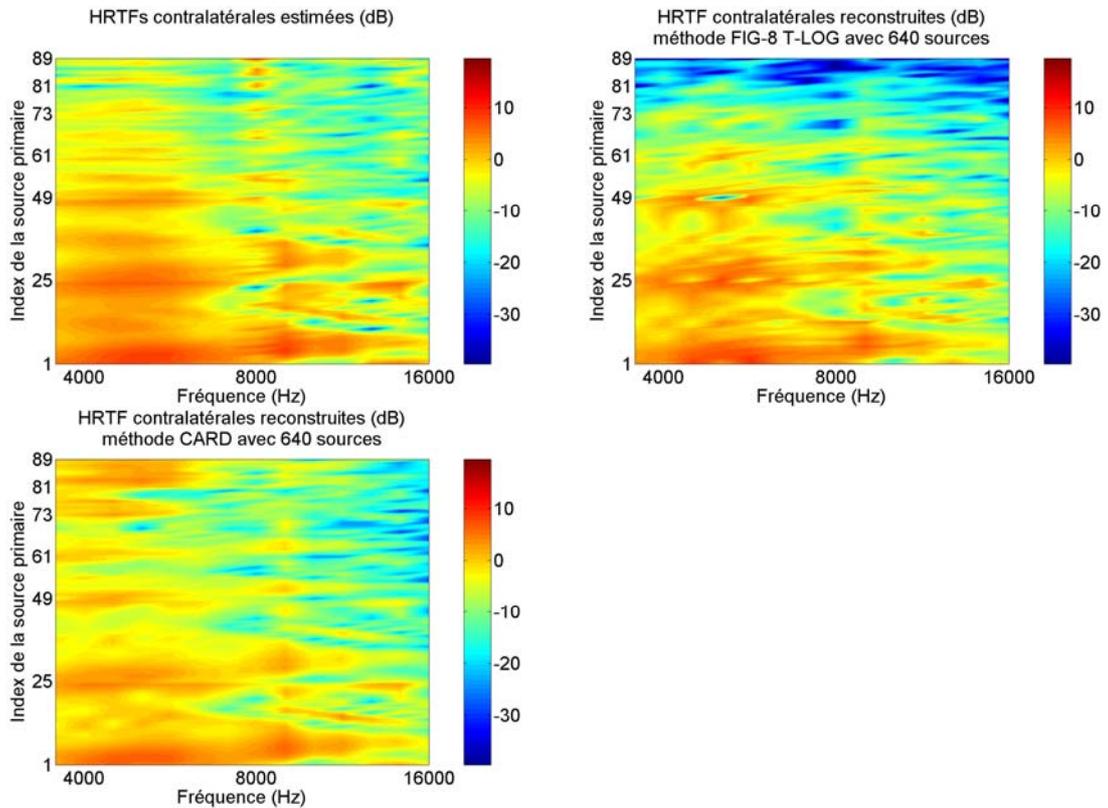


Fig. III-20 : Reconstruction des HRTF contralatérales via 640 sources secondaires - tête KU100

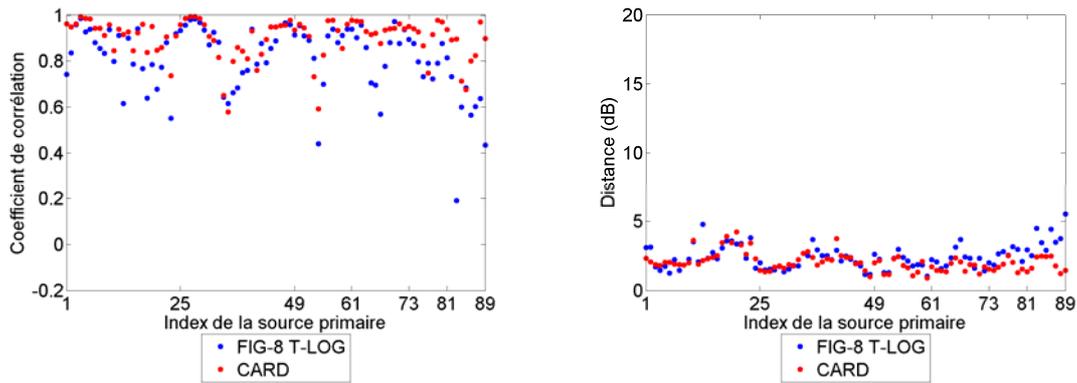


Fig. III-21 : Corrélation et distance en fonction de l'incidence - côté ipsilatéral

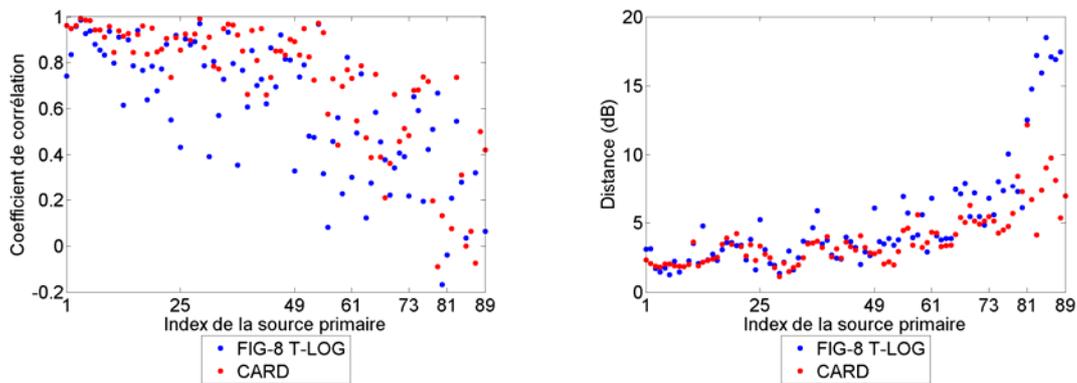


Fig. III-22 : Corrélation et distance en fonction de l'incidence - côté contralatéral

### 5.2.2 Réduction de la taille du réseau

Du côté ipsilatéral (voir Fig. III-23 et III-25), on constate le même comportement que pour les pavillons bafflés, à savoir une dégradation générale des indices de corrélation et de distance, ainsi qu'une qualité de reconstruction relativement meilleure pour la méthode FIG-8 T-LOG par rapport à la méthode CARD lorsque l'aliasing spatial est fortement effectif, c'est-à-dire pour des configurations de réseau comportant un faible nombre de sources secondaires.

Du côté contralatéral (voir Fig. III-24 et III-26), la dégradation est également clairement mise en évidence. Cependant, les indices ne permettent plus de conclure à la prédominance d'une technique de captation pour des tailles de réseau petites. En effet, plus le nombre de sources secondaires diminue, plus la corrélation est relativement meilleure pour la méthode FIG-8 T-LOG. A contrario, la distance spectrale reste plus grande pour cette même méthode et l'écart relatif avec la méthode CARD croît.

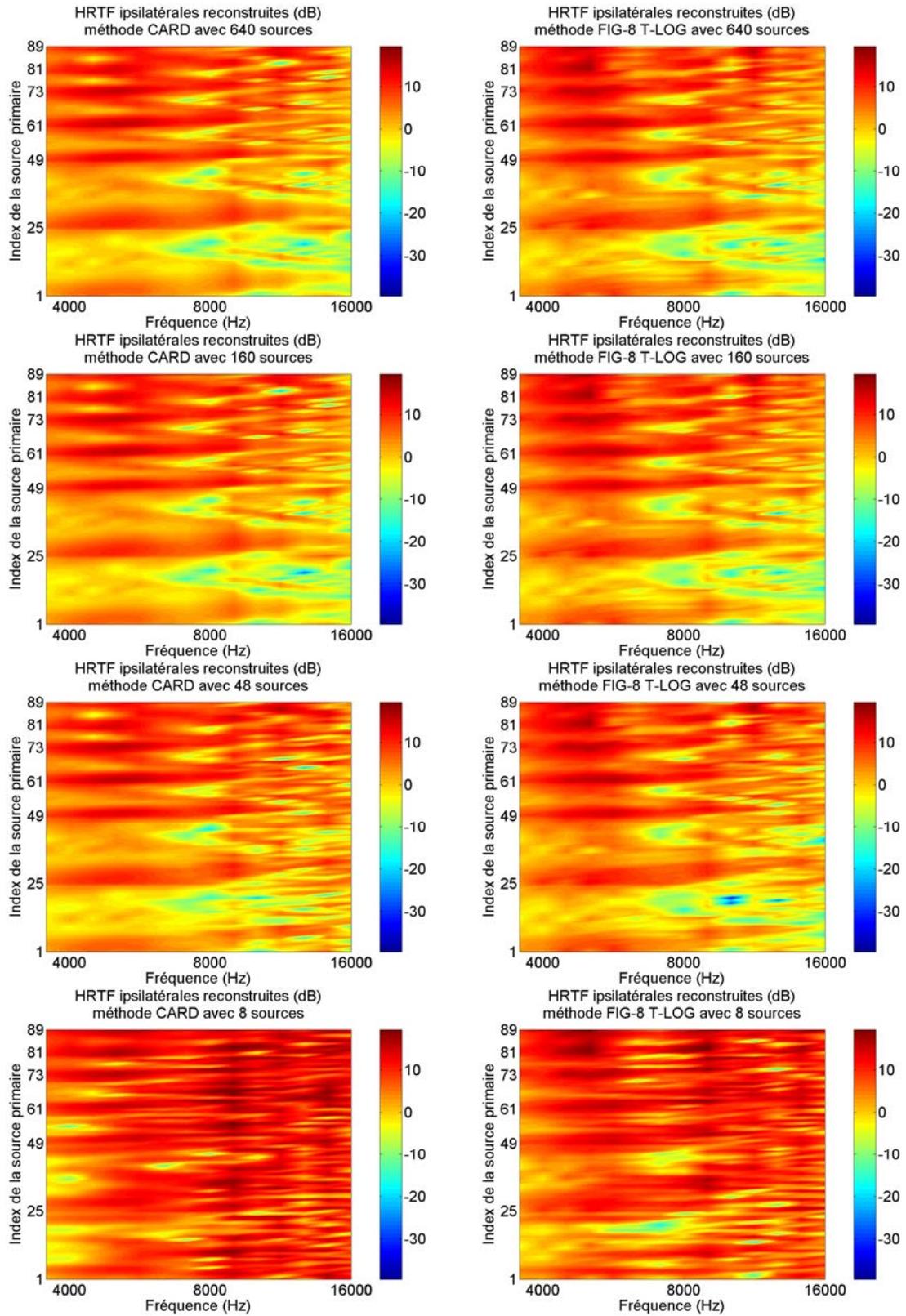


Fig. III-23 : HRTF ipsilatérales reconstruites en fonction de la taille du réseau - tête KU-100

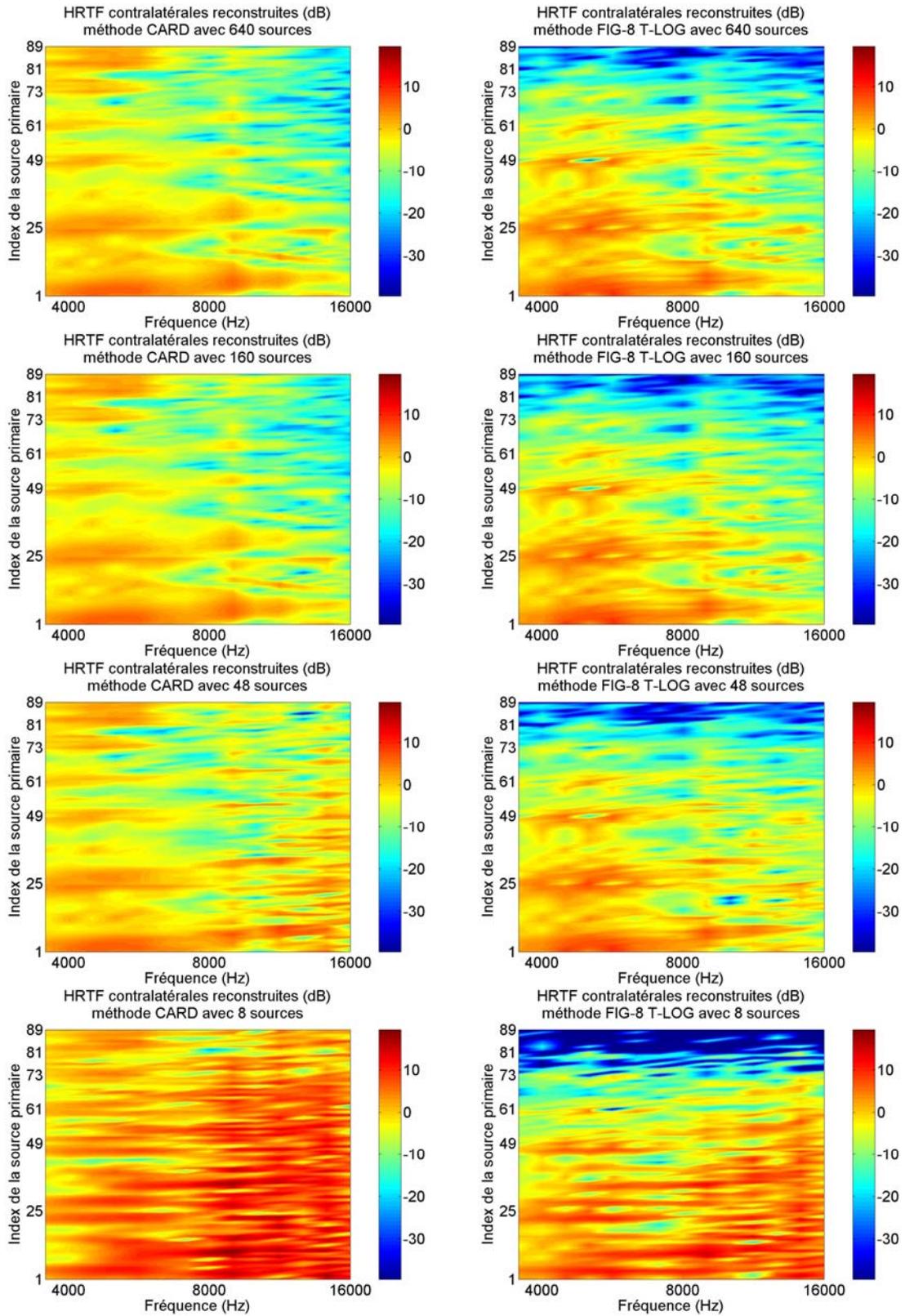


Fig. III-24 : HRTF contralatérales reconstruites en fonction de la taille du réseau - tête KU-100

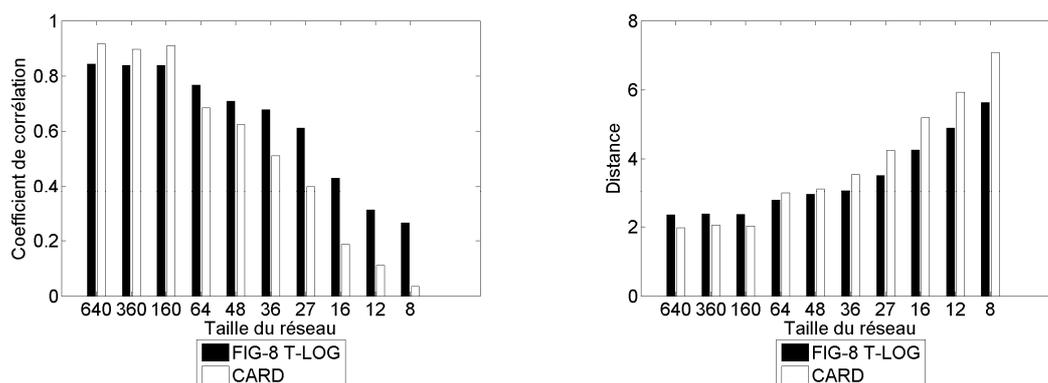


Fig. III-25 : Corrélation et distance en fonction de la taille de réseau - côté ipsilatéral

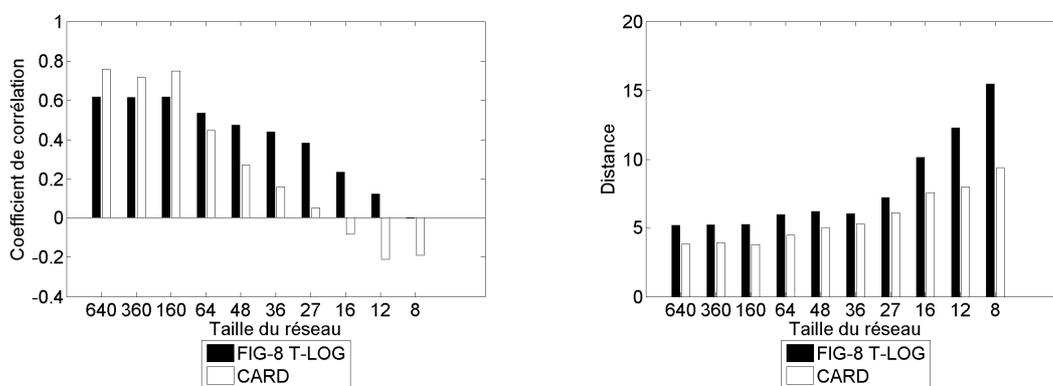


Fig. III-26 : Corrélation et distance en fonction de la taille de réseau - côté contralatéral

### 5.3 Vers une captation hybride

Pour tirer partie des avantages de chacune des méthodes de captation dans le cas d'un problème complexe comme celui de diffraction par une tête, nous proposons d'utiliser une technique de captation hybride.

La figure III-27 montre que, pour des sources primaires positionnées dans le plan médian, la captation FIG-8 T-LOG est plus robuste que la captation CARD : pour des tailles de réseau conséquentes, les deux méthodes sont comparables, mais dès que l'on diminue le nombre de sources secondaires, le comportement de la méthode FIG-8 T-LOG apparaît nettement plus fiable. Toutefois, nous avons vu précédemment qu'elle peut reconstruire de manière moins fidèle des HRTF correspondant à des angles latéraux élevés, notamment du côté contralatéral où la distance spectrale peut dépasser 15dB moyennant l'utilisation d'un réseau à 640 sources secondaires (voir Fig. III-22).

L'idée de la captation hybride est donc d'exploiter la méthode FIG-8 T-LOG à proximité du plan médian et la méthode CARD à proximité de l'axe interaural. En utilisant une simple loi de panning commandée par l'angle latéral, une approche hybride simple peut être établie. Appelons HYBRID cette nouvelle méthode.

Notons  $CAHF^{fig8}$  et  $CAHF^{card}$  les filtres holophoniques obtenus respectivement par les captations FIG-8 T-LOG et CARD, et  $\theta$  l'angle latéral (mesuré en degré). La méthode HYBRID peut être définie par les filtres holophoniques qui lui sont associés :

$$CAHF^{hybrid} = \frac{90 - |\theta|}{90} \cdot CAHF^{fig8} + \frac{|\theta|}{90} \cdot CAHF^{card} \quad (III-1)$$

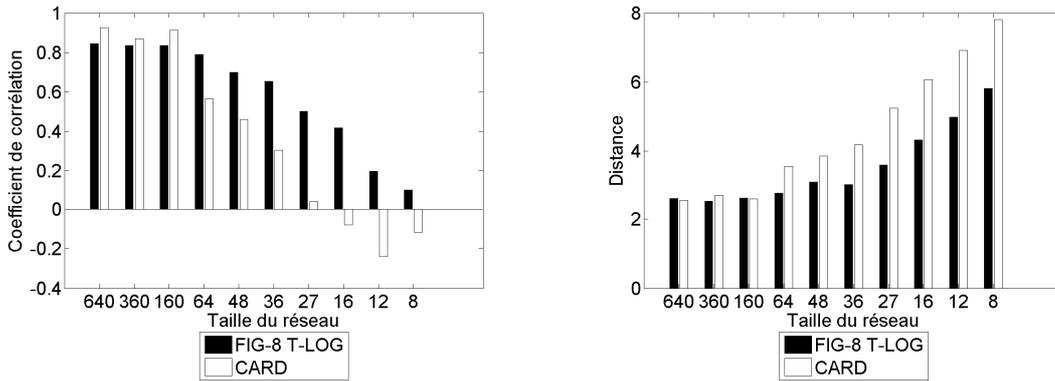


Fig. III-27 : Corrélation et distance en fonction de la taille de réseau - plan médian

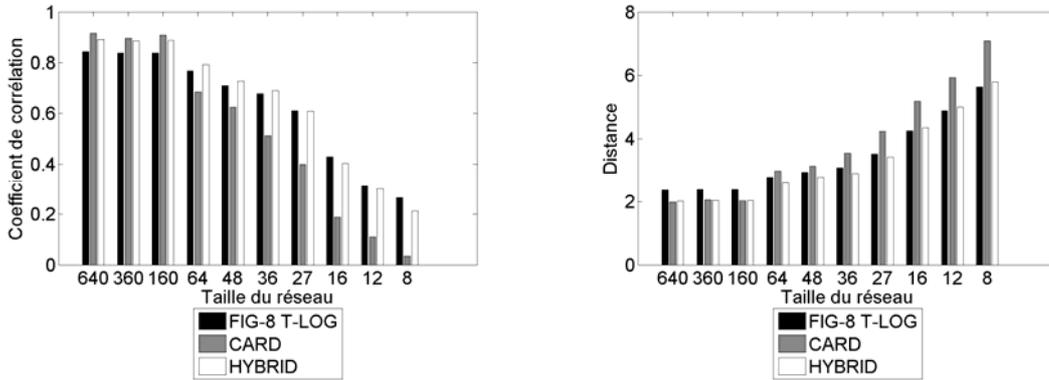


Fig. III-28 : Corrélation et distance en fonction de la taille de réseau - côté ipsilatéral

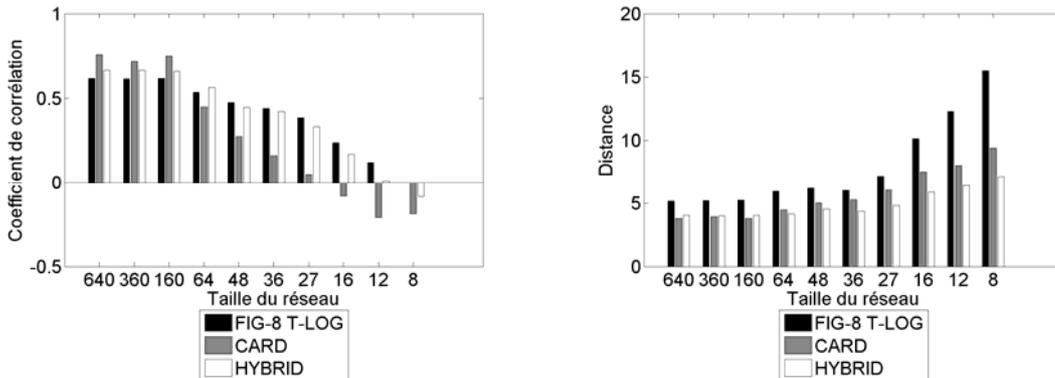


Fig. III-29 : Corrélation et distance en fonction de la taille de réseau - côté contralatéral

Les figures III-28 et III-29 montre l'intérêt de la méthode hybride, les résultats traduisant un comportement globalement plus robuste.

## **6 CONCLUSION**

Au vu de ces résultats, il apparaît nécessaire de trouver un moyen d'améliorer la reconstruction des HRTF lorsque le réseau considéré n'est composé que d'un petit nombre de sources secondaires. Dans le prochain chapitre, nous présenterons différentes idées ayant pour vocation d'atteindre cet objectif.

De plus, l'estimation des CAHF sur un modèle quelconque de tête étant extrêmement coûteux en temps, car nécessitant l'utilisation de logiciel de calcul BEM (ou autres), nous évaluerons l'impact d'une modélisation sphérique « équivalente » à la tête acoustique KU-100 sur la reconstruction des HRTF, car celle-ci donne accès à une résolution analytique du problème de diffraction, donc bien plus rapide qu'une résolution BEM.

## **COMPLEMENT DE FIGURES**

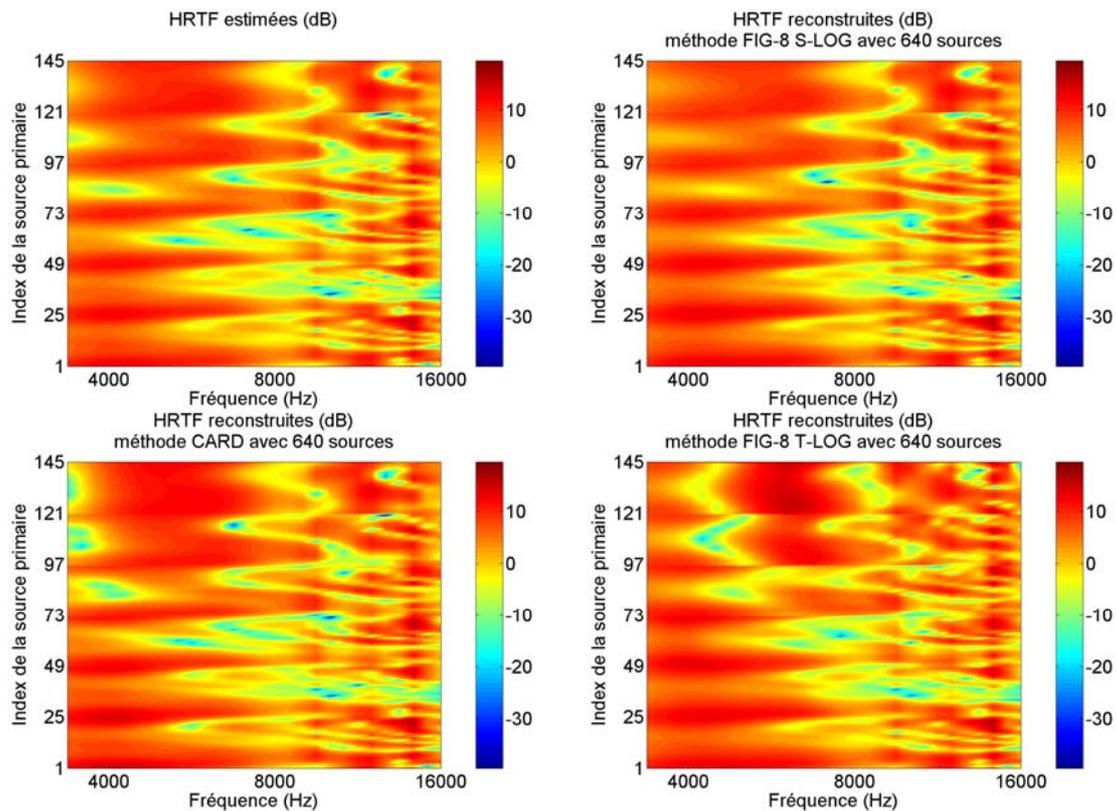


Fig. III-30 : Reconstruction des HRTF via 640 sources secondaires - pavillon Yuvi

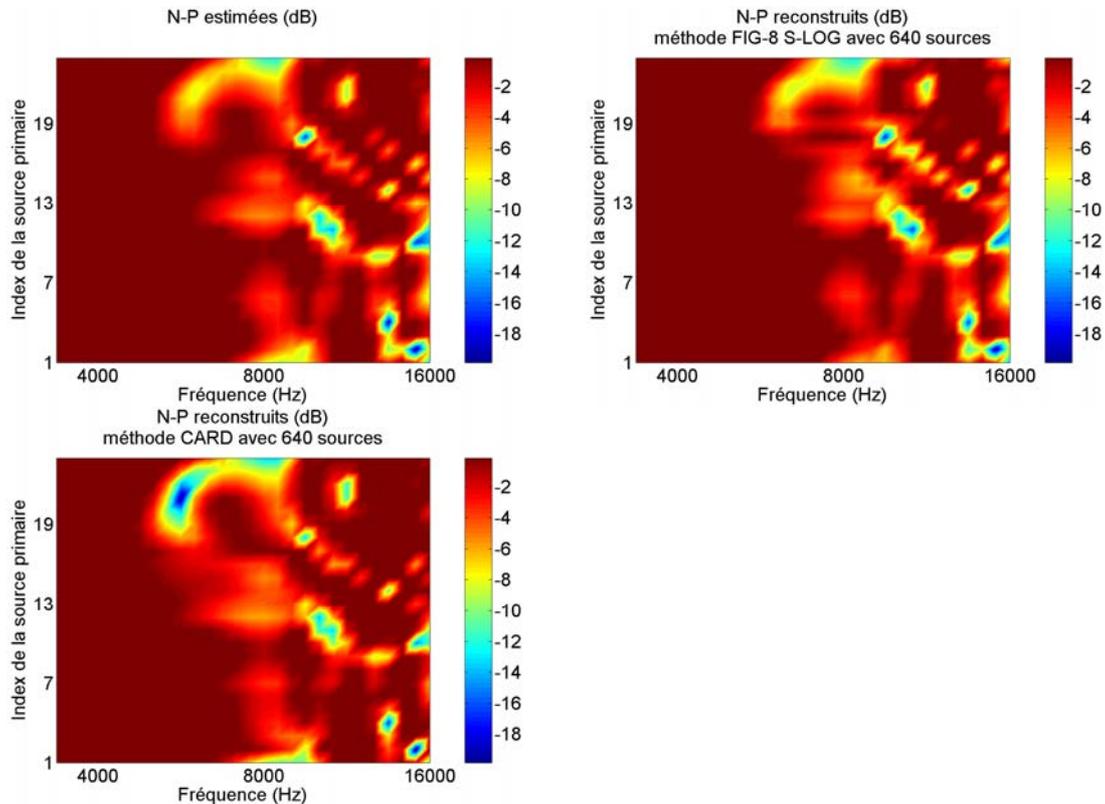


Fig. III-31 : Reconstruction des N-P via 640 sources secondaires - pavillon Yuvi

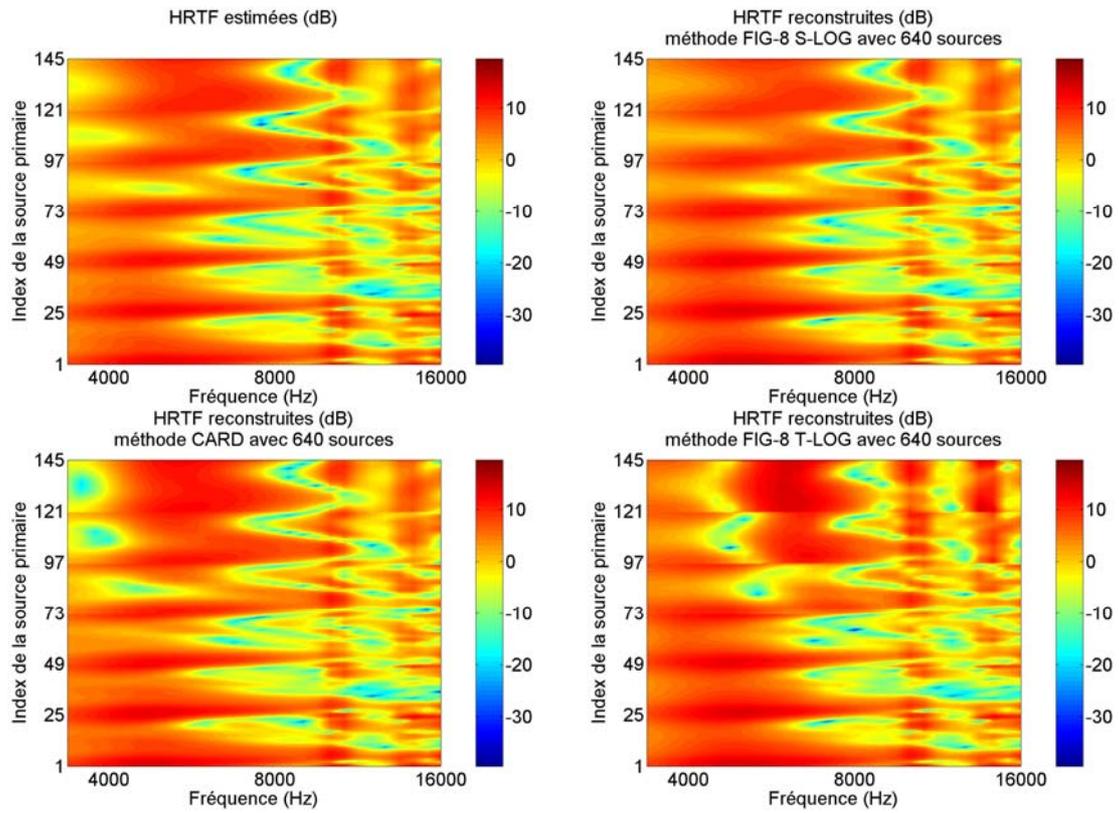


Fig. III-32 : Reconstruction des HRTF via 640 sources secondaires - pavillon DB60

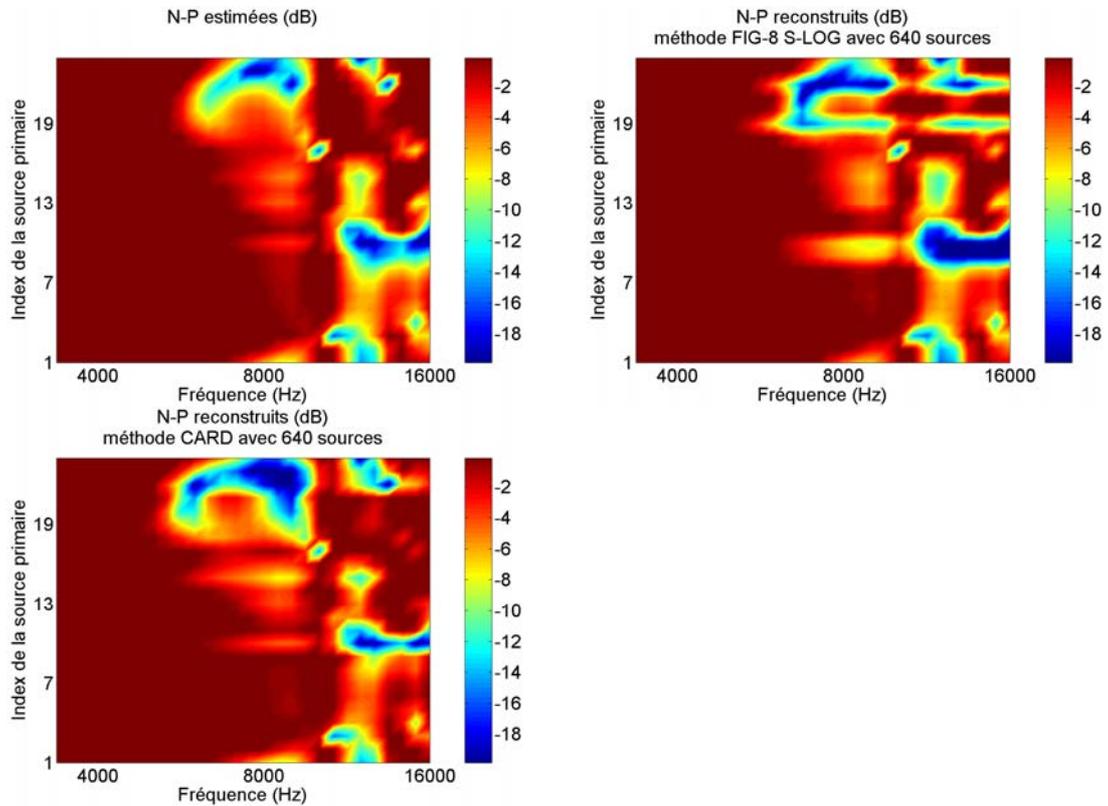


Fig. III-33 : Reconstruction des N-P via 640 sources secondaires - pavillon DB60

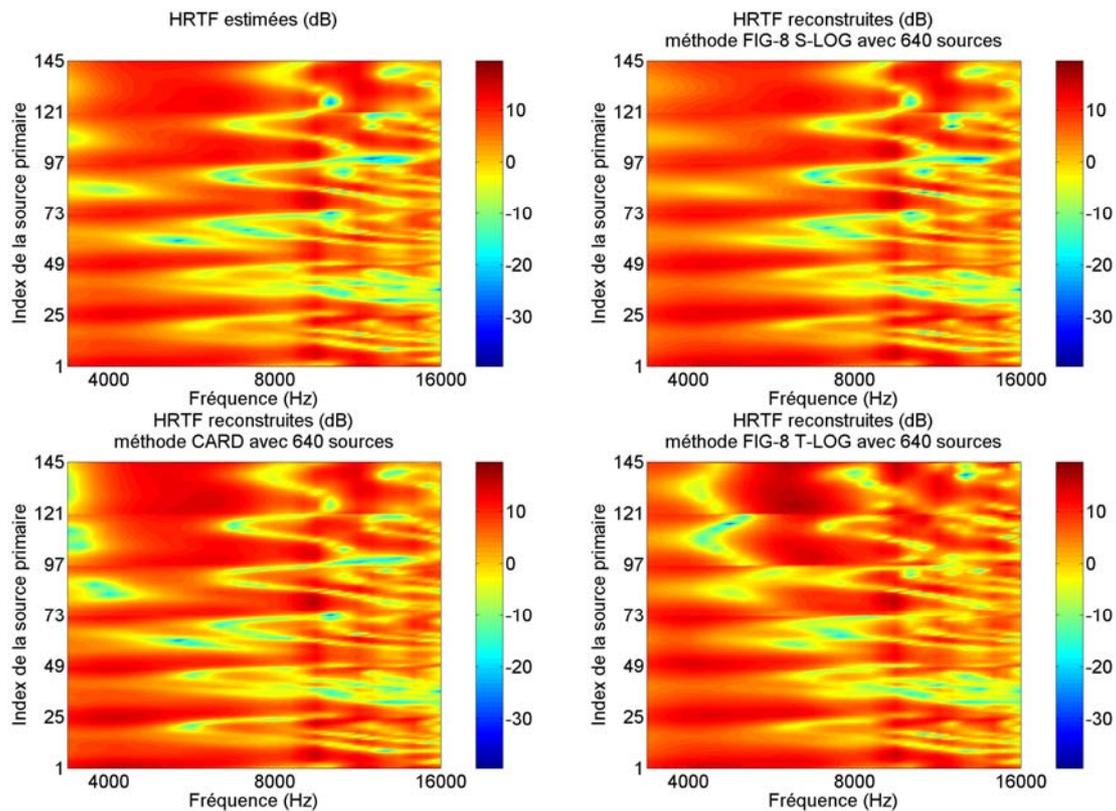


Fig. III-34 : Reconstruction des HRTF via 640 sources secondaires - pavillon DB65

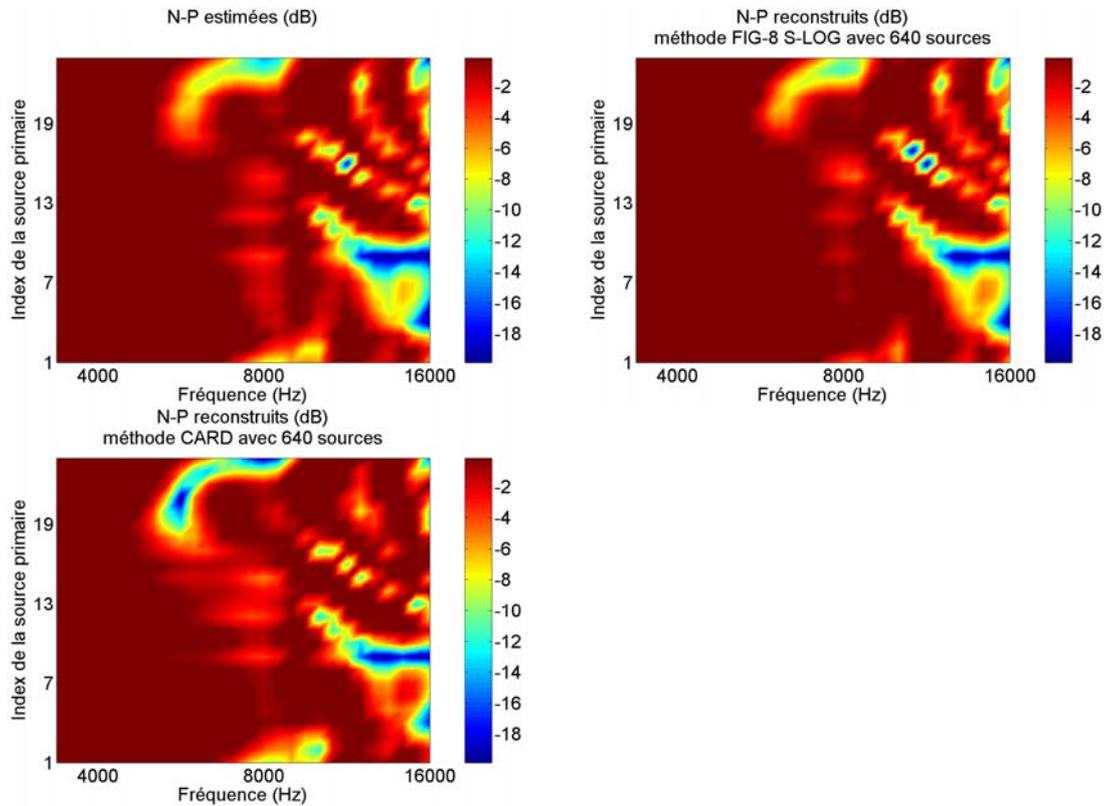


Fig. III-35 : Reconstruction des N-P via 640 sources secondaires - pavillon DB65

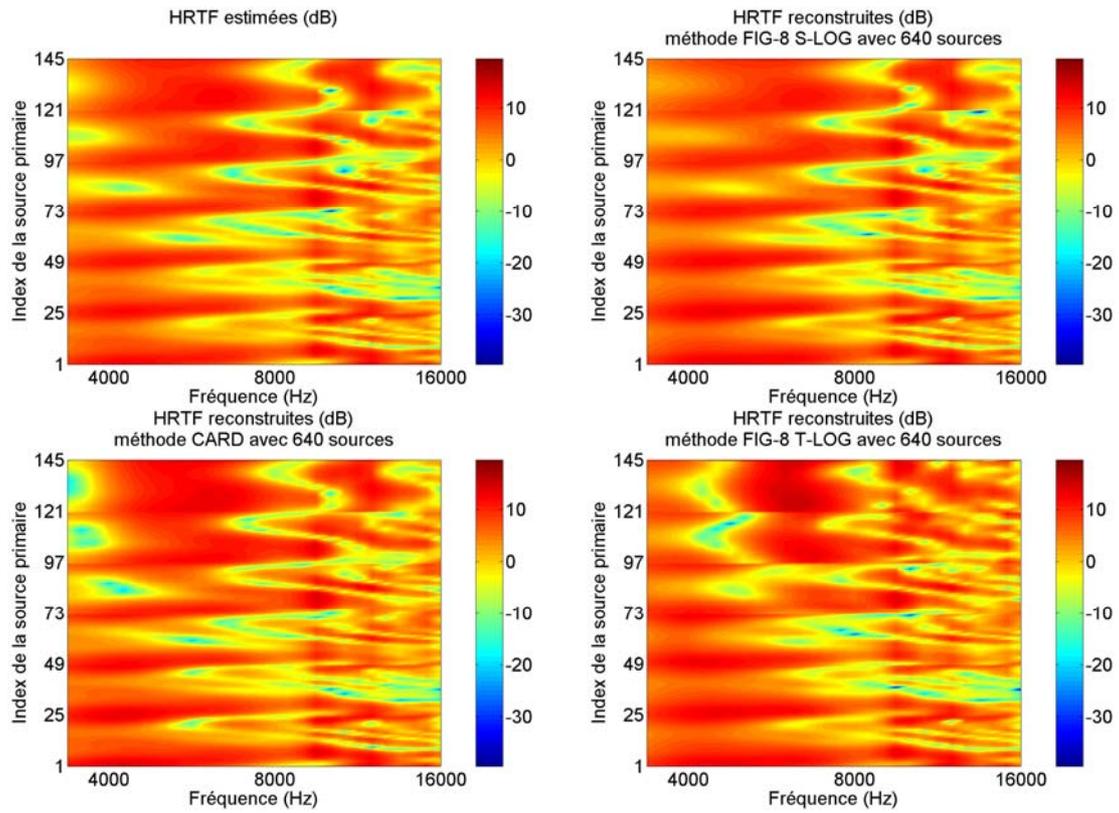


Fig. III-36 : Reconstruction des HRTF via 640 sources secondaires - pavillon DB90

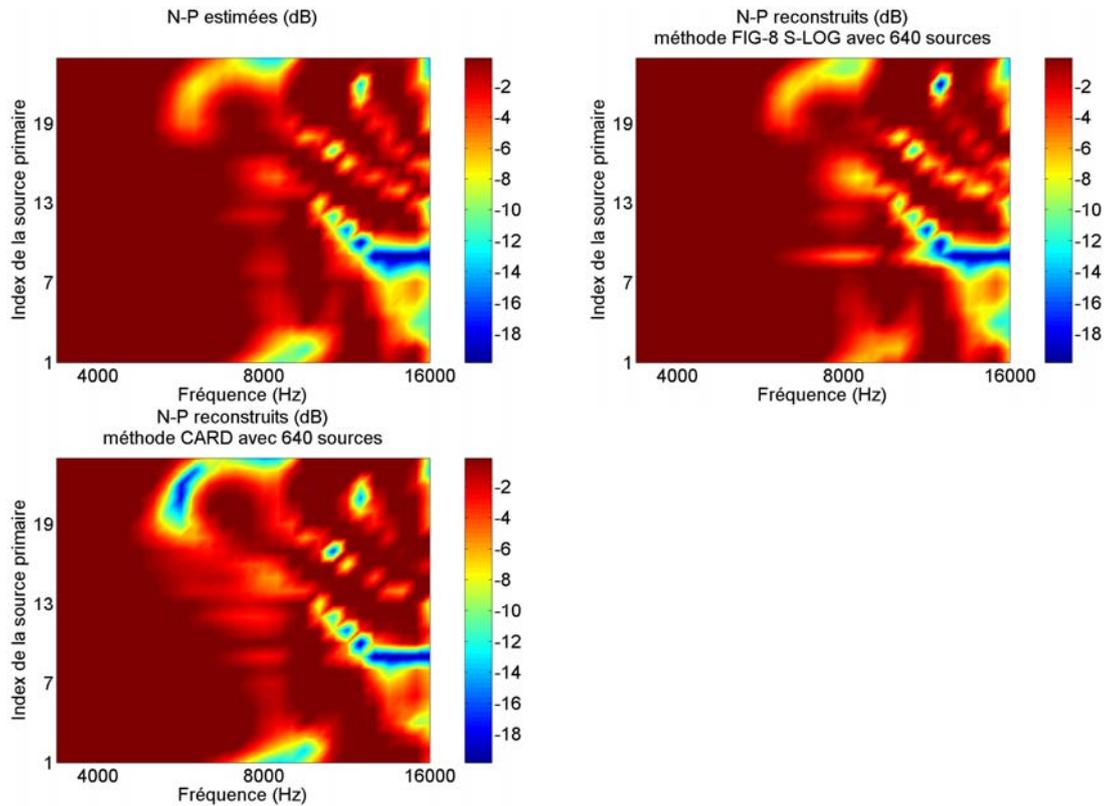


Fig. III-37 : Reconstruction des N-P via 640 sources secondaires - pavillon DB90

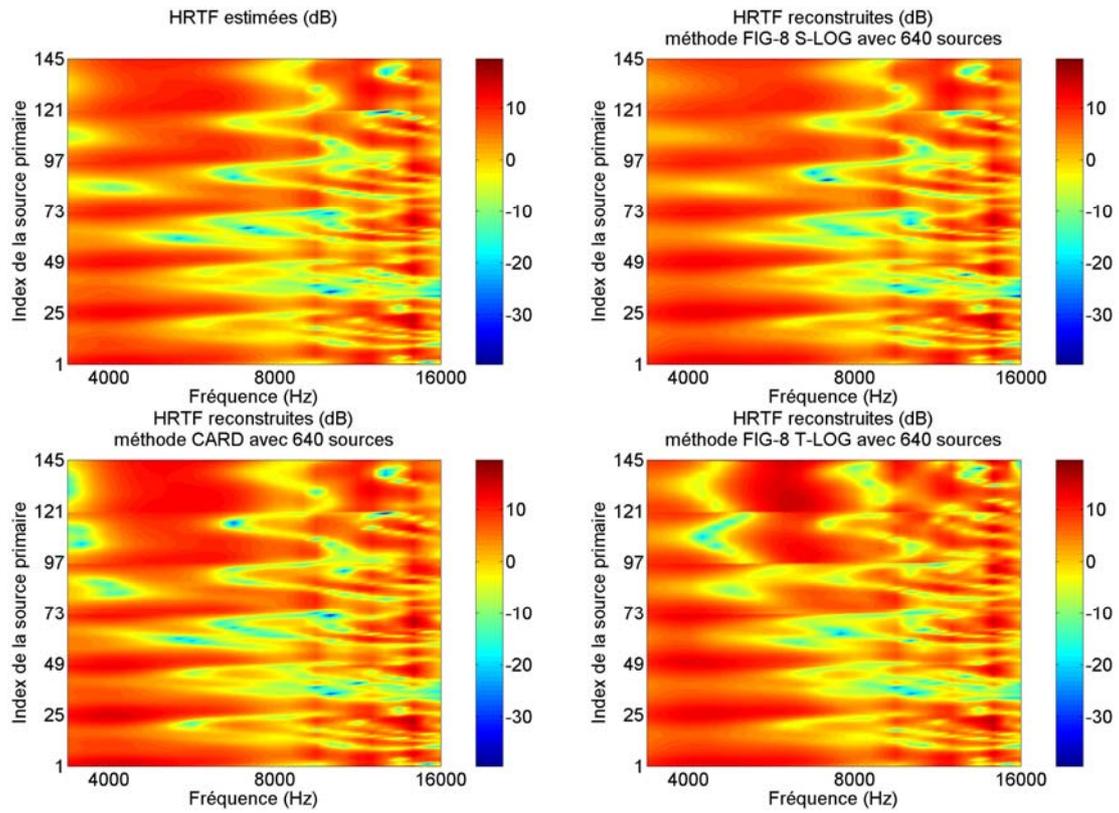


Fig. III-38 : Reconstruction des HRTF via 640 sources secondaires - pavillon DB95

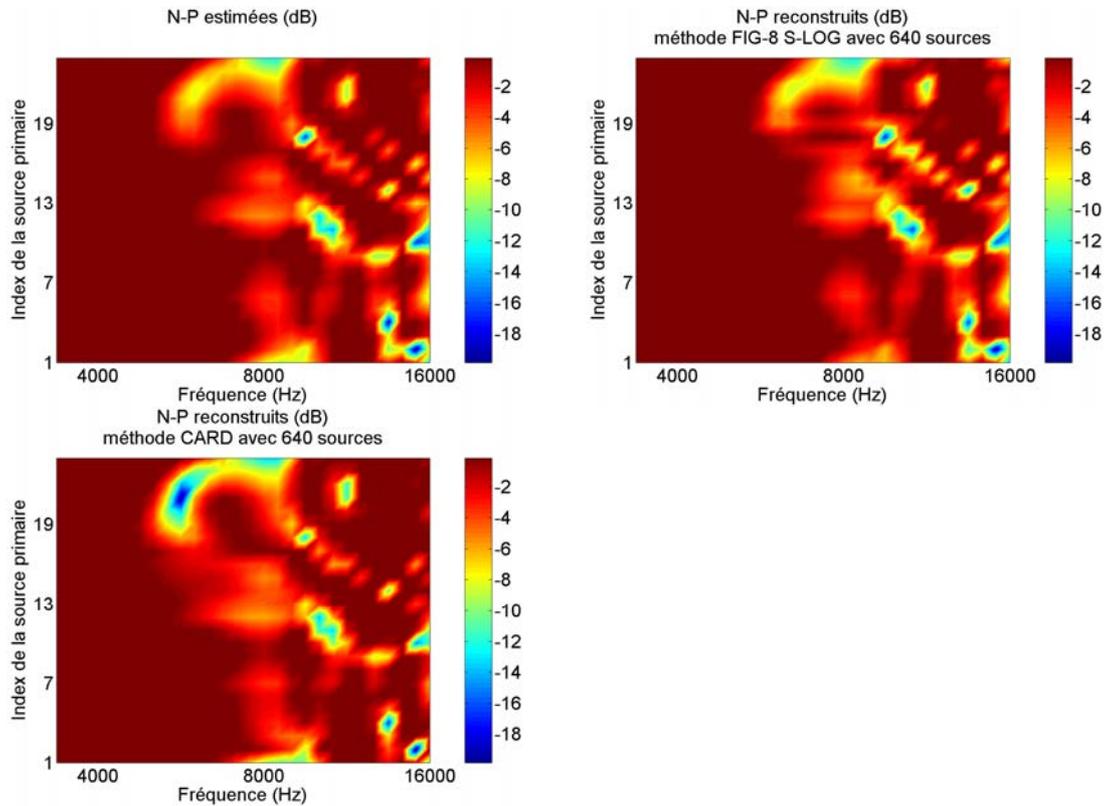


Fig. III-39 : Reconstruction des N-P via 640 sources secondaires - pavillon DB95

# IV

## Simplifications et optimisations

## 1 INTRODUCTION

Ce chapitre aborde tout d'abord une simplification couramment étudiée dans le domaine de la synthèse binaurale. Celle-ci consiste à modéliser une tête par une sphère équivalente. L'attrait du modèle sphérique s'explique par l'existence d'une solution analytique à la résolution du problème de diffraction acoustique. Nous évaluerons l'impact de ce type de simplification sur la reconstruction des HRTF de la tête KU-100.

Dans un deuxième temps seront exposées différentes optimisations destinées à améliorer le fonctionnement de réseaux composés d'un faible nombre de sources secondaires, et donc fortement sujets au phénomène d'aliasing spatial, et ce sur presque toute la bande de fréquence d'utilisation des pavillons. Ces optimisations peuvent être séparées en deux groupes, en distinguant celles qui tentent d'améliorer l'information diffusée par une disposition de réseau donnée, de celles qui ont pour but de trouver le meilleur positionnement des sources secondaires et donc la disposition du réseau elle-même.

## 2 MODELE DE TETE SPHERIQUE EQUIVALENTE

### 2.1 Sphère iso-ITD

Différentes approches géométriques permettent d'approcher la morphologie d'une tête par une sphère. Pour ce faire, il est nécessaire de choisir quelle caractéristique de la tête devra être conservée par la modélisation sphérique. Dans sa thèse, Katz [Kat98] avait envisagé la simplification de modèle numérique de tête par une sphère en étudiant l'équivalence volumique ou encore une équivalence entre le diamètre de la sphère et la distance interaurale. Cette dernière idée, bien que purement géométrique, ouvre la voie à des considérations perceptives : en effet, l'idée est d'approximer le modèle de tête par une sphère de dimension telle qu'elle permet d'approcher au mieux les indices interauraux originaux. Algazi et al. [AAD01b] concrétisent cette idée en proposant une équivalence iso-ITD : en considérant l'ITD dans le plan horizontal comme l'indice perceptivement prédominant parmi les indices liés à la tête, il détermine le rayon sphérique équivalent en minimisant l'erreur au sens des moindres carré entre l'ITD estimée du modèle de tête à simplifier, et l'ITD d'une sphère calculée selon la formule de Woodworth.

Nous nous proposons ici d'appliquer cette méthode, non plus uniquement au plan horizontal, mais à tout l'espace. La transposition à tout l'espace de la méthode d'Algazi a déjà été entreprise par Larcher [Lar01]. Cependant, alors qu'Algazi utilisait une méthode d'estimation de l'ITD par détection de seuil, et Larcher une méthode par approximation linéaire de la phase, nous choisirons d'appliquer une méthode par corrélation appliquée sur l'enveloppe énergétique des HRIR [BNK04].

## 2.2 Détermination de la sphère équivalente

### 2.2.1 Estimation de l'ITD

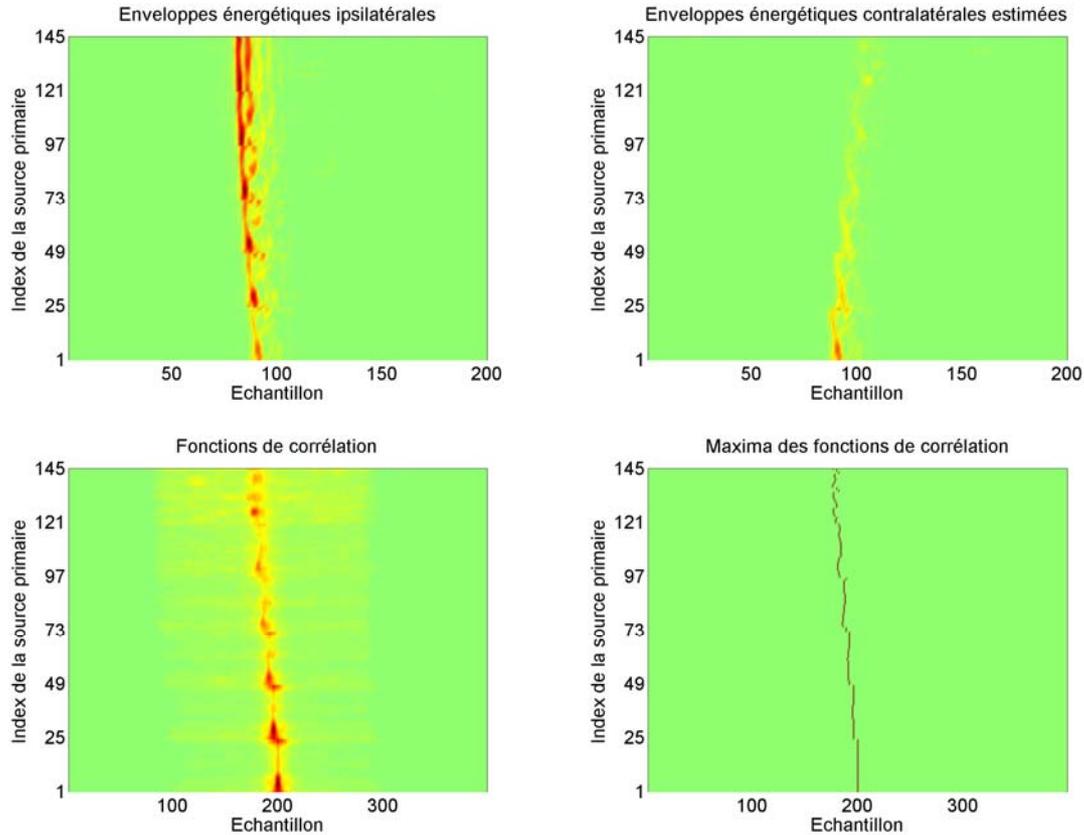


Fig. IV-1 : Illustration de l'estimation de l'ITD

Le calcul BEM des HRTF du modèle de la tête acoustique KU-100 a été reconduit sur une échelle linéaire en fréquence, de 100Hz à 16kHz par pas de 100Hz, pour pouvoir estimer les HRIR correspondantes sur une fenêtre de 200 échantillons à une fréquence d'échantillonnage de 32kHz. 145 positions de sources ont été considérées (voir Tab. III-1). L'estimation de l'ITD par corrélation a été menée sur l'enveloppe énergétique des HRIR ainsi définies (voir Fig. IV-1).

### 2.2.2 Obtention du rayon équivalent

Notons  $\hat{\tau}$  l'ITD modélisée par la formule de Woodworth en (I-9), et  $\tau$  l'ITD estimée d'une tête dont nous voulons trouver la sphère équivalente. En posant  $\alpha = \sin \theta + \theta$ , où  $\theta$  désigne l'angle latéral, on peut écrire alors  $\hat{\tau} = (r/c)\alpha$ , où  $r$  est le rayon sphérique équivalent à déterminer. Cela peut être effectué en minimisant l'erreur  $\hat{\tau} - \tau = \hat{\tau} - (a/c)\alpha$ . En définissant  $\hat{\tau}$  et  $\alpha$  comme des vecteurs colonnes contenant des valeurs estimées pour un certain nombre de valeurs discrètes de  $\theta$ , on calcule le rayon équivalent  $r$  ainsi :

$$r = c \frac{\hat{\tau}'\alpha}{\alpha'\alpha} \quad (\text{IV-1})$$

où  $c$  est la vitesse du son et l'opérateur  $t$  désigne la transposition. En appliquant ceci à la tête KU-100, on trouve un rayon équivalent de 97mm. L'erreur aux moindres carrés résiduelle entre l'ITD estimée et l'ITD modélisée est de l'ordre de  $34\mu\text{s}$ . Sachant que les HRIR sur lesquelles l'ITD a été estimée ont été échantillonnées par pas de  $32\mu\text{s}$ , l'erreur résiduelle moyenne ne représente qu'un échantillon temporel.

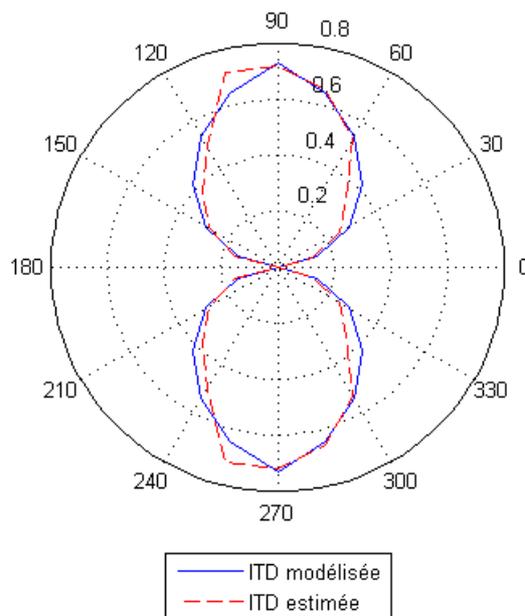


Fig. IV-2 : ITD estimées et modélisées de la tête KU-100 (en ms) – plan horizontal

## 2.3 Reconstruction des HRTF à partir du modèle sphérique

### 2.3.1 Sans aliasing spatial

Les CAHF ont été calculés de manière analytique (résolution du problème de diffraction par une sphère rigide à l'aide d'une généralisation à tout point de l'espace de l'algorithme de calcul de HRTF sphérique proposé par machin [DM98, MI68]).

Nous évaluons dans un premier temps l'impact de la modélisation sphérique de la tête KU-100 en utilisant le réseau à 640 sources secondaires : la reconstruction holophonique n'est donc pas soumise à l'effet de l'aliasing spatial. Les figures IV-3 et IV-4 semblent témoigner d'une bonne concordance des HRTF reconstruites à l'aide du modèle sphérique iso-ITD avec les HRTF originales, ainsi qu'avec les HRTF reconstruites à partir du modèle scanné de la tête KU-100.

Les indices de corrélation et de distance (voir Fig. IV-5 et Fig. IV-6) nous permettent de tirer les mêmes conclusions. Les reconstructions à l'aide du modèle KU-100 et de la sphère iso-ITD ne présentent aucune différence notable.

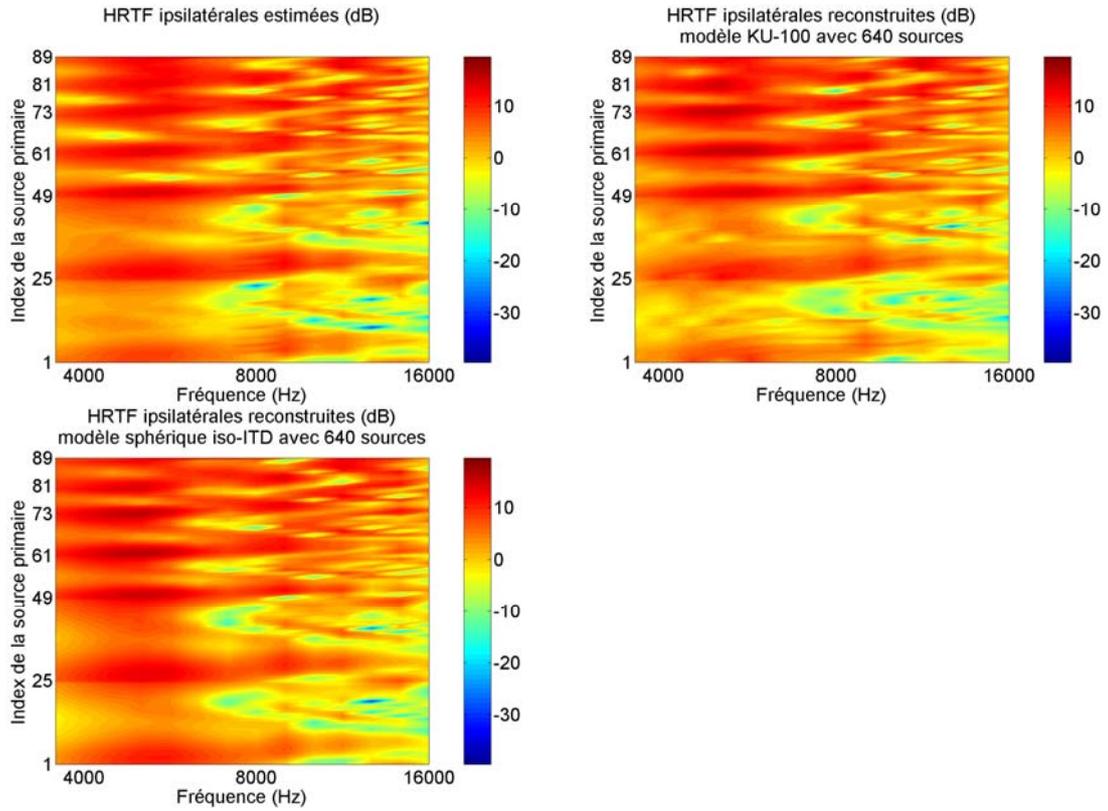


Fig. IV-3 : Evaluation de la modélisation sphérique iso-ITD - HRTF ipsilatérales

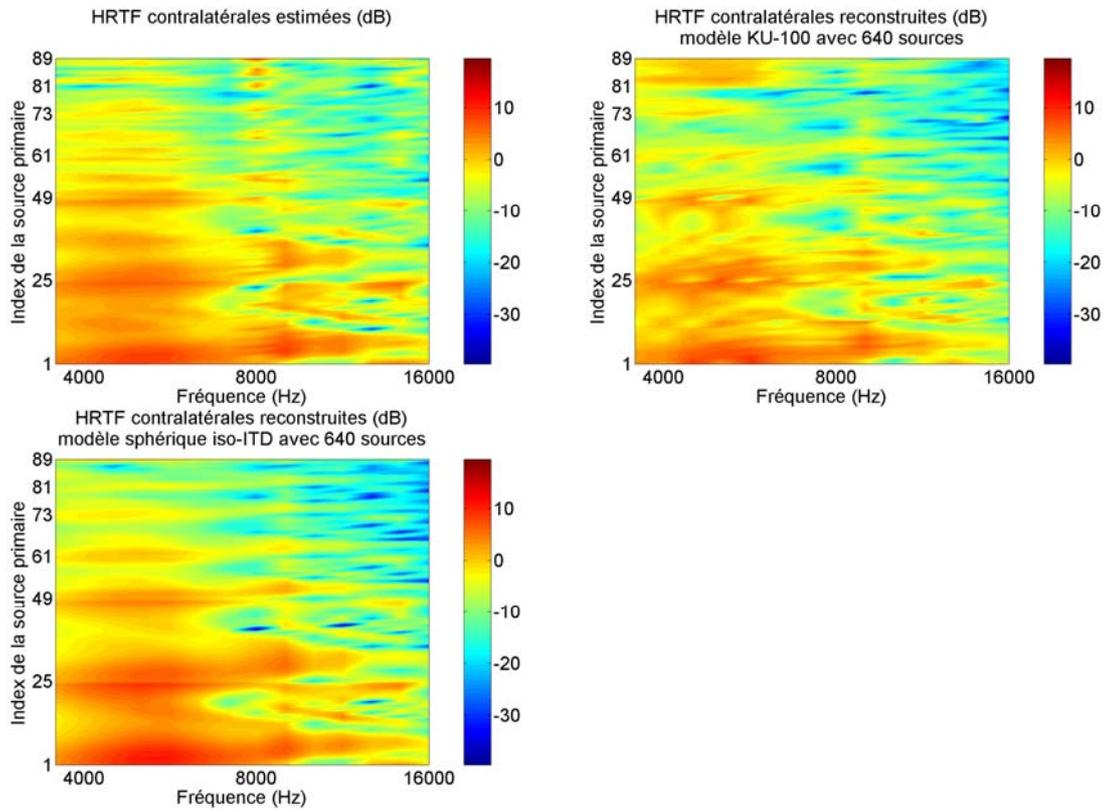


Fig. IV-4 : Evaluation de la modélisation sphérique iso-ITD - HRTF contralatérales

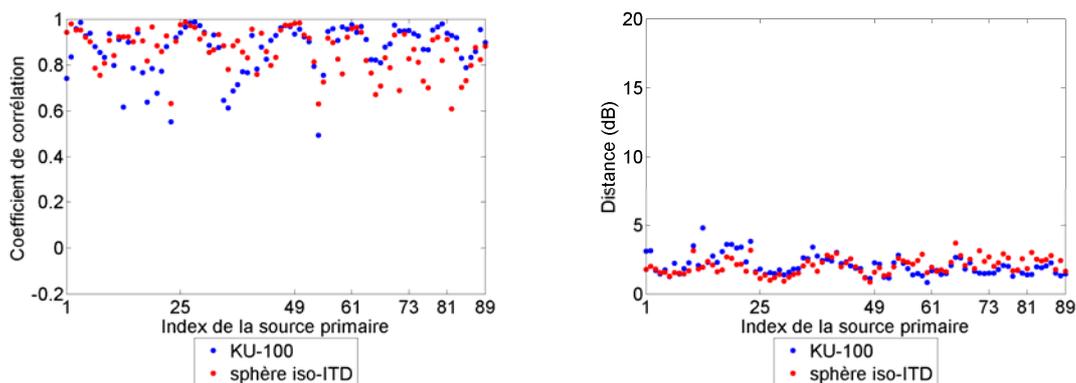


Fig. IV-5 : Corrélation et distance en fonction de l'incidence - côté ipsilatéral

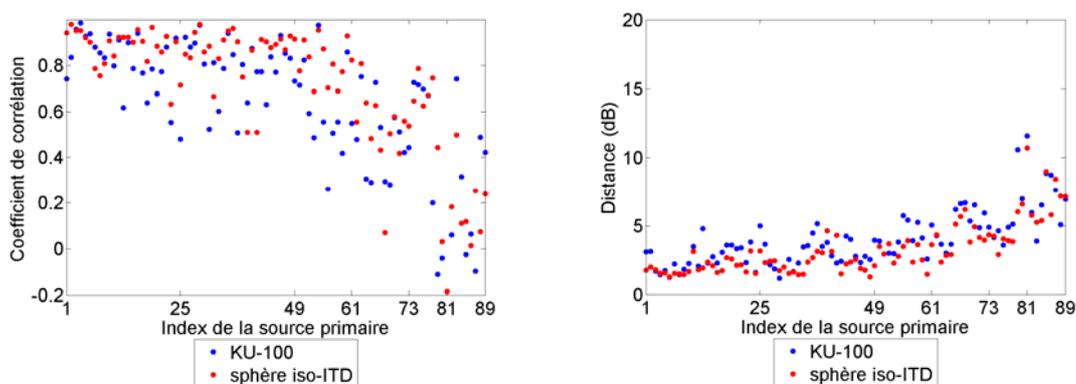


Fig. IV-6 : Corrélation et distance en fonction de l'incidence - côté contralatéral

### 2.3.2 Réduction de la taille de réseau

Lorsque l'on réduit la taille du réseau, le modèle sphérique continue d'offrir une qualité de reconstruction comparable à celle du modèle scanné (voir Fig. IV-7 et IV-8). On peut préciser que les indices de corrélation sont légèrement meilleurs moyennant l'utilisation du modèle sphérique plutôt que le modèle scanné, et ce quelque soit le côté considéré (ipsilatéral ou contralatéral) ou quelque soit la taille de réseau employée. Les indices de distance quant à eux sont comparables (voir Fig. IV-9 et Fig IV-10).

Ces résultats sont intéressants : ils montrent qu'une modélisation sphérique d'une tête rigide quelconque peut fournir des résultats tout aussi bons qu'une modélisation plus réaliste et plus précise de cette même tête, en termes de reconstruction holophonique des HRTF. De plus, et grâce à cette modélisation simplifiée, il est possible de diminuer considérablement le temps de calcul des CAHF, passant d'une vingtaine de jour à une vingtaine d'heure.

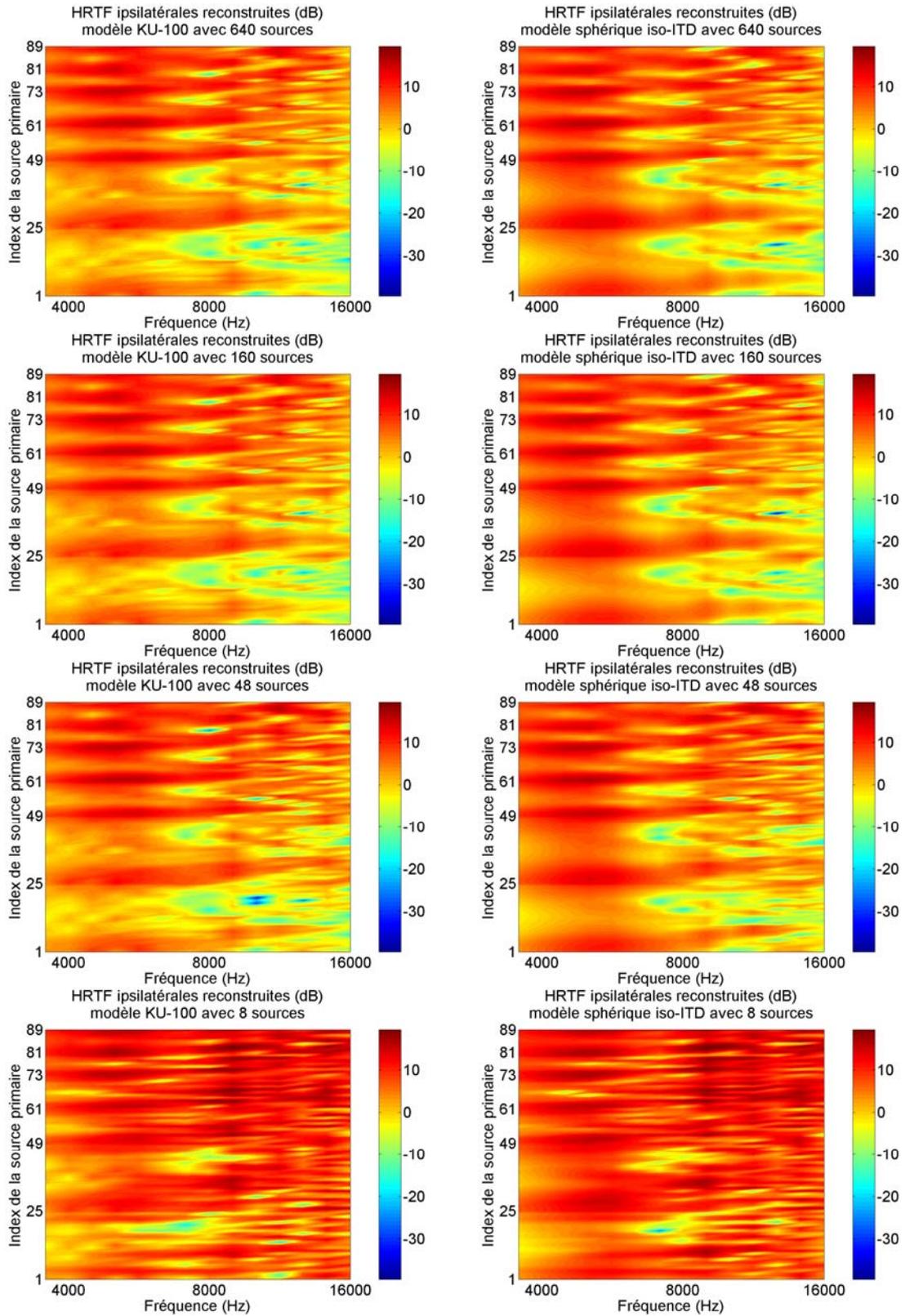


Fig. IV-7 : Evaluation du modèle sphérique en fonction de la taille du réseau - HRTF ipsilatérales

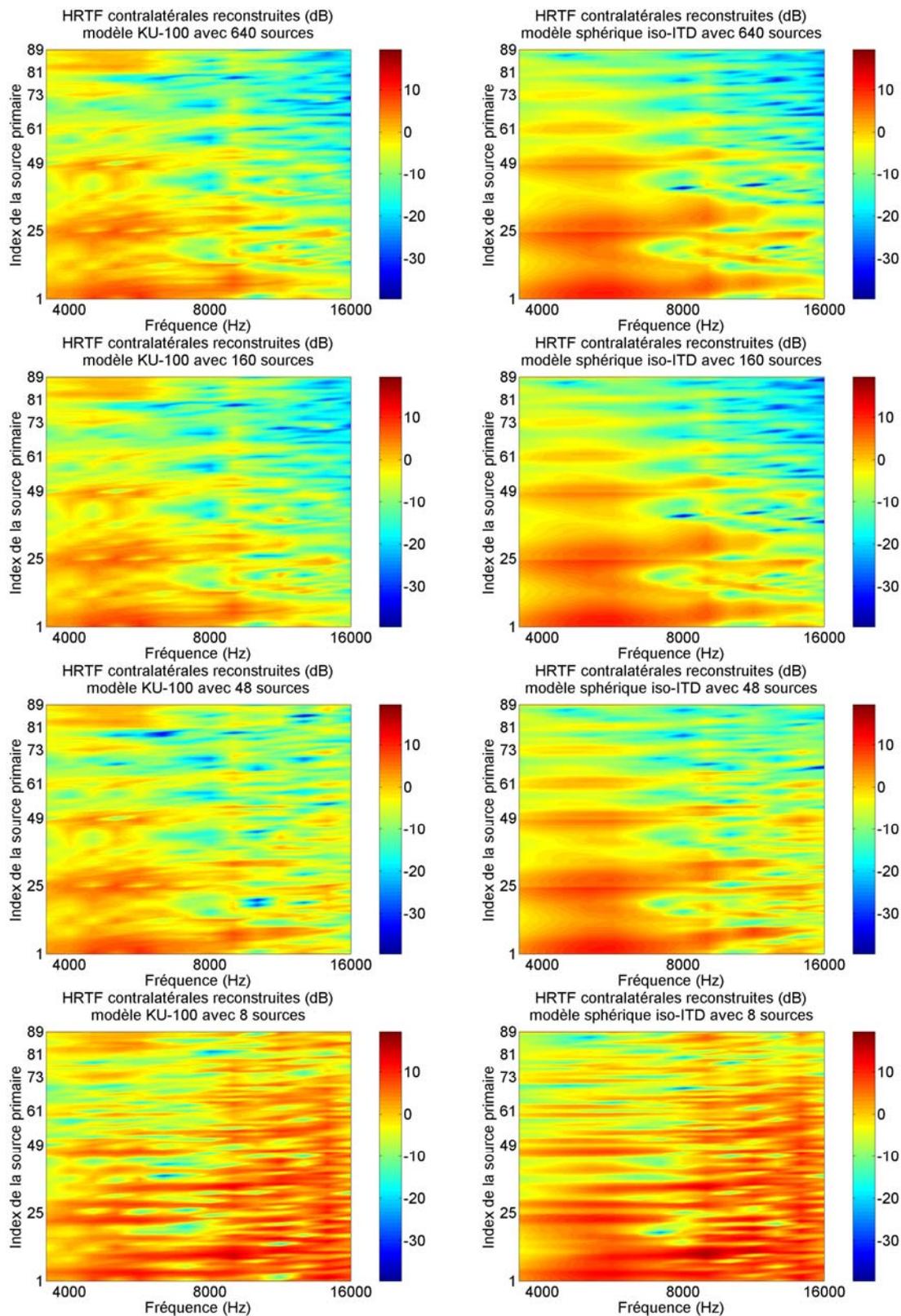


Fig. IV-8 : Evaluation du modèle sphérique en fonction de la taille du réseau - HRTF contralatéral

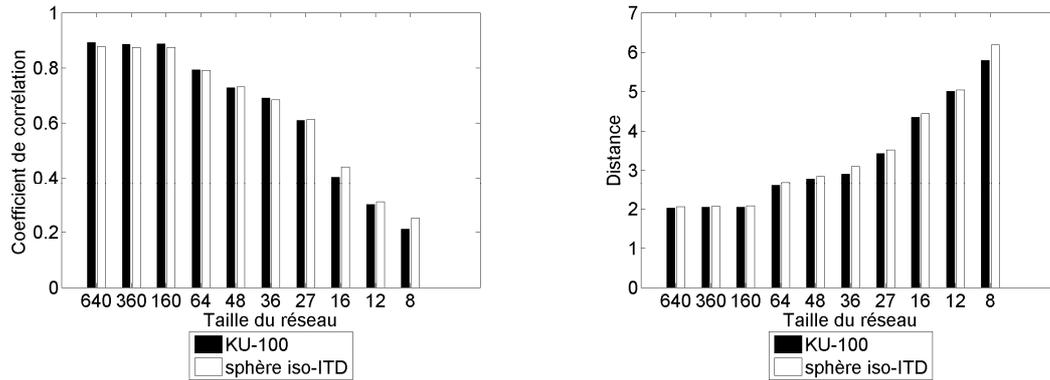


Fig. IV-9 : Corrélation et distance en fonction de la taille de réseau - côté ipsilatéral

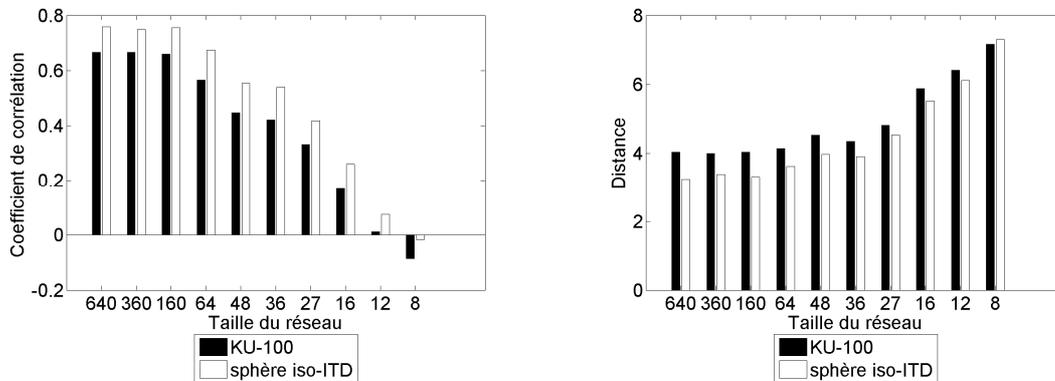


Fig. IV-10 : Corrélation et distance en fonction de la taille de réseau - côté contralatéral

### 3 OPTIMISATION DES CAHF POUR UNE DISPOSITION DONNEE

#### 3.1 Méthode de moyennage par zone

Lorsque l'on diminue le nombre de sources secondaires, la surface représentée par chacune d'elles augmente. Cependant, la captation est toujours ponctuelle et s'effectue aux lieux de ces sources secondaires. Elle peut, de ce fait, ne plus être représentative du champ capté sur chaque zone de surface associée à chaque source secondaire.

Nous proposons d'alimenter les sources secondaires par une information moyennée sur toute la zone surfacique associée à chaque source, plutôt que par la seule information captée ponctuellement à l'endroit de chaque source. Nous appellerons cette méthode AZ pour *Average Zone* (zone de moyennage).

##### 3.1.1 Illustration de la méthode

Notons  $R_N$  le réseau à  $N = 640$  sources secondaires. Nous avons vu précédemment que l'échantillonnage spatial de la surface hémisphérique, qu'il constitue, est suffisamment fin pour offrir une représentation discrète pertinente de la surface continue, même au-

delà de 16kHz. Notons  $R_r$  le réseau réduit à  $r$  sources secondaires sélectionnées parmi celle du réseau  $R_N$ , avec  $r < N$ , et correspondant à la configuration de réseau considérée. Si  $i$  est l'indice d'une source dans le réseau  $R_r$ ,  $j$  l'indice d'une source dans le réseau  $R_N$ , l'ensemble  $E_i$  des sources secondaires de  $R_N$  situées dans la zone de proximité de la  $i^{\text{ème}}$  source secondaire du réseau  $R_r$  est défini comme suit :

$$j \in E_i \Leftrightarrow \begin{cases} \text{dist}(i, j) < \text{dist}(k, j) \\ k \neq i \end{cases} \quad (\text{IV-2})$$

avec  $k = 1 \dots r$ , et où la fonction  $\text{dist}$  désigne la distance à la surface entre le lieu de la source indiquée  $i$  dans  $R_r$  et le lieu de la source indiquée  $j$  dans  $R_N$ . Les réseaux considérés ayant une géométrie hémisphérique, la distance à la surface est proportionnelle à l'écart angulaire entre les points indicés  $i$  et  $j$ .

En des termes plus simples, une source secondaire de  $R_N$  appartient à la zone de proximité de la plus proche des sources secondaires de  $R_r$ . La figure IV-11 illustre ce principe.

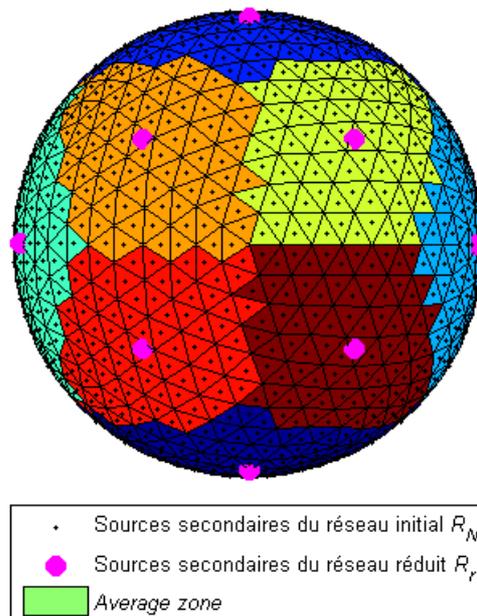


Fig. IV-11 : Illustration de la répartition de la surface en zones de proximité

### 3.1.2 Obtention des CAHF optimisés

Notons  $\text{cahf}_j^N$  le filtre holophonique non-optimisé calculé à partir de l'information ponctuelle estimée au lieu de la source secondaire  $j$  du réseau initial  $R_N$ , et  $\text{cahf}_i^{r,AZ}$  le filtre holophonique associé à la source  $i$  du réseau réduit  $R_r$  et optimisé suivant la méthode AZ. On peut écrire :

$$cahf_i^{r,AZ}(\theta, \varphi, f) = \sum_{j \in E_i} cahf_j^N(\theta, \varphi, f) \quad (\text{IV-3})$$

En définissant la matrice  $C$  de taille  $N \times r$  comme suit :

$$\begin{cases} c_{ij} = 1 \text{ si } j \in E_i \\ c_{ij} = 0 \text{ sinon} \end{cases} \quad (\text{IV-4})$$

où  $c_{ij}$  sont les éléments de  $C$ , on peut reformuler l'expression (IV-3) :

$$cahf_i^{r,AZ}(\theta, \varphi, f) = \sum_{j=1}^N cahf_j^N(\theta, \varphi, f) c_{ij} \quad (\text{IV-5})$$

La  $i^{\text{ème}}$  ligne de la matrice  $C$  peut être vue comme une directivité associée à la  $i^{\text{ème}}$  source secondaire du réseau réduit, qui permet de pondérer l'information captée sur toute la surface hémisphérique.

### 3.2 Méthode par projection

Cette méthode tente, tout comme la méthode AZ, d'optimiser les CAHF pour une disposition de réseau donnée. Cependant, celle-ci ne consiste plus à moyenniser les signaux par zone de proximité, mais s'appuie sur l'étude des PRTF pour exploiter au mieux l'information captée sur toute la surface hémisphérique. Il s'agit d'obtenir des coefficients de pondération définissant le poids de l'information captée par chaque élément de surface du réseau hémisphérique continu (ou à défaut par chaque source secondaire du réseau initial  $R_N$  défini précédemment) dans le calcul de l'information optimisée émise par chaque source secondaire du réseau réduit (soit  $R_r$  en gardant les notations précédentes).

Cette optimisation est par nature liée à la morphologie des différents pavillons, car elle se base sur la projection de l'ensemble des PRTF sur celles associées à la disposition réduite.

#### 3.2.1 Décomposition linéaire des PRTF

Soit  $H$  la matrice  $N \times M$  contenant les spectres complexes des PRTF, mesurées pour  $N$  positions et  $M$  points fréquentiels. Nous pouvons indifféremment décrire  $H$  comme une matrice de  $N$  filtres ou une matrice de  $M$  directivités. Décomposer linéairement les PRTF consiste à identifier la matrice  $C$  des fonctions spatiales, de dimension  $N \times r$ , ainsi que la matrice  $L$  des filtres de reconstruction, de dimension  $r \times M$ , telles que l'on puisse approximer  $H$  par le produit  $CL$ .

Dans notre cas, nous connaissons  $H$ , la base de données de PRTF, et  $L$  les PRTF associés aux  $r$  sources secondaires définies par la configuration de réseau considérée. Il s'agit donc de déterminer la matrice  $C$  minimisant l'erreur  $H - CL$  aux moindres carrés.

### 3.2.2 Projection des PRTF sur un ensemble réduit

Minimiser l'erreur aux moindres carrés revient à imposer une erreur de décomposition orthogonale aux vecteurs de base. Le problème de minimisation peut donc être résolu par simple projection. On obtient alors :

$$C = \frac{HL^T}{G_L^T} \quad (\text{IV-6})$$

où l'opérateur  $T$  désigne la trans-conjugaison (transposition et conjugaison simultanées) et  $G_L$  désigne la matrice de Gram associée à  $L$ , et est définie par :

$$G_L^T = L^T L \quad (\text{IV-7})$$

### 3.2.3 Calcul des CAHF optimisés

Les HRTF peuvent être décomposés en CAHF et en PRTF en appliquant les techniques holophoniques sur le réseau  $R_N$  :

$$hrtf(\theta, \varphi, f) = \sum_{i=1}^N cahf_i^N(\theta, \varphi, f) prt f_i^N(f) \quad (\text{IV-8})$$

Or les PRTF peuvent se décomposer en fonctions spatiales et filtres de reconstruction pour un nombre discret  $M$  de fréquence, où ces derniers sont les fonctions de transfert des sources du réseau  $R_r$  vers l'entrée du canal.

$$prt f_i^N(f_k) \approx \sum_{j=1}^r c_{ij} l_{jk} = \sum_{j=1}^r c_{ij} prt f_j^r(f_k) \quad (\text{IV-9})$$

où  $k = 1 \dots M$ . En utilisant l'égalité (IV-9), extrapolée à toutes les fréquences, dans l'expression (IV-6), on obtient :

$$hrtf(\theta, \varphi, f) \approx \sum_{i=1}^N [cahf_i^N(\theta, \varphi, f) \sum_{j=1}^r c_{ij} prt f_j^r(f)] = \sum_{j=1}^r [\sum_{i=1}^N cahf_i^N(\theta, \varphi, f) c_{ij}] prt f_j^r(f) \quad (\text{IV-10})$$

Par identification, on en déduit que les CAHF liés au réseau  $R_r$  et optimisés suivant la méthode PROJ, s'exprime en fonction des CAHF liés au réseau  $R_N$  par :

$$cahf_j^{r, PROJ}(\theta, \varphi, f) = \sum_{i=1}^N cahf_i^N(\theta, \varphi, f) \cdot c_{ij} \quad (\text{IV-11})$$

De la même manière que pour la méthode AZ, l'information captée sur toute la surface hémisphérique est pondérée par des directivités associées à chaque source secondaires du réseau réduit.

## 4 OPTIMISATION STATISTIQUE DE LA DISPOSITION DU RESEAU

Dans cette section, nous présentons différentes méthodes que nous avons entreprises pour optimiser la disposition même d'un réseau réduit, c'est-à-dire composé d'un petit nombre de sources secondaires. Les techniques que nous avons choisies d'évaluer sont couramment utilisées pour analyser, coder ou compresser les bases de données HRTF [LJGW00]. Ce sont des méthodes statistiques offrant une décomposition linéaire permettant de séparer automatiquement les dépendances spatiales et fréquentielles.

Nous proposons ici d'appliquer ces techniques aux PRTF des sept modèles de pavillon que nous avons calculés. Ces filtres traduisant discrètement le chemin acoustique de la surface d'acquisition holophonique au conduit auditif, nous espérons ainsi trouver un nombre réduit de points sur cette surface, où il devrait être judicieux de positionner les transducteurs. Ces optimisations sont par nature liées à la morphologie des différents pavillons associés aux filtres utilisés.

### 4.1 Description des méthodes de décomposition

#### 4.1.1 Méthodes PCA et ICA

##### a) Description succincte

Les techniques d'analyse statistiques, telles que l'analyse en composantes principales (PCA pour *Principal Components Analysis*) et l'analyse en composantes indépendantes (ICA pour *Independent Component Analysis*), fournissent la représentation compacte d'un ensemble de variables par un ensemble réduit de nouvelles variables. Cette réduction de données est obtenue en éliminant l'information partagée par les variables de départ, et en répartissant l'information résiduelle sur un nombre réduit de variables indépendantes. Les deux analyses diffèrent dans le niveau d'indépendance atteint : la méthode PCA assure une indépendance au deuxième ordre alors que la méthode ICA permet d'atteindre une indépendance aux ordres supérieurs.

PCA et ICA permettent de décomposer une matrice  $H$  contenant les spectres des PRTF en une matrice  $C$  contenant des fonctions spatiales et une matrice  $L$  contenant des filtres de reconstruction, comme décrit en IV.3.2.1. L'idée est de choisir les lieux des maxima des fonctions spatiales, issues de l'une ou l'autre des analyses, pour définir une nouvelle disposition de réseau optimisée.

##### b) Mesure de l'indépendance : relation entre PCA et ICA

Soient  $X_i$  ( $i = 1 \dots M$ )  $M$  variables centrées dont on possède  $N$  observation. Soient  $p_{X_i}$  la densité de probabilité de la variable  $X_i$  et  $p_X$  la densité de probabilité conjointe de la variable  $X = [X_1 X_2 \dots X_M]$ . On peut voir chaque variable  $X_i$  comme un vecteur colonne de dimension  $N$  et  $X$  comme une matrice de dimension  $N \times M$ . Les variables  $X_i$  sont indépendantes si et seulement si :

$$p_X = \prod_{i=1}^M p_{X_i} \quad (\text{IV-12})$$

PCA et ICA sont des démarches d'optimisation, qui requièrent le choix d'une mesure de l'indépendance des variables, qu'il s'agit de maximiser. Comon [Com94] définit les propriétés d'une telle mesure, qu'il baptise « fonction de contraste ». Une fonction de contraste  $\phi$  est une application de l'espace des densités de probabilités  $p_X$  vers l'ensemble des réels, satisfaisant trois conditions :

- $\phi(p_X)$  est invariante par permutation circulaire des variables  $X_i$ . Pour toute permutation  $A$ , on a :

$$\phi(p_{AX}) = \phi(p_X) \quad (\text{IV-13})$$

- $\phi(p_X)$  est invariante par changement d'échelle. Pour toute matrice diagonale inversible  $A$ , on a :

$$\phi(p_{AX}) = \phi(p_X) \quad (\text{IV-14})$$

- Si  $X$  possède des composantes indépendantes, alors pour toute matrice  $A$  inversible, on a :

$$\phi(p_{AX}) \leq \phi(p_X) \quad (\text{IV-15})$$

Comon définit une telle fonction de contraste à partir de la notion d'information mutuelle. La définition d'indépendance peut être reformulée en exprimant la nullité de la distance entre les deux termes de l'égalité. La divergence de Kullback-Leibler offre une définition de la distance entre densités de probabilités :

$$D(p_X, p_Y) = \int p_X(u) \log \frac{p_X(u)}{p_Y(u)} du \quad (\text{IV-16})$$

L'information mutuelle moyenne de  $X$  est définie à partir de cette distance :

$$I(p_X) = D(p_X, \prod p_{X_i}) = \int p_X(u) \log \frac{p_X(u)}{\prod p_{X_i}} du \quad (\text{IV-17})$$

Celle-ci est nulle si et seulement si les variables sont indépendantes, et strictement positive sinon. On peut ainsi définir la fonction de contraste suivante :

$$\phi(p_X) = -I(p_X) \quad (\text{IV-18})$$

L'information mutuelle peut être décomposée en plusieurs contributions :

$$I(p_X) = J(p_X) - \sum_{i=1}^M J(p_{X_i}) + I(g_X) \quad (\text{IV-19})$$

Cette équation fait intervenir la notion de négentropie  $J(p_X)$  que l'on peut définir comme une mesure de la proximité de la densité de probabilité  $p_X$  à une distribution gaussienne, au sens de Kullback.  $g_X$  représente la densité de probabilité gaussienne de mêmes moyenne et variance que  $p_X$ .

La minimisation de l'information mutuelle  $I(p_X)$  peut alors être obtenue en une ou deux étapes, suivant que l'on réalise une ACP ou une ICA :

- tout d'abord, il s'agit de trouver une transformation linéaire  $A$  telle que :

$$I(g_{AX}) = 0 \quad (\text{IV-20})$$

- puis il convient de trouver une transformation linéaire  $B$  telle que :

$$I(g_{BAX}) = 0 \quad (\text{IV-21})$$

$$J(p_{BAX}) \text{ est constante} \quad (\text{IV-22})$$

$$\sum J(p_{BAX_i}) \text{ est maximum} \quad (\text{IV-23})$$

La première étape est effectuée si et seulement si les nouvelles variables  $Y_i$ , définies par  $Y = AX$ , sont décorréelées, c'est ce que réalise l'ACP. En pratique, on pourra avoir recours à une décomposition en valeurs singulières (SVD pour *Singular Value Decomposition*), que nous présenterons ci-dessous. Dans son principe, l'ICA n'impose pas cette étape préalable. Cependant, cette contrainte de décorrélation est souvent retenue car elle permet de simplifier l'expression des fonctions de contraste. De plus, si on impose la recherche d'une transformation orthogonale pour  $B$ , alors les relations (IV-21) et (IV-22) sont vérifiées. Minimiser l'information mutuelle revient donc à maximiser la fonction de contraste définie comme suit :

$$\phi = \sum_{i=1}^M J(p_{BAX_i}) \quad (\text{IV-24})$$

Il est rare d'avoir accès à l'expression analytique de la densité de probabilité  $p_{BAX_i}$ . Pour réaliser l'ICA en pratique, nous utiliserons l'algorithme JADE (*Joint Approximate*

*Diagonalization of Eigenmatrices*) mis au point par Cardoso [CS93, Jade] qui utilise une approximation de la densité de probabilité d'une matrice à partir de ses cumulants d'ordre 4. Nous ne développerons pas plus avant l'approche de Cardoso.

**c) Mise en œuvre pratique**

Dans un premier tant, il s'agit donc d'effectuer une SVD sur la matrice  $H$  contenant les PRTF. La SVD permet de séparer une matrice en plusieurs composantes orthogonales (autrement dit décorréelées du point de vue statistique). La matrice  $H$  de taille  $N \times M$  se décompose ainsi :

$$H = U\Sigma V^t \quad (\text{IV-25})$$

où  $\Sigma$  est une matrice diagonale de taille  $N \times M$  à coefficients réels positifs classés par ordre décroissant, et où  $U$  et  $V$  sont deux matrices orthogonales de taille respective  $N \times N$  et  $M \times M$ . Les colonnes  $U_i$  de  $U$  représentent des caractéristiques spatiales de  $H$  décorréelées deux à deux, et les colonnes  $V_i$  de  $V$  des caractéristiques fréquentielles de  $H$  également décorréelées deux à deux. En notant également  $\sigma_i$  le  $i^{\text{ème}}$  élément de la diagonale de  $\Sigma$ , c'est-à-dire la  $i^{\text{ème}}$  valeur singulière, on peut exprimer  $H$  de la manière suivante :

$$H = \sum_{i=1}^{\min(M,N)} \sigma_i U_i V_i^t \quad (\text{IV-26})$$

Chaque caractéristique spatiale  $U_i$  correspond donc à une caractéristique fréquentielle  $V_i$  et à une énergie  $\sigma_i^2$ . La classification des éléments de la diagonale de  $\Sigma$  par ordre décroissant permet de facilement sélectionner un nombre  $r$  ( $r \leq \min(N,M)$ ) de ces caractéristiques représentant de façon optimale la matrice initiale  $H$  au sens des moindres carrés et contenant une certaine partie de l'énergie de celle-ci. On peut donc soit fixer  $r$  et évaluer la part d'énergie représentée, ou trouver  $r$  qui permet d'atteindre une certaine part d'énergie. Notons  $\Sigma_r$  la matrice carrée diagonale contenant les  $r$  premières lignes et colonnes de  $\Sigma$ ,  $U_r$  et  $V_r$  les matrices contenant les  $r$  premières colonnes de  $U$  et  $V$  respectivement, et  $e_r$  la part d'énergie contenue dans les  $r$  premières valeurs singulières. On peut écrire les relations suivantes :

$$H \approx U_r \Sigma_r V_r^t \quad (\text{IV-27})$$

$$e_r = \frac{\sum_{i=1}^r \sigma_i^2}{\sum_{i=1}^{\min(N,M)} \sigma_i^2} \quad (\text{IV-28})$$

La matrice  $C$  des fonctions spatiales et la matrice  $L$  des filtres de reconstruction sont définies comme suit :

$$\begin{cases} C = U_r \\ L = \Sigma_r V_r^t = U_r^t H \end{cases} \quad (\text{IV-29})$$

La transformation orthogonale  $A$  représentative de la PCA et décrite en section précédente n'est autre que  $C^t$ .

La maximisation du contraste peut alors être effectuée sur les matrices  $L$  ou  $C$  données par la SVD. Ceci conduit à une nouvelle matrice  $L'$  et une nouvelle matrice  $C'$  contenant respectivement les  $r$  filtres de reconstruction indépendants et les  $r$  fonctions spatiales indépendantes, ainsi qu'à la transformation orthogonale  $B$  décrite en section précédente :

$$\begin{cases} L' = BL \\ C' = CB^t \end{cases} \quad (\text{IV-30})$$

Les fonctions spatiales  $C$  et  $C'$  sont assimilables à des directivités possédant chacune un lobe principale plus ou moins grand. Les directions pointées par ces lobes désignent les incidences d'où provient l'information acoustique importante (par rapport aux PRTF sur lesquelles les analyses ont été conduites). Il apparaît donc judicieux de positionner les  $r$  sources secondaires du réseau réduit  $R_r$  dans ces directions.

#### 4.1.2 Méthode SUBSET

L'optimisation SUBSET ou *subset selection* (sélection de sous-groupe) permet de déterminer une sous-matrice de  $H$ , composé de  $r$  PRTF, offrant la meilleure reconstruction de  $H$  au sens des moindres carrés. Elle est réalisée en deux étapes. La première consiste à effectuer une SVD sur la matrice  $H$  des PRTF. La deuxième consiste à décomposer les fonctions spatiales  $C$  ainsi obtenues par une décomposition QR. Cette décomposition conduit à une matrice unitaire  $Q$ , une matrice triangulaire  $R$  et une matrice de permutation  $E$ , liées par l'expression suivante :

$$HE = QR \quad (\text{IV-31})$$

La matrice de permutation  $E$  est utilisée pour ordonner les lignes de  $C$  (*i.e.* les directions sur lesquelles les fonctions spatiales sont décrites) selon leur norme. Cet ordonnancement permet de sélectionner un jeu réduit de  $r$  directions qui définit un positionnement optimisé des sources secondaires du réseau réduit  $R_r$ .

#### 4.2 Application à la base de données calculée de PRTF

Si l'on voit les PRTF comme des directivités, on peut dire que les optimisations ont pour but de trouver les incidences prédominantes d'excitation des pavillons. Elles permettent donc d'associer  $r$  directions importantes à chaque pavillon. Cependant, l'objectif étant de concevoir un système holophonique réel, nous avons besoin

d'effectuer ces optimisations sur un panel de plusieurs PRTF, en espérant ainsi trouver des directions prédominantes dite « moyennes », c'est-à-dire représentatives de la morphologie générale des pavillons d'oreille humaine et non d'un pavillon en particulier.

Il existe au moins deux manières pour évaluer ces incidences moyennes à partir d'une base de données de plusieurs PRTF : soit les optimisations sont conduites sur chaque jeu de PRTF indépendamment, et les incidences sont moyennées après coup ; soit les PRTF sont concaténées selon l'axe des fréquences pour former un ensemble de directivités observées plusieurs fois par fréquence, et les optimisations sont effectuées sur cette matrice, ce qui donne directement accès aux  $r$  directions prédominantes « moyennes ». Nous choisissons d'utiliser la deuxième méthode sur la base de données composée des sept jeux de PRTF que nous avons estimés par calcul BEM.

Les trois optimisations que nous proposons commencent toutes par une décomposition en valeurs singulières. Celle-ci nous permet d'évaluer le pourcentage de variance relatif à un ordre de décomposition donné. Pour des raisons de mise en œuvre technique de l'holophonie binaurale sur un casque à plusieurs transducteurs, nous désirons trouver une configuration de réseau optimisée pour huit sources secondaires. Or un ordre de décomposition de huit apparaît suffisant pour conserver près de 99% de la variance (voir Fig. IV-12), ce qui nous conforte dans notre choix de taille de réseau.

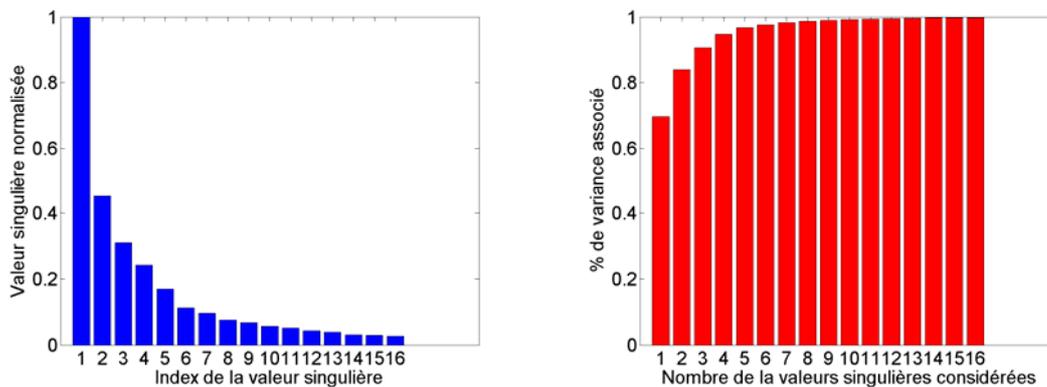
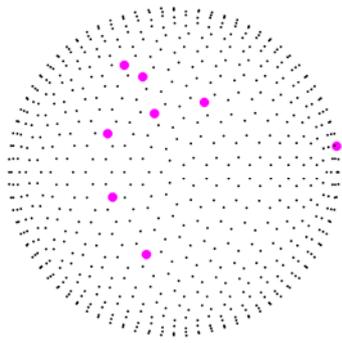


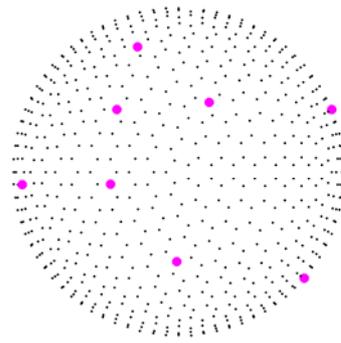
Fig. IV-12 : Valeurs singulières issues de la SVD et pourcentage de variance associé

La figure IV-13 présente les trois dispositions de réseau optimisées issues des différentes optimisations (PCA, ICA et SUBSET).

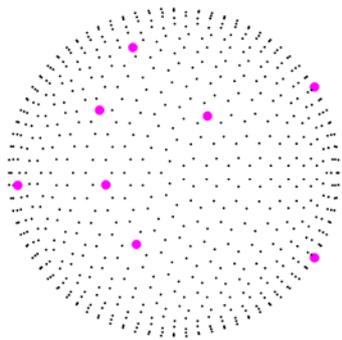
Les figures IV-15 et IV-16 illustrent respectivement les fonctions spatiales issues de la PCA et de l'ICA (un repère de lecture est donné en figure IV-14) . On remarque que les fonctions spatiales issues de l'ICA sont plus directives que celles issues de la PCA, dans le sens où elles possèdent généralement un lobe primaire très marqué et peu étendu.



Disposition optimisée  
issue de la PCA

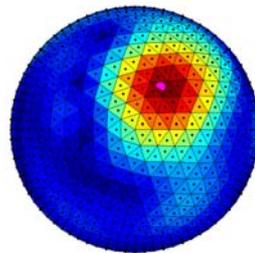


Disposition optimisée  
issue de l'ICA



Disposition optimisée  
issue de la méthode SUBSET

*Fig. IV-13 : Dispositions optimisées pour un réseau réduit à 8 sources secondaires*



*Fig. IV-14 : Repère de lecture des figures liées aux fonctions spatiales et de pondération*

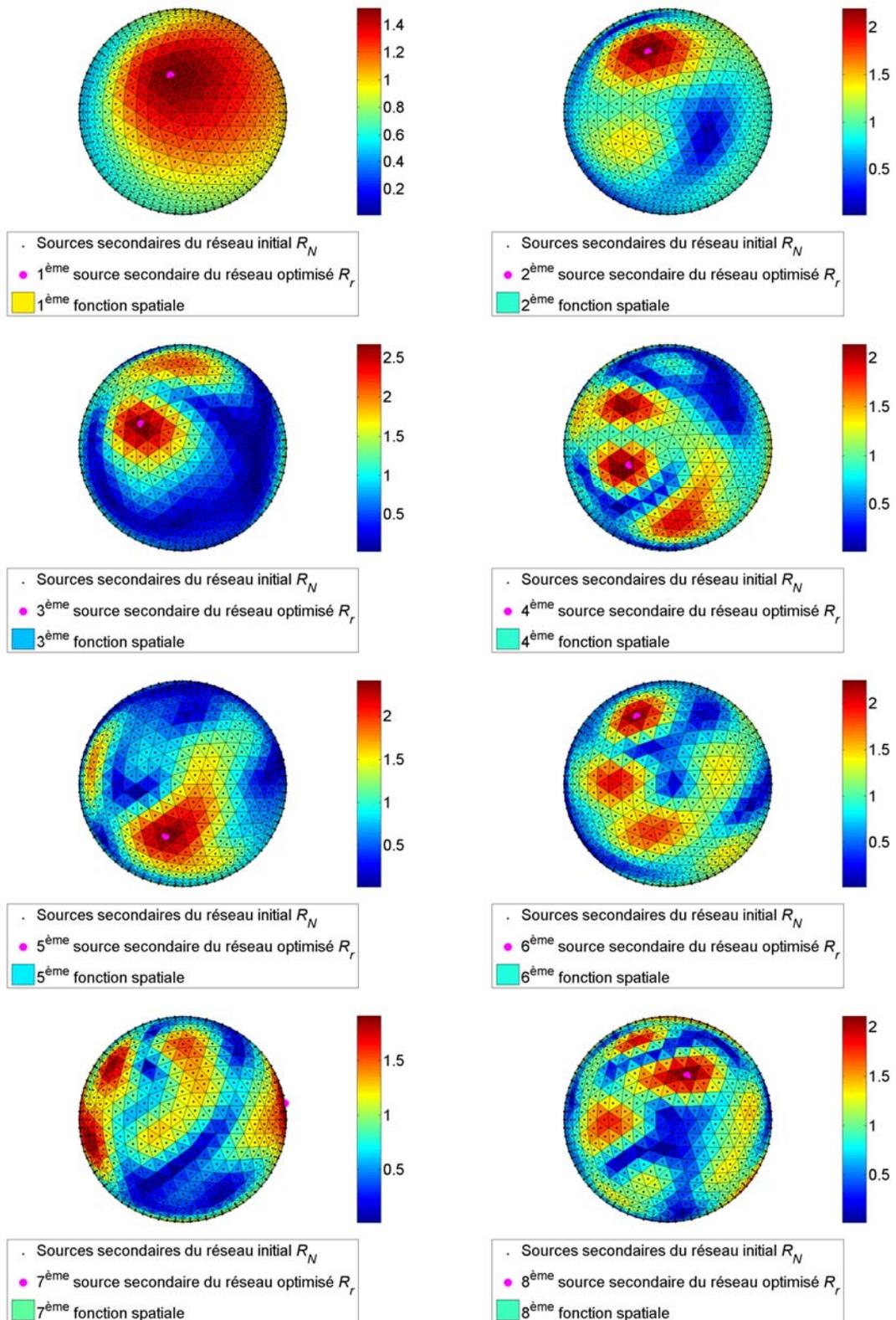


Fig. IV-15 : Fonctions spatiales issues de la PCA

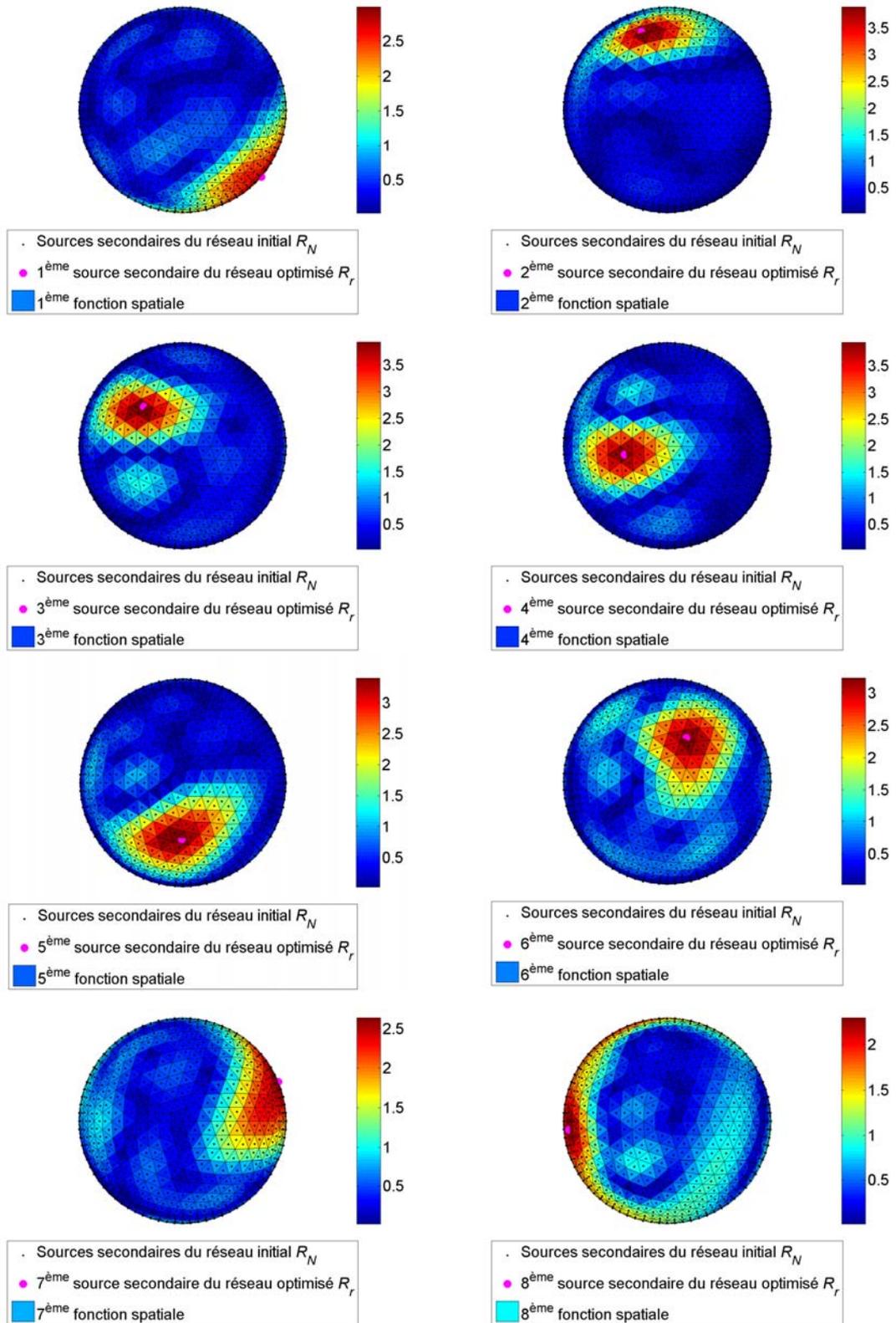


Fig. IV-16 : Fonctions spatiales issues de l'ICA

## 5 OPTIMISATION HEURISTIQUE DE LA DISPOSITION DU RESEAU

Les optimisations présentées dans la section précédente sont basées sur une analyse statistique d'un nombre réduit de jeux de PRTF. Cependant, il se peut que ces dernières ne soient pas représentatives de celles d'un auditeur donné, car dépendantes de la forme et de l'orientation des pavillons d'oreille. Si effectivement ce problème existe, ce type d'optimisation nécessiterait d'être effectué sur différents panels de pavillons, regroupés suivant des critères de proximité morphologique, pour ainsi adapter la configuration des transducteurs du casque à un auditeur : l'holophonie binaurale appliquée de cette manière perdrait tout son sens, car le problème d'individualisation des indices spectraux liés aux pavillons serait remplacé par un problème d'individualisation de la configuration du casque.

Il peut donc être intéressant d'essayer d'optimiser le positionnement des sources secondaires sur des considérations heuristiques simples. On peut par exemple tenter de répondre à deux questions.

- Pour quelles positions de source primaire les pavillons jouent-ils un rôle prédominant dans la tâche de localisation ?
- Comment la radiation d'une source est-elle simulée par les réseaux lorsque celle-ci se trouve dans une de ces positions ?

### 5.1 De l'importance du plan médian

Au premier chapitre (voir I.2.3.7), nous avons vu que les informations spectrales perceptivement prédominantes pour localiser des sources dans le plan médian sont situées en hautes fréquences et liées aux modifications spectrales engendrées par la diffraction de l'onde sonore sur les pavillons d'oreille, et que les informations spectrales perceptivement prédominantes pour localiser des sources dans le plan frontal sont situées en basses fréquences, en dessous de 3kHz, et associés au phénomène de diffraction par le corps et la tête. On peut présumer que les pavillons ont une influence essentielle lors de la tâche de localisation dans le plan médian.

D'autre part, les figures de diffraction représentant les HRTF des pavillons bafflés que nous avons estimées témoignent d'un schéma récurrent sur chaque plan sagittal : les variations spectrales induites par une variation de l'angle polaire sont plus importantes que les variations spectrales induites par une variation de l'angle latéral (voir Fig. III-8 par exemple). Cette constatation est concordante avec les travaux de Morimoto *et al.* [MII03]. Ces derniers ont évalué la précision avec laquelle trois auditeurs pouvaient localiser des sources large bande spatialisées en utilisant des HRTF construites à partir des indices spectraux mesurés uniquement dans le plan médian et des ITD et ILD indépendantes de la fréquence. Leurs résultats ont montré que, pour chaque sujet, les schémas d'erreur de localisation suivant l'angle polaire étaient similaires dans chaque plan sagittal testé (angle latéral de 0°, 30° et 60°).

Au vu de ces différents constats, il semblerait qu'il soit important de se focaliser sur la reconstruction des HRTF pour des sources primaires localisées dans le plan médian.

## 5.2 Vers une distribution circulaire

L'information captée à la surface de l'hémisphère ipsilatéral par les techniques holophoniques est quasiment toujours maximale dans la direction de la source primaire. Lorsque cette dernière est située dans le plan médian, la majorité de l'information est captée à la base de l'hémisphère (voir Fig. V-17). On peut donc envisager de répartir uniformément les sources secondaires à la base de l'hémisphère, et réduisant ainsi la géométrie du réseau à un cercle.

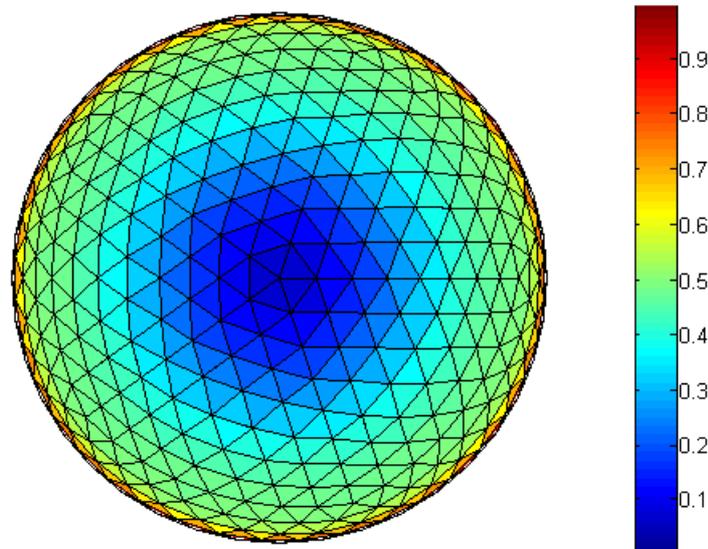


Fig. IV-17 : Activité du réseau pour des sources primaires situées dans le plan médian - 8kHz

## 6 EVALUATION DES DIFFERENTES OPTIMISATIONS

Considérons les cinq dispositions de réseau à huit sources secondaires que nous avons élaborées (voir Fig. IV-18) :

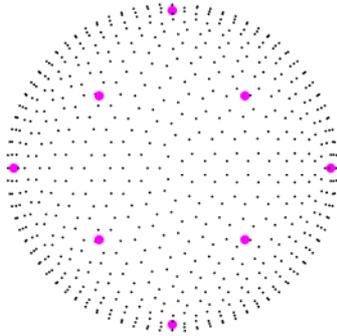
- EQUI : les sources secondaires sont uniformément réparties sur la surface hémisphérique,
- PCA : disposition issue de l'optimisation PCA,
- ICA : disposition issue de l'optimisation ICA,
- SUBSET : disposition issues de l'optimisation SUBSET,
- CIRC : les sources secondaires sont uniformément réparties sur le cercle de base de l'hémisphère.

Considérons les trois types de méthode d'obtention des CAHF :

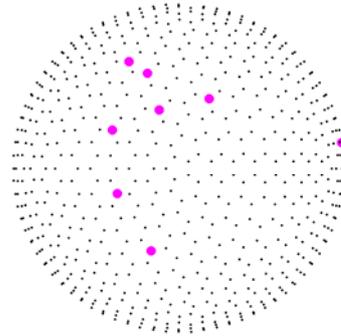
- la méthode holophonique classique,

- la méthode optimisée AZ,
- la méthode optimisée PROJ.

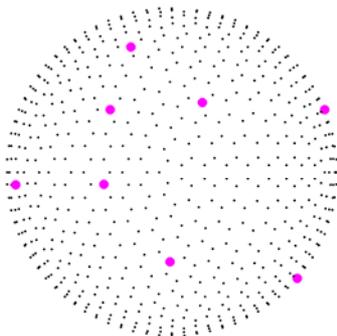
Nous allons évaluer l'efficacité des différentes configurations, c'est-à-dire des différentes dispositions couplées aux différentes méthodes d'alimentation.



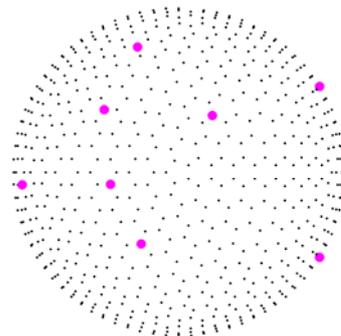
Disposition EQUI



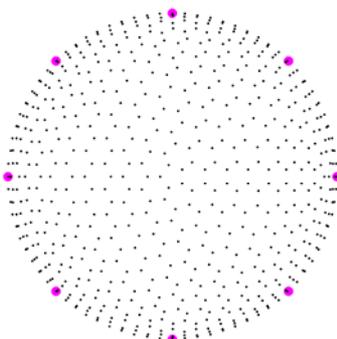
Disposition PCA



Disposition ICA



Disposition SUBSET



Disposition CIRC

*Fig. IV-18 : Récapitulatif des dispositions évaluées*

Les figures présentées en annexe de ce chapitre (voir le complément de figures) présentent toutes les fonctions de pondérations obtenues pour les différentes dispositions de sources secondaires et issues des optimisations AZ et PROJ.

On remarque que les fonctions (ou directivités) de pondération issues de l'optimisation PROJ varient fortement d'une configuration à l'autre. En effet, les directivités des configurations ICA et SUBSET ont généralement un lobe principal très marqué et dirigé vers la source secondaire associée (voir Fig. IV-19, partie de gauche), alors que celles des autres dispositions vont souvent être plus larges et peuvent même ne pas pointer vers la source secondaire associée, comme dans le cas de la disposition CIRC (voir Fig. IV-19, partie de droite).

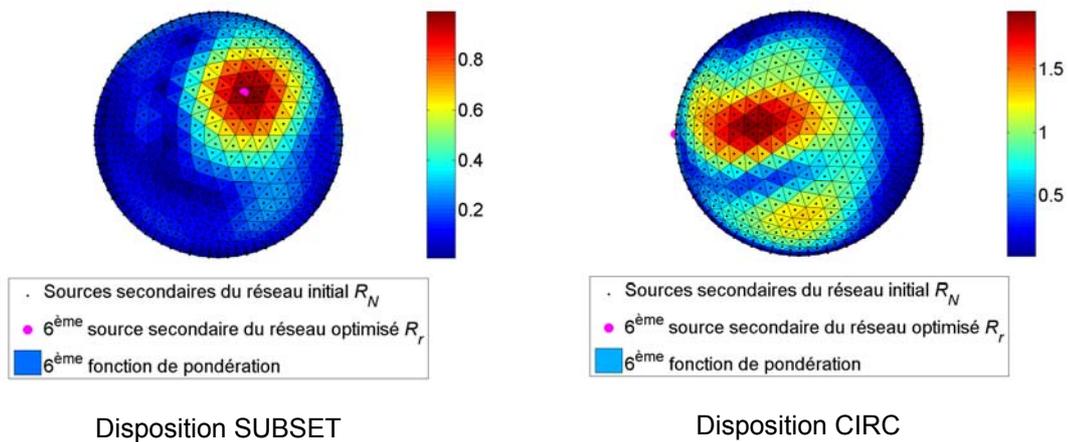


Fig. IV-19 : Exemple de fonctions de pondération issues de l'optimisation PROJ

La figure IV-20 présente les indices de corrélation et distance moyennés sur l'ensemble des reconstructions simulées, à savoir les reconstructions des HRTF des six pavillons bafflés et celle des HRTF de la tête KU-100. Les indices montrent clairement que les optimisations relatives aux CAHF ont un impact très marqué par rapport aux optimisations relatives aux positionnements des sources secondaires. L'optimisation AZ permet de passer de coefficients de corrélation inférieurs à 0.3 à des coefficients supérieurs à 0.6, et l'optimisation PROJ permet même de dépasser les 0.8 en moyenne. Une tendance équivalente est observable sur les indices de distance.

Les meilleurs résultats sont obtenus sur les configurations ICA PROJ et SUBSET PROJ, avec un coefficient de corrélation de l'ordre de 0.9 et une distance de l'ordre de 2dB : ces résultats sont pratiquement comparables à ceux fournis par l'application des méthodes holophoniques dites « classiques » sur le réseau à 640 sources secondaires.

Si l'on se focalise sur la reconstruction des HRTF pour des sources primaires localisées dans le plan médian, on s'aperçoit que la disposition CIRC fournit les meilleurs résultats, et ce quelque soit la méthode de calcul des CAHF employées (voir Fig. IV-21). En utilisant une captation ponctuelle classique, cette disposition permet d'atteindre un indice de corrélation de 0.4 dans le plan médian, alors que les autres ne dépassent pas 0.2. En couplant cette disposition à un calcul des CAHF optimisé AZ, le coefficient de corrélation avoisine les 0.8 alors que les autres dispositions ne permettent pas d'aller au-delà de 0.6. Les différences entre les différentes dispositions sont moins flagrantes lorsqu'elles sont couplées à la méthode d'optimisation PROJ, mais la distribution CIRC reste légèrement meilleure.

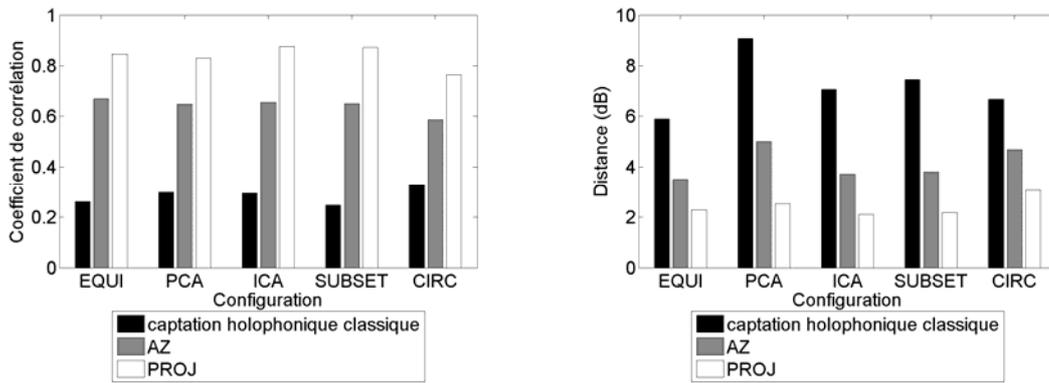


Fig. IV-20 : Evaluations des différentes configurations optimisées

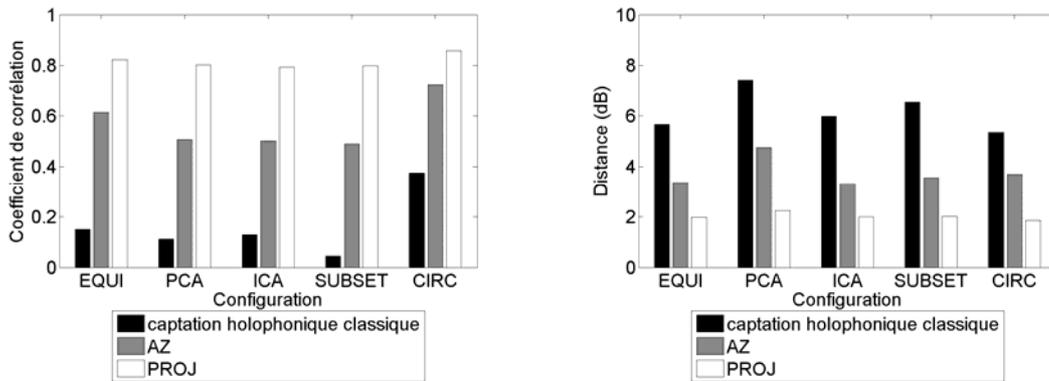


Fig. IV-21 : Evaluations des différentes configurations optimisées dans le plan médian

## 7 CONCLUSION

Les simulations de reconstruction des HRTF conduites à l'aide des différentes optimisations fournissent des résultats concluant. Hormis la distribution CIRC, l'optimisation de la position des sources secondaires semble avoir un impact mineur sur la qualité de reconstruction. Les méthodes d'optimisation des CAHF, quant à elles, influent dans une plus grande mesure sur la qualité de reconstruction. La méthode AZ fournit des résultats bien meilleurs que la méthode de captation ponctuelle, et n'étant pas liée à une morphologie de pavillon donné, elle pourra être utilisée par défaut. Bien que la méthode PROJ fournissent les meilleurs résultats quelque soit la répartition des sources secondaires dans le réseau, il se pourrait qu'elle ne soit pertinente que couplée à certaines configurations. En effet, pour les configurations EQUI, PCA et CIRC, les directivités de pondération peuvent ne pas être dirigées vers la source secondaires associée ou posséder plusieurs lobes principaux, ce qui est philosophiquement contradictoire avec les principes holophoniques ; il est alors probable qu'un léger décalage du réseau par rapport aux pavillons dégrade fortement la reconstruction.

## **COMPLEMENT DE FIGURES**

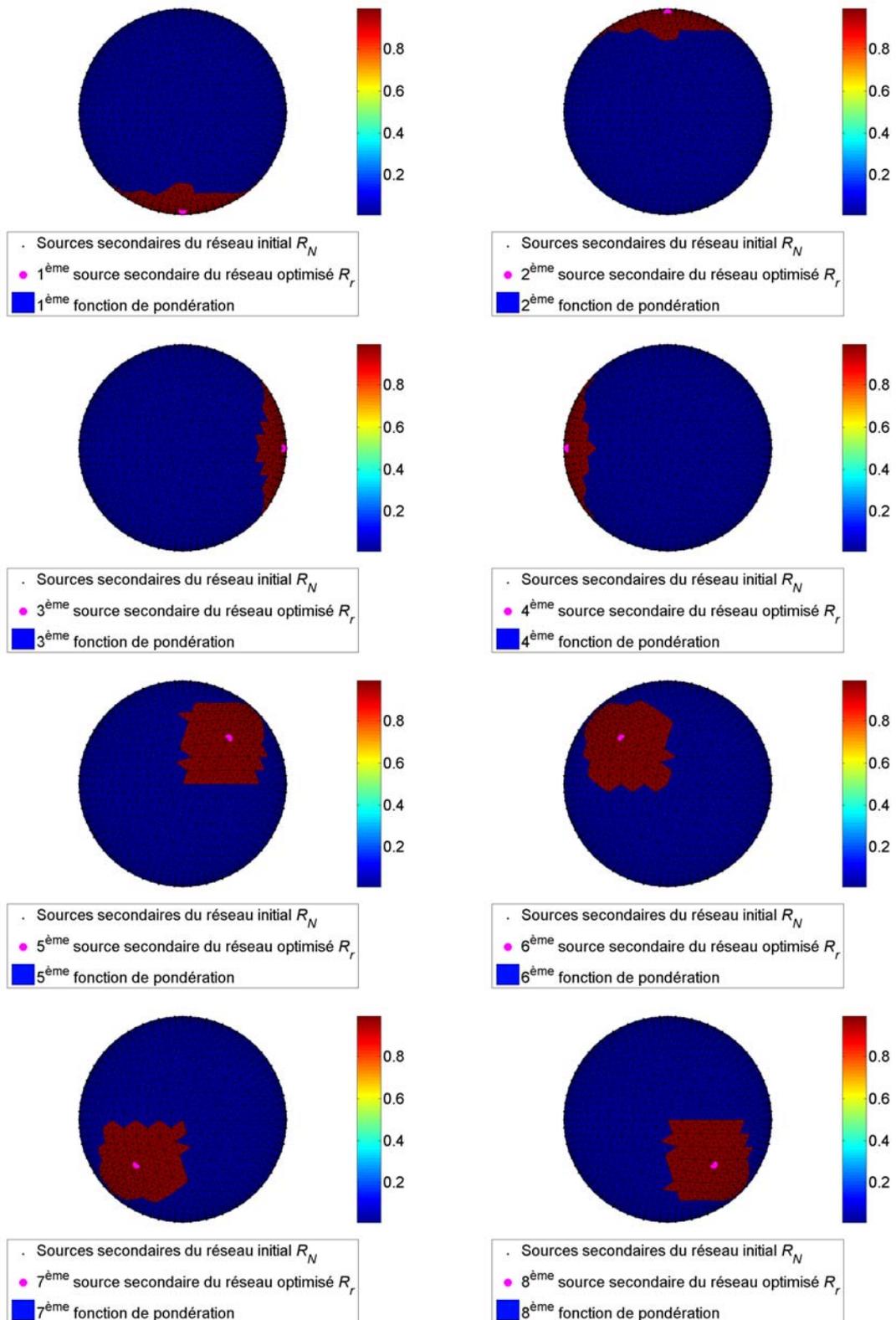


Fig. IV-22 : Fonctions de pondération - configuration EQUI - optimisation AZ

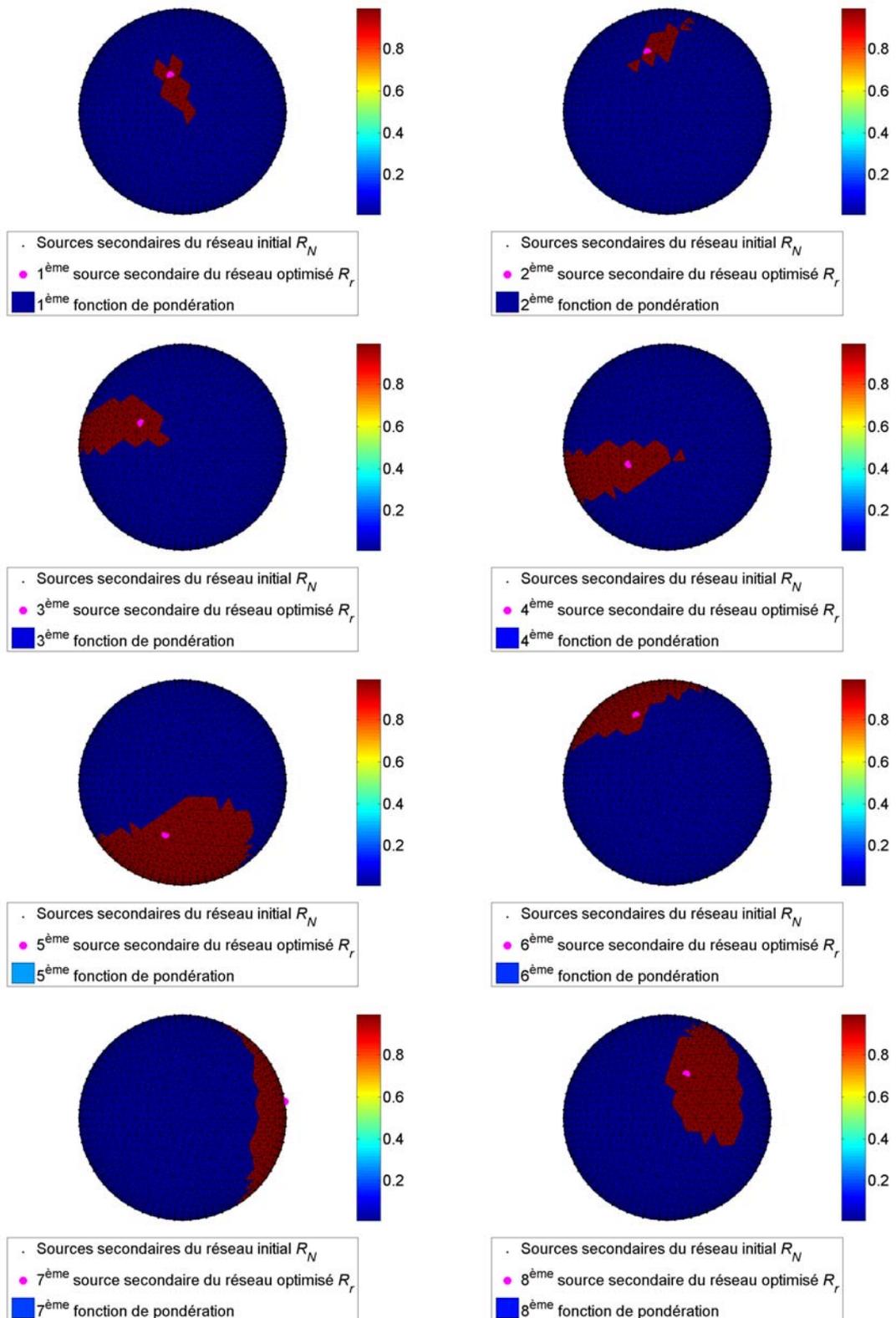


Fig. IV-23 : Fonctions de pondération - configuration PCA - optimisation AZ

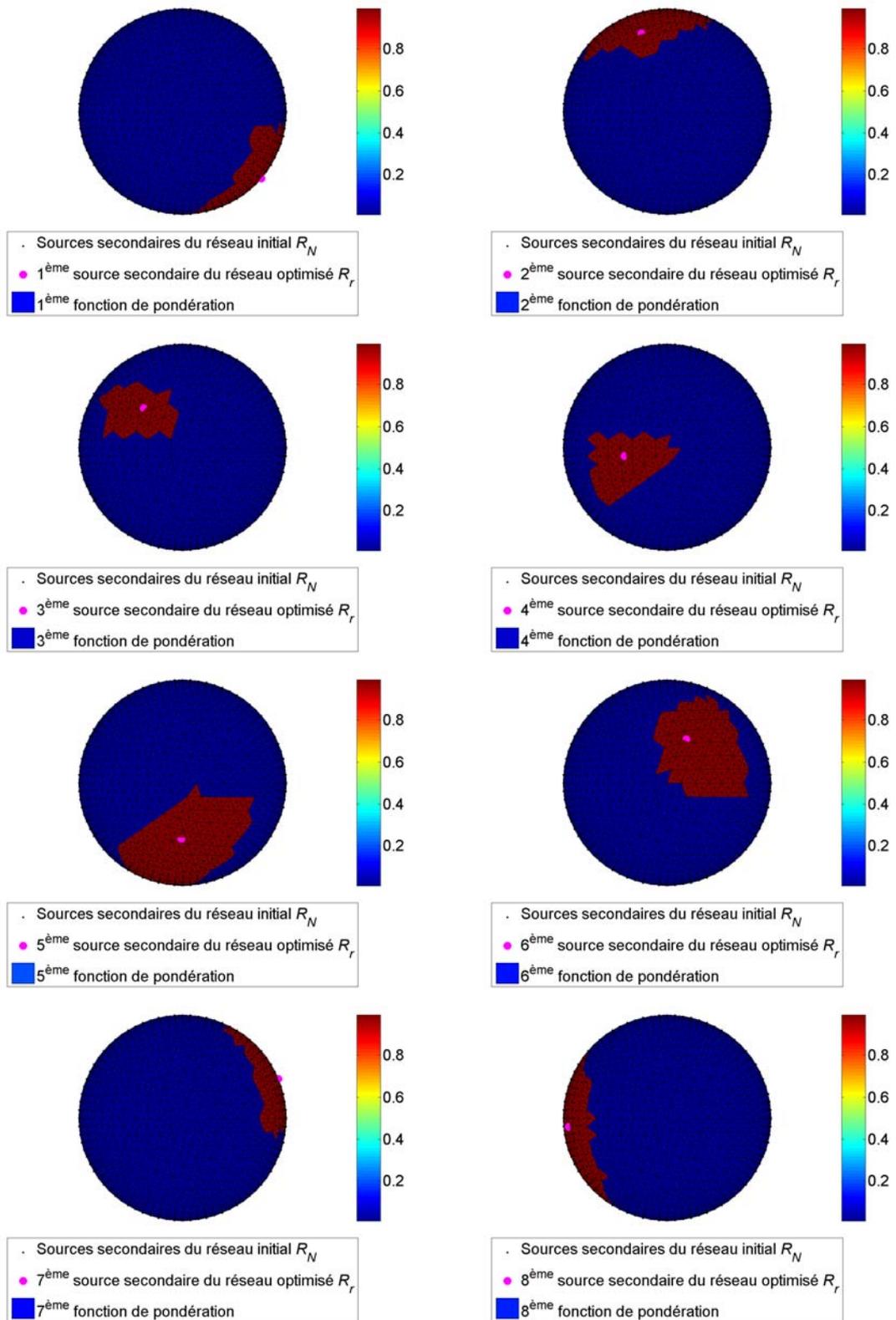


Fig. IV-24 : Fonctions de pondération - configuration ICA - optimisation AZ

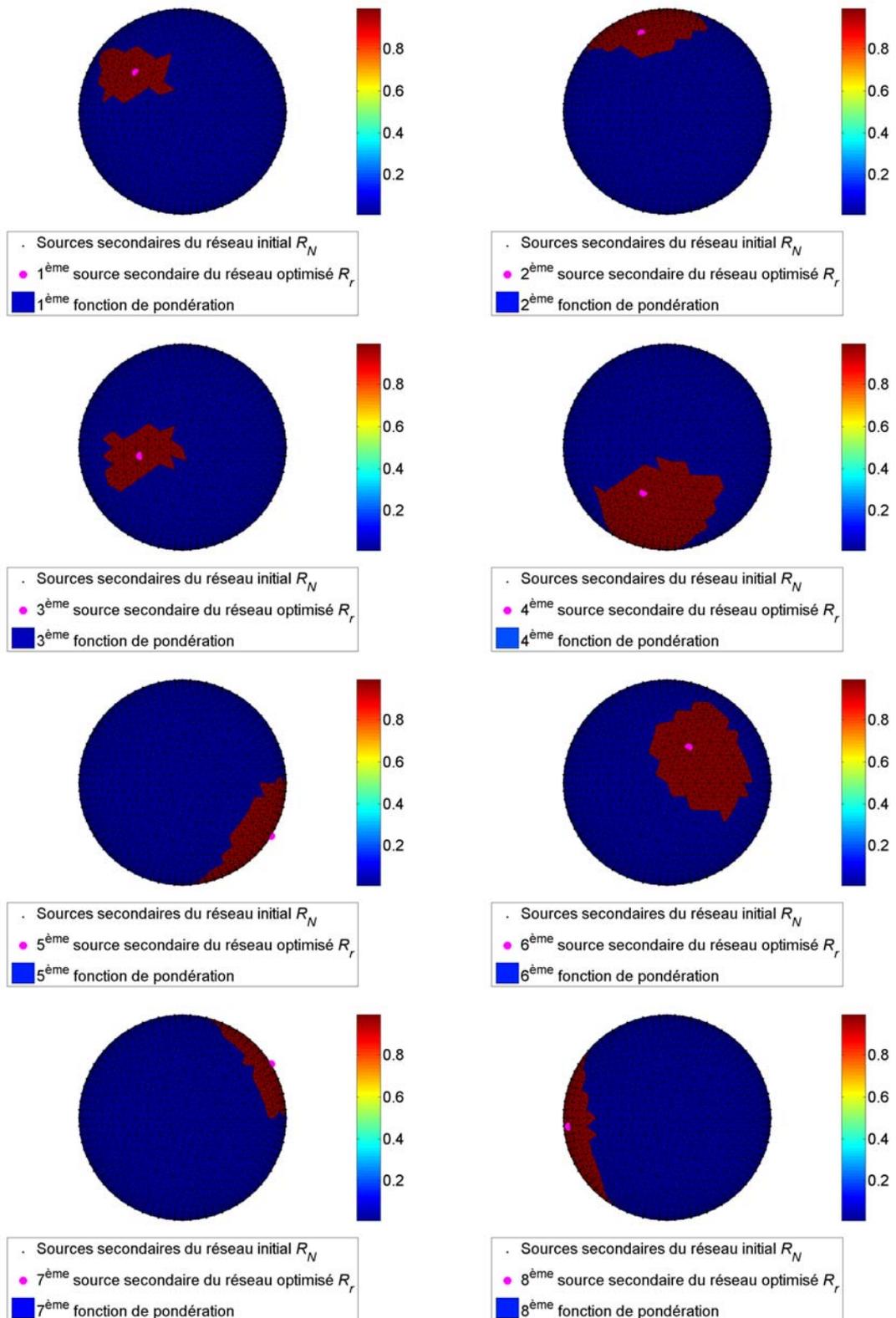


Fig. IV-25 : Fonctions de pondération - configuration SUBSET - optimisation AZ

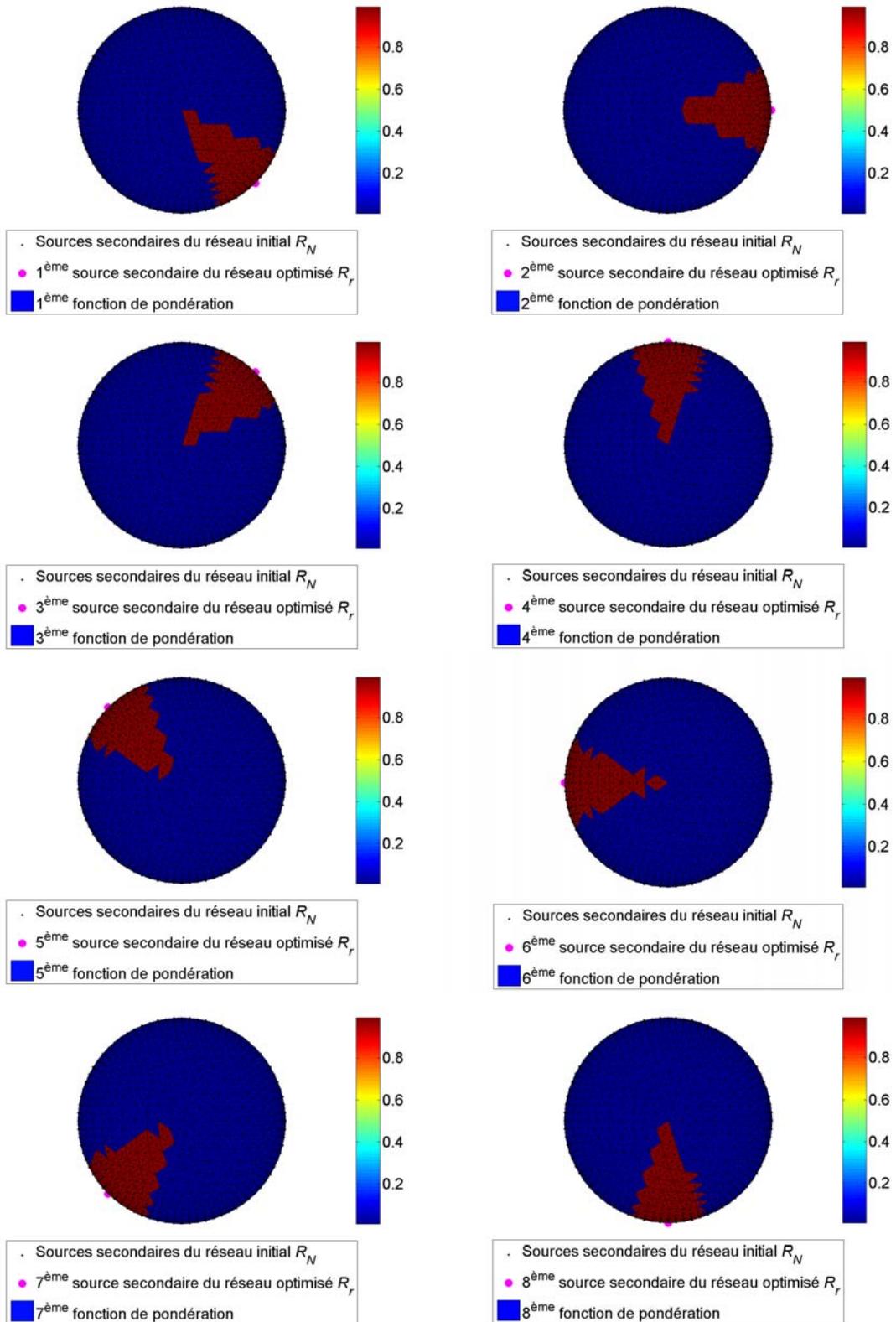


Fig. IV-26 : Fonctions de pondération - configuration CIRC - optimisation AZ

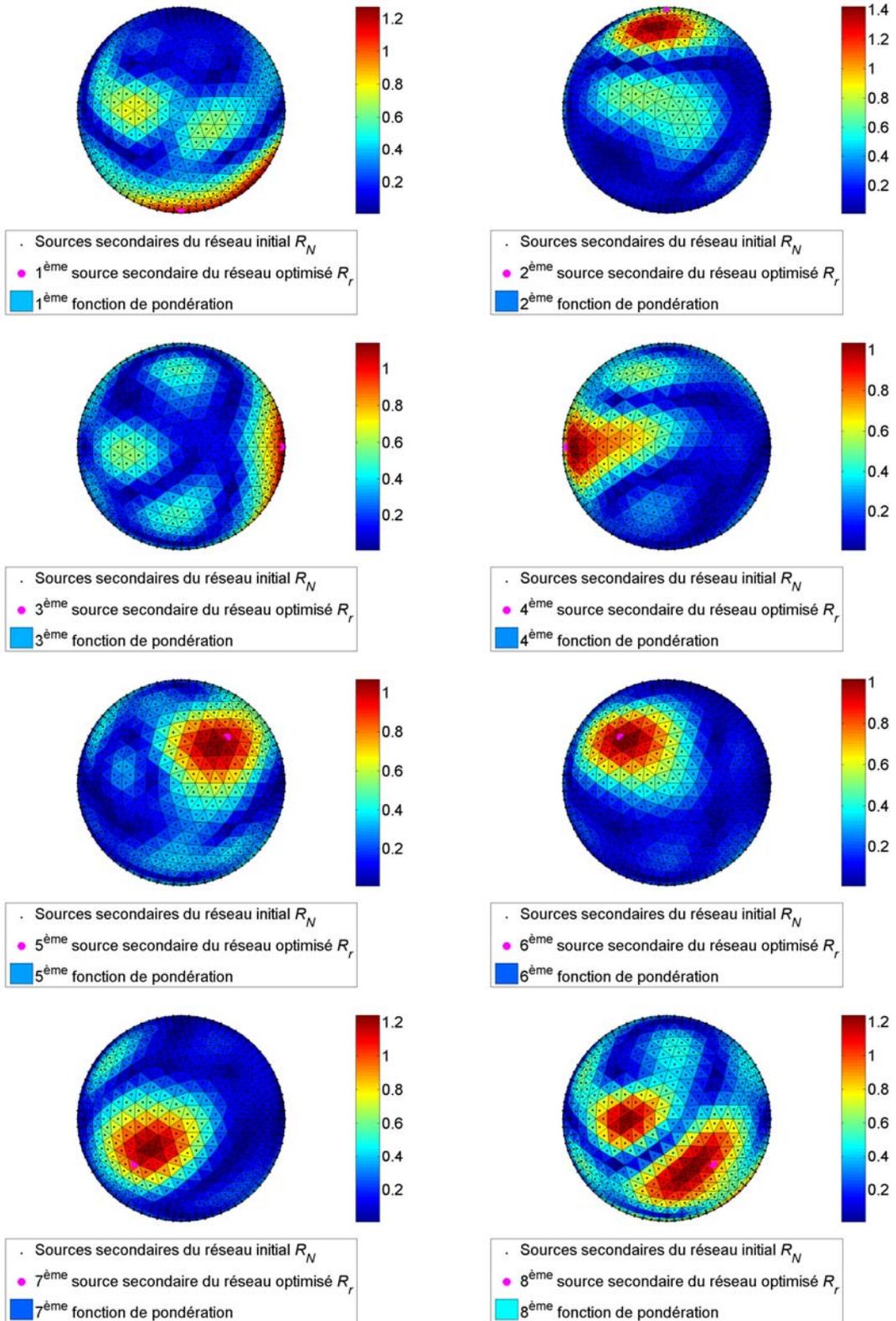


Fig. IV-27 : Fonctions de pondération - configuration EQUI - optimisation PROJ

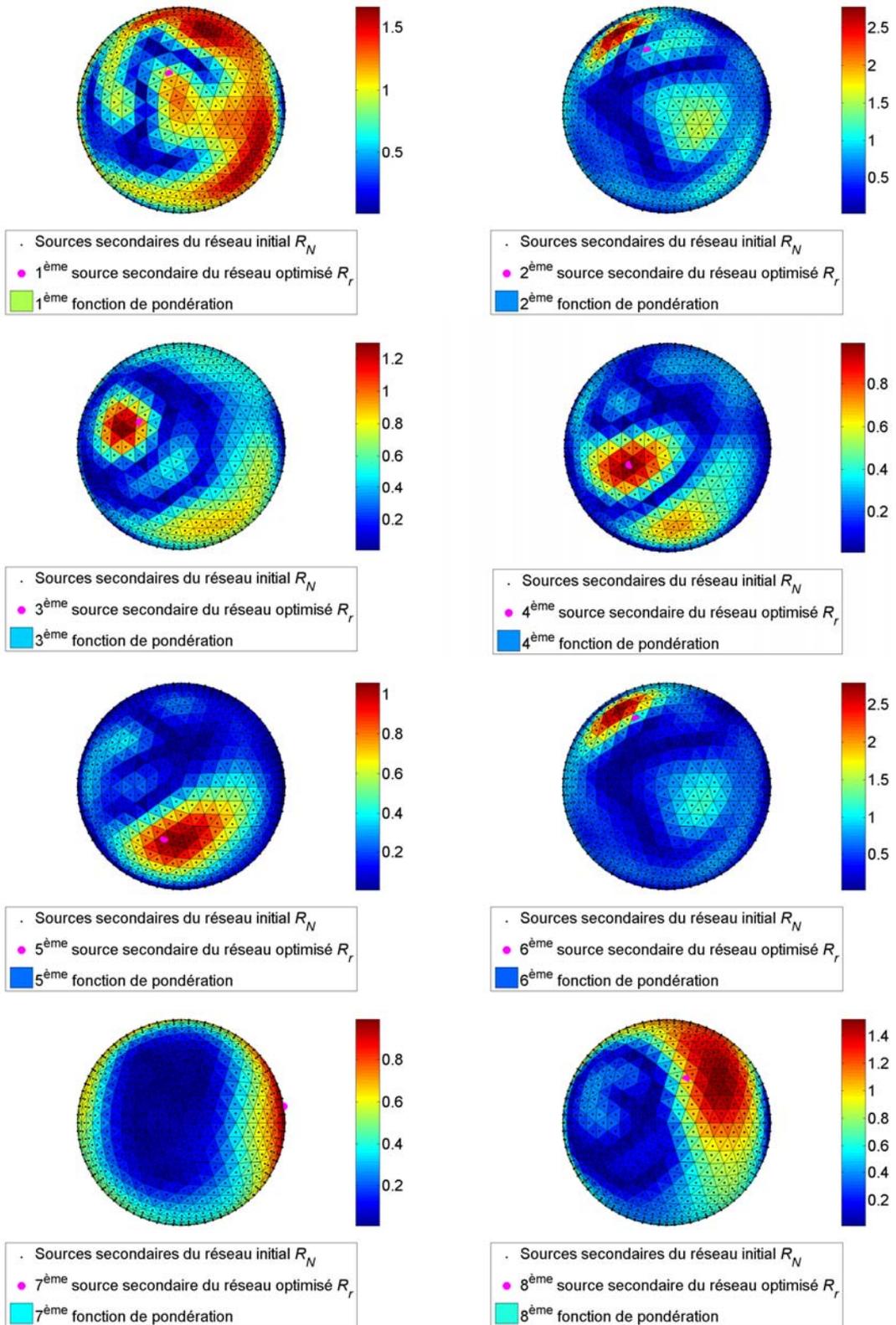


Fig. IV-28 : Fonctions de pondération - configuration PCA - optimisation PROJ

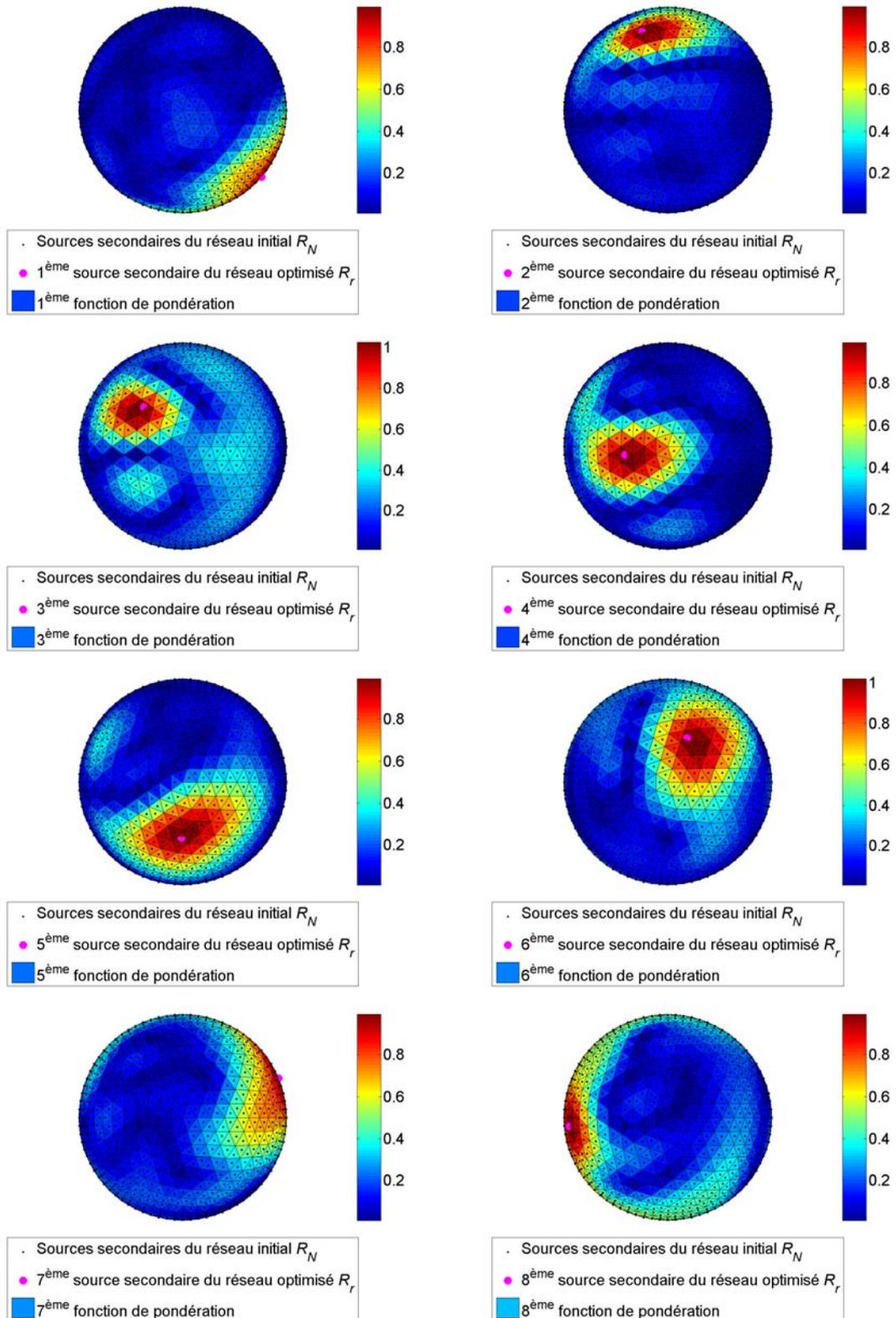


Fig. IV-29 : Fonctions de pondération - configuration ICA - optimisation PROJ

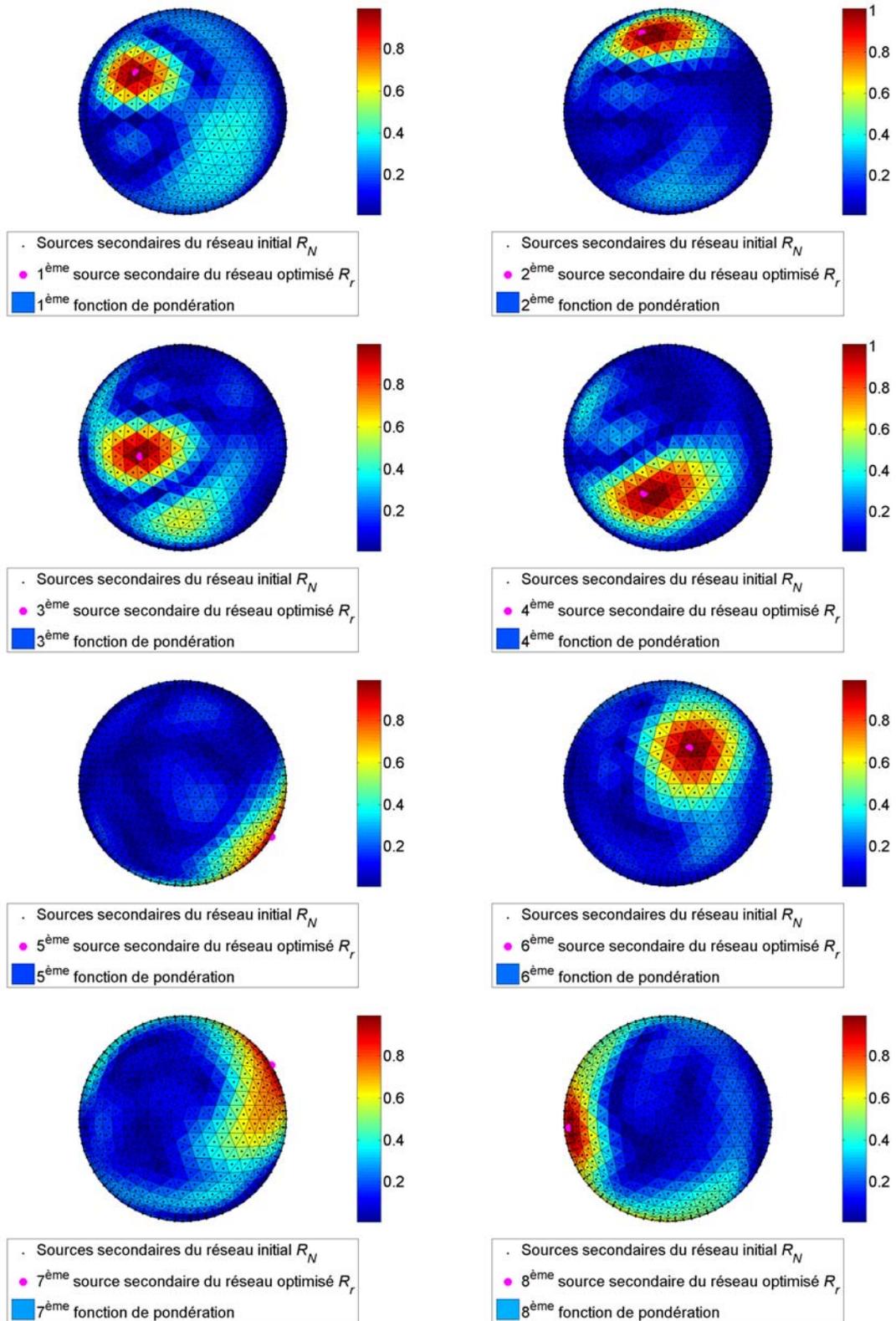


Fig. IV-30 : Fonctions de pondération - configuration SUBSET - optimisation PROJ

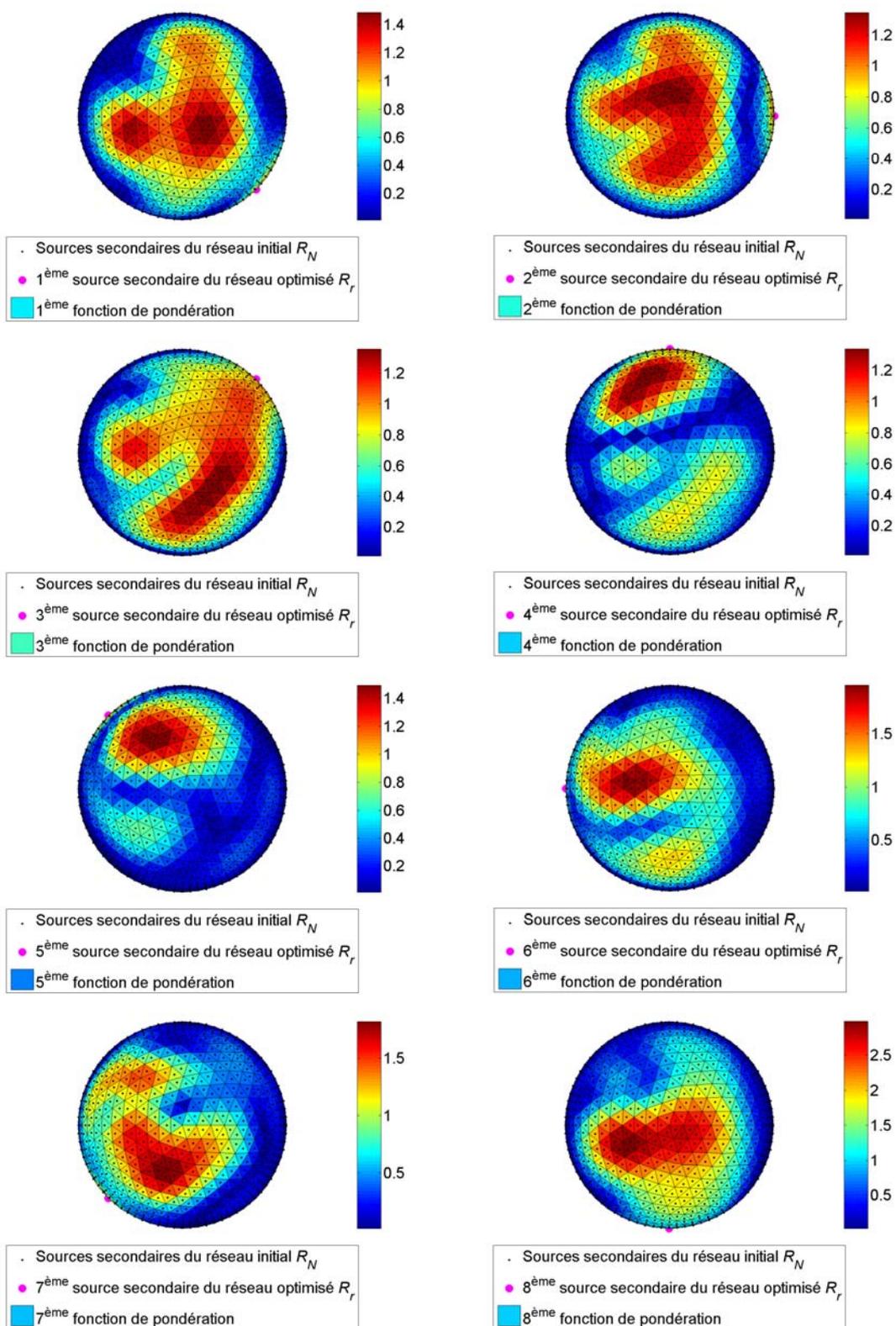


Fig. IV-31 : Fonctions de pondération - configuration CIRC - optimisation PROJ



# V

## Mise en œuvre et évaluation perceptive

## **1 INTRODUCTION**

Dans ce chapitre, nous présentons les résultats de différents tests perceptifs, conduits sur différentes configurations de réseau. Dans cette optique, un prototype de casque modulaire a été élaboré, modulaire dans le sens où le positionnement des haut-parleurs pouvait être modifié de manière aisée, pour ainsi changer facilement la configuration des réseaux de sources secondaires.

Outre le casque lui-même, différentes applications ont été mises au point pour effectuer la synthèse sonore : un moteur de synthèse temps-réel permettant la spatialisation de sources monophoniques sur différents supports, utilisant différentes technologies (binaurale, transaurale, holophonie binaurale), ainsi qu'une application dédiée à la conduite de tests de localisation dans les plans médian et horizontal. Comme cette section aborde uniquement la mise en œuvre de l'holophonie binaurale dans une optique d'évaluation perceptive de la technologie, seule la deuxième application sera brièvement introduite.

Une comparaison de l'holophonie binaurale et d'une synthèse binaurale individualisée est également présentée, le but étant d'évaluer l'intérêt de l'holophonie binaurale en tant que technologie de spatialisation, au vu des performances qu'elle est susceptible d'offrir et des contraintes matérielles qu'elle nécessite.

## **2 SYSTEME DE RESTITUTION DEDIE A L'HOLOPHONIE BINAURALE**

### **2.1 Modèle sphérique de rayon moyen**

Les systèmes holophoniques évalués ont été conçus (pour la partie filtrage) à partir d'un modèle sphérique et d'une captation HYBRID. L'objectif des tests perceptifs étant essentiellement d'évaluer les performances de localisation en terme de sensation en élévation et de confusions avant/arrière, l'individualisation du modèle sphérique à chaque sujet participant à l'étude nous est parue inutile. Cependant, pour que la spatialisation reste pertinente pour des positions latérales pour l'ensemble des individus, il était nécessaire d'utiliser une sphère de dimension moyenne, et donc d'estimer un rayon moyen.

Pour ce faire, nous avons conduit l'estimation du rayon équivalent sur une base de données HRTF, suivant la méthode décrite en IV.2.2. La base de HRTF est composée de 43 jeux de filtres provenant de la base de données LISTEN [Listen] et 37 de la base de données CIPIC [Cipic]. Le rayon sphérique équivalent moyen a été estimé à 92mm, le minimum étant 83mm, le maximum 103mm et la variance environ 4mm.

## 2.2 Compensation du phénomène d'atténuation

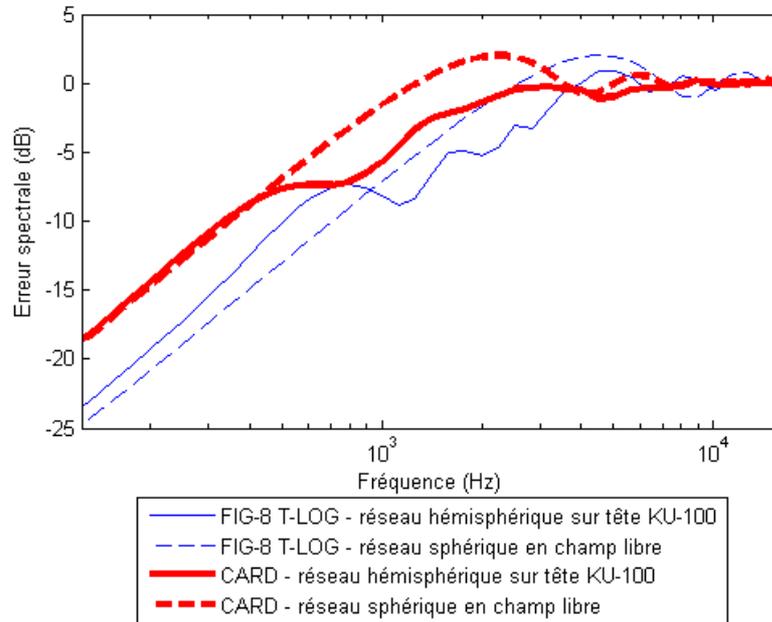


Fig. V-1 : Estimation de l'atténuation basse-fréquence

L'objectif étant d'évaluer la spatialisation de son large bande, il est nécessaire de corriger le problème d'atténuation qui se manifeste aux fréquences basses lors du calcul des CAHF. Ces derniers sont calculés à partir du modèle de tête sphérique précédemment défini et en utilisant la captation HYBRID, qui consiste à combiner une captation FIG-8 T-LOG et une captation CARD en fonction de l'angle latéral de la source primaire. Il est donc nécessaire d'estimer cette atténuation pour chaque méthode de captation, et de la corriger.

Les seules données nous permettant d'effectuer une estimation quantitative de ce phénomène sont celles issues des simulations conduites sur le modèle KU-100. A partir des  $HRTF^o$  et des  $HRTF^r$ , calculées en utilisant les techniques holophoniques classiques (c'est-à-dire sans optimisation) sur le réseau hémisphérique à 640 sources secondaires, on peut estimer cette atténuation (voir Fig. V-1). On s'aperçoit que les courbes calculées à partir de ces données possèdent de fortes variations, qui ne sont pas forcément représentative du phénomène d'atténuation, mais plutôt liées au calcul même des CAHF voire au modèle de tête utilisé. En comparant ces résultats avec ceux issus d'une estimation de l'atténuation induite par un réseau sphérique de même dimension (5cm de rayon) placé en champ libre, on observe que leurs comportements asymptotiques sont similaires, comme l'indique la figure V-1.

Nous préférons compenser l'atténuation telle estimée à partir du réseau sphérique en champ libre.

### 2.3 Casque multi-transducteur modulaire



Fig. V-2 : Prototype modulaire de casque multi-transducteurs

La figure V-2 présente le prototype de casque modulaire mis au point pour permettre l'évaluation de différentes configurations de réseau de transducteurs. On peut distinguer les haut-parleurs constituant les réseaux de source secondaire au sens de Huygens, et la structure permettant de les fixer à la position voulue.

- La structure se compose de deux coques assimilables à des hémisphères. Celles-ci sont élaborées à partir d'un maillage en plastique filaire. Elles ont été choisies ainsi pour s'affranchir d'un éventuel « effet de coque » : le maillage peut être considéré comme acoustiquement transparent, ce qui évite toute réflexion à l'intérieur de la structure. Outre le fait d'offrir des propriétés acoustiques intéressantes, le maillage permet de fixer facilement les haut-parleurs à la structure.
- Les transducteurs ont été sélectionnés les plus petits possibles pour s'approcher du caractère omidirectionnel que les sources secondaires doivent vérifier, tout en s'assurant d'une bande passante suffisante. En effet, en s'appuyant sur la

théorie des pistons vibrants [Pie99], on s'aperçoit que la directivité dépend fortement du produit  $ka$ , où  $k$  est le nombre d'onde et  $a$  le rayon du piston (voir Fig. V-3). On peut considérer que le rayonnement est omnidirectionnel pour  $ka \leq 1$ . Notre choix s'est donc porté sur des transducteurs de type « *in ear* » de 5mm de rayon que l'on peut considérer comme omnidirectionnel jusqu'à une fréquence d'environ 11kHz. L'utilisation de ce type de haut-parleur s'est avérée de plus très pratique, car ils ont également pu être utilisés lors des tests perceptifs conduits sur des HRTF individuelles : le même matériel de base a donc servi à l'évaluation de l'holophonie binaurale et du binaural individuel.

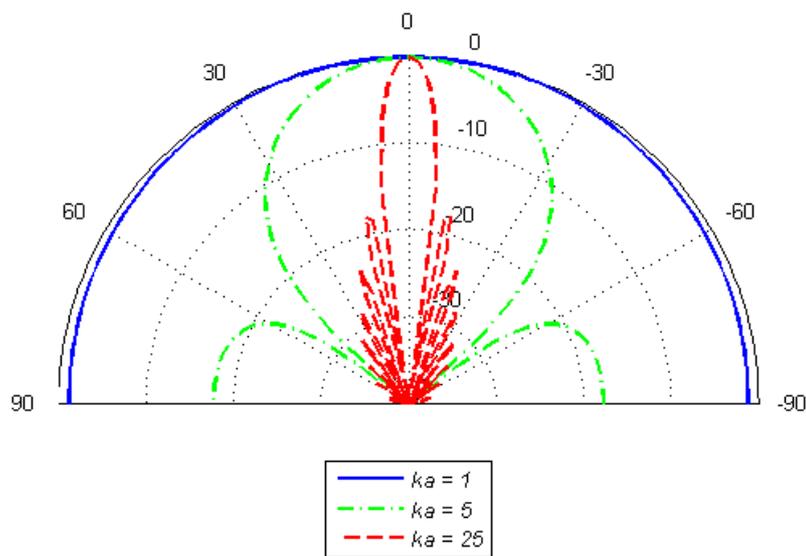


Fig. V-3 : Directivité théorique d'un piston vibrant (en dB)

## 2.4 Application de test

L'application de test a été réalisée sous l'environnement Max/MSP. Elle est principalement composée d'une interface de pointage à la souris, permettant à un utilisateur de reporter une position angulaire dans un plan.

Le traitement des sources n'est pas réalisé en temps-réel pour alléger la charge de calcul et éviter s'affranchir de tout artéfact éventuel qui pourrait être lié à une surcharge des processeurs. Les stimuli ont donc été prétraités sous l'environnement Matlab.

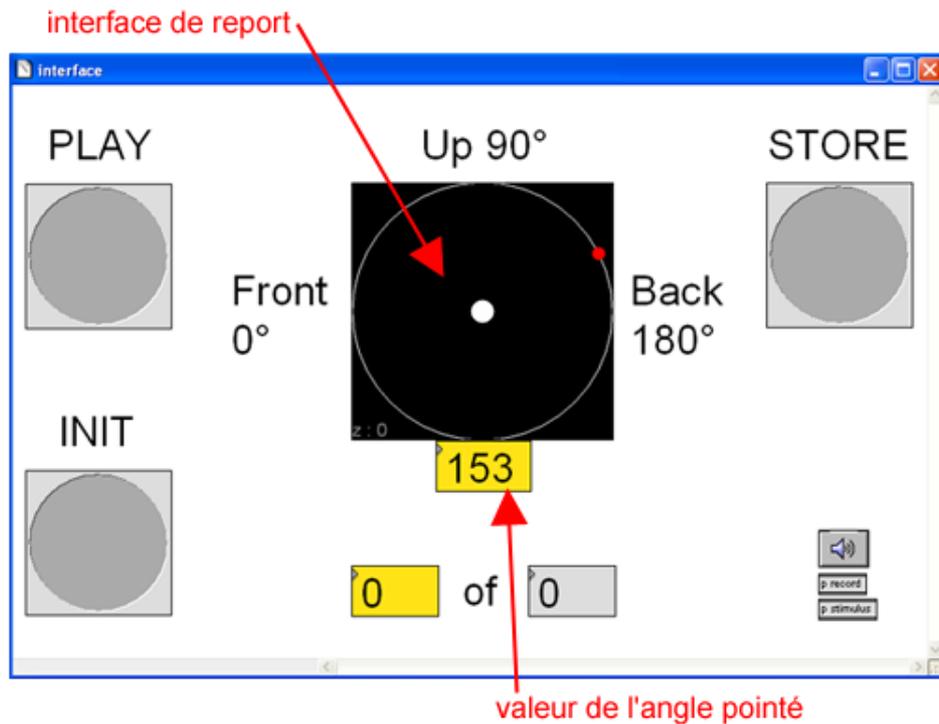


Fig. V-4 : Interface de l'application de test de localisation (plan médian)

## 2.5 Autres matériels

Contrairement aux technologies binaurales classiques, la mise en œuvre de l'holophonie binaurale nécessite de commander plus qu'une simple sortie audio stéréophonique. L'emploi d'une carte son multicanal est donc obligatoire.

Pour alimenter les haut-parleurs du prototype de casque, nous avons utilisé une carte son professionnelle RME Hammerfall Digiface permettant de gérer 24 canaux numériques, ainsi que des convertisseurs numérique-analogique RME ADI-8 Pro permettant chacun de convertir 8 voies numériques en 8 voies analogiques.

## 3 EVALUATION PERCEPTIVE

### 3.1 Evaluation subjective de différentes configurations dans le plan médian

#### 3.1.1 Configurations testées

Trois configurations de réseau (voir Fig. V-5), élaborées suivant différentes approches, ont été testées, l'objectif étant de voir si l'une d'elles est susceptible d'offrir de meilleures performances en localisation que les autres. Toutes sont composées de 8 haut-parleurs (16 au total).

- La configuration *A* correspond à une répartition quasi-uniforme des transducteurs sur la surface hémisphérique (distribution EQUI). Les CAHF associés sont calculés en utilisant l'optimisation AZ.

- La configuration **B** et les CAHF associé sont issus de l'optimisation SUBSET PROJ.
- La configuration **C** correspond à une répartition circulaire uniforme des transducteurs à la base de la surface hémisphérique (disposition CIRC). Les CAHF associés sont calculés en utilisant l'optimisation AZ.

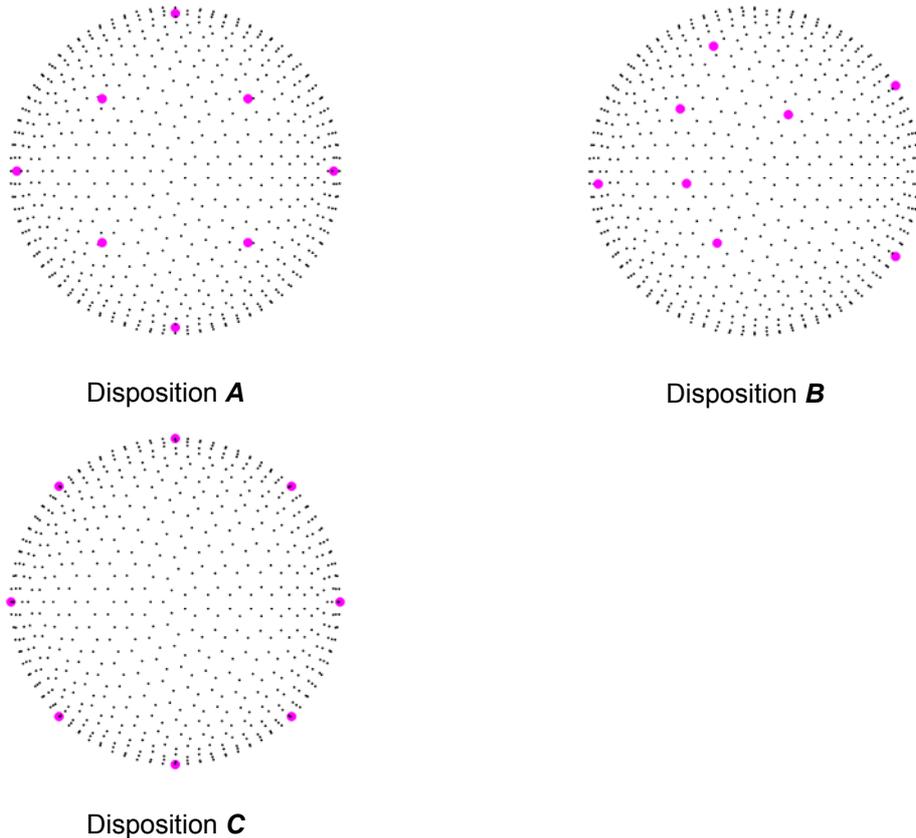


Fig. V-5 : Distributions relatives aux configurations testées

Ces trois configurations ont été choisies par rapport aux constats et remarques que nous avons faits au chapitre précédent. La configuration **A** sert en quelque sorte de « mètre-étalon ». Les simulations conduites sur la configuration **B** ont fourni les meilleurs résultats dans tout l'espace, et celles de la configuration **C** ont montré des résultats étonnants. De plus, pour la disposition **C**, la méthode AZ nous est paru plus pertinente que la méthode PROJ, de part l'incohérence entre les positions des sources secondaires et les maxima des fonctions de pondération associées.

### 3.1.2 Evaluation qualitative dans le plan médian

#### a) Protocole

Les positions de source virtuelles testées sont au nombre de 7, de  $-45^\circ$  à  $225^\circ$  par pas de  $45^\circ$  (angle polaire pour un angle latéral nul). 4 sujets participent à ces écoutes. Le stimulus de base est un burst de bruit rose modulé à 25Hz et d'une durée de 200ms. Il est diffusé par train de 5, séparé par un silence de 100ms. Le sujet choisit la position (il

a donc connaissance de la direction d'où provient le son), écoute le rendu autant de fois qu'il le désire, puis donne son avis global oralement.

### **b) Rapport**

Nous rapportons ici un résumé des différents avis donnés par les différents sujets. L'objectif est d'évaluer qualitativement les sensations fournies par les différentes configurations.

- Configuration **A** : difficulté de localisation, très peu de différenciation avant/arrière, mauvaise sensation verticale à l'avant.
- Configuration **B** : difficulté de localisation, très peu de différenciation avant/arrière, mauvaise sensation verticale à part pour certaines positions (qui varient suivant le sujet testé).
- Configuration **C** : localisation floue, différenciation avant/arrière marquée, bonne sensation verticale à part pour certaines positions (qui varient suivant le sujet testé).

### **c) Conclusion**

Les écoutes qualitatives tendent à indiquer que la configuration **C** fournit les meilleures sensations de positionnement dans le plan médian. On peut supposer qu'elle offrira donc les meilleures performances en localisation.

## **3.1.3 Test de localisation dans le plan médian**

### **a) Protocole de test**

Les 4 mêmes sujets participent à ces tests. 9 positions sont testées, de  $-30^\circ$  à  $210^\circ$  par pas de  $30^\circ$ . Le stimulus utilisé est le même que lors de l'évaluation informelle. L'ordre de positionnement est ici aléatoire, et le sujet ignore où sont positionnés les sons et ignore également quels sont les positionnements possibles. Le sujet peut rejouer le son par train de 5 autant de fois qu'il le souhaite avant d'indiquer la direction perçue sur une interface graphique logicielle. Chaque position est présentée 5 fois.

### **b) Protocole d'analyse**

Les données issues du test sont prétraitées suivant une méthode dérivée de celle proposée par Wightman et Kistler [WK89] :

- correction des confusions de jugement pour chaque position reportée,
- calcul des centroïdes et de la dispersion pour chaque position cible, après correction des confusions,
- moyenne effectuée sur les centroïdes.

Ce protocole, établi pour l'analyse de test de localisation où les positions cibles peuvent décrire tout l'espace, peut être simplifié lorsque l'on se limite à des sources localisées dans un plan. Le calcul des centroïdes et des dispersions ne nécessite plus de passer par une représentation géométrique vectorielle, et peut être remplacé par un calcul de moyennes et d'écart-types. De plus, nous choisissons de ne corriger que les confusions avant/arrière, en ramenant par simple symétrie la position reportée dans le même hémisphère que la position cible, si cela est nécessaire.

Nous noterons *E* l'erreur moyenne entre les centroïdes et les valeurs cibles, *Std* l'écart-type moyen et *F/B* le taux de confusion avant/arrière.

**c) Résultats**

Les figures V-6 à V-8 présentent les positions reportées par rapport aux positions cibles, par configuration et par individu. Les croix rouge indiquent chaque jugement, la ligne bleu foncé relie les centroïdes et la zone bleu clair est définie par l'écart-type. Chaque graphique correspond aux résultats d'un individu testé sur une configuration. Le tableau V-1 reporte de même les résultats par individu et par configurations, ainsi que les résultats moyens par configuration.

		<i>A</i>	<i>B</i>	<i>C</i>
<i>1</i>	E (°)	23	22	15
	Std (°)	17	14	25
	F/B (%)	29	42	27
<i>2</i>	E	19	17	12
	Std	16	15	14
	F/B	7	11	13
<i>3</i>	E	39	24	19
	Std	21	17	20
	F/B	38	24	22
<i>4</i>	E	29	27	11
	Std	24	21	18
	F/B	27	60	18
<i>Moyenne</i>	E	28	23	<b>14</b>
	Std	20	17	19
	F/B	25	34	<b>20</b>

Tab. V-1 : Résultats moyens par individu et par configuration – plan médian

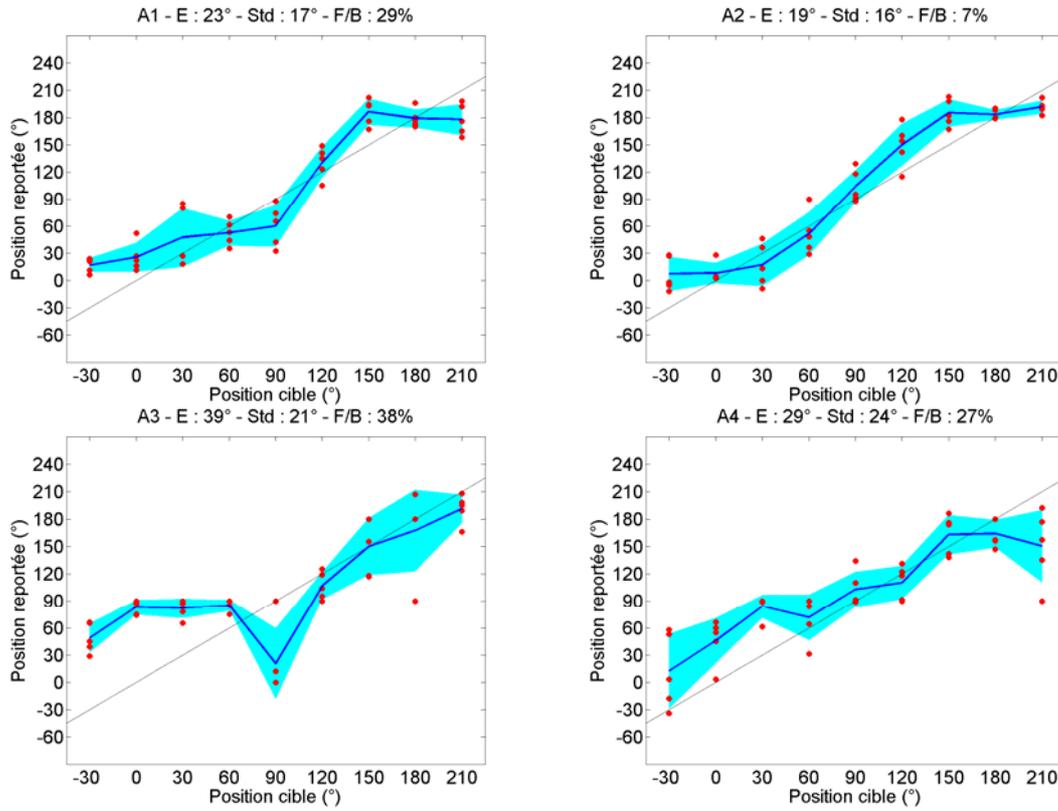


Fig. V-6 : Report des jugements après correction des confusions – configuration A – plan médian

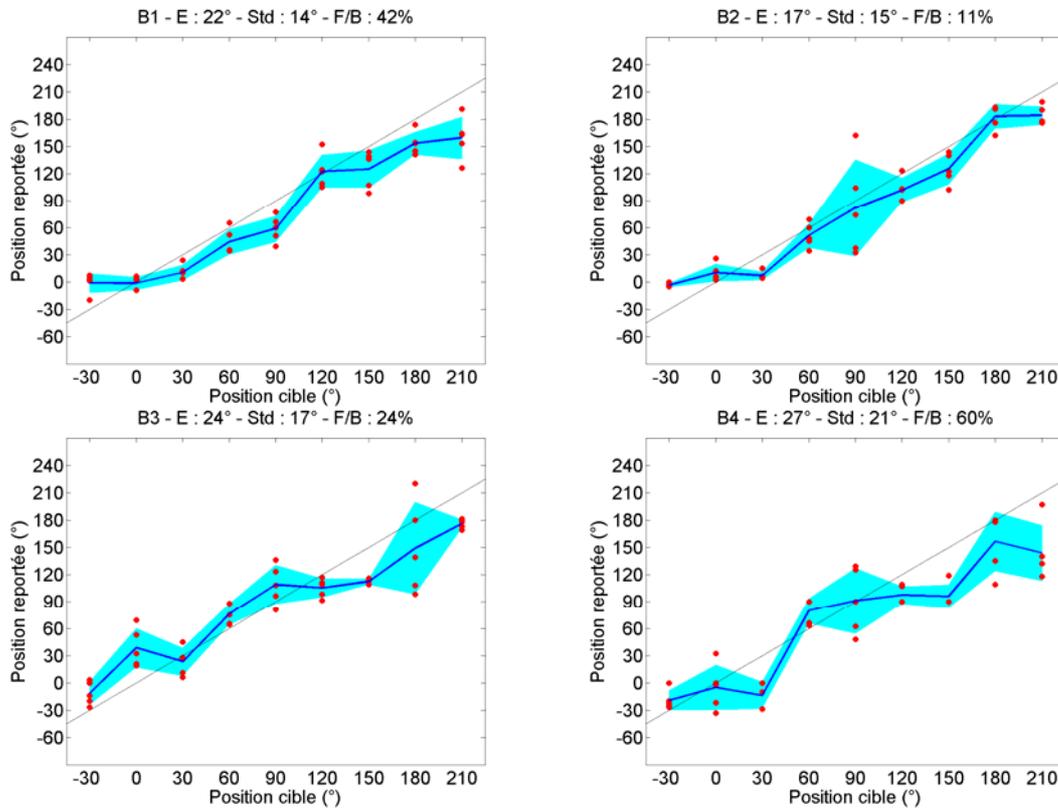


Fig. V-7 : Report des jugements après correction des confusions – configuration B – plan médian

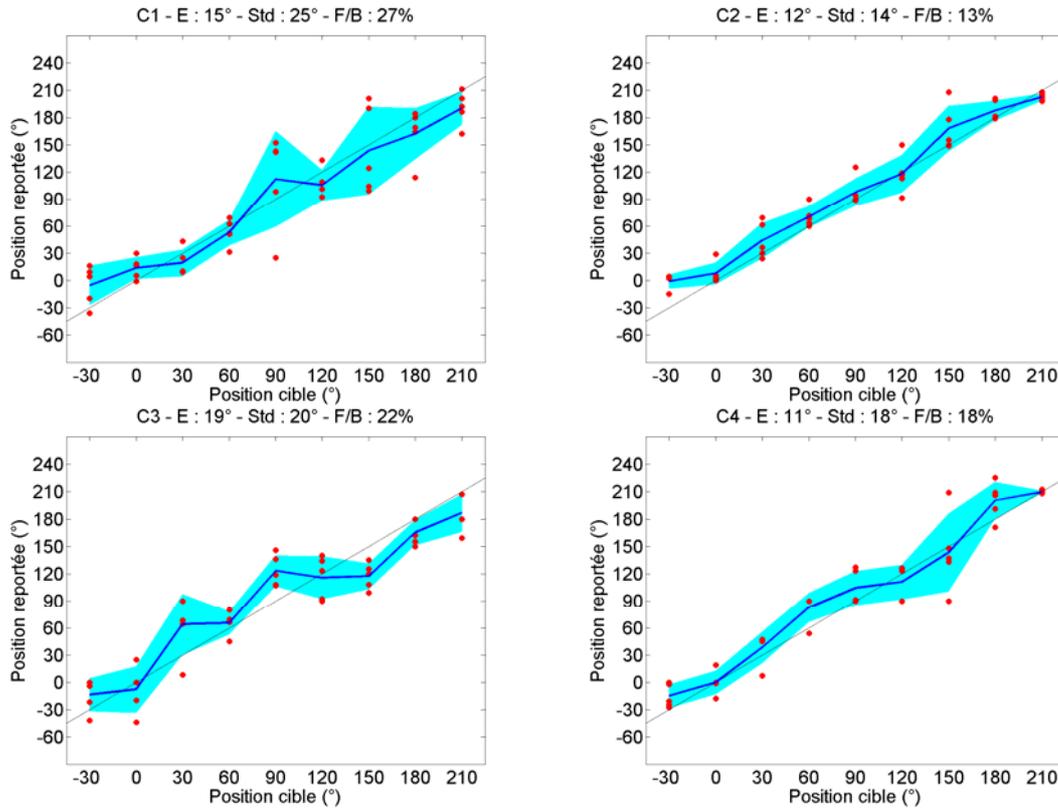


Fig. V-8 : Report des jugements après correction des confusions – configuration C – plan médian

La configuration **C** apparaît clairement comme la meilleure des trois. L'erreur moyenne de localisation et le taux de confusion avant/arrière moyen sont nettement plus petits, cette tendance étant également vérifiée pour chaque individu. L'écart-type cependant reste comparable pour toutes les configurations.

### c) Conclusion

Comme le suggérait le test d'écoute qualitatif, le test de localisation indique que la configuration **C** offre les meilleures performances de localisation dans le plan médian, en se détachant nettement des deux autres, que ce soit en terme d'erreur de localisation ou de confusions avant/arrière.

## 3.2 Évaluation subjective de la configuration C dans le plan horizontal

### 3.2.1 Évaluation qualitative dans le plan horizontal

Les positions de source virtuelles testées sont au nombre de 8, de  $-135^\circ$  à  $180^\circ$  par pas de  $45^\circ$  (azimuth pour une élévation nulle). Le reste du protocole est identique à celui présenté en V.3.1.1.

Les remarques importantes sont une latéralisation trop rapide, surtout dans l'hémisphère arrière, ainsi qu'une sensation désagréable pour les sources positionnées strictement à droite ou à gauche, les sons étant souvent perçus dans l'oreille ipsilatérale.

Il est bien entendu nécessaire de quantifier les performances en localisation dans le plan horizontal, mais il semble d’ores et déjà nécessaire de modifier la configuration C pour améliorer le confort d’écoute de sources positionnées latéralement.

### 3.2.2 Test de localisation dans le plan horizontal

12 positions sont testées, de  $-150^\circ$  à  $180^\circ$  par pas de  $30^\circ$ . Le reste du protocole est identique à celui présenté en V.3.1.2. La figure V-9 présente les positions reportées par rapport aux positions cibles par individu. Le tableau V-2 reporte les résultats par individu, ainsi que les résultats moyens.

Bien que les graphiques témoignent effectivement d’un problème de latéralisation plus ou moins rapide dans l’hémisphère arrière, les résultats moyens sont bons. En effet, une erreur moyenne de l’ordre de  $10^\circ$  et un écart-type de l’ordre de  $10^\circ$  sont des indicateurs très correctes. De plus, le taux de confusion reporté dans le plan horizontal est similaire à celui reporté dans le plan médian.

	E (°)	Std (°)	F/B (%)
1	11	11	22
2	7	10	11
3	13	13	29
4	13	12	18
Moyenne	11	12	20

Tab. V-2 : Résultats par individu pour la configuration C – plan horizontal

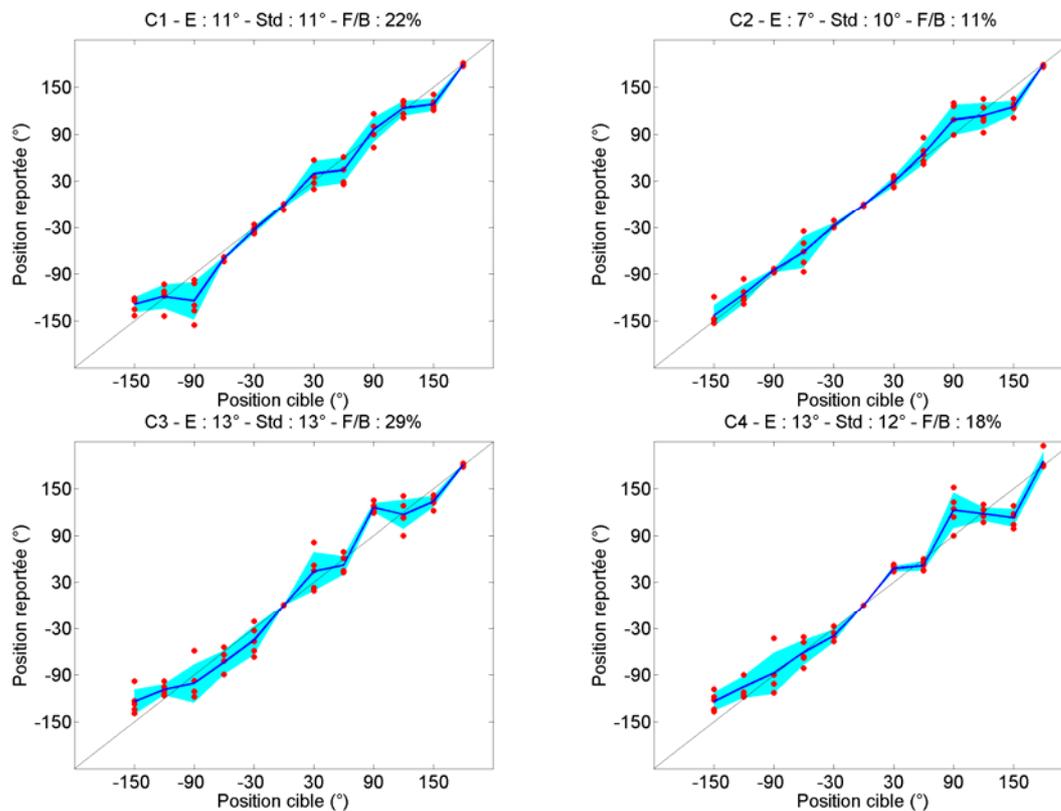


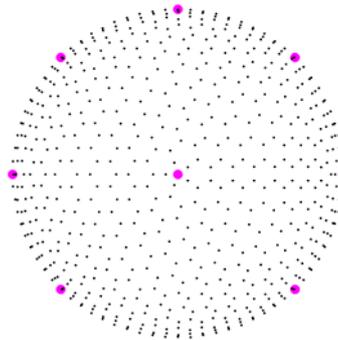
Fig. V-9 : Report des jugements après correction des confusions – configuration C – plan horizontal

### 3.3 Modification de la configuration **C**

La configuration **C** a été élaborée sur des considérations psychoacoustiques et physiques de manière à optimiser l'excitation des indices de localisation pavillonnaires pour des sources virtuelles positionnées dans le plan médian. Cependant, un front d'onde ipsilatéral ne peut être reconstruit, même partiellement, par cette géométrie de réseau. Il est probable que la sensation étrange rapportée lors des tests d'écoute qualitative, soit la conséquence de ce manque de cohérence entre le champ reconstruit par le réseau de transducteur et le champ qui aurait été induit par une source latérale.

Pour tenter de résoudre ce problème, nous élaborons une quatrième configuration en déplaçant un haut-parleur vers le sommet de l'hémisphère. Ainsi un front d'onde latéral sera mieux reconstruit du côté ipsilatéral. Le transducteur que nous choisissons de déplacer est celui dont la contribution est prédominante pour des sources virtuelles positionnées à des élévations très basses. En effet, ces positions de source étant rarement exploitées en réalité virtuelle par rapport au reste de l'espace, il est attendu que le déplacement de ce haut-parleur précisément est un impact moindre sur les performances de localisation que le déplacement d'un des 7 autres haut-parleurs. La figure V-10 présente schématiquement la nouvelle configuration **D**.

Tout comme pour la configuration **C**, nous couplons cette nouvelle disposition à un calcul des CAHF suivant la méthode AZ.



Disposition **D**

Fig. V-10 : Distribution relative à la configuration **D**

### 3.4 Evaluation subjective de la configuration **D**

#### 3.4.1 Ecoutes informelles

Les sensations fournies par la configuration **D** semblent équivalentes à celles fournies par la configuration **C**, hormis deux améliorations notables : les sujets ne font plus rapport ni d'une latéralisation trop rapide ni d'une sensation étrange lorsque les sources sont placées strictement à droite ou à gauche.

La modification apportée à la configuration **C** semble remplir son rôle, à savoir améliorer le confort d'écoute et fournir une sensation plus homogène. Cependant, il est

nécessaire d'évaluer quantitativement l'impact de cette transformation. Les tests de localisation dans le plan médian et dans le plan horizontal ont donc été reconduits sur la configuration **D**.

### 3.4.2 Tests de localisation

Les résultats issus des tests effectués sur la configuration **D** sont globalement comparables à ceux issus des tests effectués sur la configuration **C** (voir Fig. V-11, Fig. V-12, Tab. V-3). Certaines différences sont néanmoins remarquables dans le plan médian : l'erreur moyenne est légèrement plus grande, passant de 14 à 18°, et le taux de confusion avant/arrière deux fois plus petit, passant de 20 à 10%.

	Plan médian			Plan horizontal		
	<i>E</i> (°)	<i>Std</i> (°)	<i>F/B</i> (%)	<i>E</i>	<i>Std</i>	<i>F/B</i>
1	19	14	7	10	12	18
2	16	18	2	7	9	7
3	21	18	7	12	12	31
4	17	19	18	13	8	27
Moyenne	18	17	9	11	10	21

Tab. V-3 : Résultats par individu pour la configuration **D** – plans médian et horizontal

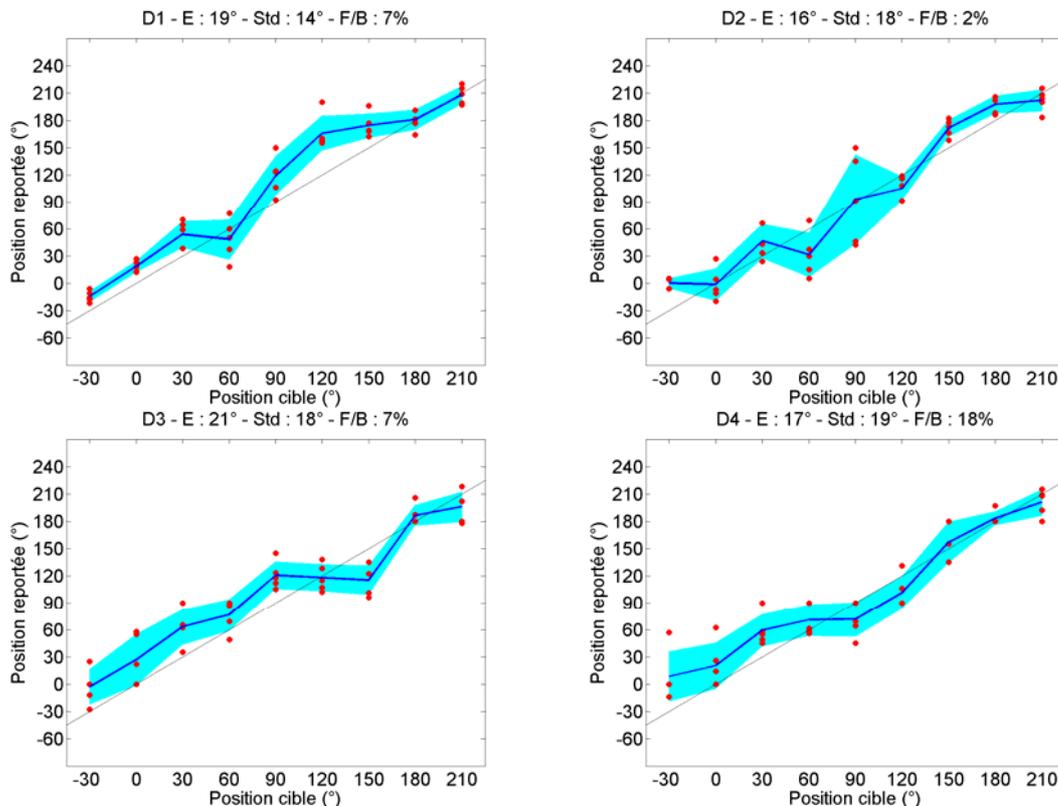


Fig. V-11 : Report des jugements après correction des confusions – configuration **D** – plan médian

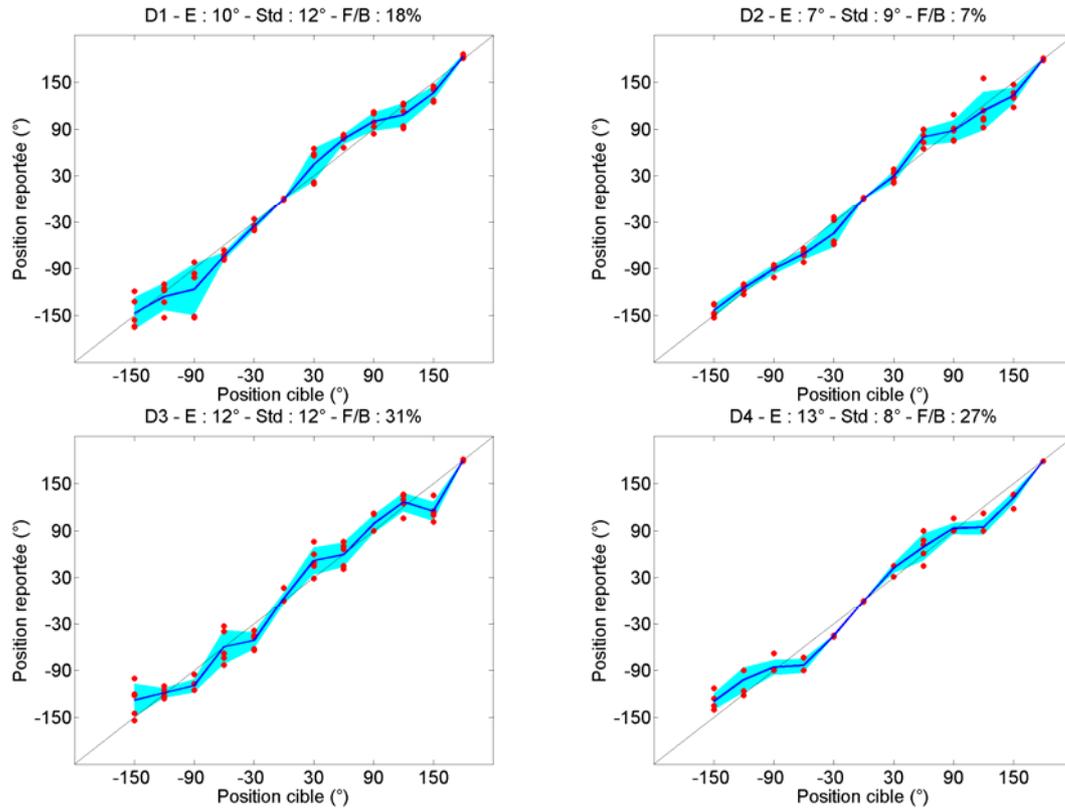


Fig. V-12 : Report des jugements après correction des confusions – configuration **D** – plan horizontal

Il est difficile de fournir une explication à ces variations, car le déplacement d'un haut-parleur n'est pas le seul facteur qui intervient. En effet, la phase de positionnement du prototype modulaire est délicate, et il est impossible de garantir la répétabilité de cette étape. Il est probable que, sur un même test, générant un volume de données suffisant pour effectuer des analyses statistiques robustes, deux positionnements différents du casque fournissent des résultats différents.

Quoiqu'il en soit, au vu des résultats des différents tests présentés dans cette section, la configuration **D** est celle qui, sans conteste, offre les meilleures performances globales tout en garantissant une sensation d'écoute homogène.

### 3.5 Holophonie binaurale vs. synthèse binaurale individuelle

Dans l'optique de confronter l'holophonie binaurale à une synthèse binaurale individualisée, nous avons réalisé une seconde série de tests de localisation sur 4 sujets possédant leur propre HRTF et sur les deux systèmes en concurrence. Nous reportons ici les résultats de ces tests conduits en utilisant le même protocole que précédemment (voir Fig. V-13 à V-16 et Tab. V-4).

		Plan médian		Plan horizontal	
		Holophonie binaurale	HRTF individuelles	Holophonie binaurale	HRTF individuelles
1	<i>E</i>	24	20	9	11
	<i>Std</i>	<b>25</b>	44	10	10
	<i>FB</i>	18	24	13	18
2	<i>E</i>	16	15	7	9
	<i>Std</i>	18	16	9	16
	<i>FB</i>	<b>2</b>	16	7	20
3	<i>E</i>	18	16	13	10
	<i>Std</i>	14	16	10	6
	<i>FB</i>	38	<b>24</b>	31	<b>13</b>
4	<i>E</i>	30	27	8	13
	<i>Std</i>	<b>20</b>	30	11	11
	<i>FB</i>	16	20	<b>22</b>	36
Moyenne	<i>E</i>	22	20	9	11
	<i>Std</i>	<b>19</b>	27	10	11
	<i>FB</i>	19	21	18	22

Tab. V-4 : Comparaison des résultats par individu – holophonie binaurale vs. HRTF individuelles

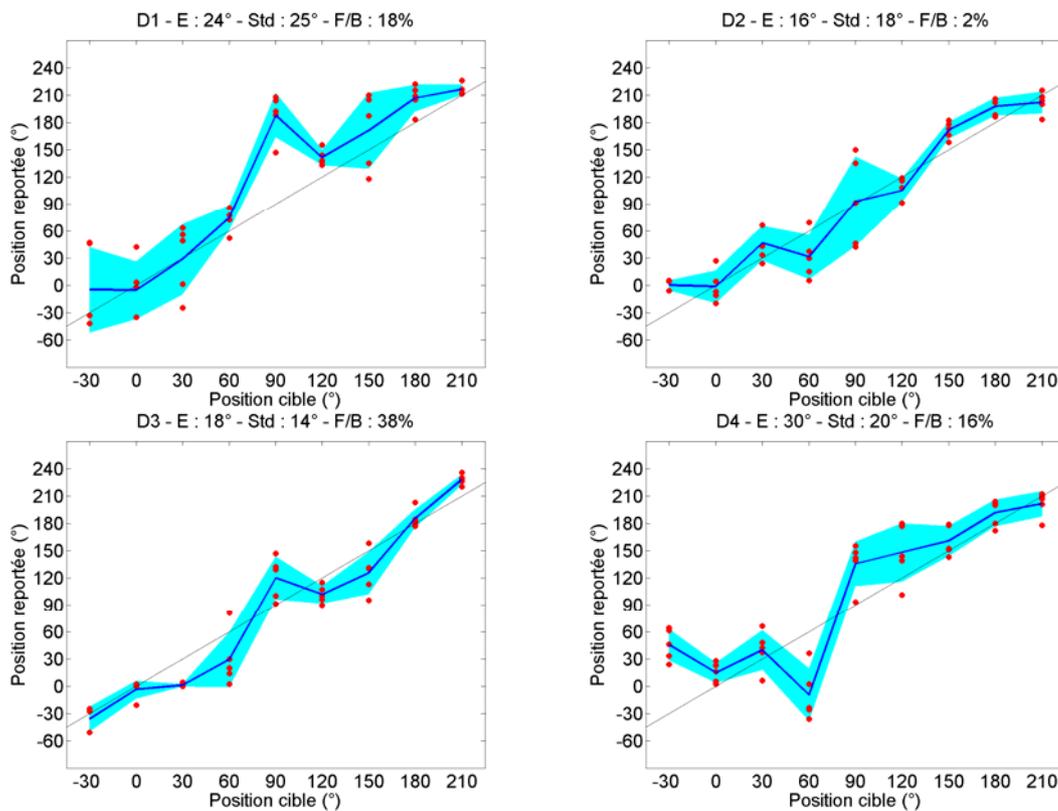


Fig. V-13 : Report des jugements – holophonie binaurale – plan médian

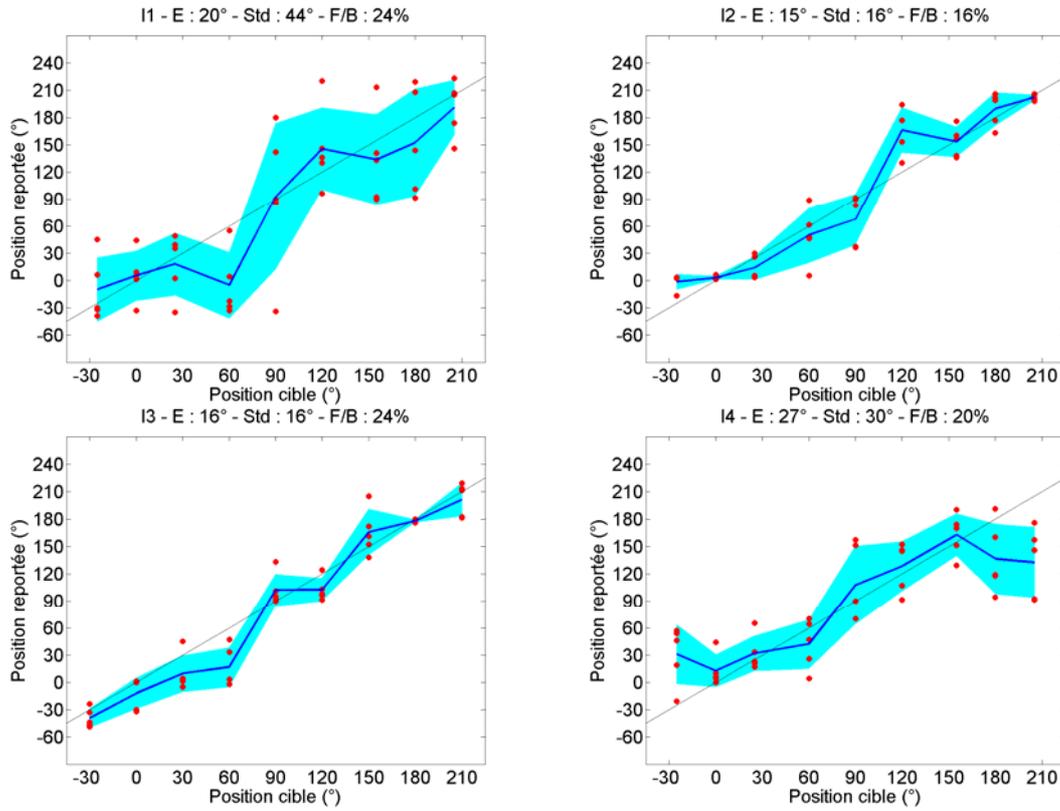


Fig. V-14 : Report des jugements – HRTF individuelles – plan médian

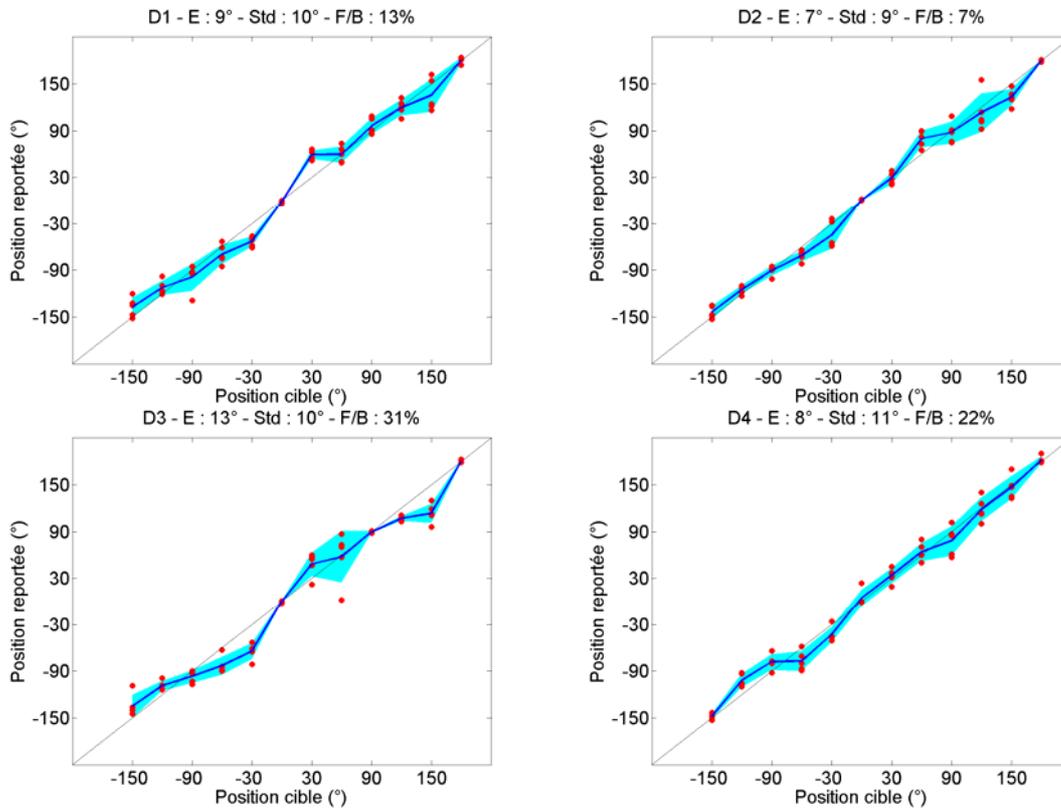


Fig. V-15 : Report des jugements – holophonie binaurale – plan horizontal

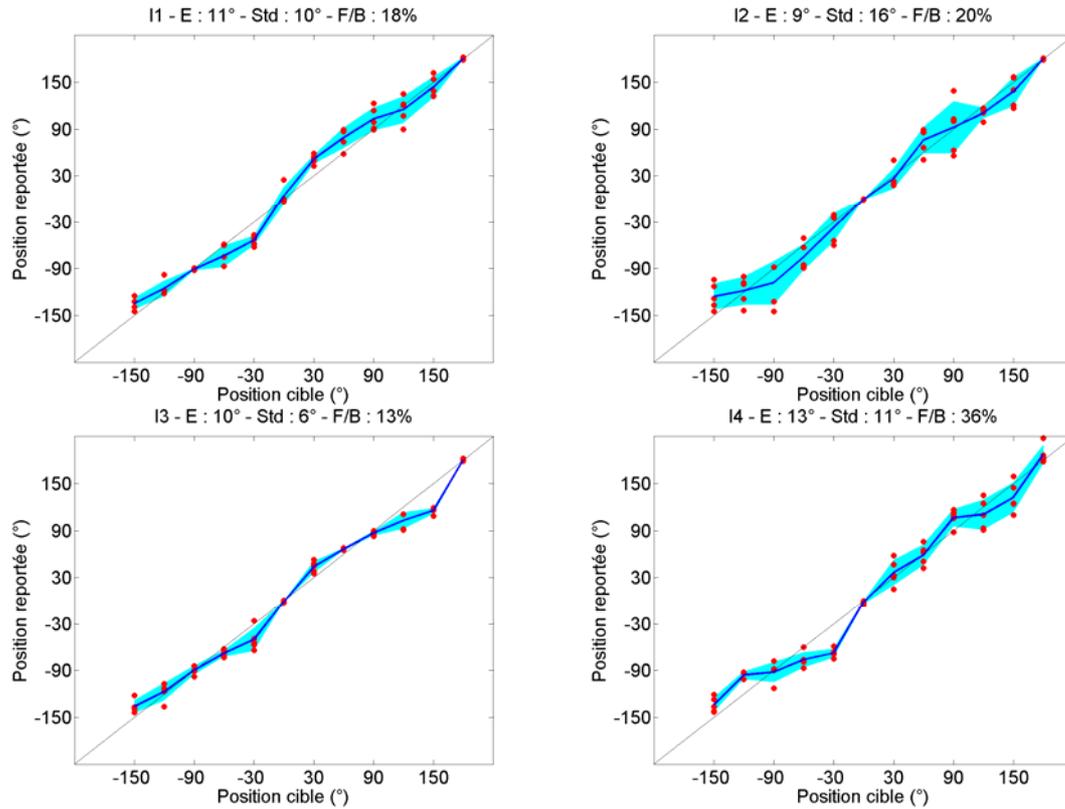


Fig. V-16 : Report des jugements – HRTF individuelles – plan horizontal

Peu de différences sont à noter si l'on compare les deux technologies dans le plan horizontal, alors que les filtres utilisés sur le système holophonique sont générés à partir d'un modèle de tête sphérique de rayon moyen, c'est-à-dire que les indices binauraux classiques ne sont pas individualisés. En ce qui concerne le plan médian, les biais de localisation d'un individu sont différents, suivant que l'on considère l'une ou l'autre des technologies. Par rapport au système holophonique, par exemple, le sujet 1 localisera derrière lui les sons positionnés au-dessus lui, et le sujet 4 aura du mal à localiser l'élévation des sons dans l'hémisphère avant (confusions haut-bas). Par rapport au système binaural personnalisé, le sujet 1 ne détectera aucune élévation dans l'hémisphère avant et le sujet 4 aura tendance à localiser les sons trop en hauteur dans l'hémisphère arrière.

Si l'on compare les résultats par individu, on ne peut conclure d'aucune supériorité d'une technique par rapport à l'autre, car l'une ou l'autre peut offrir de meilleures performances, tout dépend du sujet considéré. Si l'on compare les résultats moyens, on les deux technologies semblent être équivalente.

#### 4 CONCLUSION

Les tests de localisations que nous avons menés montrent qu'un système d'holophonie binaurale semble pouvoir exciter les indices de localisation pavillonnaires d'un auditeur aussi bien qu'un système binaural individuel. Les statistiques étant conduites sur seulement quatre sujets, il est difficile de conclure directement à l'équivalence des deux

systèmes, mais les résultats permettent de justifier l'existence même de l'holophonie binaurale comme alternative pertinente à la synthèse binaurale pour des applications de spatialisation sonore grand public.

L'ensemble des travaux menés sur l'holophonie binaurale, de la définition du concept jusqu'aux évaluations perceptives présentées dans ce chapitre, ont conduit à la publication d'un article [GK08] (voir Publications, Article C).



# VI

## Implémentation de la spatialisation

## 1 INTRODUCTION

L'implémentation de la spatialisation pour un système d'holophonie binaurale est plus complexe que celle d'un système binaural classique, basé sur de HRTF, car le nombre de canaux gérés est plus important. Dans la littérature, différentes optimisations ont été proposées pour simplifier la spatialisation binaurale dynamique multi-source, celles-ci étant généralement effectuées en décomposant linéairement les HRTF en un nombre réduit de filtres de reconstruction et de fonctions spatiales. On parle alors d'implémentation multicanale de la synthèse binaurale, par opposition à l'implémentation bicanale classique qui consiste à filtrer un son monophonique par deux filtres relatifs à une position de l'espace.

Dans ce chapitre, nous proposons de transposer ces optimisations à un système holophonique binaural composé de  $K$  transducteurs. Nous présenterons également une implémentation simplifiée dédiée à un système conçu suivant la configuration **D**.

## 2 TRANSPOSITION DE L'APPROCHE BICANALE

### 2.1 Synthèse binaurale bicanale

La synthèse binaurale bicanale consiste à filtrer le signal de chaque source que l'on souhaite positionner à une position de l'espace par les HRTF droite et gauche correspondant à la direction appropriée. Les signaux stéréophoniques ainsi obtenus et relatifs à chaque source sont alors sommés pour donner le signal binaural représentatif de la scène sonore synthétisée qui sera diffusé aux tympans droit et gauche de l'auditeur.

Pour spatialiser  $N$  sources, *i.e.* traiter  $N$  flux monophoniques, l'implémentation bicanale fera intervenir  $2N$  filtres pour une synthèse binaurale statique, voire  $4N$  filtres pour une synthèse binaurale dynamique (pour mettre en mouvement des sources sonores et éviter tout artéfact audible, on utilise généralement deux briques de traitement qui tournent en parallèle). La figure VI-1 présente schématiquement l'implémentation bicanale de la synthèse binaurale.

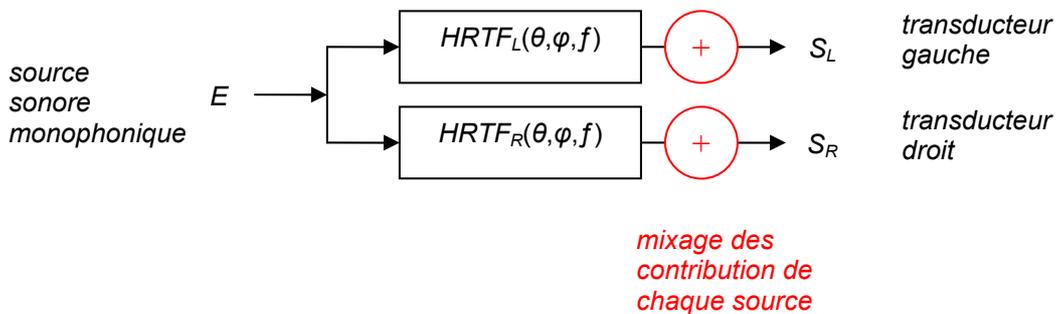


Fig. VI-1 : Implémentation bicanale de la synthèse binaurale

## 2.2 Application à la synthèse holophonique binaurale

Comme pour la synthèse binaurale, il s'agit de faire correspondre un filtre à une position de l'espace et un transducteur, mais au lieu de n'y avoir qu'un seul haut-parleur par oreille, il y en a  $K$ , et au lieu d'utiliser les HRTF, on utilise les CAHF. Le nombre de filtres mis en jeu est donc de  $2NK$  (voire  $4NK$  pour une synthèse dynamique). Pour garder l'analogie avec la synthèse binaurale, on parlera d'implémentation  $K$ -bicanale de la synthèse holophonique binaurale. La figure VI-2 illustre le traitement.

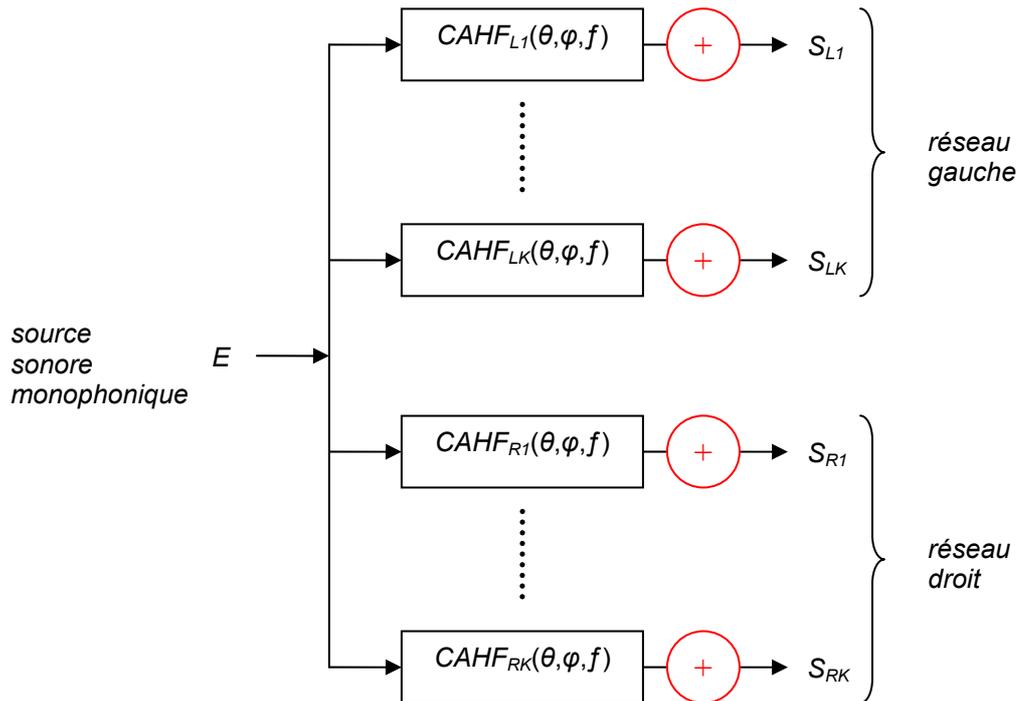


Fig. VI-2 : Implémentation  $K$ -bicanale de la synthèse holophonique binaurale

## 3 TRANSCRIPTION DE L'APPROCHE MULTICANALE

### 3.1 Synthèse binaurale multicanale

Dans le cas de la synthèse binaurale multicanale, les dépendances spatiales et fréquentielles des HRTF sont séparées grâce à une décomposition linéaire des HRTF [JWL98, JLP99]. Elles s'écrivent alors comme une combinaison linéaire de  $r$  filtres de reconstruction  $L_i(f)$  et de  $r$  fonctions spatiales  $C_i(\theta, \varphi)$  ( $i = 1 \dots r$ ):

$$HRTF(\theta, \varphi, f) = \sum_{k=1}^r C_k(\theta, \varphi) L_k(f) \quad (\text{VI-1})$$

La décomposition proprement dite peut être réalisée par les méthodes que nous avons décrites au chapitre IV, à savoir une PCA, une ICA, ou encore une *subset selection* qui consiste à sélectionner les filtres de reconstruction par une décomposition QR (méthode SUBSET) puis à calculer les fonctions spatiales associées par projection (méthode PROJ).

Dans le cas de figure de l'implémentation multicanale de la synthèse binaurale, le signal de chaque source à spatialiser est pondéré par les coefficients  $C_k$  issus de la décomposition linéaire des HRTF. Ces coefficients ont pour particularité de ne dépendre que de la position où l'on souhaite placer la source considérée, et non de la fréquence. Le nombre de ces coefficients dépend du nombre  $r$  de vecteurs de base que l'on a conservés pour reconstruire les HRTF. Les signaux pondérés par la  $k^{\text{ème}}$  fonction spatiale sont additionnés avant d'être filtrés par le  $k^{\text{ème}}$  filtre de reconstruction. Puis les signaux résultants sont sommés pour former les signaux binauraux droit et gauche.

Contrairement à l'approche bicanale, l'ajout d'une source supplémentaire ne nécessite pas la mise en jeu de deux filtres additionnels. Les  $r$  filtres de reconstruction sont partagés par toutes les sources présentes. De plus, l'approche multicanale simplifie l'implémentation dynamique de la spatialisation, car il est alors possible de faire varier les coefficients  $C_k$  sans apparition d'artéfact audible.

Alors que la synthèse binaurale bicanale dynamique fait intervenir  $4N$  filtres, où  $N$  est le nombre de sources à spatialiser, la synthèse binaurale multicanale dynamique ne nécessite que  $2r$  filtres, où  $r$  est pour ainsi dire l'ordre de la décomposition. L'implémentation multicanale devient donc intéressante lorsque  $N > r/2$ .

La figure VI-3 décrit le fonctionnement de l'approche multicanale de la synthèse binaurale (pour un seul côté).

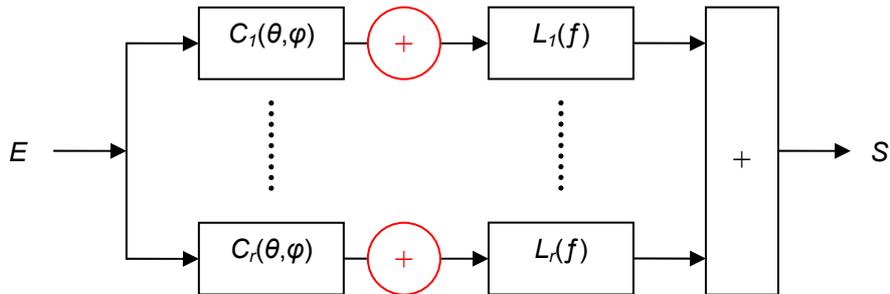


Fig. VI-3 : Implémentation multicanale de la synthèse binaurale

### 3.2 Application à la synthèse holophonique binaurale

L'approche multicanale peut être transposée à la synthèse holophonique binaurale de deux manières. La première consiste à considérer les CAHF comme  $K$  jeux de HRTF et à décomposer linéairement les CAHF comme on le ferait avec les HRTF. On parlera alors d'implémentation  $K$ -multicanale. La deuxième exploite la décomposition des HRTF en CAHF et PRTF. La transposition de l'approche multicanale à un système

d'holophonie binaurale est alors obtenue directement par identification. On parlera d'implémentation multicanale- $K$ .

### **3.2.1 Approche $K$ -multicanale**

Tout comme pour les HRTF, les CAHF sont décomposés linéairement en fonctions spatiales et filtres de reconstruction :

$$CAHF^i = \sum_{k=1}^r C_k^i(\theta, \varphi) L_k^i(f) \quad (VI-2)$$

L'appellation  $K$ -multicanale est donc représentative du traitement : au lieu d'appliquer la décomposition aux seules HRTF droites et gauches, on l'effectue sur les  $K$  paires de CAHF droits et gauches.

### **3.2.2 Approche multicanale- $K$ : paradigme des haut-parleurs virtuels**

Lorsqu'une matrice  $H$  de HRTF est décomposée en sélectionnant les filtres de reconstruction  $L$  parmi les HRTF ou par le biais de la méthode « subset selection », où les filtres de reconstruction sont alors sélectionnés de manière optimale parmi les HRTF, les fonctions spatiales  $C$  peuvent être vues comme un ensemble de lois de panning entre des haut-parleurs virtuels placés aux positions définies par les filtres de reconstruction.

La transposition vers l'holophonie binaurale s'effectue alors simplement en « virtualisant » les haut-parleurs à l'aide des CAHF équivalents.

## **4 IMPLEMENTATION SIMPLIFIEE : DECOMPOSITION DES CAHF**

Les précédentes implémentations de l'holophonie binaurale sont essentiellement des transpositions de techniques utilisées dans le cadre de la synthèse binaurale. La charge de calcul et de mémoire relatif à l'holophonie binaurale reste supérieure d'un facteur  $K$  à celle relatif à la synthèse binaurale : en effet, d'un point de vue traitement du signal, ces transpositions d'une technologie à l'autre consistent à considérer  $K$  canaux stéréo au lieu d'un seul.

Une simplification peut être envisagée pour fortement alléger le processus de spatialisation dédié au casque multi-transducteurs. Celle-ci consiste à réduire les CAHF à un filtrage binaural global, couplé à des directivités associées à chaque haut-parleur. Le traitement de spatialisation est alors effectué comme indiqué à la figure VI-4, où  $H_L$  et  $H_R$  sont les filtres représentatifs du flux d'énergie acoustique éclairant chacun des pavillons d'oreille, contenant les indices interauraux classiques, à savoir l'ITD et l'ILD, ainsi que les indices spectraux non liés au pavillons, et où  $C_{Lk}$  et  $C_{Rk}$  ( $k = 1 \dots K$ ) sont les directivités associées à chaque transducteur et qui permettent, en quelque sorte, de diriger le flux d'énergie et donc de contrôler l'incidence de l'onde arrivant sur les pavillons.

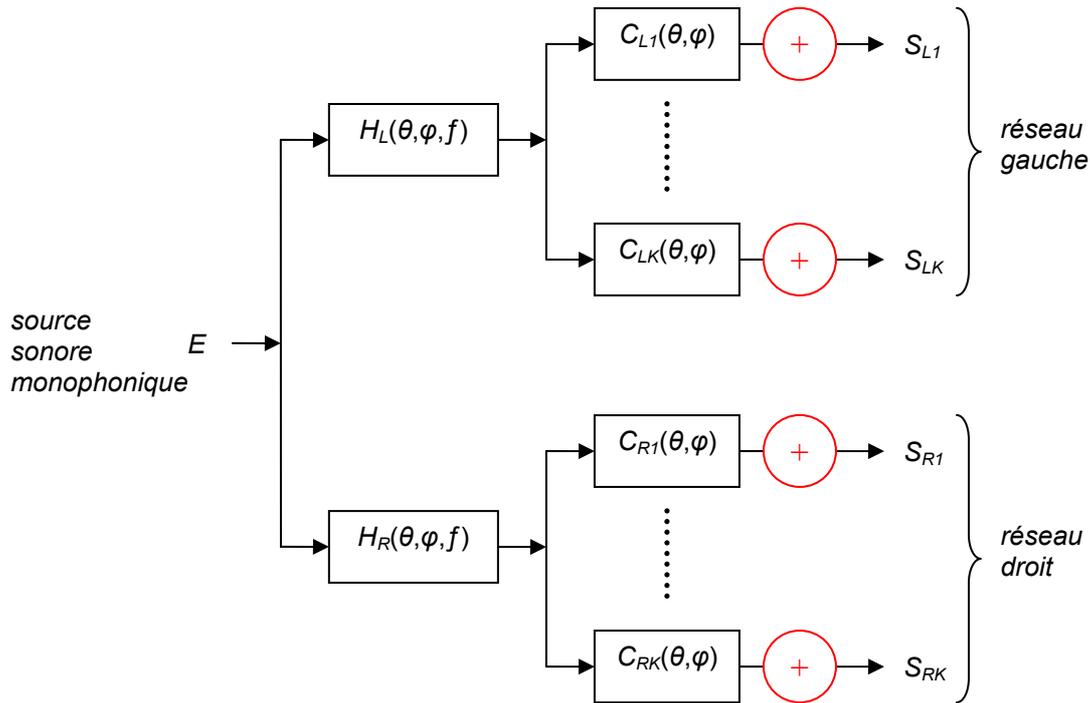


Fig. VI-4 : Implémentation simplifiée de la synthèse holophonique binaurale

Pour chaque position de source primaire, les CAHF peuvent être représentés par une matrice à deux dimensions de taille  $r \times M$ , où  $r$  est le nombre de transducteurs du réseau et  $M$  le nombre de points fréquentiels sur lesquels ils sont estimés. En effectuant une ACP sur chacune de ces matrices et en ne conservant que la composante principale, on peut les séparer chacune en un filtre et  $r$  coefficients : le filtre est alors représentatif des modifications que subit l'onde avant d'atteindre les pavillons, et les coefficients témoignent de la répartition de l'énergie au sein du réseau.

#### 4.1 Décomposition des CAHF associés à la configuration $D$

Les CAHF de la configuration  $D$  sont calculées pour une disposition de transducteur comme décrite à la figure V-10 et en effectuant une optimisation AZ à partir d'un modèle de tête sphérique de 92mm de rayon. La décomposition décrite ci-dessus aboutit aux filtres et aux directivités présentées en figure VI-5.

L'axe interaural étant un axe de révolution du modèle de tête sphérique, les filtres issus de la décomposition sont pratiquement invariants par rapport à l'angle polaire. C'est pourquoi ils sont moyennés pour chaque cône de confusion et représentés uniquement en fonction de l'angle latéral. Ils témoignent bien de l'effet de masquage de la tête et ressemblent fortement aux HRTF d'une tête sphérique.

Les directivités sont toujours maximales dans la direction du transducteur considéré et sont de formes simples. Elles semblent décrire des lois de panning d'intensité simples : il devrait être envisageable de les modéliser à partir de fonctions sinusoïdales.

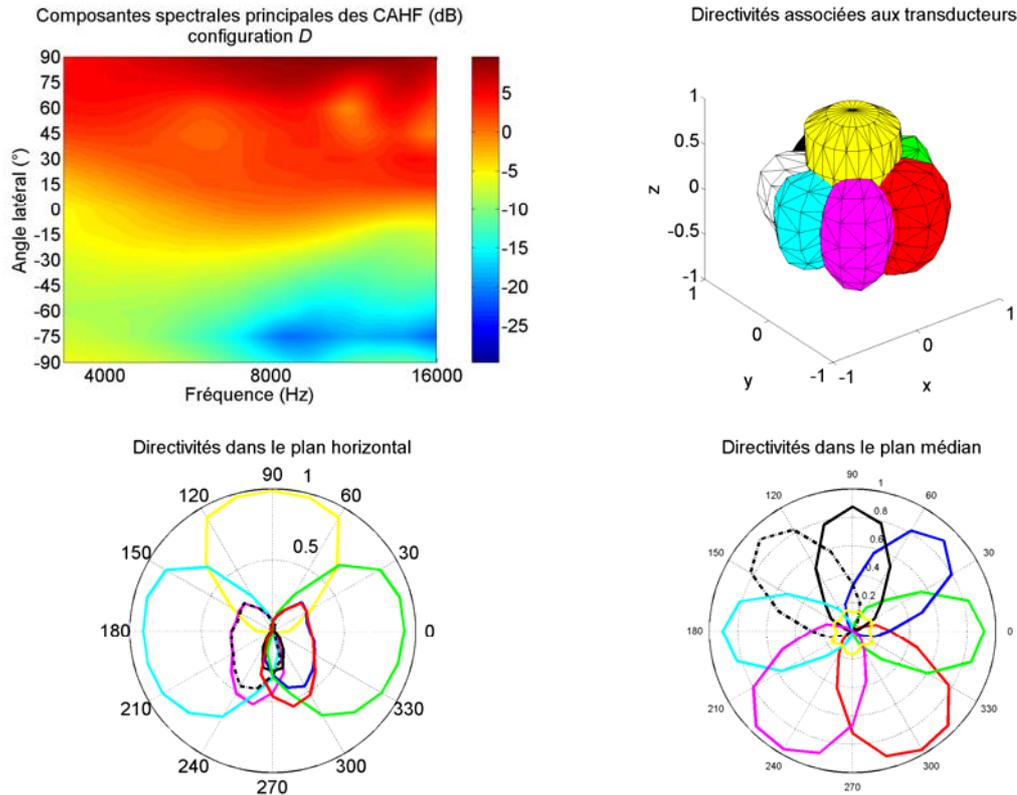


Fig. VI-5 : Filtres et directivités issus de la décomposition

## 4.2 Définition de l'implémentation P-BIN

A partir des remarques précédentes, il semble possible de simplifier l'implémentation ci-dessus en ne considérant non plus les filtres et directivités issus de l'ACP, mais une représentation heuristique basée sur les HRTF de la tête sphérique et des directivités simplifiées (voir Fig. VI-6). On parlera d'implémentation P-BIN pour *Panned Binaural* (binaural balancé).

## 4.3 Evaluation rapide de l'implémentation P-BIN

De manière à évaluer rapidement l'implémentation P-BIN, nous allons simuler la reconstruction des HRTF de la tête KU-100 à partir de cette méthode et la comparer à l'implémentation utilisée lors des tests perceptifs, à savoir une implémentation  $K$ -bicanale ( $K = 8$ ) alimentée par des filtres holophoniques optimisés AZ. Les directivités utilisées sont les directivités simplifiées présentées ci-dessus, et les filtres sont les HRTF de la sphère iso-ITD du modèle KU-100, à savoir une sphère de 97mm de rayon.

Les indices de corrélation et de distance (voir Fig. VI-7) semblent indiquer que l'implémentation P-BIN, à partir des HRTF de la tête sphérique iso-ITD, offre une qualité de reconstruction similaire à une implémentation  $K$ -bicanale, à partir de CAHF optimisés AZ et calculés sur le modèle de tête sphérique iso-ITD.

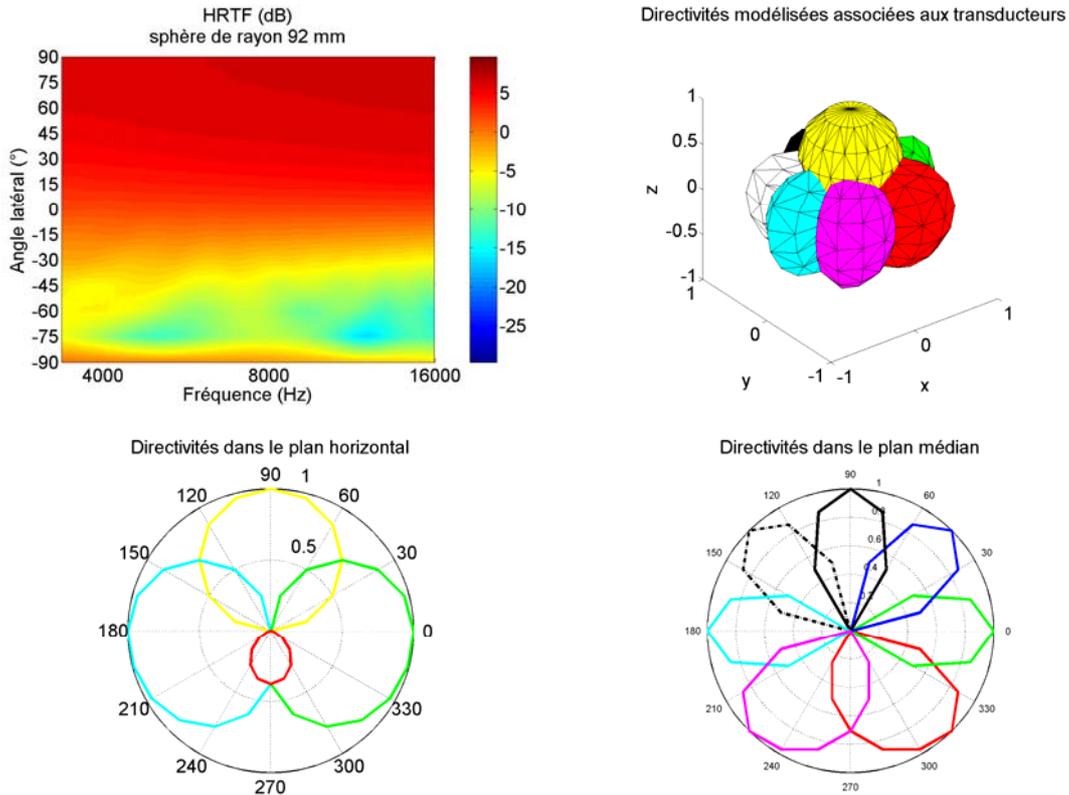


Fig. VI-6 : Filtrés et directivités simplifiés

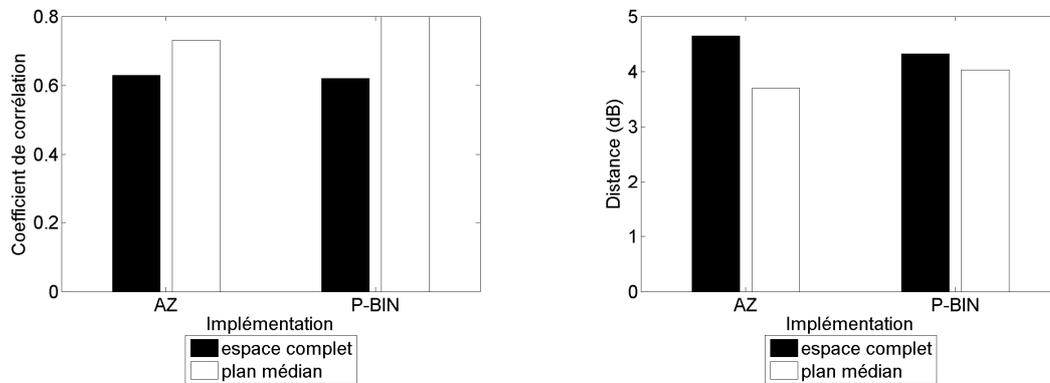


Fig. VI-7 : Corrélation et distance en fonction de l'implémentation

## 5 CONCLUSION

Nous venons de voir qu'il existe plusieurs façons d'implémenter le traitement de spatialisation dédié à l'holophonie binaurale. Le choix entre une implémentation type bicanale ou type multicanale dépend essentiellement des capacités de calcul disponibles et de la volonté ou non de maîtrise quant à la charge de calcul requise (une implémentation de type multicanale permet une meilleure maîtrise dans le sens où le nombre d'opérations de filtrage liées au traitement de spatialisation reste constant, et ce quelque soit le nombre de sources sonores considérées).

L'implémentation simplifiée que nous avons présentée est philosophiquement différentes des autres, dans le sens où elle ne correspond pas à une transposition de méthodes classiquement employé en synthèse binaurale, mais dérive d'une simplification extrême du concept d'holophonie binaurale et d'une ultime décomposition des CAHF : un filtrage contenant les indices de localisation modélisés liés à la tête et une loi panoramique pour définir un angle d'incidence de l'onde recrée par le réseau sur le pavillon, espérant ainsi activer les indices primordiaux de localisation liés au propre pavillon de l'auditeur.

Cette implémentation ouvre la voie à une autre manière d'envisager la spatialisation sur casque multi-transducteur, faisant la transition entre une approche physique, l'holophonie binaurale, et une approche heuristique, la synthèse binaurale panoramique, que nous avons brièvement introduite (implémentation P-BIN), et qu'il conviendrait de développer.



# Conclusion et perspectives

L'utilisation de HRTF générique lors de la synthèse binaurale ne permet généralement que d'offrir une spatialisation médiocre, sujette à divers artéfacts notables comme un fort taux de confusions avant/arrière et une mauvaise restitution de l'élévation des sources virtuelles. Une synthèse binaurale haute qualité nécessite donc l'utilisation de HRTF personnelles. Les techniques les plus robustes pour estimer les filtres d'un auditeur sont la mesure en chambre sourde ou le calcul numérique, mais sont irréalisables pour des applications grand public : en effet, il faudrait soit que l'utilisateur possède le matériel et le savoir-faire nécessaires à l'acquisition de ses propres HRTF, soit qu'il est accès à des établissements capables de les lui fournir. Une alternative est alors d'essayer de proposer des outils d'individualisation permettant de générer des HRTF personnalisées, proches des HRTF personnelles de l'auditeur. La conception de ce genre d'outils est un sujet de recherche majeur au sein de la communauté scientifique travaillant sur la synthèse binaurale, et différentes approches sont envisagée. Elles s'appuient sur le fait que les HRTF traduisent des effets de réflexion et diffraction de l'onde sonore sur les différentes parties du corps humain, et qu'elles sont donc liées à certaines grandeurs morphologiques. Les quatre méthodes principales sont les suivantes :

- l'approche que l'on peut qualifier de « sélective » : sélection d'un jeu de filtres au sein d'une base de données de HRTF à partir d'un critère de proximité entre les grandeurs anthropométriques,
- l'approche statistique : reconstruction d'un jeu de filtres à partir d'analyses statistiques sur une base de données de HRTF et sur la base de grandeurs morphologiques associée,
- l'approche signal : transformation d'un jeu de filtres génériques par warping fréquentiel,
- l'approche dite « structurelle » : synthèse d'un jeu de filtres à partir d'un modèle géométrique adaptable.

Le problème de ces techniques est qu'il est nécessaire d'estimer les paramètres morphologiques d'un utilisateur, soit en les mesurant, soit en les réglant empiriquement. L'obtention de filtres HRTF adaptés à un auditeur demande une démarche rigoureuse de la part de ce dernier. Il est probable que des utilisateurs grand public fassent abstraction de la phase d'individualisation si celle-ci est trop complexe ou trop longue.

Partant du constat que les pavillons d'oreille sont en grande partie responsable de la localisation suivant les axes avant/arrière et haut/bas, et que c'est essentiellement suivant ces composantes spatiales que se démarque une synthèse binaurale individualisée (HRTF individuelles ou adaptées) d'une synthèse binaurale générique, l'idée développée lors de cette thèse consiste à utiliser les pavillons de l'auditeur lors de la phase d'écoute pour ainsi s'affranchir de l'étape d'adaptation de la composante des HRTF qui leur est liée. Le principe est d'exciter les indices de localisation liés aux pavillons en reconstruisant localement l'onde sonore incidente. Pour ce faire, nous avons choisi d'appliquer des techniques holophoniques, et avons ainsi défini le concept d'holophonie binaurale : deux réseaux de transducteurs montés sur un casque sont

positionnés de manière à englober chacun des pavillons, et sont alimentés par des signaux calculés suivant les principes holophoniques. Nous avons ainsi défini un nouveau type de filtre, les CAHF, dépendant de la morphologie du corps humain (hormis les pavillons), et permettant de contrôler, via un casque multi-transducteur adapté, le positionnement de sources sonores virtuelles.

Nous avons tout d'abord étudié la validité du concept d'holophonie binaurale en effectuant différentes simulations via l'utilisation de méthode de calcul numérique BEM. L'influence sur la qualité de reconstruction des HRTF, des différentes techniques de calcul des CAHF (deux techniques généralement employées sur des systèmes holophoniques classiques, à savoir FIG-8 S-LOG et CARD, et une technique originale, FIG-8 T-LOG), ainsi que de la taille de réseau considéré et du modèle d'auditeur utilisé (pavillons bafflés, modèle de tête complet) a été mise en évidence, et trois problèmes majeurs ont été révélés.

- La méthode FIG-8 S-LOG est apparue être la meilleure solution pour le calcul des CAHF, mais son exploitation dans le cas d'un modèle de tête quelconque est difficile, la présence de la tête dans le champ de la source primaire générant un champ diffracté complexe.
- Le temps de calcul BEM des CAHF dans le cas d'un modèle de tête complet quelconque est extravagant, dépassant trois semaines sur un cluster de trois machines de bon gabarit.
- La réduction de la taille des réseaux utilisés (par « taille », nous entendons le nombre de sources secondaires qui composent les réseaux) entraîne une dégradation de la qualité de reconstruction des HRTF, car plus le nombre de sources secondaires est faible, plus la fréquence d'aliasing spatial est basse. Les techniques holophoniques ne peuvent donc pas être appliquées telles quelles sur un casque composé de par exemple seulement 16 haut-parleurs (8 par réseau).

Une solution non triviale pour résoudre le premier problème aurait été par exemple de déterminer une distribution de sources primaires élémentaires générant un champ équivalent au champ diffracté. Nous avons préféré élaborer une méthode de calcul hybride des CAHF, suffisamment satisfaisante, tirant partie des avantages des deux autres méthodes de captation (CARD et FIG-8 T-LOG).

Le deuxième point étant lié à la lourdeur du calcul BEM, et des techniques de calcul numériques en général, lié à la résolution d'un problème acoustique aussi complexe, nous avons choisi d'évaluer l'impact sur la reconstruction des HRTF d'une simplification d'un modèle de tête quelconque en un modèle sphérique iso-ITD : l'avantage d'utiliser un modèle sphérique réside dans le fait que le problème acoustique de diffraction peut alors être résolu analytiquement, et donc bien plus rapidement que par le calcul BEM (une journée au lieu d'une vingtaine).

L'amélioration de la reconstruction des HRTF avec un faible nombre de sources secondaires (8 par réseau en l'occurrence) représente la plus grande part de travail effectué lors de cette thèse. Deux stratégies complémentaires ont été étudiées.

- La première stratégie a consisté à optimiser les signaux alimentant les haut-parleurs (autrement dit les CAHF) pour une distribution de sources secondaires donnée. Une application directe des principes holophoniques implique d'utiliser

l'information acoustique présente au lieu ponctuel d'une source secondaire pour calculer le filtre associé. Or lorsque le nombre de sources secondaires est faible, la part de surface associée à chacune d'elle augmente. L'information ponctuelle peut ne plus être représentative de l'information globale sur cette zone. L'idée a donc été d'utiliser toute l'information présente à la surface décrite par le réseau pour calculer les CAHF, et non plus l'information mesurée en un seul point, en la pondérant de manière adéquate. Deux points de vue ont été abordés pour calculer ses fonctions de pondération : le premier a été de simplement fenêtrer spatialement la surface décrite par le réseau. Seule la transposition du cas unidimensionnel d'une fenêtre carrée a été étudiée, conduisant alors à une simple moyenne de l'information acoustique mesurée sur la part de surface associée à chaque source secondaire (optimisation AZ). Il est tout à fait envisageable de transposer d'autres fenêtrages, de type gaussiens par exemple. La deuxième approche a été de calculer les fonctions de pondération par rapport au récepteur qu'est le pavillon d'oreille. En utilisant une base de données de plusieurs jeux de PRTF, filtres définissant les fonctions de transfert de la surface du réseau vers l'entrée du canal auditif, il a été possible d'établir des fonctions de pondération permettant d'améliorer la reconstruction des HRTF (optimisation PROJ). Par nature cette démarche génère des fonctions de pondération dépendantes de la morphologie des pavillons auxquels sont associés les jeux de PRTF considérés.

- La deuxième stratégie a consisté à optimiser la disposition des sources secondaires au sein du réseau. Là encore, deux approches ont été considérées. A l'instar de l'optimisation PROJ, la première idée a été de se placer du point de vue du pavillon d'oreille. Celui-ci pouvant être assimilé à un récepteur possédant une certaine directivité, différentes analyses statistiques ont été conduites sur plusieurs jeux de PRTF et ont permis de définir des directions prédominantes (optimisation PCA, ICA et SUBSET). La deuxième idée a été de se baser sur des considérations perceptives pour définir de manière heuristique une configuration de réseau améliorée (optimisation CIRC) se focalisant sur une reconstruction des ondes dans le plan médian.

Les simulations de reconstruction de HRTF ont montré que l'amélioration apportée par l'optimisation des CAHF est en général plus importante que celle induite par une optimisation de l'emplacement des sources secondaires, hormis pour la configuration CIRC. La technique de calcul AZ permet d'atteindre un niveau de corrélation d'environ 60% entre les HRTF cibles et les HRTF reconstruites, et n'étant pas liée à une base de données de pavillons, elle peut être utilisée par défaut comme méthode de calcul des CAHF. La méthode PROJ permet de dépasser les 90% de corrélation, mais est cependant dépendante de la base de données de PRTF considérée.

Pour éprouver les résultats théoriques issus des simulations, nous avons voulu évaluer perceptivement différentes configurations. Un prototype modulaire de casque a été construit, permettant de repositionner les transducteurs à volonté. Trois configurations différentes ont été testées : une première configuration de référence (disposition équirépartie des transducteurs et un calcul des CAHF suivant la méthode AZ), une deuxième configuration optimisée entièrement liée à la base de données de PRTF que nous avons considérée (disposition SUBSET et calcul PROJ), et une troisième configuration privilégiant la reconstruction des HRTF dans le plan médian (disposition CIRC et calcul

AZ). Les résultats des tests de localisations ont bien montré l'amélioration apportée par les deuxième et troisième configurations par rapport à la première, mais contrairement aux résultats des simulations, ils ont indiqué que la troisième configuration est meilleure que la deuxième. Des différences morphologiques entre les pavillons des auditeurs testés et des modèles utilisés pour générer la base de données de PRTF sur laquelle s'appuient les optimisations SUBSET et PROJ peuvent expliquer ces résultats, notamment les différences de positionnement angulaire des pavillons (positionnement non normalisé). Ce type d'optimisation nécessiterait certainement d'être adapté à un auditeur donné, ce qui sous-entend un processus d'individualisation peut-être encore plus complexe que l'individualisation des HRTF. Il semblerait donc qu'une disposition de réseau et un calcul des CAHF liés à une base de données de PRTF ne soient pas la meilleure solution en situation réelle.

Nous avons ensuite effectué des tests de localisation pour confronter un système d'holophonie binaurale sur une version modifiée de la troisième configuration (disposition CIRC+ et calcul AZ) à une synthèse binaurale utilisant des HRTF individuelles mesurées en chambre anéchoïque. Les résultats sont concluants et montrent que l'holophonie binaurale est une alternative prometteuse à la synthèse binaurale individuelle.

Pour finir, nous avons abordé la mise en œuvre du traitement de filtrage lié à la spatialisation sonore sur un tel système. Nous avons notamment établi deux transpositions à l'holophonie binaurale de l'encodage multicanal couramment utilisé en synthèse binaurale (approches  $K$ -multicanale et multicanale- $K$ ). Une dernière optimisation a été rapidement abordée, dont l'objectif était essentiellement de réduire l'implémentation de l'holophonie binaurale à un filtrage par réseau et un panoramique d'amplitude entre les transducteurs d'un réseau (implémentation P-BIN). Plus qu'une simplification de traitement, cette optimisation ouvre la voie à un nouveau concept, celui d'une synthèse binaurale panoramique.

Différents points restent à améliorer ou à étudier, voici les deux qui nous ont paru être les plus importants. Le premier serait de trouver un moyen pour améliorer le calcul des CAHF dans le cas de problèmes complexes (par exemple une source primaire en présence d'un objet diffractant). Comme nous l'avons déjà mentionné, une solution serait de déterminer une distribution de sources primaires rayonnant un champ équivalent au champ diffracté pour ainsi pouvoir évaluer une approche FIG-8 S-LOG de la captation holophonique, qui correspond à une généralisation de l'approche de Rayleigh. Dans le cas où une résolution BEM est envisagée, une distribution de sources équivalentes est donnée implicitement par les matrices  $A$  et  $B$  de l'expression (III-17), la théorie de l'holophonie et du calcul par éléments de frontière étant similaire, mais cela implique d'avoir accès à un code de calcul BEM. Une autre solution serait de totalement se détacher de la théorie de l'holophonie (du moins pour le calcul des CAHF), et de faire appel à la théorie des problèmes inverses : essayer de retrouver le champ à émettre par une distribution de sources monopolaires à la surface du réseau en connaissant le champ à tenter d'approcher à l'intérieur du volume du réseau. Cette approche mériterait certainement d'être considérée. Le second point serait de continuer les simulations sur des configurations optimisées telles que SUBSET PROJ et CIRC, pour tenter de quantifier l'impact d'un mauvais positionnement du réseau par rapport au pavillon, ou évaluer la dégradation de reconstruction des HRTF liée à l'utilisation d'un modèle de pavillon ne faisant pas partie du panel sur lequel se basent les optimisations, et essayer

alors d'expliquer à partir de mesures objectives les différences que nous avons mises en évidence entre les résultats théoriques et les résultats pratiques.

Bien que certaines questions restent sans réponse, les travaux menés au cours de cette thèse ont permis de définir et d'optimiser le concept d'holophonie binaurale, et d'aboutir à la conception d'un système de spatialisation sur casque multi-transducteur utilisant des CAHF génériques dont l'efficacité est comparable à celle d'un système de synthèse binaurale utilisant des HRTF individuelles. L'objectif principal était le prototypage d'un tel système de restitution dans l'optique d'un futur développement commercial, mais une autre application de l'holophonie binaurale a été implicitement validée, à savoir la composition/décomposition des HRTF à partir des principes holophoniques. De manière comparable à la méthode de composition/décomposition structurelle d'Algazi *et al.* [ADMT01], les HRTF sont décomposées en une composante liée à la tête (et au corps), les CAHF, et une composante liée aux pavillons d'oreille, les PRTF (champ proche dans notre cas). La méthode holophonique est certes plus complexe à mettre en œuvre car au lieu de substituer une HRTF par une série de deux filtres, elle fait intervenir autant de série de deux filtres que de sources secondaires considérées. Cependant, alors que la décomposition structurelle offre principalement une solution pour reconstruire des HRTF ipsilatérales (une autre étude d'Algazi *et al.* présente une modélisation des HRTF contralatérales à partir des HRTF ipsilatérales [ADA99]), la méthode holophonique offre une solution directe pour reconstruire les HRTF contralatérales, car les PRTF sont alors liées à la position des sources secondaires, et non plus à la position de la source primaire. L'approche P-BIN peut alors également permettre de simplifier la composition des HRTF en remplaçant les CAHF par un seul filtre lié à la tête et un jeu de coefficients.

L'intérêt de l'holophonie binaurale est donc double : elle peut être utilisée directement pour alimenter un système de spatialisation des sons sur un casque dédié, et indirectement comme outil de couplage/découplage et d'analyse des dépendances morphologiques des HRTF.



# Références bibliographiques

- [AAD01a] V.R. Algazi, C. Avendano and R.O. Duda. Elevation localization and head-related transfer function analysis at low frequencies. *J. Acoust. Soc. Am.*, 109(3), 2001.
- [AAD01b] V.R. Algazi, C. Avendano and R.O. Duda. Estimation of a spherical-head model from anthropometry. *J. Acoust. Soc. Am.*, 49 :472-478, 2001.
- [AD02] V.R. Algazi and R.O. Duda. Approximating the head-related transfer function using simple geometric models of the head and torso. *J. Acoustic. Soc. Am.*, 112(5) :2053-2064, 2002.
- [ADA99] C. Avendano, R.O. Duda and V.R. Algazi. Modeling the contralateral HRTF. *Presented at the 16<sup>th</sup> AES International Conference*, Rovaniemi, 1999.
- [ADMT01] V.R. Algazi, R.O. Duda, R.P. Morison and D.M. Thompson. Structural composition and decomposition of HRTFs. *Proc. IEEE Workshop*, New Paltz, 2001.
- [ADTA01] V.R. Algazi, R.O. Duda, D.M. Thompson and C. Avendano. The CIPIC HRTF Database. *Proc. IEEE Workshop*, New Paltz, 2001.
- [ASS90] F. Asano, Y. Suzuki and T. Sone. Role of spectral cues in median plane localization. *J. Acoustic. Soc. Am.*, 88 :159-168, 1990.
- [Beg94] D.R. Begault. *3-D sound for virtual reality and multimedia*. Academic Press, Cambridge, Massachusetts, 1997.
- [BKW04] A. Blum, B.F.G. Katz and O. Warusfel. Eliciting adaptation to non-individual HRTF spectral cues with multi-modal training. *Proc. CFA/DAGA*, Strasbourg, 2004.
- [BHM90] R.A. Butler, R.A. Humanski, and A.D. Musicant. (1990). Binaural and monaural localization of sound in two-dimensional space. *Perception*, 19 :241–256, 1990.
- [Bla97] J. Blauert. *Spatial Hearing, the psychophysics of human sound localization*. MIT Press, first published in 1974, re-edited in 1997.
- [BNK05] S. Busson, R. Nicol and B.F.G. Katz. Subjective investigations of the interaural time difference in the horizontal plane. Presented at the 118<sup>th</sup> AES Convention, Barcelona, 2005.
- [BNW04] S. Busson, R. Nicol and O. Warusfel. Influence of the ears canal location on spherical head model for the individualized interaural time difference. *Proc. CFA/DAGA*, Strasbourg, 2004.
- [Bus05] S. Busson. *Individualisation d'indices acoustiques pour la synthèse binaurale*. PhD thesis, Université de la Méditerranée - Aix-Marseille II, 2005.

- [But86] R.A. Butler. The bandwidth effect on monaural and binaural localization. *Hear. Res.*, 21 :67-73, 1986.
- [BVV93] A.J. Berkhout, D. de Vries and P. Vogel. Acoustic control by wave field synthesis. *J. Acoust. Soc. Am.*, 93 :2764-2778, 1993.
- [BWA01] D.R. Begault, E.M. Wenzel and M.R. Anderson. Direct comparison of the impact of head-tracking, reverberation and individualized head-related transfer functions on the spatial perception of a virtual speech source. *J. Audio Eng. Soc.*, 49(10) :904-916, 2001.
- [Car96] S. Carlile. *Virtual auditory space: Generation and applications*. Springer-Verlag, 1996.
- [Cha96] N. Chateau. *Localisation de sources sonores multiples dans l'hémisphère supérieur*. PhD thesis, Université de la Méditerranée - Aix-Marseille II, 1996.
- [Cipic] CIPIC HRTF database.  
[http://interface.cipic.ucdavis.edu/CIL\\_html/CIL\\_HRTF\\_database.htm](http://interface.cipic.ucdavis.edu/CIL_html/CIL_HRTF_database.htm)
- [Com94] P. Common. Independent component analysis, a new concept? *Signal Processing*, 36(3), 1994.
- [CS93] J.F. Cardoso and A. Souloumiac. Blind beamforming for non Gaussian signals. *Proc. IEEE*, 140(6) :362-370, 1993.
- [DAA99] R.O. Duda , C. Avendano and V.R. Algazi. An adaptable ellipsoidal head model for the interaural time difference. *Proc. IEEE ICASSP*, Phoenix, 1999.
- [DM98] R.O. Duda and W.L. Martens. Range dependence of the response of a spherical head model. *J. Acoust. Soc. Am.*, 104(5) :3048-3058, 1998.
- [FBV04] J. Fels, P. Buthmann and M. Vörländer. Head-related transfer functions of children. *Acta Acoustica united with Acoustica*, 90 :918-927, 2004.
- [Fra03] P. di Francescantonio. *VNoise theoretical manual*.  
<http://www.sts-soft.com/VNoise.aspx>
- [Gen84] K. Genuit. *A model for the description of outer-ear transmission characteristics*. PhD thesis, Rhenish-Westphalian Technical University, 1984.
- [GK07a] R. Greff and B.F.G. Katz. Perceptual evaluation of HRTF notches versus peaks for vertical localisation. *Proc. ICA*, Madrid, 2007.
- [GK07b] R. Greff and B.F.G. Katz. Round robin comparison of HRTF simulation results: Preliminary results. *Presented at the 123<sup>rd</sup> AES Convention*, New York, 2007.
- [GK08] R. Greff and B.F.G. Katz. Circum-aural transducer arrays for binaural synthesis. *Proc. Acoustics'08*, Paris, 2008.
- [Han94] H.L. Han. Measuring a dummy head in search of pinna cues. *J. Audio Eng. Soc.*, 42 :15-36, 1994.

- [HK74a] J. Hebrank and D. Wright. Spectral cues used in the localization of sound sources on the median plane. *J. Acoust. Soc. Am.*, 56 :1829-1834, 1974.
- [HK74b] J. Hebrank and D. Wright. Are two ears necessary for localization of sound sources on the median plane? *J. Acoust. Soc. Am.*, 56 :935–938, 1974.
- [Jade] JADE algorithm.  
<http://www.tsi.enst.fr/~cardoso/Algo/Jade/jade.m>
- [JCCS04] C. Jin, A. Corderoy, S. Carlile and A. van Shaik. Contrasting monaural and interaural spectral cues for human sound localization. *J. Acoust. Soc. Of Am.*, 115(6) :3124-3141, 2004.
- [JLLCC00] C. Jin, P. Leong, J. Leung, A. Corderoy and S. Carlile. Enabling individualized virtual auditory space using morphological measurements. *International Symposium on Multimedia Information Processing*, Sydney, 2000.
- [JLP99] J.-M. Jot, V. Larcher and J.-M. Pernaux. A comparative study of 3D audio encoding and rendering techniques. *Presented at the 16<sup>th</sup> AES Conference*, Rovaniemi, 1999.
- [JLW95] J.-M. Jot, V. Larcher and O. Warusfel. Digital signal processing issues in the context of binaural and transaural stereophony. *Presented at the 98<sup>th</sup> AES Convention*, Paris, 1995.
- [JWL98] J.-M. Jot, S. Wardle and V. Larcher. Approaches to binaural synthesis. *Presented at the 105<sup>th</sup> AES Convention*, San Francisco, 1998.
- [Kah00] Y. Kahana. *Numerical modelling of the head-related transfer function*. PhD thesis, University of Southampton, 2000.
- [Kat98] B.F.G. Katz. *Measurement and calculation of individual head-related transfer functions using a boundary element model including the measurement and effect of skin and hair impedance*. PhD thesis, Pennsylvania State University, 1998.
- [Kat01a] B.F.G. Katz. Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation. *J. Acoust. Soc. Of Am.*, 110(5), 2001.
- [Kat01b] B.F.G. Katz. Boundary element method calculation of individual head-related transfer function. II. Impedance effects and comparisons to real measurements. *J. Acoust. Soc. Of Am.*, 110(5), 2001.
- [KB07] B.F.G. Katz and D.R. Begault. Round robin comparison of HRTF measurement system: Preliminary results. *Proc. ICA*, Madrid, 2007.
- [KFCS82] L.E. Kinsler, A.R. Frey, A.B. Coppens and J.V. Sanders. *Fundamentals of acoustics*. John Wiley & Sons, 1982.
- [KN00] Y. Kahana and P.A. Nelson. Spatial acoustic mode shapes of the human pinna, *Presented at the 109<sup>th</sup> AES Convention*, Los Angeles, 2000.

- [KNPC99] Y. Kahana, P.A. Nelson, M. Petyt and S. Choi. Numerical modeling of the transfer functions of a dummy-head and of the external ear. *Presented at the 16<sup>th</sup> AES International Conference*, Rovaneimi, 1999.
- [Lar01] V. Larcher. *Techniques de spatialisation des sons pour la réalité virtuelle*. PhD thesis, Université Pierre et Marie Curie - Paris VI, 2001.
- [LB02] E.H.A. Langendijk and A.W. Bronkhorst. Contribution of spectral cues to human sound localization. *J. Acoust. Soc. Am.*, 112 :1583, 2002.
- [Listen] LISTEN HRTF database.  
<http://recherche.ircam.fr/equipes/salles/listen/>
- [LJGW00] V. Larcher, J.-M. Jot, J. Guyard and O. Warusfel. Study and comparison of efficient methods for 3D audio spatialization based on linear decomposition of HRTF data. *Presented at the 108<sup>th</sup> AES Convention*, Paris, 2000.
- [LJV98] V. Larcher, J.-M. Jot and G. Vandernoot. Equalization methods in binaural technology. *Presented at the 105<sup>th</sup> AES Convention*, San Francisco, 1998.
- [MI68] P. Morse and K.U. Ingrad. *Theoretical acoustics*. Mc Graw Hill, 1968.
- [Mid92] J.C. Middlebrooks. Narrow-band sound localization related to external ears acoustics. *J. Acoust. Soc. Am.*, 92 :2607-2624, 1992.
- [Mid99a] J.C. Middlebrooks. Individual differences in external-ear transfer functions reduced by scaling in frequency. *J. Acoust. Soc. Am.*, 106(3) :1480-1492, 1999.
- [Mid99b] J.C. Middlebrooks. Virtual localization improved by scaling non-individualized external-ear functions in frequency. *J. Acoust. Soc. Am.*, 106(3) :1493-1509, 1999.
- [MII03] M. Morimoto, K. Iida and M. Itoh. Upper hemisphere sound localization using head-related transfer functions in the median plane and interaural differences. *Acoust. Sci. & Tech.*, 24 :267–275, 2003.
- [MMG89] J.C. Middlebrooks, J.C. Makous and D.M. Green. Directional sensitivity of sound-pressure levels in the human ear canal. *J. Acoust. Soc. Am.*, 86 :89-107, 1989.
- [Mor01] M. Morimoto. The contribution of two ears to the perception of vertical angle in sagittal planes. *J. Acoust. Soc. Am.*, 109 :1596–1603, 2001.
- [MSHJ95] H. Moller, M.F. Sorensen, D. Hammershoi and C.B. Jensen. Head-related transfer functions of human subjects. *J. Audio. Eng. Soc.*, 43(5) :300-321, 1995.
- [Nic99] R. Nicol. *Restitution spatialisée sur une zone étendue - application à la téléprésence*. PhD tesis, Université du Maine, 1999.
- [NKTH97] P.A. Nelson, O. Kirkeby, T. Takeuchi and H. Hamada. Sound fields for the production of virtual acoustic images. *Journal of Sound and Vibration*, 2 :386-396, 1997.

- [OP84a] S.R. Oldfield and S.P.A. Parker. Acuity of sound localization: A topography of auditory space. I. Normal hearing conditions. *Perception*, 13 :581-600, 1984.
- [OP84b] S.R. Oldfield and S.P.A. Parker. Acuity of sound localization: A topography of auditory space. II. Pinna cues absent. *Perception*, 13 :601-617, 1984.
- [OP86] S.R. Oldfield and S.P.A. Parker. Acuity of sound localization: A topography of auditory space. III. Monaural hearing conditions. *Perception*, 15 :67–81, 1986.
- [Per03] J.M. Pernaux. *Spatialisation du son par les techniques binaurales : application aux services de télécommunications*. PhD thesis, Institut National Polytechnique de Grenoble, 2003.
- [Pie91] A.D. Pierce. *Acoustics - An introduction to its physical principles and applications*. Published by Acoust. Soc. Am., second printing, 1991.
- [PMOCM00] J. Plogsties, P. Minnaar, S.K. Olesen, F. Christensen, H. Moller. Audibility of all-pass components in head-related transfer function. Presented at the 108<sup>th</sup> AES Convention, Paris, 2000.
- [PN97] S. Perret and W. Noble. The effect of head rotations on vertical plane sound localization. *J. Acoust. Soc. Of Am.*, 102(4) :2325-2332, 1997.
- [Rad94] M. Radeau. Auditory-spatial interaction and modularity. *Cahiers de Psychologie Cognitive*, 13 :3-51, 1994.
- [RDY04] V.C. Raykar, R. Duraiswami and B. Yegnanarayama. Extracting frequencies of the pinna spectral notches in measured head related impulse response. *Technical report CS-TR-4609*, Perceptual Interface and Reality Laboratory, University of Maryland, 2004.
- [RR05] S.G. Rodriguez and M.A. Ramirez. Extracting and modelling approximated pinna-related transfer functions from HRTF data. *Proc. ICAD*, 2005.
- [Sas97] D. Sastrapradja. *Sound localization simulation via wavefront recreation*. Master's thesis, Pennsylvania State University, 1997.
- [Sha82] E.A.G. Shaw. *Localization of sound: Theory and applications*. Amphora press, 1982.
- [SM94] W.H. Slattery and J.C. Middlebrooks. Monaural sound localization: Acute versus chronic unilateral impairment. *Hear. Res.*, 75 :38–46, 1994.
- [The80] G. Theile. *Localization in the superposed sound field*. PhD thesis, Technical University of Berlin, 1980.
- [The86] G. Theile. On the standardization of the frequency response of high-quality studio headphones. *J. Audio Eng. Soc.*, 34(12), 1986.

- [WAKW93] E.M. Wenzel, M. Arruda, D.J. Kistler and F.L. Wightman. Localization using nonindividualized head-related transfer functions. *J. Acoust. Soc. Am.*, 94(1) :111-123, 1993.
- [Wei84] S. Weinrich. Sound-field calculations around the human head. Technical report, Technical University of Denmark, 1982.
- [WK89] F.L. Wightman and D.J. Kistler. Headphone simulation of free-field listening. II. Psychophysical validation. *J. Acoust. Soc. Am.*, 85(2) :868-878, 1989.
- [WK92] F.L. Wightman and D.J. Kistler. The dominant role of low-frequency interaural time differences in sound localization. *J. Acoust. Soc. Am.*, 91 :1648-1661, 1992.
- [WK97] F.L. Wightman and D.J. Kistler. Monaural sound localization revisited. *J. Acoust. Soc. Am.*, 101 :1050–1063, 1997.
- [WK99] F.L. Wightman and D.J. Kistler. Resolution of front-back ambiguity in spatial hearing by listener and source movement. *J. Acoust. Soc. Am.*, 105(5) :2841-2853, 1999.
- [WO04] M.M. van Wanrooij A.J. van Opstal. Contribution of head shadow and pinna cues to chronic monaural sound localization. *J. Neurosci.*, 24 :4163–4171, 2004.
- [ZDD02] D.N. Zotkin, R. Duraiswami and L.S. Davis. Customizable auditory displays. *Proc. ICAD*, Kyoto, 2002.
- [ZDDRM02] D.N. Zotkin, R. Duraiswami, L.S. Davis, V. Raykar, A. Mohan. Virtual audio system customization using visual matching of ear parameters. *Proc. ICPR*, Quebec, 2002.

# Publications



**ARTICLE A – PRESENTE AU 19<sup>EME</sup> CONGRES ICA**





## PERCEPTUAL EVALUATION OF HRTF NOTCHES VERSUS PEAKS FOR VERTICAL LOCALISATION

PACS: 43.66.Pn

Greff, Raphaël<sup>1</sup>; Katz, Brian F.G.<sup>2</sup>

<sup>1</sup>A-Volute; 4120 route de Tournai, Douai, France; [raphael.greff@a-volute.com](mailto:raphael.greff@a-volute.com)

<sup>2</sup>LIMSI – CNRS; BP 133, Orsay, France; [brian.katz@lmsi.fr](mailto:brian.katz@lmsi.fr)

### ABSTRACT

Head-Related Transfer Functions (HRTF) vary with frequency, azimuth, and elevation. Binaural cues such as the interaural time and level differences are the primary cues for estimation of the azimuth and distance of a sound source. Monaural information such as the spectral content of the HRTF is known to be an important cue for estimation of elevation, localisation in the median plane, resolution of front-back confusion, etc.

HRTF spectral cues can be represented by an ensemble of peaks and notches in the frequency spectrum. Some previous studies indicate that the spectral notches are responsible for localisation of elevated sound sources in the median plane, while others discuss the importance of spectral peaks. This paper presents results of an experiment aimed at studying this question. A listening test using modified HRTF spectral cues has been carried out to evaluate the localisation bias resulting from spectral peak and notch modifications.

Subjects performed auditory localisation tasks using three HRTF sets comprising a reference (selected from the LISTEN database), a “peaks only” and a “notch only” set (both based on the reference set). The modifications were made by collapsing the peaks or notches to the average spectrum. The average spectrum was obtained using spectral smoothing.

### INTRODUCTION

Humans have the ability to localize sound sources, to determine their range, azimuth, and elevation. Interaural Time Difference (ITD) and Interaural Level Differences (ILD) are known to provide primary cues for localisation in the horizontal plane, i.e. to determine the azimuth of a sound source [1, 2]. However these differences are not sufficient to uniquely locate sounds in 3D space, resolving positions on the so called “cones of confusion” defined by constant ITD and ILD cues. For a source located in the median plane, for example, ITD and ILD are both zero, and spectral cues are the only information available for source localisation [3]. These spectral features are due to diffraction and reflection of the incident sound on the body, shoulder, head, and pinna. The prominent features contributed by the pinna are the sharp spectral notches and peaks at higher frequencies. There is evidence to support the hypothesis that the spectral notches due to the pinna are important cues for vertical localisation [4, 5, 6]. At the same time, the physics that creates a spectral notch at one frequency also creates a spectral peak at another.

This article presents an experiment whose aim is to investigate the influence of spectral notches versus the influence of spectral peaks for vertical localisation. The experiment consists of an auditory localisation test in the median plane for three HRTF sets comprising a reference set (A) and two other sets derived from this reference, one where the pattern of notches is kept (B), the other where the pattern of peaks is kept (C). For each subject, reference A is chosen from the LISTEN database [8] based on subjective rated spatial impression. An identical signal is presented to left and right ears using only the left ear data component of the HRTF set, forcing the ITD and ILD to be zero.

## HRTF PROCESSING

The aim of the processing applied to HRTF set A is to determine the boundary spectrum used to distinguish spectral notches from spectral peaks. "Notches only" set B is obtained by collapsing peaks to this boundary spectrum. In a similar manner "peaks only" set C is obtained by collapsing the notches. An important question is how to determine this boundary set. Smoothing by its nature appears to be a potentially good approach. By definition, smoothing attenuates differences, and therefore will diminish the amplitude of notches and peaks.

For the experiment, notches and peaks need to be separated in a hearing sense, i.e. according to the sensibility of the human auditory system. Therefore, data was warped on the frequency bark scale and transformed into decibels before applying smoothing. Different algorithms of smoothing processing exist, and the choice is somewhat a subjective task. The RLOESS method with a span taking in account the whole sampled magnitude spectrum seems to be a good method for keeping the general pattern of the data. Figure 1 shows the result of this processing for a given HRTF position.

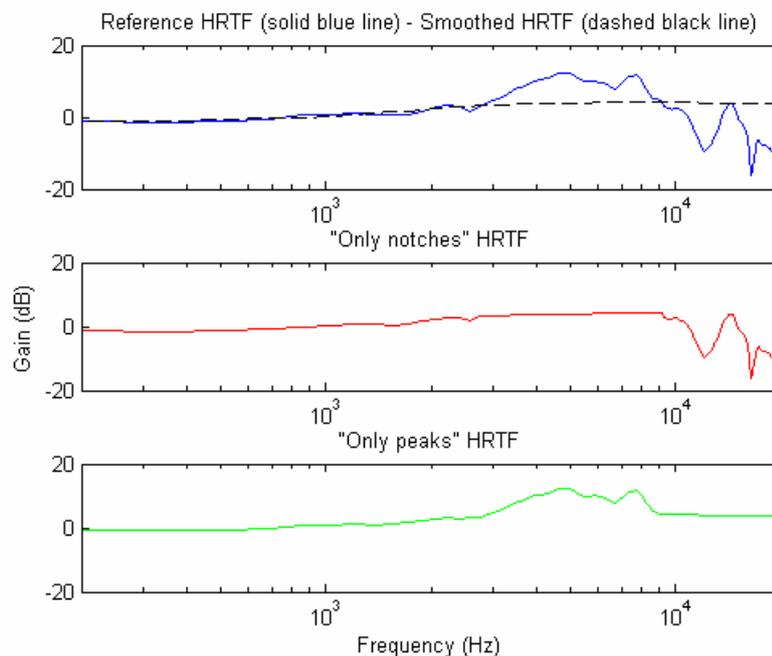


Figure 1.-Reference A HRTF with boundary spectrum. Processed B and C HRTF sets.

## EXPERIMENT

An auditory localisation test was carried by 8 subjects, repeated 3 times. For each subject, the best matching HRTF set from the LISTEN database was chosen. The reference set A was processed to obtain the "notches only" B set and the "peaks only" C set. A total of 19 positions on the median plane were chosen, equally spaced from  $-45^\circ$  to  $225^\circ$ , where  $0^\circ$  corresponds to the front direction horizontal plane,  $90^\circ$  to above and  $180^\circ$  to the rear. Each position was presented from each set (A, B, and C), resulting in a total of 171 sounds for the subjects to localize (3 sets  $\times$  19 positions  $\times$  3 repetitions).

The sound stimulus used was a 200 ms pink noise burst modulated at 20 Hz. Even though an HRTF selection procedure was used, due to the absence of individual HRTFs, a novel approach was chosen in the hopes of improving subjects performance in the localisation task. This approach consisted of coupling the stimulus to localise (SL) with a spatially referenced stimulus (SR) corresponding to the frontal direction ( $0^\circ$ ) for the same HRTF set. The test stimulus S was constructed by coupling SR and SL with a 200 ms silence between SR and SL, with SR always preceding SL. The stimulus couple S was repeated 5 times. Subjects were told the position of the reference stimuli SR. As such, it was hoped that they could use the reference position to calibrate the HRTF and thereby correctly determine position of the sound to localise; SL. Subjects indicated the position of the target sound SL on a visual interface.

## RESULTS AND DISCUSSION

Two analyses were carried out on the responses of the subjects. The first was an evaluation of the median plane localisation bias between the reported angle and the target angle. The second was an evaluation of the localisation difference, between either the “only notches” set B and the reference set A, or between the “only peaks” set C and the reference set A.

### Absolute localisation error

Analysing the error between reported angle and target angle for each subject shows that half of the 8 subjects were not able to properly localise sounds even with the reference set A, supposedly the best match. Two patterns emerge in the data of these subjects, either a random pattern, or a concentration pattern at angles 0° (front) and 180° (rear). Figure 2 shows an example of the random pattern.

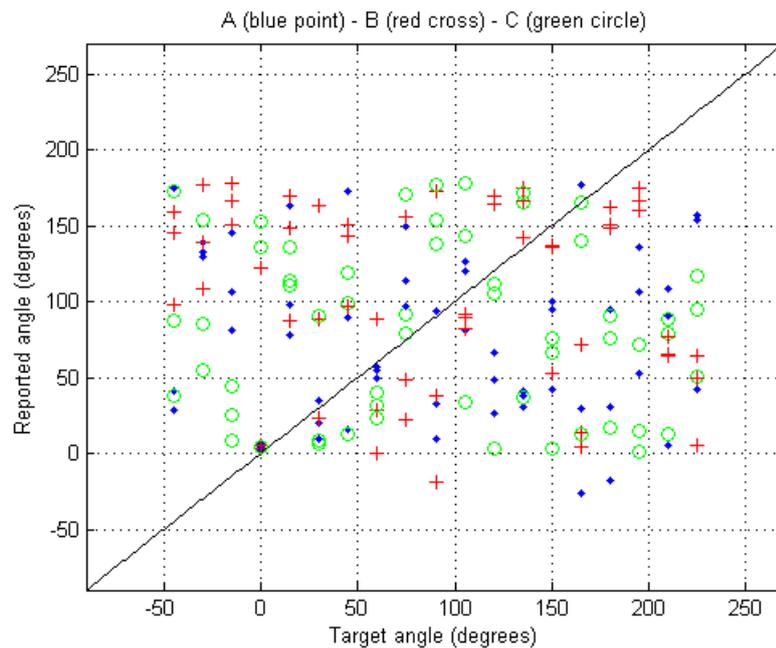


Figure 2.-Example of random localisation response pattern.

It was decided to reject the data of these “poor localising” subjects and to concentrate further analysis on the 4 subjects that were able to localise with set A. It is understood that 4 subjects is a small selection, but it permits the extraction of some basic characteristic patterns.

Another phenomenon appeared on the remaining data, a phenomenon that subjects reported after the test. Subjects explain that even if they felt an elevation impression, often, they were not able to decide between the front and the back hemisphere. It is uncommon to deal with front-back confusion in the median plane, but this subjective report indicated that a consideration should be made to correct this confusion in order to be able to exploit the data. A recent study [7] shows that there exists frontal plane symmetry in the magnitude spectrum of HRTF data that can explain front-back confusions even in the median plane. Figure 3 highlights the problem of front-back confusion.

It was therefore decided to take the symmetric position of reported data when target and reported positions were not in the same front-back hemisphere.

As subjects performed localisation tasks with different reference sets, statistics are done only per position for both analyses. Figure 4 shows the mean reported angles by position for each of the three HRTF sets A, B and C. Table I indicates the mean unsigned error for each set.

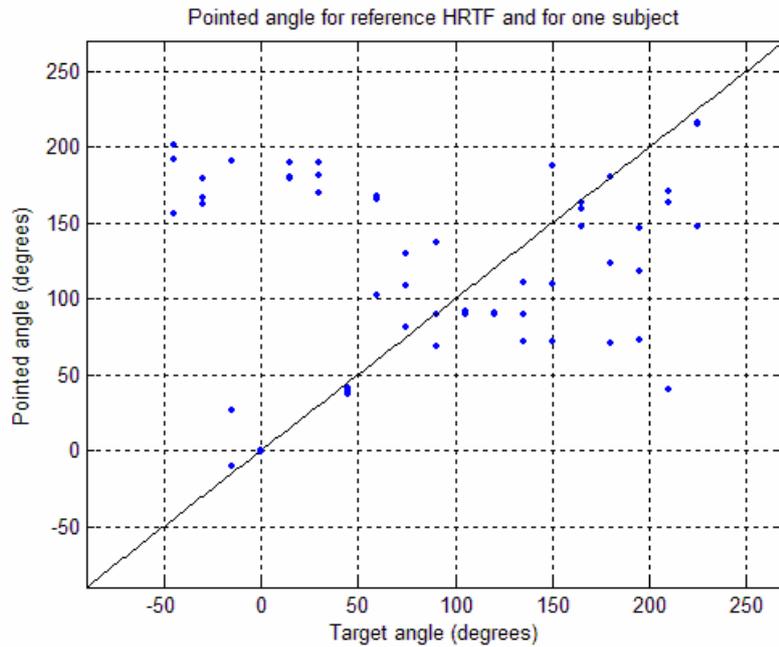


Figure 3.-Example of front-back confusion in the median plane.

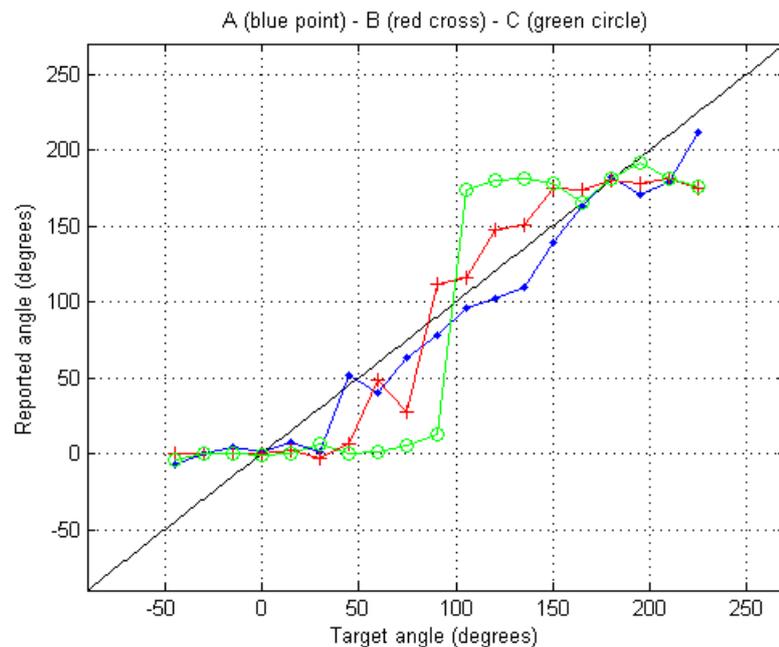


Figure 4.-Mean reported angle vs.. target position for each HRTF set.

Table I.-Mean unsigned error relative to the target angle.

A	B	C
16°	23°	35°

There appears to be a good correlation between the “notches only” set and the reference set, with elevation localisation accuracy being similar. In contrast, vertical localisation performance with the “peaks only” set seems to rather poor, with reported positions being collapsed to the horizontal plane, concentrated directly front or directly rear.

**Relative localisation error**

Figure 5 shows the mean signed error angles versus position for the two HRTF sets B and C relative to the reported position using the reference HRTF set A. Table II indicates the mean unsigned error for each sets.

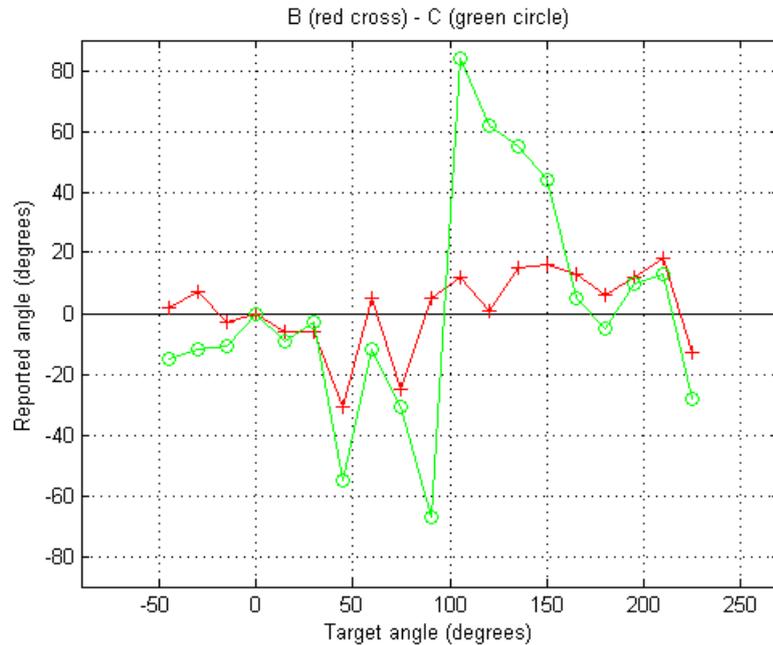


Figure 5.- Mean signed error angle vs. position relative to position using reference set.

Table II.-Mean unsigned error relative to the reference set.

B	C
10°	27°

**CONCLUSIONS**

It is understood that the experiment reported in this article contained only 4 reliable subjects. Half of the test subjects were rejected due to poor median plane localisation performance using non-individual HRTFs (even though the subjects were able to chose the HRTF based on perceived spatial quality). Despite this limitation, significant pattern have been observed.

Results of the absolute error analysis are coherent with those of the relative error analysis. Localisation tasks in the median plane using a “notches only” HRTF set and a reference set are comparable. A bias in localisation was observed for elevated sound source between the “peaks only” HRTF set and the reference set. These results appears to indicate that peaks alone are not sufficient cues for vertical localisation, and that notches provide the prominent cues necessary for median plane localisation.

**FUTURE WORK**

The protocol of the experiment should be improved, allowing subjects to use their own measured HRTF sets, and also maintaining both left and right ear HRTF content to avoid the absence of asymmetrical monaural content due to the asymmetry of human morphology.

**References:**

[1] J. Blauert: Spatial hearing: The Psychophysics of Human Sound Localization. The MIT Press, Cambridge, Massachusetts (1996)  
 [2] D. R. Begault: 3-D Sound for Virtual Reality and Multimedia. Academic Press, Cambridge, Massachusetts (1994)  
 [3] J. Hebrank, D. Wright: Spectral cues used in the localisation of sound sources on the median plane. Journal of the Acoustical Society of America **56** (1974) 1829–1834

- [4] V. C. Raykar, R. Duraiswami, B. Yegnanarayana: Extracting frequencies of the pinna spectral notches in measured head related impulse response. Technical report CS-TR-4609, Perceptual Interfaces and Reality Laboratory, University of Maryland (2004)
- [5] S. G. Rodriguez, M. A. Ramirez: Extracting and modelling approximated pinna-related transfer functions from HRTF data. Proceedings of the 2005 International Conference on Auditory Display (2005)
- [6] M. Morimoto, M. Itoh, K. Iida: 3-D sound image localization by interaural differences and the median plane HRTF. Proceedings of the 2002 International Conference on Auditory Display (2002)
- [7] P. Vovor: Application de techniques d'apprentissages statistiques à la prédiction d'HRTF. Mémoire pour le Master Science et Technologie, Université Pierre et Marie Curie, France Telecom Recherche et Développement (2005).
- [8] Listen Project – Information Society Technologies Program - IST-1999-20646: <http://listen.gmd.de/>. LISTEN HRTF Database: <http://www.ircam.fr/equipes/salles/listen/>

**ARTICLE B – PRESENTE A LA 123<sup>EME</sup> CONVENTION AES**





---

# Audio Engineering Society Convention Paper

Presented at the 123rd Convention  
2007 October 5–8 New York, NY, USA

*The papers at this Convention have been selected on the basis of a submitted abstract and extended precis that have been peer reviewed by at least two qualified anonymous reviewers. This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42<sup>nd</sup> Street, New York, New York 10165-2520, USA; also see [www.aes.org](http://www.aes.org). All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.*

---

## Round Robin Comparison of HRTF Simulation Results: Preliminary Results

Raphaël Greff<sup>1</sup>, Brian F.G. Katz<sup>2</sup>.

<sup>1</sup> A-Volute, Douai, 59000, France.  
[raphael.greff@a-volute.com](mailto:raphael.greff@a-volute.com)

<sup>2</sup> LIMSI – CNRS, Orsay, 91403, France.  
[brian.katz@limsi.fr](mailto:brian.katz@limsi.fr)

### ABSTRACT

Variability in experimental measurement techniques of the HRTF is a concern that numerical calculation methods can hope to avoid. Numerical techniques such as the Boundary Element Method (BEM) allow for the calculation of the HRTF over the full audio spectrum from a geometrical model. While numerical calculations are not prone to the same errors as physical measurements, other problems appear which cause variations: geometry acquisition and modeling of real shapes as meshes can be performed in different ways.

An on-going international round-robin study, “Club Fritz”, gathers HRTF data measured from different laboratories on a unique dummy head. This work presents preliminary results of numerical simulations based on an acquired geometrical model of this artificial head.

### 1. INTRODUCTION

In reviewing the literature, it is apparent that no two HRTF measurement systems are identical. This variability induces differences between acquired HRTFs from a single test specimen when measured with multiple systems.

In order to compare these variations, a measurement campaign project, “Club Fritz”, have been initiated, whereby the HRTFs of a single and unique dummy head, Neumann KU-100, have been measured by various institutions [1]. This paper presents the

extension of this study to numerical simulations using BEM techniques.

Whereas for the study on measurement systems the aim is to compare the impact of variation in measurement system where physical aspects are taken into consideration (architecture of the system, room, etc.), the goal is here to evaluate the impact of using different simulation systems that are not prone to problems due to physical features.

To partly avoid the problem of using different geometric models of the Neumann KU-100 dummy head, a single reference mesh have been created allowing calculation

upto 20kHz, respecting the 6 nodes per wavelength rule [2, 3]. From this reference mesh, each participating institution can extract suitable meshes according to the limitations of their simulation systems.

It is expected that variations caused by extracting a mesh from another are less significant than variations caused by different methods of model geometry acquisition.

So the parameter of this round robin comparison is not only the simulation system but also the final mesh used to perform calculation.

## 2. MESH MODELING

The reference mesh of the Neumann KU-100 artificial head has been obtained in two steps: an acquisition step and a post-processing step.

The first step has been performed with a medical MRI scanner. This produces a multitude of image slices. Slice pixel resolution is 0.63mm. Distance between two slices is 0.8mm. At first, the head was scanned intact. It appeared that metallic parts induce serious artifacts as shown by the DICOM image in Figure 1; pinna shape could not be extract from this scan. To obtain exploitable slices, almost all metallic parts of the head were removed, including the microphones. As such, the exact location of microphone diaphragm has been lost. Figure 2 shows an example of a DICOM image after removing all dismountable metallic pieces of the artificial head.

The second step has been performed in the following way: each slice has been manually post-processed to obtain closed contours. These contours have been used to obtain a closed volume. A first mesh has been extracted from this volume. This resulting mesh was not perfect, with multiple holes on the surface of the right pinna. This is due to the scanner resolution being somewhat comparable to the pinna thickness. As it was known that the contributors of this preliminary study were using symmetry in their calculation procedure, it was decided to use only the left half of the head. From this, symmetry was used to reconstruct the whole model (see Figure 3).

The mesh was then processed to validate the rule of 6 nodes per wavelength for a frequency of 20kHz. Figure 3 shows a photo of the real dummy head and the resulting reference model.

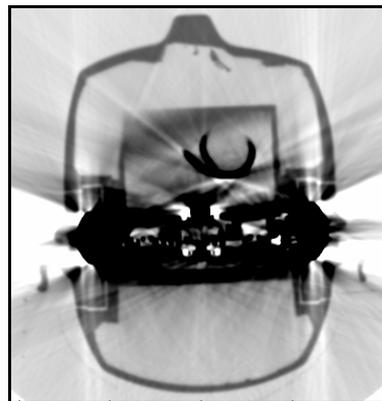


Figure 1 DICOM image before removing metallic part of the dummy head.

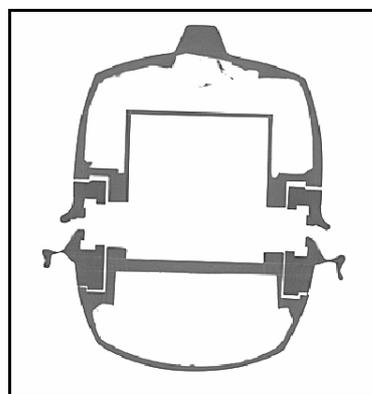


Figure 2 DICOM image after removing metallic part of the dummy head.



Figure 3 Real Neumann KU-100 dummy head and reference mesh model.

### 3. CONTRIBUTORS

For the present study two participating institutions have conducted simulations. Table 1 shows the systems used by these two institutions including information about the final meshes on which calculation have been performed.

Institution name	<i>LIMSI – CNRS Groupe Perception Située</i>	<i>University of York Intelligent Systems Group</i>
Software	<i>VNoise [5]</i>	<i>PAFEC [6]</i>
Hardware	<i>Cluster of 3 machines each with 3GHz Dual Core processors and 4Gb of RAM</i>	<i>“White Rose Grid” [7], cluster of 24 machines each with 4 2.2GHz Dual Core processors and 8Gb of RAM</i>
Mesh size	<i>10967 nodes and 21693 panels</i>	<i>6217 nodes and 12054 panels</i>
Job duration	<i>25 minutes per job, where a job computes all positions for a single frequency using all processors</i>	<i>2.5 hours per job, where a job computes all positions for a single frequency using one processor only</i>
Frequency range	<i>From 100Hz to 15kHz in 100Hz steps</i>	<i>From 100Hz to 15kHz in 100Hz steps</i>
Theoretical minimal duration time	<i>2 days and 4.5 hours</i>	<i>4 hours running simultaneously 4 jobs per machine over 24 machines</i>

Table 1 Contributors and simulation systems overview.

Contributor 1 has extracted from the reference mesh a new one that conforms to the classical rule of discretisation: 6 nodes per wavelength till 10kHz or 4 nodes per wavelength till 15kHz. The “4 to 6 nodes per wavelength” rule is known to be a sufficient discretisation to be sure to obtain valid results when using BEM calculation method [2, 3, 4].

Because of a limitation of the maximum number of nodes, contributor 2 derived from the reference mesh an “optimized” mesh using a novel method. The resulting

mesh has a very high discretisation at the pinna (conforming to 6 nodes per wavelength for 20kHz) and a lower discretisation for the rest of the head (conforming to 6 nodes per wavelength for 4kHz). The idea of contributor 2 is that this kind of discretisation can be sufficient to achieve a valid full range calculation. Figure 4 shows the different degrees of discretisation that exist in the “contributor 2” mesh.

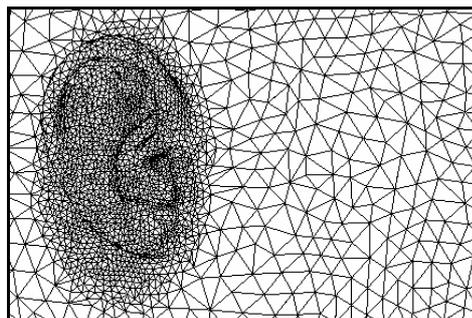


Figure 4 Zoom on “optimized” mesh.

### 4. RESULTS

The results from this preliminary study should help to verify the quality of the model of the dummy head by comparing simulation to measurement results and to have a first idea of the variations caused by the different simulation systems. Simulations have been conducted from 100Hz to 15kHz in steps of 100Hz.

#### 4.1. Magnitude differences

Data from both contributors are compared to one of the data sets resulting from the “Club Fritz” HRTF measurements. Figure 5 and 6 shows the spectral response of left ear for two position of the source. Numerical simulation results appear to match well. Simulated results seem to exhibit a shift or dilation for frequencies above 5 kHz as compared to the measured data, patterns are coherent. The fact that the reference measurement is not free-field equalized may explain some of the differences, specially at low frequencies.

A potential reason for the observed mismatch between measured and numerical simulation results is that the point where the pressure level was calculated is not at the same location as where the microphone diaphragm was (see Figure 7). Regarding results of the “Club Fritz” measurement study [1], this frequency shift is concordant with some of the differences observed

between measurements performed with internal Neumann microphones and those performed with other microphones, where the ear canals were blocked. Further investigation is required to validate this hypothesis. The low frequency patterns of simulated data seem to match well with data of other measurements (internal or external microphones).

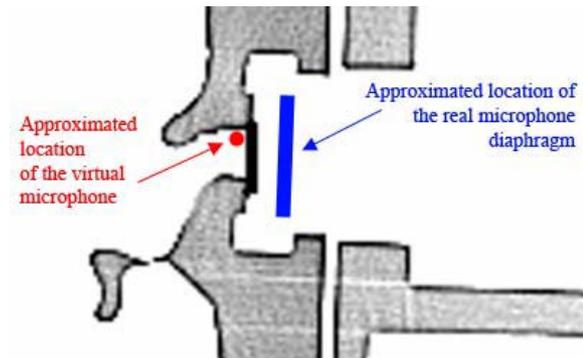


Figure 7 Difference in locations of the real internal microphone and the virtual microphone placed in front of the blocked canal (imposed during mesh generation).

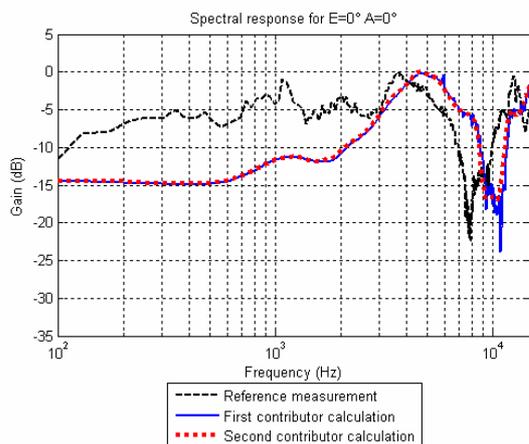


Figure 5 HRTF spectral responses for frontal position of the source (left ear).

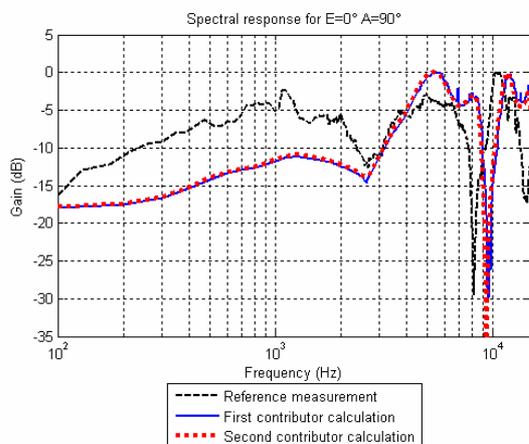


Figure 6 HRTF spectral responses for direct left position of the source (left ear).

#### 4.2. Interaural time differences

Figure 8 shows estimation of the ITD from the measured and simulated HRTFs. A good match is obtained between both simulations. Data indicate a maximal difference of 0.1ms between simulations and the reference measurement and may be explain by the difference in locations of the real internal microphone and the virtual microphone; whatever, this is acceptable, since difference in results of the measurement study [1] reaches a value as high as 0.2ms for a given azimuth position.

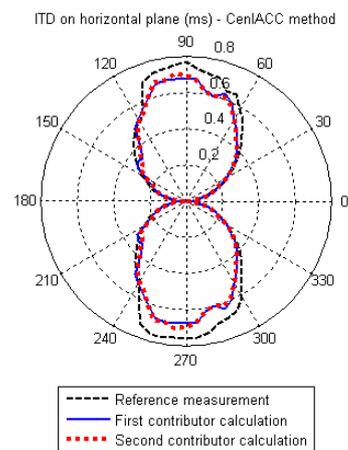


Figure 8 Estimation of interaural time differences at 0° elevation.

## 5. FUTURE WORKS

The first work will be to solve the problem of the frequency shift using an iterative procedure for refining the model geometry. Results obtained here will be reflected in a revised discretization, and a new simulation will be conducted; this process will be repeated until a satisfying model for the internal microphone configuration is reached. Then, this study will be open to new contributors.

In addition, contributors will be allowed to use their own acquisition and/or modeling procedures in order to study variations induced by different acquisition (ex. laser scanner vs. MRI) and modeling techniques, and also other kind of simulation systems (different BEM codes as well as other numerical methods).

## 6. ACKNOWLEDGEMENTS

The authors would like to thank Dr. Carl Hetherington and Mr. Tony Tew, Audio Lab, Intelligent Systems Group, Department of Electronics, University of York, for participating in this preliminary study of the “Club Fritz” simulation extension, and Dr. Angel Osorio Sainz, LIMSI – CNRS, Groupe Perception Située, Département Communication Homme-Machine, for performing the MRI scan. Thanks also to the Space Human Factors Engineering Project of NASA's Human Research Program for lending us the Neumann KU-100 dummy head and to A-Volute company which is financing the PhD thesis of first author.

## 7. REFERENCES

- [1] B. F.G. Katz, D. R. Begault: Round robin comparison of HRTF measurement systems – preliminary results. Proceedings of the 19<sup>th</sup> International Congress on Acoustics (2007).
- [2] J.-M. Pernaux: Spatialisation du son par les techniques binaurales – application aux services de télécommunications. PhD thesis, Institut National Polytechnique de Grenoble, France Telecom R&D, Laboratoire Interaction par la parole et le Son (2003).
- [3] Y. Kahana: Numerical modelling of the head-related transfer function. PhD thesis, University of Southampton (2000).
- [4] B. F.G. Katz: Measurement and calculation of individual head-related transfer functions using a boundary element model including the measurement and effect of skin and hair impedance. PhD thesis, Pennsylvania State University (1998).
- [5] Scientific Technical Software (STS): VNoise. [www.sts-soft.com/VNoise.aspx](http://www.sts-soft.com/VNoise.aspx).
- [6] PACSYS Ltd: PAFEC. [www.vibroacoustics.co.uk/](http://www.vibroacoustics.co.uk/).
- [7] University of Leeds, Sheffield and York: White Rose Grid. <http://www.wrgrid.org.uk/>.



**ARTICLE C – PRESENTE A LA CONFERENCE ACOUSTICS'08**





## Circumaural transducer arrays for binaural synthesis

R. Greff<sup>a</sup> and B. F G Katz<sup>b</sup>

<sup>a</sup>A-Volute, 4120 route de Tournai, 59500 Douai, France

<sup>b</sup>LIMSI-CNRS, B.P. 133, 91403 Orsay, France

raphael.greff@a-volute.com

Binaural cues such as the interaural time and level differences are the primary cues for estimation of the lateral position of a sound source, but are not sufficient to determine elevation or the exact position on the “cone of confusion”. Spectral content of the head-related transfer function (HRTF) provides cues that permit this discrimination, notably high frequency peaks and notches created by diffraction effects within the pinnae. For high quality binaural synthesis, HRTFs need to be individualized, matching the morphology of the listener. Typical means for this are to measure or calculate the HRTF of the listener, but these lengthy and costly methods are not feasible for general public applications. This paper presents a novel approach for HRTF individualization, separating the head and torso effect from that of the pinnae. The head/torso component is numerically modeled while the pinnae component is created using a multiple transducer array placed around each pinna. The philosophy of this method consists in trying to excite the correct localization cues provided by the diffraction of the reconstructed wave front on the listener’s own pinnae. Simulations of HRTF reconstruction with various array sizes and preliminary auditory localization tests are presented.

## 1 Introduction

High quality binaural synthesis requires individualized HRTF, *i.e.* matching the morphology of the listener. Measuring or calculating the HRTF of a listener are common but lengthy and costly methods that are not feasible for general public applications. In order to simplify the estimation procedure of individualized HRTF, a great deal of research efforts aim to model the relation between morphology and the HRTF and provide a fast individualization based on anthropometric measurements [1,2]. However, accurate estimations of morphological parameters may only be guaranteed using accurate measurement tools, essentially for estimating pinnae dimensions.

In this paper, a novel approach for HRTF individualization is proposed, by separating the head and torso effect from that of the pinnae using a multiple transducer array placed around each pinna. The goal is to reduce the individualization of the HRTF to the modeling of the head/torso component, and to try to excite the correct localization cues provided by the diffraction of the reconstructed wave front on listener’s own pinnae.

## 2 Binaural holophony

The concept of binaural holophony consists in using holophonic techniques to reconstruct the sound fields in each pinna area. In practice, this should be done by using circum-aural transducer arrays mounted on a headset and by feeding them correctly in order to reconstruct the desired binaural signal at listener’s eardrums.

To characterize array signals, one needs first to understand the theory of general holophony and see how this is used for reproducing sound fields.

### 2.1 Kirchhoff-Helmholtz integral

The basic principle of holophony is based on the Huygens’s principle. Huygens stated that a wave front of a propagating wave at any instant conforms to the envelopes of spherical waves emanating from every point on the wavefront at the prior instant. The mathematical foundation of this illustrative description is given by the Kirchhoff-Helmholtz integral [4] which expresses that the acoustical pressure  $p$

within a source-free volume  $V$  is derived from the knowledge of the acoustical pressure  $p_0$  and its gradient over the boundary  $S$  enclosing the considered volume.  $\forall \vec{r} \in \Omega$ :

$$p(\vec{r}) = \iint_S \left[ \vec{\nabla} p_0 \cdot \vec{n} - \frac{\vec{R}}{R} \cdot \vec{n} (1 + jkR) \frac{p_0}{R} \right] \frac{e^{-jkR}}{4\pi R} dS_0 \quad (1)$$

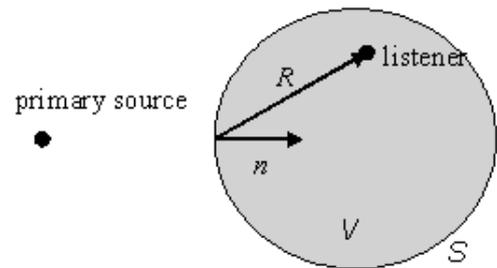


Fig.1 Parameters used in the expression of the Kirchhoff-Helmholtz integral.

### 2.2 Spatial sampling

Though the Kirchhoff-Helmholtz integral provides a very attractive solution for sound field reproduction, in practice only discontinuous array of secondary sources are available, which raises the problem of spatial sampling. Discrete arrays can not correctly sample incident waves whose wavelength is small compared to the transducer spacing  $\Delta$ . Such spatial aliasing occurs above the so-called “spatial aliasing frequency” ( $c$  is the speed of sound):

$$f_a = \frac{c}{2\Delta} \quad (2)$$

In practical situations, the integral in Eq.(1) should be replaced by a summation. This gives a straightforward way of reproducing a sound field. At the recording stage, the listening area is surrounded by a microphone array composed by pressure and velocity microphones which record the primary sound field due to external sources. For the reproduction stage, loudspeakers are substituted for the microphones by replacing pressure microphones by dipole sources and velocity microphones by monopole sources. Each loudspeaker is fed with the signal recorded by its associated microphones (assuming the use of some weighting factors due to the spatial distribution of the secondary sources).

## 2.3 Single directivity array

It has been remarked that the two elementary transducers of the secondary source are highly redundant, so that only one of them is necessary [3,4]. In practice, only monopole loudspeakers are used and fed by figure-of-eight or even cardioid microphones. Figure-of-eight solution is an exact solution if one is able to record the contribution of each primary source separately. The recorded contribution of a primary source is taken into account only if this source is viewed by the positive pole of the microphone. For planar array, this selection is done by construction [4], but for complex array geometries, the logical weighting can only be performed if the exact location of each primary source is known. For this reason cardioid microphones are often preferred when recording real complex sound fields using complex array geometry.

Nevertheless, the location of a sound source relative to an arbitrary field point is given by the active intensity at this point. Therefore, we may try to generalize the logical weighting - when using figure-of-eight microphones - to any arbitrary primary sound field, *i.e.* created by unknown sources.

In the next section, three kinds of holophonic recording solutions are employed for investigating the concept of binaural holophony:

- *CARD*: cardioid recording,
- *FIG-8 S-LOG*: figure-of-eight recording, considering the contribution of each primary sources separately,
- *FIG-8 T-LOG*: figure-of-eight recording, considering the total sound field.

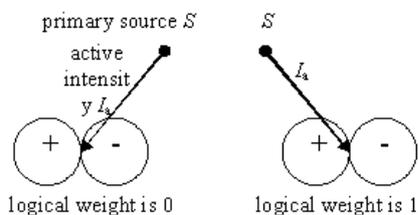


Fig.2 Illustration of logical weighting.

## 3 Simulations

The array used here has a hemispherical geometry. Based on Listen and CIPIC anthropometric databases, we evaluated that a radius of 50mm is sufficient to surround all of the pinnae, and 90% of them when centring the array on ear canal entrance. Simulations are only performed with centred positioning. The primary array is composed of 640 equally distributed secondary sources to avoid spatial aliasing in the considered frequency range (3.5 to 16 kHz).

The goal here is to compare the original HRTF to a reconstructed HRTF using holophonic techniques for different array sizes. By introducing two kind of filters, the transducer driving filters (CAHF for Circum-Aural Holophonic Filters) and the transfer functions representing the path from a considered transducer to the entrance of the ear canal (PRTF for near-field Pinnae-Related Transfer

Functions), the reconstructed *HRTF* can be expressed as follows (where  $N$  is the total number of transducers and  $i$  the index of the considered transducer):

$$HRTF_r = \sum_{i=1}^N CAHF[i].PRTF[i] \quad (3)$$

For estimating the quality of reconstruction, two metrics are introduced, correlation and distance. Both are calculated on a logarithmic scale and based on decibel spectral values of the HRTF.

### 3.1 Baffled pinnae

Simulations have been performed on 6 different baffled pinnae meshes (Cortex, DB60, DB65, DB90, DB95, Yuvi), modelled by Yuvi Kahana [5]. In this case, the three kinds of holophonic recording solutions previously mentioned can be used, because all sources are known (the primary source and its image). CAHF are calculated analytically due to the simplicity of the problem. Original HRTF and PRTF are estimated via BEM techniques, respecting the sampling rule of 6 nodes per wavelength [5,6].

For all pinnae models, results indicate a good match between original HRTF and reconstructed HRTF using *CARD* and *FIG-8 S-LOG* solutions (cf. Fig.3). *FIG-8 T-LOG* is dependant on lateral angle: quality of reconstruction decreases with lateralization of the primary source (cf. Fig.4), because contributions of the primary source and those of the image source tends to oppose.

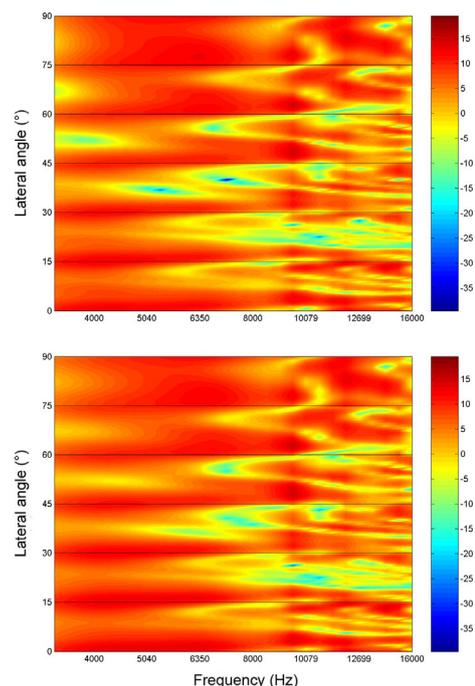


Fig.3 Cortex pinna - original HRTF (upper) and reconstructed HRTF using an array of 640 transducers with *FIG-8 S-LOG* recording method (lower).

If the number of secondary sources is too small, spatial aliasing occurs which degrades the quality of reconstruction. By decreasing the number of transducers from 640 to 8 (keeping a homogeneous repartition), the correlation decreases from more than 0.9 to less than 0.2

and spectral distances increases from approximately 1 dB to more than 5 dB (FIG-8 S-LOG method).

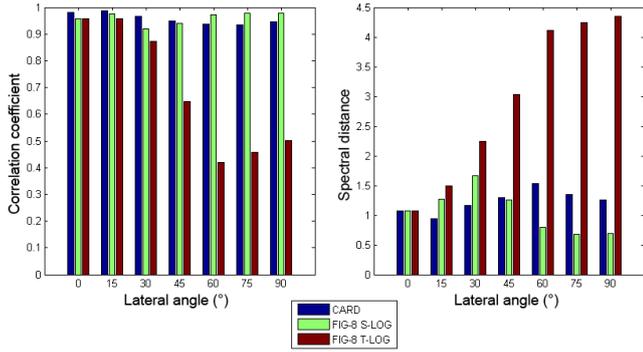


Fig.4 Cortex pinna – correlation coefficient (left) and spectral distance (right) per sagittal plane.

### 3.2 Head model

Simulations have been done on a model of the Neumann KU-100 dummy head, obtained via a MRI scan [7]. In this case, only CARD and FIG-8 T-LOG solutions have been calculated, because of the complexity of the problem. Moreover, all transfer functions have been estimated via BEM calculations. Because of the size of the model, a sampling of 6 nodes per wavelength is only respected up to 10 kHz, but simulation are still conducted up to 16 kHz.

For the ipsilateral HRTF, results are equivalent to those of the baffled pinnae simulations. For the contralateral HRTF, quality of reconstruction decreases with respect to lateralization for both recording methods, but CARD solution seems to be more robust.

Decreasing the number of secondary sources degrades the reconstruction for both recording method, wherever the primary source is located. For small array sizes (under 20 secondary sources), FIG-8 T-LOG method offers better results than CARD for median plane HRTF, but can not reconstruct the contralateral HRTF properly.

### 3.3 Equivalent spherical head

BEM calculations dedicated to CAHF estimation are very time-consuming (the current calculation required more than three weeks on a cluster of three dual core processors with a total of 9 GB RAM). So it is interesting to evaluate the degradation of HRTF reconstruction using an equivalent spherical head model which permits an analytical resolution of the problem. A similar approach from that Algazi proposed [8] has been used to find the radius of the iso-ITD sphere relative to the KU-100 dummy head. CAHF have then been calculated for this sphere.

No relevant difference has been noticed compared to KU-100 results. The spherical head approximation seems to offer an easy way to design a binaural holophonic system.

## 4 Optimizations

It was desired, for practical reason, to arrive at a solution for a binaural holophonic system using a maximum of 16 channels, *i.e.* 8 transducers per pinna. Precedent simulations

have shown that raw applications of holophony principles provide bad results when spatial aliasing is occurring. Transducer driving filters and transducer positioning need to be optimized.

### 4.1 Non pinnae-related optimization

When the number of secondary sources decreases, the area associated to each microphone (or transducer) increases. The pressure (and pressure gradient) recorded at a single point may not be representative of what is happening in the entire region. A simple solution for obtaining a more relevant recording is to average the whole area information.

Applying this approach to the precedent simulations provides a significant enhancement of the reconstruction quality. For example, correlation coefficient increases from less than 0.2 to over 0.5.

### 4.2 Pinnae-related optimization

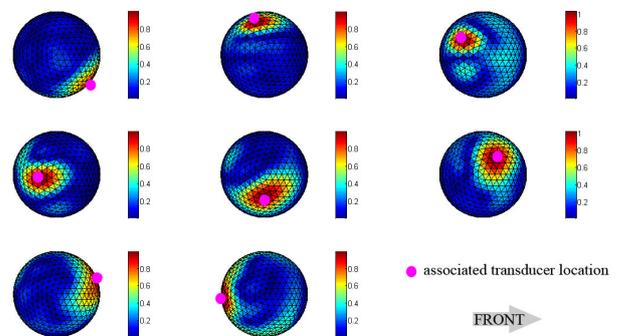


Fig.5 Optimized directivities given by “subset selection” analysis.

A way of optimizing both the location and the associated CAHF of a transducer is to apply statistical analysis on a database of PRTF (corresponding to the 640 secondary source locations of the primary array). Three types of optimizations have been investigated: PCA, ICA and “subset selection” [9]. The PRTF database is composed from that of 7 pinnae (Cortex, DB60, DB65, DB90, DB95, Yuvi and KU-100). The considered number of transducer is 8 per pinnae.

The optimization procedure consists first in separating the PRTF database into a reduced number of filters and directivities. Optimized locations of the transducers are derived from the directivity patterns (directions pointed by each maximum of directivity). As associated filters may not correspond to the PRTF at those locations - for PCA and ICA - the directivity patterns can not be used directly to calculate the optimized CAHF. Projecting whole PRTF database on the PRTF associated to the optimized locations permits one to obtain the desired optimized associated directivities. These directivities are then used as masks to derive the optimized CAHF associated with the optimized array. Naming  $CAHF_{prim}$  the driving filters of the primary array of 640 secondary sources and  $D_{opt}$  the optimized directivities, the optimized driving filters  $CAHF_{opt}$  of the optimized array (of 8 transducers in this case) are obtained as follows:

$$CAHF_{opt}[k] = \sum_{i=1}^N D_{opt}[i, k] \cdot CAHF_{prim}[i] \quad (4)$$

where  $N$  is the total number of the primary array,  $i$  the index of the secondary source related to the primary array and  $k$  the index of the transducer related to the optimized array.

Results indicate that all three methods highly increase the quality of HRTF reconstruction. However, “subset selection” seems to be the best solution (correlation coefficient and spectral distance reach values that exceed 0.9 and fall under 1 dB respectively). Moreover, it is the only optimization that produces directly the optimized directivities (cf. Fig.5): the filters issued from the “subset selection” analysis correspond exactly to the PRTF pointed by each directivity pattern.

## 5 Subjective evaluations

A modular hand-made headset prototype has been constructed to perform subjective evaluations of different array configurations. A preliminary listening test has been conducted to choose the transducer dimensions. The protocol of this test is not presented here, but results indicated that it is preferable to use transducers as small as possible.

### 5.1 Configuration comparison

Three array configurations have been tested, composed of 8 transducers per pinnae. In configuration **A** transducers are equally distributed and CAHF are calculated using the first optimization (average per area). For configuration **B**, the associated filters are obtained through the “subset selection” optimization method. Configuration **C** is based on psychoacoustic hypothesis: as pinnae spectral cues are predominant cues in the median plane [10,11], transducers are placed at grazing incidence where recorded energy is maximal when primary sources are located on median plane.

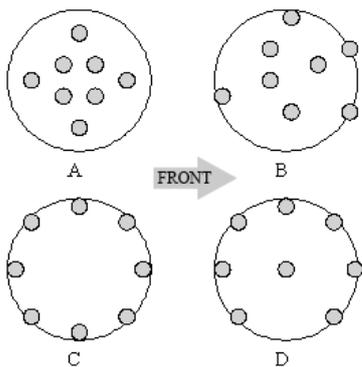


Fig.6 Tested array configurations (schematic representation).

Localization tests were performed only for virtual sources in the median plane. Perceived direction was reported on a visual interface. The stimulus used was a 200ms pink noise burst modulated at 25Hz. 4 subjects participated to the test. Each tested location was repeated 5 times.

Results, reported in Table 1, show that configuration **C** offers the best performance. “Subset selection” optimized configuration **B** has good simulation results, where an exact positioning of the headset is guaranteed, but compared to configuration **C** it seems not to be the best way for exciting pinnae cues when headset placement is approximate. Further simulations must be conducted to evaluate the degradation of reconstruction relative to off-centred positioning of the array. However, localization accuracy is better using **B** compared to **A**, which is coherent with simulations results.

(min./max)	<b>A</b>	<b>B</b>	<b>C</b>
Mean error $E$ ( $^{\circ}$ )	28 (19/39)	23 (17/27)	<b>14</b> (11/19)
Standard deviation $Std$ ( $^{\circ}$ )	20 (16/24)	<b>17</b> (14/21)	19 (14/25)
Front-back confusions $FB$ (%)	25 (7/38)	34 (11/60)	<b>20</b> (13/27)

Table 1 Localization test results for all configurations (**A**, **B**, **C**) - median plane.

$E$ ( $^{\circ}$ )	$Std$ ( $^{\circ}$ )	$FB$ (%)
11 (7/13)	12 (10/13)	20 (11/29)

Table 2 Localization test results for configuration **C** – horizontal plane.

Horizontal plane localization tests have then been conducted on retained configuration **C**. Even if results reported in Table 2 are good, informal listening tests and discussions with tested subjects indicate a strange sensation for lateral source. It is considered that this may be due to the fact that all transducers are placed at grazing incidence so that a lateral wavefront can not be recreated. We decided to test a fourth configuration (**D**) obtained by relocating one transducer in order to improve the coherence of recreated lateral wavefronts. Results (cf. Table 3) are comparable to those of the configuration **C** and informal discussions indicate that the strange sensation has disappeared.

(min./max)	$E$ ( $^{\circ}$ )	$Std$ ( $^{\circ}$ )	$FB$ (%)
Median plane	18 (16/21)	17 (14/19)	9 (2/18)
Horizontal plane	11 (7/13)	10 (8/12)	21 (7/31)

Table 3 Localization test results for configuration **D** – median and horizontal plane.

### 5.2 CAHF vs. individualized HRTF

Localization tests have been conducted on 4 other subjects, using the binaural holophonic system (configuration **D**) and also individualized HRTF. Performances are compared per listener.

Results show that both spatialization systems are comparable (cf. Table 4 and Table 5). In our experiment, individualized binaural accuracy is worth than what is

reported in the literature [12], but generally the rate of tested median plane localisation is less important, median plane localization task being the most difficult task.

		Median plane		Horizontal plane	
		Binaural holophony	Personal HRTF	Binaural holophony	Personal HRTF
1	<i>E</i>	24	20	9	11
	<i>Std</i>	<b>25</b>	<b>44</b>	10	10
	<i>FB</i>	18	24	13	18
2	<i>E</i>	16	15	7	9
	<i>Std</i>	18	16	9	16
	<i>FB</i>	<b>2</b>	<b>16</b>	<b>7</b>	<b>20</b>
3	<i>E</i>	18	16	13	10
	<i>Std</i>	14	16	10	6
	<i>FB</i>	<b>38</b>	<b>24</b>	<b>31</b>	<b>13</b>
4	<i>E</i>	30	27	8	13
	<i>Std</i>	<b>20</b>	<b>30</b>	11	11
	<i>FB</i>	16	20	<b>22</b>	<b>36</b>
Mean	<i>E</i>	22	20	9	11
	<i>Std</i>	19	27	10	11
	<i>FB</i>	19	21	18	22

Table 4 Comparison of binaural holophony and personal HRTF results.

## 6 Conclusion

Simulations and subjective evaluations show that it is possible to excite pinnae cues using multiple transducers well-placed around listener's pinnae, and using transducer driving filters related only to a simple spherical head model. Comparison of the proposed system to individualized HRTF don't permit to conclude directly that the proposed binaural holophonic system offers localization performances as accurate as individualized HRTF system do, because statistics on 4 subjects may not be representative of the entire population. However, it justifies the use of binaural holophony for general public application as a promising alternative to common HRTF-based binaural systems.

Moreover, the holophonic approach offers a way to decompose the HRTF into head/torso component - CAHF - and pinnae component - PRTF (exact physical decomposition in comparison to Algazi's heuristic signal approach [13]). This allows us to study separately the link between pinnae morphology and associated spectral modifications. Pinnae-related optimizations presented above offer also another point of view for characterizing this relation and open the way of fast HRTF composing.

## Acknowledgments

Special thanks to Yuvi Kahana for access to his pinnae models. Funding was provided by the ANRT CIFRE program in collaboration with A-Volute.

## References

- [1] C. Jin, P. Leong, J. Leung, A. Corderoy and S. Carlile, "Enabling individualized virtual auditory space using morphological measurements", *International Symposium on Multimedia Information Processing, Sydney, Australia* (2000)
- [2] D. N. Zotkin, R. Duraiswami and L. S. Davis, "Customizable auditory displays", *International Conference on Auditory Display, Kyoto, Japan* (2002)
- [3] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis", *Journal of the Acoustical Society of America* 93, 2764-2778 (1993)
- [4] R. Nicol, "Restitution sonore spatialisée sur une zone étendue : application à la téléprésence", *PhD thesis, Université du Maine* (1999)
- [5] Y. Kahana, "Numerical modelling of the head-related transfer function", *PhD thesis, University of Southampton* (2000)
- [6] B. F.G. Katz, "Measurement and calculation of individual head-related transfer functions using boundary element model including the measurement and effect of skin and hair impedance", *PhD thesis, Pennsylvania State University* (1998)
- [7] R. Greff, B. F.G. Katz, "Round robin comparison of HRTF simulation results: preliminary results", *Audio Engineering Society Convention, New York, USA* (2007)
- [8] V. R. Algazi, C. Avendano and R. O. Duda, "Estimation of a spherical-head model from anthropometry", *Journal of the Audio Engineering Society* 49(6), 472-478 (2001)
- [9] V. Larcher, "Techniques de spatialisation des sons pour la réalité virtuelle", *PhD thesis, Université Paris VI* (2001)
- [10] J. Hebrank, D. Wright, "Spectral cues used in the localisation of sound sources on the median plane", *Journal of the Acoustical Society of America* 56, 1829-1834 (1974)
- [11] F. Asano, Y. Suzuki and T. Sone, "Role of spectral cues in median plane localization", *Journal of the Acoustical Society of America* 88, 159-168 (1990)
- [12] J.-M. Pernaux, "Spatialisation du son par les techniques binaurales : application aux services de telecommunications", *PhD thesis, Institut National Polytechnique de Grenoble* (2003)
- [13] V. R. Algazi, R. O. Duda, R. P. Morrison, D. M. Thompson, "Structural composition and decomposition of HRTF", *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, USA* (2001).