Technical and perceptual issues on head-related transfer functions sets for use in binaural synthesis

Toledo, Daniela

Publication date:
2011

Citation for published version (APA):
Toledo, D. (2011). *Technical and perceptual issues on head-related transfer functions sets for use in binaural synthesis*. Acoustics, Department of Electronic Systems, Aalborg University.

Technical and Perceptual Issues on Head-related Transfer Functions Sets for Use in Binaural Synthesis

Acoustics - Department of Electronic Systems
Aalborg University

November 5, 2010

Ph.D. thesis by Daniela Toledo

Technical and Perceptual Issues on Head-Related Transfer Functions Sets for Use in Binaural Synthesis

ISBN: 978-87-92328-68-7

Address correspondence to:
Section of Acoustics, Department of Electronic Systems
Aalborg University, Frederik Bajers Vej 7-B5, DK-9220 Aalborg
Phone: +45 9940 8710, E-mail: acoustics@es.aau.dk

Ph.D. thesis defended on April 28, 2011
Ph.D. degree awarded on June 8, 2011
Aalborg University, Denmark

Assessment Committee:
Associate Professor Søren Krarup Olesen, Aalborg University, Denmark (chairman)
Professor Emeritus Jens Blauert, Ruhr-University Bochum, Germany
Professor Yôiti Suzuki, Tohoku University, Japan

Supervisor: Professor Henrik Møller, Aalborg University, Denmark

Preface

This thesis has been submitted to the Faculty of Engineering, Science and Medicine at Aalborg University in partial fulfillment of the requirements for the award of the Ph.D. degree. The research was carried out at the University's Section of Acoustics in the period from June 2006 to August 2009.

I would like to thank all the colleagues at the Section of Acoustics for the inspiring environment. Special thanks to Henrik Møller and Dorte Hammershøi for offering me the PhD position that led to the research presented here.

Some of the investigations conducted for this Thesis, and other pilot studies that are not included here, made used of a database of HRTFs, ATFs and HTTFs measurements on human subjects which was available from previous works at the Section of Acoustics. The group of colleagues that conducted those measurements is kindly acknowledged and mentioned where appropriate throughout the Thesis. Claus Vestergaard Skipper is acknowledged for his help with the mechanical work and mounting of the arc of loudspeakers that was used in several of the studies reported here, and Henrik Zimmermann for helping me translating the summary to Danish. Brian Katz and Durand Begault are thanked for their invitation to participate in the round robin of HRTFs measurement systems, which materialized in part of Chapter 2.

I would also like to thank my friends in Denmark and in Argentina, who supported me during the process of working towards a Ph.D. Special thanks to my partner Kristian, whose artistic mind helped my development in all aspects of life.

I dedicate this work to the memory of my parents and brother: Oscar, Beatriz and Roberto.

Aalborg University, November 2010

Summary

Head-related transfer functions (HRTFs) are the core of binaural synthesis, a technology in which a sound recorded under anechoic conditions is filtered with pairs of HRTFs. The HRTFs provide the cues required so that the virtual (synthesized) sound source is localized in virtual 3D space. The success of binaural synthesis in providing a realistic and/or convincing virtual sonic experience relies largely on technical and perceptual issues inherent to the HRTFs and their use, which are the focus of this Thesis.

A first study analyzed the issues of calibration, DC correction and low frequency control in measured HRTFs. The issues were seen to be interconnected, as a proper calibration was a requirement for correct low frequency control and DC correction, and DC correction helped controlling the low frequencies and avoiding audible ripples that affected sound perception.

As measured HRTFs are usually implemented as minimum phase filters (with a corresponding interaural time difference), a second study compared two methods of minimum phase decomposition. The methods were equivalent in terms of success rate as a function of zero padding applied to the signals. Extensive zero padding was needed for the contralateral signals of HRTFs from directions to the sides and below the horizontal plane, to prevent the decomposition methods from failing.

In a third study, a listening test confirmed that removing high Q-factor all-pass sections from HRTFs did not have audible consequences. In the listening test, potentially audible high Q-factor all-pass sections were presented alone and with their minimum phase HRTFs counterparts, both under binaural and diotic presentations.

The spectral features that cue elevation in the mid-sagittal plane were investigated in a fourth study. The results suggested that: a) the first peak, particularly its Q-factor and its high frequency slope, would cue directions high-front, above and high-back; b) the first notch and its global Q-factor would cue front, back and back-low directions, and would provide redundant information for the above and high directions; and c) the second peak, particularly its center frequency, would provide some redundant information to also disambiguate back and back-low directions. These findings showed agreement with the current knowledge on the topic.

A localization experiment with real sound sources under anechoic conditions showed that some subjects presented strong biases in their localization performance in the elevation dimension: they tended to localize sound sources towards a particular hemisphere, or had degraded performance in a specific range of space. Results from these *biased localizers* were presented and discussed in a fifth study.

The last study showed the computed ratio between non-individual to individual HRTFs, from those HRTFs which evoked the same perception of sound source direction. The results provided theoretical evidence of possible perceived quality degradation when sound is synthesized with spectrally non-matching non-individual HRTFs.

The results of the six studies presented in this Thesis can be used for the advancement of binaural synthesis, on the basis of a better understanding of the technical and perceptual issues related to the HRTFs and their use.

Resumé (Summary in Danish)

Head-related transfer functions (HRTF'er) er kernen i binaural syntese, hvor en lyd, optaget i et lyddødt rum, er filtreret ved hjælp af HRTF'er for at få de nødvendige cues, der muliggør lokalisering af en virtuel (syntetiseret) lydkilde i et virtuel 3D rum. Succesen for at binaural syntese giver en realistisk og/eller overbevisende virtuel lydoplevelse beror i vid udstrækning på de tekniske og perceptuelle problemstillinger, som er tilknyttet til HRTF'er og deres brug. Nærværende afhandlingen sætter fokus på disse problemstillinger.

Den første undersøgelse analyserer forskellige problemstillinger med tilknytning til kalibrering, DC korrektion og respons ved lave frekvenser. Disse problemstillinger er relaterede, da kalibrering er nødvendig for at kontrolere lave frekvenser og korrigere DC, og DC korrektion sikrer kontrol af lave frekvenser således, at hørbar ringen undgås og ikke forstyrrer lydopfattelsen.

Da målte HRTF'er ofte er implementeret ved hjælp af minimum phase filtre (med en interaural tidsforskel), vil den anden undersøgelse sammenligner to metoder til minimum phase dekomposition. Disse metoder har tilsvarende succesrate som funktion af længden af zero padding der blev tilføjet til signalet. Omfattende zero padding var nødvendigt til kontralaterale signaler for HRTF'er med retninger til siderne og under det horizontal plan, for at hindre at metoderne fejlede.

En tredje undersøgelse viste, at høj Q-faktor all-pass sektionerne kan fjernes fra HRTF'er uden hørbare konsekvenser. I et lytteforsøg, blev potentielt hørbare høj Q-faktor all-pass sektionerne præsenteret for forsøgspersonerne i to former: alene og sammen med deres minimum phase dele, begge under binaural og diotic forhold.

De spektrale karakteristika af HRTF'er, som giver relevante cues til lokalisering i midsagittal plane, blev undersøgt i en fjerde undersøgelse. Resultaterne viste at: a) den første top, hvor især Q-faktoren og den store hældningskoefficient, giver anledning til cues med retninger høj-front, over og høj-bag; b) det første dyk, og tilsvarende globale Q-faktor, giver cues til front, bag og bag-lav retninger og giver samtidig redundant in-

formation om over og høje retninger; og c) den anden top, hvor især centerfrekvensen, giver redundant information og tydeliggører bag og bag-lav retninger. Disse resultater er i overensstemmelse med den nuværende viden om emnet.

Et lokaliseringsforsøg med rigtige lydkilder i et lyddødt rum viste, at nogle lyttere havde stærke preferencer i deres lokaliseringsperformance, særlig i den vertikale dimension: de havde tendens til at lokalisere i bestemt retninger, eller havde degrederet lokalisering i et bestemt område af rummet. Resultaterne fra disse lyttere er præsenteret og beskrevet i en femte undersøgelse.

Den sidste undersøgelse viste hvordan individuelle og ikke-individuelle HRTF'er, som var årsag til samme lydkilde lokalisering, kunne relateres. Resultaterne giver en teoretisk baggrund til at undersøge degraderet lydkvalitet, når lyden er syntetiseret ved ikke-individuelle HRTF'er med spektrale karakteristika, som ikke matcher lytterens egne HRTF'er.

Resultaterne fra de seks undersøgelser, som står i denne afhandling har praktisk betydning for implementering af binaural syntese, og muliggør en bedre forståelse af de tekniske og perceptuelle problemstillinger, som er relateret til HRTF'er og hvordan de bruges.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Head-related transfer functions (HRTFs) contain the amplitude and phase transformations that affect the sound in its path from a sound source to the eardrum of a listener. These transformations are responsible for the human ability of localizing sound sources and, if measured and implemented as filters in the context of binaural synthesis, they should theoretically provide the listener with realistic virtual sonic environments. The potential of binaural techniques is powerful, but in the current state of the technology there are several issues that are still unaddressed and prevent the technology to fully achieve its theoretical expectation.

HRTFs have been around in research and commercial applications for a considerable time: according to Paul [2009], in the early 20's the terms dummy-head an binaural recording technique were already in use. It was around that time that the first dummy-heads were built, as reported by Firestone [1930], to increase the fidelity of recorded signals. General technological advancements led to attempts to investigate the transformations imposed by the human shape on incoming sounds, but it was not until the 60's and 70's that strong effort was systematically put in binaural hearing research and applications. For example, Shaw & Teranishi [1968] measured the transfer function of an isolated pinnae and explained its modes of resonance, and Mellert *et al.* [1974] measured the transfer function of the external ears by an impulse response method, among other relevant studies. These investigations were related to several experiments -see Blauert [1969/70] and Hebrank & Wright [1974b], among others- which characterized the evoked localization of sound. Watkins [1978] reported the use of transfer functions as filters to create virtual sounds that were presented through headphones, bypassing the pinnaes of the listeners. The concept was, however, not new as the picked up sound from the manikins in the 20's was also presented through headphones.

Even though HRTFs have been subject of research for as long as described previously, there is still a lack of common understanding about how to obtain valid measured HRTFs. The lack of a consensual standard or protocol is seen not only when comparing different publications, but it is also explicit in the results of a current round robin of measured HRTFs as reported by Katz & Begault [2007]. The question of how to obtain valid measured HRTFs is of a broad importance, since it is relevant for either human (individual and non-individual) and dummy-head HRTFs. It relates not only to the actual measurement but also to the post-processing required so that HRTFs are consistent with their theoretical basis.

The implementation of HRTFs as filters in actual applications that use binaural synthesis has also undergone a significant amount of research. A widely used model is in the form of minimum phase filters with linear phase components (delays) that keep the interaural time difference or ITD information. This raises two questions: the first one relates to the methods for decomposing HRTFs and the second one relates to the perceptual validity of such a model. Regarding the first question, the literature reveals that it is actually difficult to decompose a signal into its minimum phase representation even though the theory behind different methods is well understood. Each method has limitations which more or less affect the results depending on the nature of the signal. There is abundant literature about how to decompose signals of different nature (EEG, seismic, etc.), but an assessment within the context of HRTFs is lacking. Regarding the second question, some investigations have shown that low Q-factor all-pass sections in HRTFs can be discarded unless they pass a threshold of audibility, in which case they have to be accounted for as part of the ITD. Whether high Q-factor all-pass sections can also be discarded has not been demonstrated, and there are only hypotheses regarding the topic. More specifically, it has been hypothesized by Møller *et al.* [2007] that high Q-factor all-pass sections in HRTFs are inaudible since they are centered at those frequencies where deep notches are present in the magnitude response. A dedicated study to confirm this hypothesis is required, since it deals with the validity of HRTFs implementation as a minimum phase filter and ITD.

The aforementioned questions relate to the technical issues of obtaining and implementing HRTFs in binaural synthesis, but there are other questions that remain unaddressed which have a much more basic nature. These other questions have remained unanswered since the fundamentals of spatial hearing were proposed more than a century ago, in spite of repeating the seminal experiments more carefully and with increased precision as technology allowed. The role of spectral features in sound source localization is a field that has been thoroughly studied, and yet it embraces many of those unanswered questions. It is known that HRTFs can be separated into temporal (ITD or delay) and spectral components, both of which are relevant in sound localization as just mentioned in the context of minimum phase decomposition. Spectral components from different subjects are characterized by noticeable differences due to anthropometric variations (Møller *et al.* [1995a]), particulary in the shape of the pinna. Even though the importance of the individual spectral features is widely accepted, it remains unclear whether relevant cues relate to broad frequency ranges or to more specific features. This is a relevant issue in the context of binaural synthesis: for many applications the use of non-individual HRTFs is required but the localization performance is degraded if the spectrum of the HRTFs used does not match the spectrum of the listener's own HRTFs. The magnitude of this degradation can be observed, even though the works cannot be directly compared, in Wightman & Kistler [1989b], Wightman & Kistler [1997], Wenzel *et al.* [1993], Møller *et al.* [1999], Møller *et al.* [1996a], Minnaar *et al.* [2001] and Langendijk & Bronkhorst [2000]. Understanding the role of spectral features would bring not only more insight into the processes involved in spatial hearing, but they will have consequences in the application of the technology. For example, it will help determining the *'range of distortion'* that a listener allows to his or her own HRTFs and still evoke the same sound source direction, and it will help clarifying whether an enhanced set of non-individual HRTFs that evokes the *correct* localization for a large sample of listeners can be found at all.

As said, current research and commercial applications of binaural synthesis make use of non-individual HRTFs. This is due to the impractical -if not impossible, depending on the end application- task of measuring and implementing individual HRTFs. Since the goal of the technology is to provide a realistic virtual sound experience in 3D space, a major emphasis is given to sound localization performance. It is believed that if all the technical steps are carefully controlled and the spectral shape of the HRTFs does not differ much from that of the listener, the goal has to be achieved. However, there are sound quality and behavioral issues that have not been addressed before and that relate to the potential success of the technology as a preferred system. For instance, binaural synthesis assumes that listeners can accurately localize real sound sources. The evidence is poor, and the literature mentions a duality between *good localizers* and *poor localizers* (Begault [2000]). Some studies have proposed training as an option to decrease the rate of front-back confusion when using non-individual HRTFs (Zahorik *et al.* [2006]), but the issue is that it is unclear what can be expected from the technology in terms of sound localization performance since it is not completely understood how subjects localize sound sources. On the other hand, an analysis beyond localization performance is needed, to investigate if using HRTFs with spectra that do not match that of the listener could introduce a degraded perception of sound quality. These issues are important as they relate to the potential of binaural synthesis to succeed and conquer the consumer market or at least compete with multichannel reproduction systems.

The efforts made on the advancement of binaural technology have historically followed two leads: commercial applications and hearing research. These are not completely independent in as much as the formers also benefit from new findings in the latter. The problems that have been outlined here are relevant for both leads. Moreover, they are also relevant for the two approaches that can be found in the implementation of binaural techniques: to allow for individual or, at least, individualized HRTFs or to work towards a set of non-individual HRTFs that provides good localization performance for a large population of people. This means that the issues that were chosen for investigation throughout this Thesis are of broad character and of general advancement in the field. Even though the work for this Thesis was framed according to previous work at Aalborg University and the approach to the problems kept consistency with methodologies already used, the technical and perceptual issues related to HRTFs that this Thesis covers are necessary steps to better understand binaural synthesis and how to use it.

1.2 Objectives

This Thesis focuses on investigating some of the key points that would help the construction of better sets of HRTFs and, if possible, the construction of enhanced sets of non-individual HRTFs -i.e. non-individual sets which evoke the *correct* sound source direction for a large sample of listeners and that has not quality issues from technical aspects such as measurement and choice of filter representation. A major emphasis is given to investigating whether spectral cues to sound localization can be identified and parameterized. The objectives of the work and different questions addressed are:

- Investigate those technical aspects associated to measuring HRTFs which can compromise the validity of the HRTFs for being used in binaural synthesis. Is it necessary/possible to work towards a suggested standard protocol to ensure good quality HRTFs measurements?

- Investigate different methods to decompose HRTFs into minimum phase and all-pass components, as a step forward towards a reliable method that can be used with large HRTFs databases.

- Investigate the perceptual consequences of removing high Q-factor all-pass sections from HRTFs, to verify the hypothesis proposed by Møller *et al.* [2007] that there are no perceptual consequences in doing so, since the all-pass sections are centered at frequencies with deep notches in the HRTFs frequency response.

- Investigate whether spectral features in HRTFs that cue sound localization in the mid-sagittal plane (MSP) can be identified and parameterized. How much mismatch in the spectra of HRTFs can a listener allow without compromising localization? Is it feasible to construct a non-individual set of HRTFs that provides the necessary spectral cues to sound localization for a large sample of users?

- Investigate the localization ability, with real sound sources and in the MSP, of subjects that present strong performance biases. Can localization errors be explained? Under which circumstances is sound direction *correctly* evoked? Is the conceptual duality *poor localizers* vs. *good localizers* appropriate in the context of binaural technology?

- Investigate issues that relate to the perceived quality of sound when spectrally non-matching non-individual HRTFs are used in binaural synthesis. Would the spectral mismatch be audible as quality degradation?

1.3 Original contribution

Different issues related to the enhancement of binaural synthesis through a better understanding of HRTFs characteristics have been investigated, which led to a range of original contributions. In general, the perceptual investigations led to more relevant original contributions than the technical studies. In the following, a list of contributions that are inherent to this Thesis, and which are ordered by relevance, is presented.

- Three potential spectral cues to sound source localization from HRTFs (whose impulse responses had been windowed to the first few samples) were identified and parameterized: the Q-factor of the first peak, the global Q-factor of the first notch and the center frequency of the second peak. The first parameter seemed to disambiguate directions above the horizontal plane (front-high, back-high and above), the second parameter seemed to disambiguate front, back and back-low regions, and the third parameter seemed to convey redundant information on back and back-low directions. The parameters found were seen to take ranges of frequency and values without compromising sound localization. The ranges for the parameters seemed to be individual for each subject. Parameters were seen to change smoothly in value and frequency as different regions of directions were evoked. Therefore, results suggested that determining ranges for the parameters that relate to constrained directions would not be relevant. As the spectral cues were defined in a broad range of frequencies covering more than a single spectral feature, it was seen that most of them had to be consistent with the particular direction that was to be evoked. The results suggested that it would be unlikely to achieve a set of non-individual HRTFs that evoked a better localization performance for a great sample of listeners, but supported two other approaches: *individualizing* non-individual HRTFs sets or selecting a few distinctive and representative sets. (Chapter 5).

- It was found that sound externalization of virtual sound sources reproduced binaurally through headphones did not necessarily correlate with localization performance -i.e. a sound source could be consistently localized to its target direction and still be perceived inside the head. The assessment of sound being perceived as externalized seemed to be inversely related to the awareness of the subjects about sound externalization: listeners that were actively asked about whether they were externalizing showed a much higher rate of inside-the-head answers than those who were not asked about it. (Chapter 5).

- Localization of real sound sources in an anechoic environment from five *biased localizers* was reported. These were *poor localizers* which presented strong biases in their responses: they perceived all the sound sources either in the rear or frontal hemispheres, or had degraded localization in specific ranges of frequency. The cases were reported descriptively as observations on human sound localization, since it is not clear what the basis for the strong biases is, nor how common/rare these cases are: only a few of them were reported previously in the literature (Itoh *et al.* [2007], Makous & Middlebrooks [1990]) and in the context of broader experiments, for which the cases were overlooked. For the cases of *biased localizers* that were tested with virtual sound sources synthesized with individual HRTFs, it was seen that the general trends exhibited in the real sound sources localization performance were maintained, even though there was no correlation between distributions. It is suggested that the validation of binaural synthesis should be done by correlating localization of real and virtual sound sources between them, instead of correlating each of them with target locations. It appears that binaural technology should be aimed at evoking those directions that subjects can actually perceive with real sound sources instead of evoking fixed nominal directions for which HRTFs are measured or in which loudspeakers are positioned. (Chapter 6).

- The ratio between non-individual and individual HRTFs that were matched as evoking the same direction was computed for 10 subjects. The results were compared with the evidence from the literature (Moore & Tan [2003]), and it was concluded that the spectral mismatch found would be potentially audible in terms of perceived quality of sound. The spectral mismatch was generally negligible in the low- and mid-frequency ranges, but large in the higher frequency range. However, in several cases the mismatch was characterized by ripples and slopes that occupied broad ranges of frequency, starting from frequencies well below 1 kHz. The spectral mismatch that subjects allowed so that localization was preserved seemed to exhibit subject and direction dependencies. Based on these observations, the following hypothesis was proposed: *Provided that the non-individual HRTFs used in binaural synthesis perceptually match the individual HRTFs of the listener in sound localization terms, it is hypothesized that the perception of sound quality would be degraded when the ratio of non-individual to individual HRTFs spectra deviates from a flat response in the low, mid and mid-high frequency ranges. It is further hypothesized that a direction dependent effect in the perception of sound quality degradation will be found.* (Chapter 7).

- The perceptual validity of HRTFs represented by a minimum phase function and a linear delay as ITD was completely confirmed since the last psychoacoustical test that was required for that confirmation was conducted: to test if high Q-factor all-pass sections could be removed without audible consequences in binaural synthesis. Candidate HRTFs all-pass sections were selected from a large database and they were used in a listening experiment where 12 subjects participated. The results showed that high Q-factor all-pass sections from HRTFs were audible if tested alone, but inaudible when they were combined with their associated minimum phase HRTFs. The results held for both binaural and diotic reproduction. The results confirmed the hypothesis proposed by Møller *et al.* [2007] within the topic, which stated that the all-pass sections would be inaudible when presented with their minimum phase counterparts because the ringing of the all-pass sections was centered at frequencies for which deep notches were present in the magnitude of the HRTFs. (Chapter 4).

- Three technical issues involved in measuring HRTFs were systematically investigated: DC correction, low frequency control and calibration. A case study using a dummy-head showed that these issues were closely interrelated and that they were crucial to ensure that measured HRTFs were valid for binaural synthesis. The protocol was extended to the measurement of HRTFs from 25 human subjects, whose results were also shown. (Chapter 2).

- Two methods of minimum phase decomposition were compared: Hilbert transform and homomorphic filtering. The assessment consisted of decomposing a large database of measured HRTFs and looking into the success rate and the computational time. It was found that the success rate depended on the choice of length N used in the FFT computation of the algorithms. While 63.28% of the HRTFs could be decomposed with their original length, the rest of the signals needed zero padding to achieve longer N values. The most problematic cases corresponded to the contralateral signals for directions to the sides, which were seen to present more zeros outside the unit circle. It was also shown that more than 80% of the pooled zeros from the database lied very close to the unit circle, which could potentially compromise the ability of the algorithms to decompose a signal into minimum phase. Hilbert transform and homomorphic filtering were comparable in terms of success rate as a function of length N, but homomorphic filtering proved to be less computationally demanding. (Chapter 3).

1.4 Overview of the Thesis

This Thesis relates two aspects of HRTFs: the technical validity that allows to use them in binaural synthesis and the perceptual characteristics that make them useful in binaural synthesis. The Thesis is organized as follows.

Chapter 2 presents the protocol used for measuring dummy-head and human HRTFs throughout the presented investigations. The issues of DC correction, low frequency control and calibration are explicitly analyzed and evaluated in a case study of dummy-head HRTFs measurements. The consequences of not controlling these issues are shown. HRTFs measurements from 25 subjects, for 15 sound sources in the MSP, are also reported.

Chapter 3 presents the theory and implementation of Hilbert Transform, homomorphic filtering and a z-domain method for computing minimum phase representations of HRTFs. Two of the methods are used to obtain minimum phase HRTFs from a large database, and the results are compared. In this Chapter, the locations of zeros in the z-plane for measured HRTFs are also analyzed.

Chapter 4 presents the 3-AFC listening test that was designed to assess the audibility of high Q-factor all-pass sections. The theory is very briefly covered as the concepts are those presented in Chapter 3. The choice of signals to be tested is explained, the experimental design is presented and the results are analyzed and discussed.

Chapter 5 presents listening tests conducted on 10 subjects, from which groups of individual and non-individual HRTFs that evoked similar directions were obtained. These HRTFs were analyzed in search of spectral features that cued sound localization and three potential features were identified and parameterized. The Chapter discusses not only spectral features in HRTFs but also externalization of virtual sound sources and the effect of solving or discarding front-back confusion cases in behavioral data.

Chapter 6 focuses on biased localization performance with real sound sources, for which results from 5 subjects are presented. Three of these subjects were further invited to localization tests with individual HRTFs binaural synthesis and the results are also shown. The Chapter takes a descriptive approach, since these cases have received little attention in the literature before. A discussion in terms of *good* and *poor localizers* is presented.

Chapter 7 analyzes the potential audibility, in terms of perceived sound quality, of the spectral filtering that non-individual HRTFs impose to the individual HRTFs of the listener. The Chapter focuses on a theoretical approach, were the computation of the filtering imposed is compared to previous studies on sound quality. A working hypothesis is defined together with an outline of a possible course to investigate it.

Chapter 8 presents a general discussion of the investigations part of this Thesis, where the studies are framed in a more comprehensive approach. The Chapter ends with a general conclusion and proposals for future work.

1.5 Coordinate systems

For simplicity, a general coordinate system was chosen for Chapters 2, 3 and 4, while a modified one was used for Chapters 5, 6 and 7. This decision relied on the nature of the HRTFs analyzed in these three latter Chapters.

The coordinate system chosen for Chapters 2, 3 and 4 is the same that was used by Algazi [1998]. The origin is the center of the head, and azimuth angles are computed between a vector to the sound source and the mid-sagittal plane (MSP). Azimuth angles vary from 90° (left side) to -90° (right side), with 0° to the front and 180° to the back. Elevation angles are computed between the horizontal plane and the projection of the source into the MSP. Elevation angles in the frontal hemisphere vary from -90° (down to the front) to 90° (above), with 0° directly to the front. Similarly, elevation angles in the rear hemisphere vary from -90° (down to the back) to 90° (above), with 0° directly to the back. All directions are given in (azimuth θ, elevation ϕ). As said, in this coordinate system $(0^\circ,0^\circ)$ is directly to the front and $(180^\circ,0^\circ)$ is directly to the back of the subject.

While the coordinate system described above is useful when directions in the whole sphere around the subject are considered, it is impractical for referring to HRTFs from one vertical plane only. That is the case of the HRTFs analyzed in Chapters 5, 6 and 7, where 15 HRTFs from the MSP are considered. In such a case, using the above described coordinate system would result in HRTFs ranging from $(0^\circ,-90^\circ)$ to $(0^\circ,90^\circ)$ in the frontal hemisphere, and HRTFs ranging from $(180^\circ,-90^\circ)$ to $(180^\circ,90^\circ)$ in the rear hemisphere. It was instead decided to refer to these HRTFs by their elevation angle only, as the azimuth angle is implied by stating that they belong to the MSP. Therefore, in Chapters 5, 6 and 7 HRTFs are referred to as ranging from -90° to 270° in elevation, which covers the whole plane (360°) from front-below to back-below. In this modified system, elevation at 0° is to the front of the subject and elevation at 180° is to the back of the subject.

Chapter 2

Issues on dummy-head HRTFs measurements

2.1 Introduction

HRTFs have been extensively measured for various purposes, as the literature reveals. Even though some studies have mentioned possible sources of errors and how to control them, there is still not a standard procedure to ensure the quality of measured HRTFs. For example, it is reported by Algazi [1998], Zotkin *et al.* [2006] and Hammershøi & Møller [2005] that the low frequencies in measured HRTF need to be controlled. Considering that the dimensions of a person are small compared to the wavelength at low frequencies, it is expected that HRTFs would decrease asymptotically until they reach 0 dB -i.e. unity gain- at DC. This is not the case in measured HRTFs: the limitations of the equipment used in the measuring chain result in a wrong -and random- value at DC and the effect can be seen well within the audio frequencies. Moreover, improper microphone calibration can also affect the results. These issues are relevant particularly if the HRTFs are used in perceptual studies or applications, since informal listening tests suggest that the effect of a wrong low frequency range affects the sound quality in binaural synthesis.

In this chapter, issues associated to calibration, DC correction and low frequency response are analyzed in measured HRTFs of the commercially available dummy-head Neumann KU 100. These measurements were conducted as a contribution to an international round robin of HRTFs measurement systems (see Katz & Begault [2007] for preliminary results) and are presented here as a case study. The measurement of human HRTFs that were conducted for the experiments reported in Chapters 5, 6 and 7 are presented as a generalization of the measurement protocol.

The Chapter is organized as follows: firstly, the relevant literature is reviewed and some basic concepts such as free field transfer function and diffuse field equalization are covered. The theoretical framework for HRTFs measurements is also briefly introduced. Secondly, the generalities of the HRTFs measurement procedure that was followed is presented. In following sections, the issues of calibration, DC correction and low frequency response are analyzed. These issues are not independent from each other but are treated separately for simplicity. A discussion follows where the relationship among the issues is exposed. As part of the discussion, HRTFs measurements from human subjects conducted by following the proposed measurement protocol are presented, along with brief recommendations for ensuring the validity of HRTFs measurements. Finally, some concluding remarks are made.

Part of the work presented in this Chapter has already been reported by Toledo & Møller [2009].

2.2 Previous works

The use of manikins for sound recording was already reported in the decade of 1920, according to Paul [2009]. Attempts to measure sound at the ears of dummy-heads did not come much later (Firestone [1930], Mills [1958]). As soon as the technology allowed, small microphones were positioned in the ears of dummy-heads and human subjects to measure head-related transfer functions. These measurements were conducted for a variety of purposes: to characterize phase and magnitude of HRTFs (for example, Mellert *et al.* [1974], Searle *et al.* [1975], Mehrgardt & Mellert [1977], Butler & Belendiuk [1977], Hiranaka & Yamasaki [1983], Humanski & Butler [1988], Middlebrooks *et al.* [1989], Wightman & Kistler [1989a], Middlebrooks & Green [1990], Carlile & Pralong [1994], Han [1994], Macpherson [1994], Møller *et al.* [1995a], Brungart & Rabinowitz [1999], and others as reviewed by Shaw [1974]), to characterize the anthropometrical based features that would allow developing and optimizing dummy-heads (for example, Burandt *et al.* [1991], Burkhard & Sachs [1975]), to present and/or validate different HRTFs modeling methods with an analytical, non-perceptual, approach (for example, Middlebrooks & Green [1992], Chen *et al.* [1995], Brown & Duda [1998], Middlebrooks [1999], Minnaar *et al.* [2005], Kahana [2000], Katz [2001], Algazi *et al.* [2002a], Fels *et al.* [2004]), to develop computational models of sound localization (for example, Chung *et al.* [2000]), to show the application of new measurement procedures (for example, Pralong & Carlile [1994], Zotkin *et al.* [2006], Majdak *et al.* [2007]) and to synthesize binaural signals to be presented in listening tests -mainly localization studies but also perceptual investigations of different character, including perceptual validation of modeled HRTFs- (for example, Wightman & Kistler [1989b], Asano *et al.* [1990], Kistler & Wightman [1992], Algazi *et al.* [2001b], Shimada *et al.* [1994], Bronkhorst [1995], Blauert *et al.* [1998], Kulkarni *et al.* [1999], Huopaniemi *et al.* [1999], Langendijk & Bronkhorst [2000], Jin *et al.* [2000], Brungart & Scott [2001], Begault *et al.* [2001], Langendijk & Bronkhorst [2002], Pernaux *et al.* [2002], Jin *et al.* [2004], Best *et al.* [2005], Minnaar *et al.* [2005], Iida *et al.* [2007]).

In such different contexts, measuring HRTFs was sometimes a secondary activity to the main study being reported. As a consequence, the technical details of the HRTFs measurement procedure followed were, in some cases, not even described. In those studies which did report the details, large variability was found in the choice of measurement method, directions measured, placement of the microphones, environment, etc. In other words, methods and measurement conditions varied largely across studies and there is not enough information to thoroughly compare procedures and results.

There were some authors, however, who mentioned very precise concepts and methodologies involved in the measurement pocedures. Algazi [1998] and Algazi *et al.* [2001a] reported the need to control the low frequency behavior of measured HRTFs: at DC, HRTFs were expected to approach 0 dB but that was not seen in the measurements due to noise and the offset voltages of the equipment used. Furthermore, it was acknowledged that the HRTFs pressure division exacerbated the problem resulting in a DC value without physical meaning. In a later study, Algazi *et al.* [2002a] measured pinnaless dummy-head HRTFs to validate numerical methods of modeling HRTFs. They reported that, since the loudspeakers used in the measurements did not radiate sound at low frequencies, the signal-to-noise ratio below 500 Hz was poor and the resulting HRTFs had to be ignored below that frequency. Brown & Duda [1998] measured HRTFs from 3 human subjects to evaluate a model for binaural synthesis. They discussed that the pressure division that defines HRTFs could not be done at DC, and that in order to force the response towards the free field reference as 0 Hz was approached, it was decided to set the value of DC to unity in the frequency domain. Begault [2000] discussed the methodology for measuring HRIRs and storing them for their implementation as filters. He acknowledged that if the frequency response of the equipment used in the setup -i.e. microphones and loudspeakers- deviated from flat, the HRTFs would be affected in some way and therefore it had to be accounted for. Riederer [1998] mentioned that background noise and reflections from the setup limited the signal-to-noise ratio of the HRTFs measurements. It was considered that the signal-to-noise-ratio was severely affected in contralateral sides. Other sources of inaccuracies were the DC offsets introduced by the electrical equipment. To remove the DC offset, Riederer recommended measuring with two subparts with opposite polarities, so that differentiating the subresults would remove the DC offset errors -but not the DC component. Riederer stated that the method did not provide improvements in the frequency response analysis, since the information around 0 *Hz* was disregarded. Zotkin *et al.* [2006] reported that low frequencies in HRTFs measurements were prone to inaccuracies due to the poor low frequency response of the loudspeakers usually used for the measurements, the truncation required to window out reflections from the setup and the poor anechoic characteristics of any measurement environment (including anechoic chambers) in that range, leading to poor signal-to-noise ratio. However, the authors acknowledged the possibility of using simplified models as that reported by Algazi *et al.* [2002b] to analytically represent the HRTFs at low frequencies. Hammershøi & Møller [2005] also covered different issues inherent to HRTFs measurements. Of special importance for this Chapter is their discussion on low frequency control of HRTFs, as several of the points that will be analyzed here were already outlined in that work. For example, the authors showed how an incorrect DC value produced by offset voltages from the measuring equipment affected a broad range of frequencies. They proposed correcting the DC value either in frequency or time domain. The authors also mentioned the possibility of perceptual consequences if large differences in the DC value at both ears existed: an unpleasant feeling of sub- or super-pressure in the ears.

The issue of ensuring the validity of measured HRTFs is relevant in many planes: for example, wrongly measured HRTFs can lead to a wrong subjective perception, as mentioned already, thus breaking the fundamentals of binaural techniques. They can also lead to wrong objective evaluations of, for example, modeled HRTFs. As there are researchers who cannot measure HRTFs and have to rely in measurements made by others, it is important that the quality of the measurements can be checked. Fortunately, the publicly available databases of HRTFs such as KEMAR (reported by Gardner & Martin [1994] and used by Lopez Poveda & Meddis [1996] and Larcher *et al.* [1998]), IRCAM (IRCAM [2002]) and CIPIC (reported by Algazi *et al.* [2001a] and used by Zotkin *et al.* [2004], Raykar *et al.* [2005], Nicol *et al.* [2006], Xu *et al.* [2009]) seem to be well documented in their measurement procedure. Measurements conducted by Wightman and Kistler, even though not publicly available, have also been used by others (Wenzel *et al.* [1993], Wenzel & Foster [1993], Kulkarni *et al.* [1995], Kulkarni *et al.* [1999], Grantham *et al.* [2005]).

In this context, it is clear that some standardization is needed if an experiment from one laboratory is to be replicated at another place. An antecedent to a comparative study of HRTFs measured at different laboratories was reported by Shaw [1974], where investigations dating from the year 1933 to the year 1972 were gathered in search of average transformation curves. A small round robin reported by Blauert *et al.* [1998] was the starting point for a set of *golden rules* for HRTFs measurement procedures but still large cross-laboratory differences are seen in more contemporary studies. Those differences have lead, for example, to a current round robin of HRTFs measurement systems (Katz & Begault [2007]). The dummy-head measurements for the case study presented here were a collaboration to that round robin. The HRTFs measured from human subjects, on the other hand, were used in the listening tests reported in Chapters 5, 6 and 7. It is the goal of this investigation to present a systematic analysis of some of the issues that can be faced when measuring HRTFs.

2.3 Background concepts

The Neumann KU 100 used in the case study is a dummy-head with built-in diffuse field equalization (see the documentation of the dummy-head in Neumann [2009]). Therefore, some basic concepts are reviewed in the following -e.g. HRTF, free field transfer function, diffuse field equalization, etc. Even though these definitions are clearly stated in the literature -see, for example, Blauert [1997], Møller [1992], Hammershøi & Møller [2005]- it seems that they are not always followed.

If the human anthropometry is considered as a linear time-invariant system, the transformations that it imposes over an impinging sound can be expressed as a transfer function. By definition, two terms are then necessary to obtain the transfer function: the output and the input to the system. In the broad sense, HRTFs are defined as the complex pressure division of the sound incoming at the ears of a subject ($P2$ or output of the system) to the sound at the position of the center of the head when the subject is absent ($P1$ or input to the system).[1] This can be expressed as:

$$HRTF = \frac{P2}{P1} \tag{2.1}$$

As one $P2$ measurement exists for each ear, HRTFs are defined in pairs which are angle dependent. It is normal practice to state to which coordinate system the HRTFs are being referred to. Angles are usually given in (azimuth θ, elevation ϕ), which is the nomenclature also used in this work (see Chapter 1, 1.5, for more information on the coordinate systems used in this Thesis).

The microphone position for the measurement of $P2$ can vary: it can be at the entrance of the blocked or open ear canal, at the eardrum or at some known position in the ear canal. A review on the choice of measurement point is given by Hammershøi & Møller [1996].

[1] Note that the names $P2$ and $P1$ are arbitrarily chosen for consistency with previous work done at Aalborg University's Acoustics Section (see Møller [1992] and Møller *et al.* [1995a], among others).

The definition of HRTF as in Eq. 2.1 has received other names in the literature: free field transfer function (Blauert [1997]), sound pressure transformation from the free field to the eardrum or to the outer ear (Shaw [1974]), external ear transfer function (Mehrgardt & Mellert [1977]), transfer function from free sound field to ear canal entrance or to the eardrum (Mehrgardt & Mellert [1977]).

Other transfer functions within the context of binaural techniques have been defined by Blauert [1997]: interaural transfer function and monaural transfer function. The former relates the sound pressure measured at both ears of the subject, where the reference sound pressure is that at the ear facing the sound source. Monaural transfer functions relate the sound pressure at the ears of a subject to the sound pressure measured at the same position but with the sound source located at a reference position -as a rule, it corresponds to the position to the front with coordinates $(0^\circ,0^\circ)$. Monaural transfer functions can also be referenced to a diffuse field (Møller [1992]). In that case, the reference is the average of the transfer functions from all directions.

If Eq. 2.1 is considered from a practical point of view, $P2$ and $P1$ are ideal transfer functions that have to be obtained from real measurements which are defined here as M_{P2} and M_{P1}, respectively. These M_{P2} and M_{P1} measurements also contain the transfer functions of the measurement setup. In the case study presented here, the following transfer functions are included:

- Transfer function of a computer-based *MLS* system (Olesen *et al.* [2000]).
- Transfer function of a RME ADI-8 DS AD/DA converter.
- Transfer function of a Pioneer A-616 power amplifier.
- Transfer function of 3-inch loudspeakers VIFA M10MD-39 (15 different units were used).
- Transfer function of the microphones used for M_{P2} and M_{P1} measurements.
- Transfer functions of two Brüel & Kjær 2607 measuring amplifiers.

A complete explanation of the measurement chain will be given in 2.4.1. Equation 2.1 is approximated as:

$$HRTF = \frac{M_{P2}}{M_{P1}} \tag{2.2}$$

If the same setup is used for both M_{P2} and M_{P1} measurements, the transfer functions listed above are canceled out in Equation 2.2. That is the case in the investigation presented here, except for the transfer functions of the microphones. Following a requirement of the round robin for which these measurements were done, the internal microphones of the dummy-head Neumann KU 100 were used for M_{P2} measurements. The pressure field microphone Brüel & Kjær 4136 was chosen for the reference measurement M_{P1}. In situations where the microphones are the only difference in the setups used for M_{P2} and M_{P1} measurements, HRTFs are obtained by compensating for the transfer characteristics of the respective microphones:

$$HRTF = \frac{M_{P2} \cdot H_{mic.P1}}{M_{P1} \cdot H_{mic.P2}} \Rightarrow HRTF = \frac{H_{P2}}{H_{P1}} \tag{2.3}$$

Which is equivalent to Eq. 2.1:

$$HRTF = \frac{P2}{P1} = \frac{H_{P2}}{H_{P1}} \tag{2.4}$$

In this case study, however, there was an extra factor besides the microphones. As mentioned before, the dummy-head Neumann KU 100 has a built-in diffuse field equalization, where the equalization filter is defined as the average of the transfer functions from all directions. M_{P2} can be expressed as:

$$M_{P2} = H_{P2} \cdot H_{setup} \cdot H_{diffuse} \tag{2.5}$$

Taking the microphone transfer functions out from H_{setup}, the ratio $\frac{M_{P2}}{M_{P1}}$ becomes:

$$\frac{M_{P2}}{M_{P1}} = \frac{H_{P2} \cdot H_{setup} \cdot H_{mic.P2} \cdot H_{diffuse}}{H_{P1} \cdot H_{setup} \cdot H_{mic.P1}} \tag{2.6}$$

As said, the setup was the same in the measurements presented here and therefore H_{setup} canceled out in Eq. 2.6. By calibrating the microphones and compensating for them as in Eq. 2.3, an approximated HRTF was obtained:

$$HRTF_{app} = \frac{H_{P2} \cdot H_{diffuse}}{H_{P1}} \tag{2.7}$$

In other words, in order to measure HRTFs as in Eq. 2.1 while the internal microphones of the dummy-head were being used, the inverse of the equalization filters $H_{diffuse}$ would have had to be applied. As the exact characteristics of $H_{diffuse}$ were unknown and were not compensated for in this case study, it is actually inaccurate to refer to the measurements as HRTFs. They are, however, referred to as HRTFs for simplicity.

2.4 HRTFs Measurement Procedure

The protocol followed to measure M_{P2} and M_{P1}, and the processing required to obtain H_{P2} and H_{P1} are given in this section. The issues that will be developed in further sections are also pointed.

2.4.1 Measurement Setup

The measurements were made in an anechoic chamber, the setup is depicted in Figure 2.1. The dummy-head stood in the center of an arc with 15 loudspeakers placed along it. The separation between loudspeakers was 22.5° and the distance from each loudspeaker to the point in the center of the head was 1.5 *m*. Sound sources were 3-inch VIFA M10MD-39 loudspeakers mounted in hard plastic balls. The whole setup was covered with absorbent material to avoid reflections as much as possible.

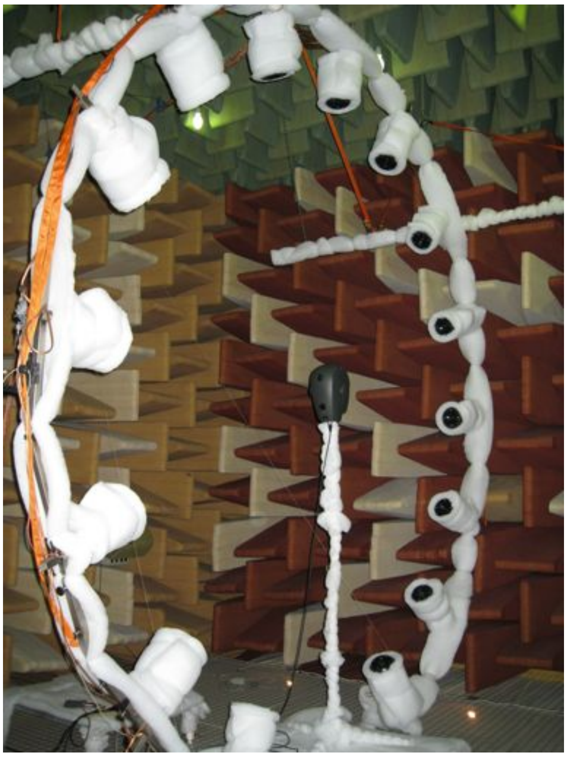

Figure 2.1: Setup used for the HRTF measurements conducted in the context of this thesis.

The head was rotated in 30° steps by means of a turntable Brüel & Kjær type 3921 (not shown in Fig.2.1). The rotation was done with respect to the head's stand, which was not coincident with the vertical axis crossing the center of the head -angular and distance errors were introduced by this procedure, being the maximum absolute angular error equal to 1° and the maximum absolute distance error equal to 5 *cm*. The position of the dummy-head was controlled by lasers crossing the mid-sagittal plane (MSP) and the interaural axis. A total of 85 HRTFs were measured, where the azimuth was sampled in 30° steps and the elevation was sampled in 22.5° steps.

The measuring equipment was placed in a control room next to the anechoic chamber. An in-house developed two-channel computer-based *MLS* system (reported by Olesen *et al.* [2000]) was used for the transfer function measurements. The computer was equipped with a digital sound card RME HDSP 9632 and generated digital signals that were fed to a RME ADI-8 DS AD/DA converter. Analog signals were then fed to a power amplifier Pioneer A-616 calibrated to provide 0 *dB* gain. The output of the power amplifier was sent to a switch box, controlled through the parallel port of the PC, which diverted the signal to the desired sound source. The balanced 5-pin XLR output of the dummy-head was used to provide external polarization -a phantom power supply Neumann BS 48-i2 was used- and to obtain the output signals. Internal microphones were calibrated for their sensitivity at 1 *kHz* (see Section 2.5). The balanced outputs were converted into unbalanced and delivered to two measuring amplifiers Brüel & Kjær 2607. The output from the measuring amplifiers fed the signals to the RME ADI-8 DS AD/DA converter. Digital signals went back to the PC for the transfer function computation. The results were impulse responses of 2048 samples length, at a sampling frequency of 48 *kHz*. Post-processing included compensating for a phase inversion that the Brüel & Kjær 2607 imposed on M_{P2} measurements (see 2.4.2). In the case of M_{P1} measurements, only the gain factor of the measuring amplifier was accounted for.

2.4.2 Frequency response of the setup

The assumption of a flat frequency response and deviations from nominal gains were verified with measurements. The Pioneer A-616 power amplifier presented negligible deviations from the nominal 0 *dB* gain. The measuring amplifiers showed deviations from the nominal gain settings of the order of 0.2 *dB* and they were compensated for in all measurements in the post-processing stage. In the case of M_{P2} measurements, the whole transfer function of the measuring amplifiers was deconvolved, in order to correct a phase inversion introduced by them. In the case of M_{P1} measurements, the phase needed not to be corrected: both the microphone Brüel & Kjær 4136 and the measuring amplifier produce a phase inversion, canceling each other.

Calibration of the internal Neumann KU 100 microphones was done with a sound level calibrator Brüel & Kjær 4230. It was not a straightforward procedure, as it is explained below in Section 2.5. Since measurements of M_{P2} were done over two different days, two calibration values were obtained for each internal microphone.

The reference microphone Brüel & Kjær 4136, on the other hand, was a $\frac{1}{4}$-inch microphone with a flat frequency response in the range 20 Hz $-$ 40 kHz. It was also calibrated for its sensitivity at 1 kHz.

2.4.3 Post-processing

Original measurements had 2048 data points. A rectangular window was applied to the raw M_{P1} and M_{P2} measurements and data points from 55 to 310 were used. The choice of window is debatable, as it is well known that rectangular windows *smear* the frequency content of the signal due to the Gibbs phenomenon (Oppenheim & Schafer [1975]). The Fourier transform of a truncated signal can be thought of as the convolution of the Fourier transform of the infinite version of the signal and the Fourier transform of a rectangular window. The spectral shape of the rectangular window is a *'sinc()'* function, which is characterized by a major lobe and side lobes. If the window was infinite, its spectral shape would resemble a delta function and the true Fourier transform could be obtained. However, since the window is finite, in the computation of the discrete-time Fourier transform not only the frequencies of the wide main lobe are present but also those of the side lobes. Other windows than rectangular ones could be chosen where the side lobes are diminished, but at the expense of an even wider main lobe. Conceptually, then, the choice of window becomes a compromise between its length and the width of its major lobe. These are conflicting requirements since windows are to be as short as possible and the major lobe is to be as narrow as possible. Knowing that spectral smearing can be then reduced but not avoided, rectangular windows were chosen in spite of their characteristics. A discussion of windows applied to HRTFs filters, with examples, was also reported by Sandvad & Hammershøi [1994]. Focusing on the post-processing of the measured M_{P1} and M_{P2} signals, it can be said that the windows applied ensured that all the impulse responses had died out. However, they could not exclude the first reflections from the setup. These were seen to come from sources in the arc neighboring the one being used in the measurement.

The files were further processed to account for the gain and phase of the measuring amplifiers and the sensitivity of the microphones. The frequency response of the measuring amplifier was deconvolved from M_{P2} measurements as mentioned in Section 2.4.2 -this was done by a division in the frequency domain.

The low frequencies of M_{P1} and M_{P2} measurements presented different transfer characteristics, and it was hypothesized that there was a low frequency gain inherent to the

Neumann KU 100 internal microphones. In order to confirm this, the low frequency investigation reported in Section 2.7 was conducted. As a result of that investigation, M_{P2} measurements were filtered with inverse filters that equalized the low frequency response of the internal microphones.

2.4.4 Computation of HRTFs

Once M_{P2} measurements were filtered to equalize their low frequency range, the approximated free field HRTFs were computed as a complex pressure division (division in the frequency domain) according to Equation 2.7. The results were low pass filtered. HRIRs were computed from the inverse Fourier Transform. HRIRs were circularly shifted 60 samples to ensure causalty. All HRIRs were shifted the same amount of samples in order to keep the interaural time difference information. Finally, HRIRs were DC corrected in time domain to provide a meaningful value at 0 Hz and minimize the effects of truncation, as explained in Section 2.6. Figures with all the measured HRTFs can be found in Appendix A. Only the HRTFs from the frontal direction $(0°, 0°)$ will be used in the following, as a representative case.

2.5 Calibration

In the context of microphones, the amount of electrical output for a certain amount of sound pressure presented to a microphone is expressed as sensitivity. The units commonly used are V/Pa (volts output per Pascal of pressure applied) or $dB\ re.\ 1\ V/Pa$ (decibels relative to 1 volt per Pascal). Microphone sensitivities are determined through calibration. Calibration can be performed either in the field or in a laboratory, and there are several methods that can be used -comparison method, substitution method, calibration by the use of a pistonphone or sound level calibrator, among others. A review of the different methods can be found in Brüel&Kjær [1996]. Nominal sensitivities are usually given for 1 kHz, which is useful if microphones have a flat frequency response but of limited help if they do not.

The requirement of calibration is widely accepted since it ensures that measurements are correctly done and the equipment involved is accurate. It also accounts for environmental variabilities, enabling measurements to be compared. This latter issue is particularly relevant in round robins like the one for which the measurements on the Neumann KU 100 dummy-head were done: the head is sent to a variety of laboratories at different locations in the world with different environmental conditions. One way to ensure that the measurements from these laboratories can be compared is through calibration of the microphones.

While the previous applies to any electroacoustical measurement, calibration is an even more critical issue in the context of HRTFs as they result from the ratio of two measurements. As explained in 2.3, if the microphones used for M_{P1} and M_{P2} measurements are not the same, their sensitivities and frequency response will be different. It can be the case that, even if the same microphone is used for both measurements, the sensitivity changes due to environmental conditions. Therefore, proper calibration has to be conducted in order to cancel the $H_{mic.P2}$ and $H_{mic.P1}$ terms as in Eq. 2.3, unless they are equal -i.e. the same microphone is used under the same environmental conditions.

Even though the reviewed concepts are well established and are considered as standard procedure, a proper calibration is not always straightforward. This was the case when attempting to calibrate the internal microphones of the Neumann KU 100 dummy-head. In the following, HRTFs obtained with different calibration values are compared.

2.5.1 Neumann KU 100 internal microphones

The internal microphones of the dummy-head under study consisted of two pressure transducers with nominal sensitivity at 1 kHz of 20 mV/Pa $\pm 1 dB$. The microphone capsules were hosted in ear adapters that contained ear channels. The ear adapters were attached to a cylindrical enclosure that included built-in filters. Since the microphone capsules could not be detached from the ear adapters nor the cylindrical enclosure, only calibration with a sound level calibrator was possible. The manufacturer of the dummy-head recommends the Neumann PA100 adapter for calibration, which is an accessory part to the dummy-head. That adapter fits into a $\frac{1}{2}$-inch Brüel & Kjær adapter, which in turn fits into a 1-inch Brüel & Kjær calibrator. However, the Neumann PA100 was not available when the measurements were done. An alternative calibration procedure was followed.

2.5.2 Calibration Procedure

The calibration was done by means of a Brüel & Kjær 4230 sound level calibrator and the procedure started on the day of the measurements, when the sensitivity of the microphones at 1 kHz was approximated by combining a $\frac{1}{2}$-inch adapter with a $\frac{1}{4}$-inch adapter (both Brüel & Kjær). Since the Neumann KU 100 internal microphones were slightly smaller than $\frac{1}{4}$-inch, their insertion in the $\frac{1}{4}$-inch adapter was sealed to avoid leakages.

The Neumann PA100 adapter was received at a later day and used to verify the calibration done the day of the measurements. The difference between the two calibration procedures -i.e. ($\frac{1}{2}$-inch Brüel & Kjær adapter + $\frac{1}{4}$-inch Brüel & Kjær adapter) vs. ($\frac{1}{2}$-inch Brüel & Kjær adapter + Neumann PA100 adapter) was computed. In average, the difference between the two calibration procedures amounted to 2 *dB*. This difference includes an insertion loss of 1.7 *dB* reported by the manufacturer of the Neumann PA100.

The uncertainty in the calibration method ranges from 0.07 *dB* to 0.3 *dB* according to Brüel&Kjær [1996].

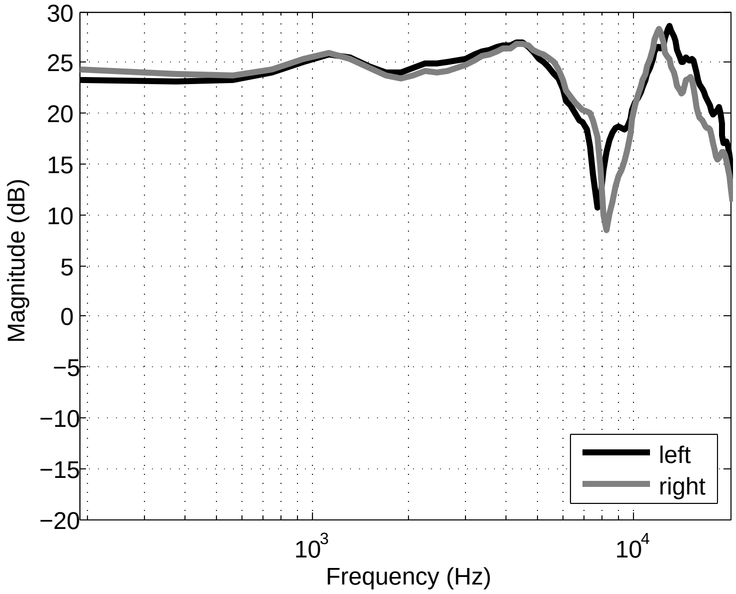

Figure 2.2: Measured HRTFs for direction $(0°, 0°)$, without microphone calibration.

2.5.3 Results and analysis

Figure 2.2 shows the measured HRTFs for direction $(0°, 0°)$, where M_{P1} and M_{P2} have not been post-processed with their corresponding calibration values and therefore the overall gain is wrong. It can be seen that the whole frequency responses are shifted upwards, and they do not decrease asymptotically until reaching 0 *dB* at DC. Figure 2.3 shows the processed HRTFs where both M_{P1} and M_{P2} have been calibrated with the values obtained by combining a $\frac{1}{2}$-inch Brüel & Kjær adapter with a $\frac{1}{4}$-inch Brüel & Kjær adapter. It can be seen that the responses are still shifted upwards, meaning that there are still added gain factors that should be accounted for. Figure 2.4 shows the same HRTFs as in Figure 2.3, but the calibration of M_{P2} measurements have been corrected by 2 *dB* according to the findings mentioned before (difference between a $\frac{1}{4}$-inch Brüel

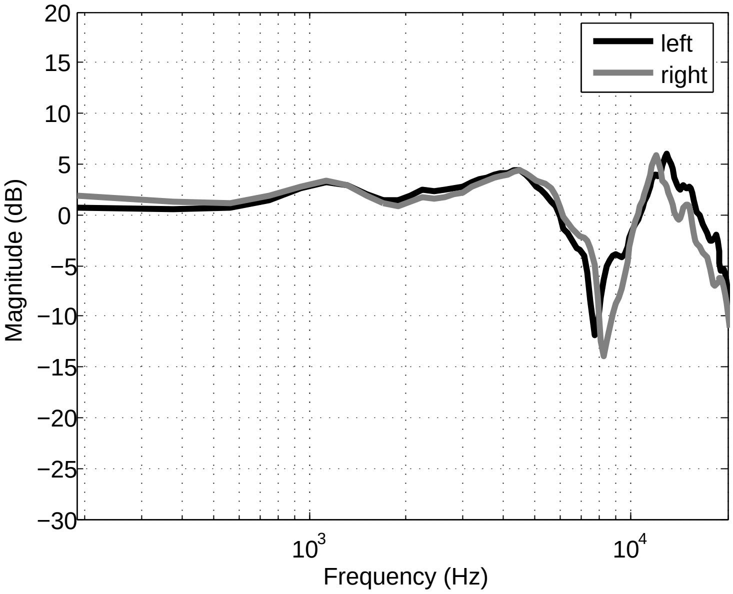

Figure 2.3: Same as Fig. 2.2 but with microphone calibration. Calibration of internal microphones done by combining a $\frac{1}{2}$-inch with a $\frac{1}{4}$-inch Brüel & Kjær adapters.

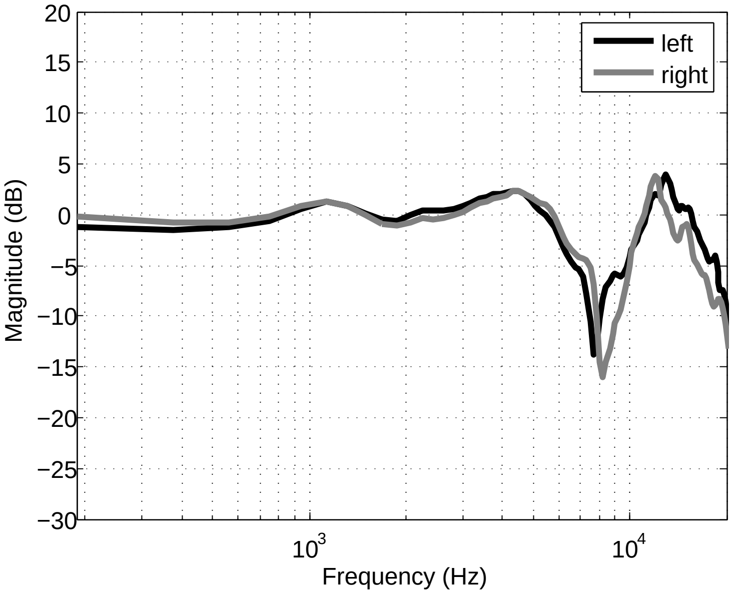

Figure 2.4: Same as Fig. 2.3 but the calibration of the internal microphones was corrected by 2 *dB* -which was the average difference between using a $\frac{1}{4}$-inch Brüel & Kjær adapter and the Neumann PA100 adapter.

& Kjær adapter and the Neumann PA100 adapter, see 2.5.2). It can be seen that the low frequency responses are closer to the expectation but there are still some differences between right and left side. Furthermore, there are deviations from the frequency response reported by the manufacturers for the same direction (Neumann [2009]).

The calibration procedure described is not free from errors: they could arise as a result of the already mentioned uncertainty in the calibration method and also from the calculation of the difference between the two calibration procedures -with and without the Neumann PA100 adapter. However, it was hypothesized that the low frequency behavior seen in Figure 2.4 was due to the choice of calibration method: if the transfer characteristics of the Neumann KU 100 internal microphones were not flat, calibration with a sound level calibrator would not be the proper choice as the method assumes a flat frequency response. A non-flat frequency response was plausible, since the built-in diffuse field equalization circuit could not be by-passed when calibrating the internal microphones. A dedicated investigation was conducted and it is reported in Section 2.7.

2.6 DC Correction

At low frequencies, the dimensions of a person become much smaller than a wavelength. Hence, ideal HRTFs are expected to decrease asymptotically until they reach 0 dB or unity gain at DC. This is not the case in measured HRTFs mainly due to two factors: limitations of the measurement setup and restrictions on the length of the HRTFs filters.

2.6.1 Limitations of the measurement setup

Sound is not reproduced nor measured at DC and this holds for both M_{P1} and M_{P2} measurements. If values are obtained at DC in these measurements, they obey the offset voltage properties of the acquisition equipment used. This issue has already been pointed by Hammershøi & Møller [2005] and Algazi [1998], among others. Moreover, the DC value in HRTFs results from the ratio of two measurements with non-zero DC value. The ratio, therefore, results in a meaningless -wrong and more or less random-value at DC. While the DC value is usually not plotted in HRTFs figures as it corresponds to 0 Hz, its value can be checked by inspecting the first coefficient of the HRTF.

2.6.2 Length of the HRTFs filters

The length of measured HRIRs is often decided so as to avoid possible reflections from the setup. These short HRIRs measurements are often implemented as FIR filters which seem to be perceptually valid: Sandvad & Hammershøi [1994] have shown that HRIRs as FIR filters of 72 taps of length (sampled at 48 *kHz*) were long enough to convey all the needed cues to sound localization. Nevertheless, these short filters impose a poor frequency resolution and noticeable consequences are seen in the low frequency range -which is then represented by too few taps. For example, Figure 2.5 shows the impulse responses (IRs) of two FIR filters of 1024 taps, corresponding to direction $(0°, 0°)$. Some reflections can be clearly seen around 0.01 seconds, which affect the whole frequency responses as seen in Figure 2.6. In turn, Figure 2.7 shows the obtained impulse responses if the FIR filters are constructed with only 256 samples of the HRIR measurements -i.e. a rectangular window of length $N = 256$ is applied. The corresponding frequency response of the filters is shown in Figure 2.8. Even though the responses are smoother than in Figure 2.6 due to the lack of reflections, the low frequencies are farther from 0 *dB* than in Figure 2.6. Moreover, some ripples can be seen in the low frequency range -around and above 200 *Hz*.

The plotted responses in Figures 2.5 to 2.8 are 4096 samples long but the FIR filters from which they were obtained are much shorter. Even though the FIR filters are determined for a few limited frequencies (determined by the actual measurement points used), they are still filtering those frequencies in-between and unfortunately ripples appear.

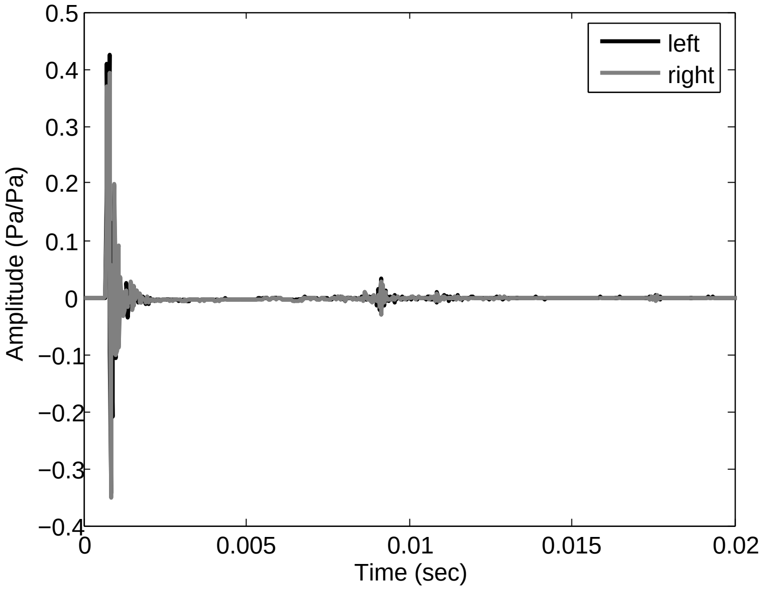

Figure 2.5: IRs of the FIR filters of HRTFs corresponding to direction $(0°, 0°)$, constructed from 1024 measured samples. IR were computed with $N = 4096$ samples.

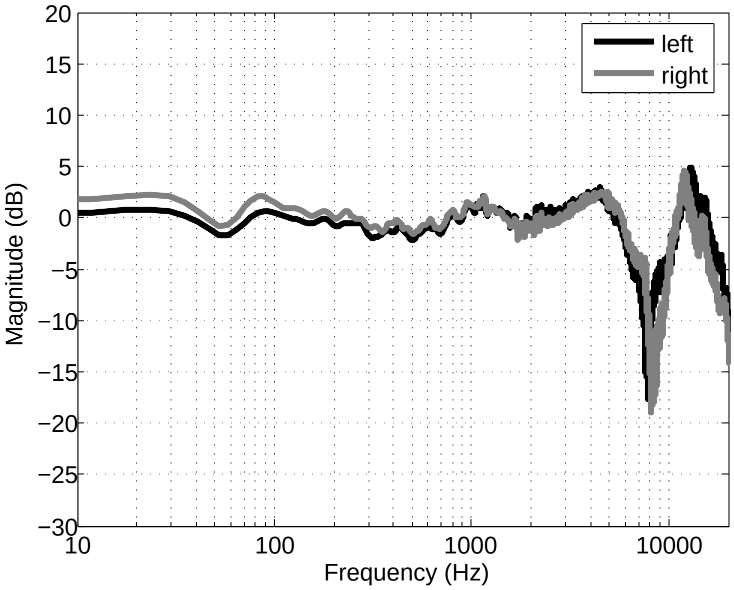

Figure 2.6: Frequency responses of the FIR filters shown in Fig. 2.5.

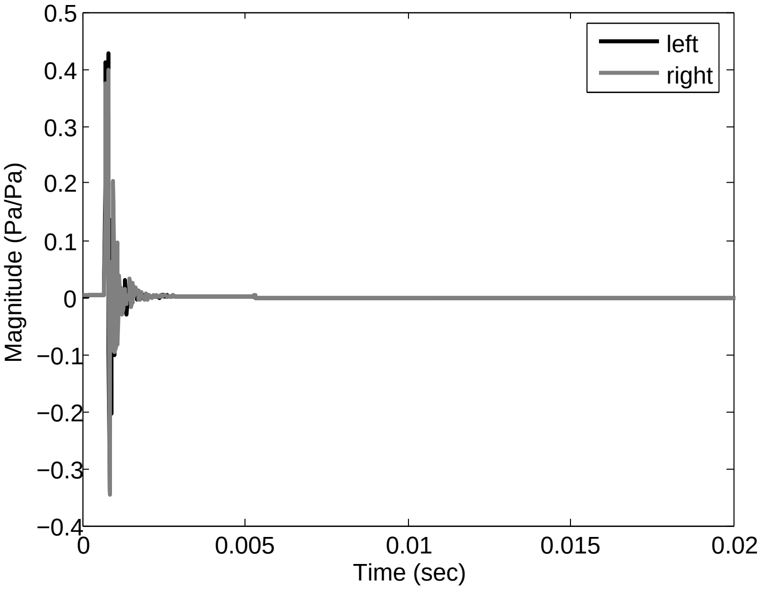

Figure 2.7: Same as in Fig. 2.5, but FIR filters were constructed from 256 samples of the HRIRs measurements.

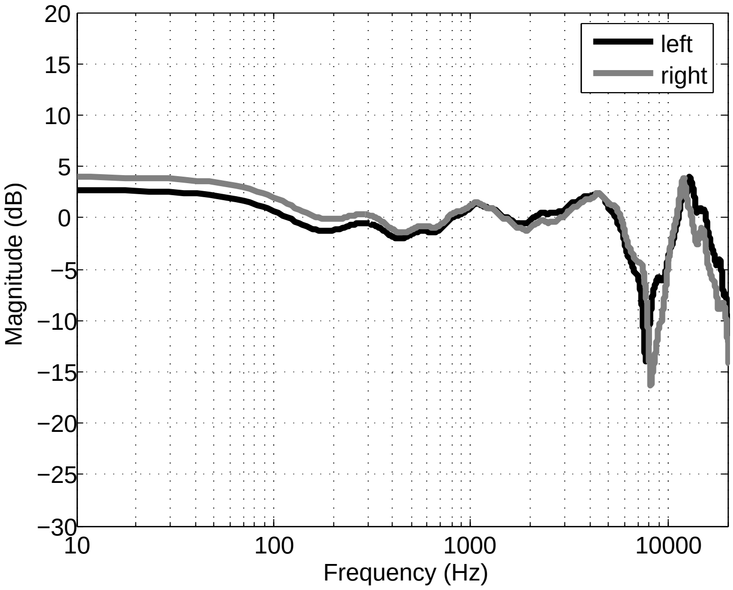

Figure 2.8: Frequency responses of the FIR filters shown in Fig. 2.7.

Informal listening tests suggest that the ripples seen in the low frequency range affect the sound quality in binaural synthesis, as already reported by Hammershøi & Møller [2005] and checked by informal listening conducted by the author of this Thesis.

2.6.3 Results and analysis

In Figure 2.9, the DC value of the filters was corrected to equal unity gain. This was done in the time domain, by ensuring that the sum of all taps was equal to 1. This procedure accounts for the two aforementioned problems: limitations of the setup and low frequency ripples due to the length of the filters.

Regarding the limitations of the setup, correcting DC ensures that HRTFs asymptotically reach 0 dB at DC -which is a theoretically valid procedure and gives a meaningful value at 0 Hz. This is perhaps not so graphically evident in the context of the Neumann KU 100 dummy-head, where a flat frequency response is expected well entered the mid-frequency range. It will become a much more intuitive concept with measured HRTFs from human subjects, as discussed later.

Regarding the low frequency ripples, it can be seen in Figure 2.9 that they are minimized by controlling DC. Figure 2.10 shows a zoom in the low frequency range were the ripple control can be seen more clearly. Informal listening tests also showed an improvement in the perceived quality of the signals synthesized with such corrected HRTFs. It has to be pointed, however, that the quality problem was only seen when large differences were present between the DC values at both ears. If the interaural difference at DC is small, and both values are around 0 *dB*, the quality does not seem to be affected.

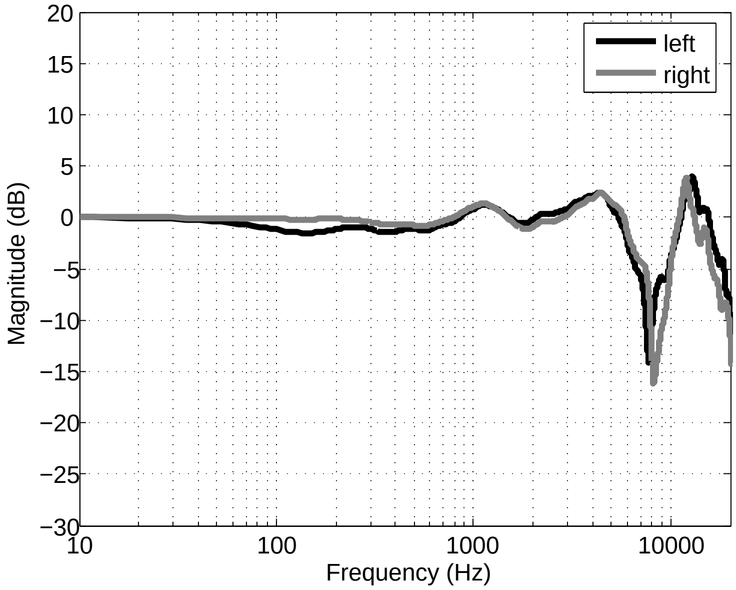

Figure 2.9: Same as in Fig. 2.8, but the DC value of the filters was corrected to equal unity gain. This was done in the time domain, by ensuring that the sum of all taps was equal to 1.

2.7 Low frequency compensation

According to the experiences gathered by working with binaural technology at Aalborg University, the combination of a proper calibration and DC control ensures well-behaved HRTFs in the low frequency range. That means that HRTFs decrease asymptotically until reaching 0 *dB* at DC, as it has been stated throughout this Chapter. Moreover, in the MSP both left and right signals are expected to be equal at low frequencies -disturbances due to positioning errors or asymmetries are possible but at higher frequencies. The Neumann KU 100 dummy-head, however, presented differences in the low frequency range -see Figure 2.9 and the signal differences in the range up to 500Hz- possibly due to the diffuse field equalization filters. As the impossibility of

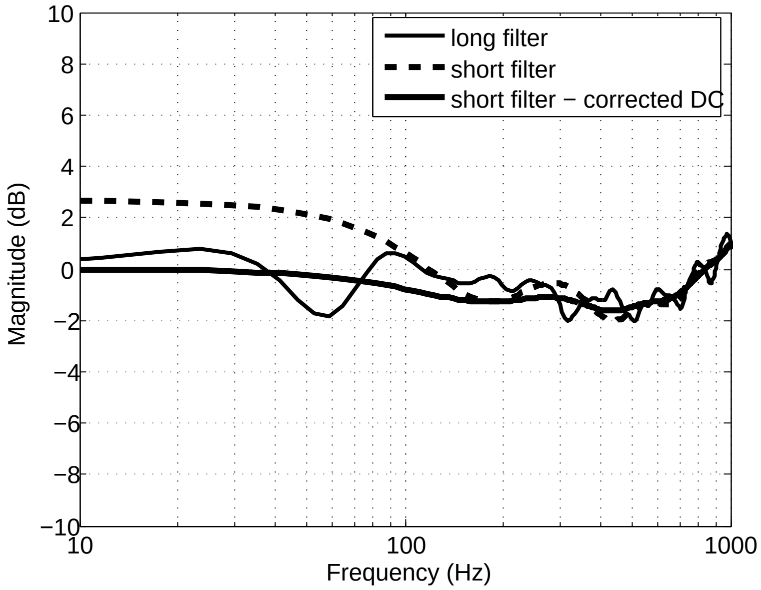

Figure 2.10: Comparison of the low frequency response of FIR filters with and without the DC value corrected, corresponding to the left side of the HRTFs for direction $(0^\circ, 0^\circ)$.

proceeding with a calibration by the method of substitution or comparison -see Section 2.5- made the frequency response of the microphones unknown, an alternative procedure was followed to investigate the response of the Neumann KU 100 dummy-head at low frequencies.

2.7.1 Measurement of the low frequency characteristics of the Neumann KU 100 internal microphones

The low frequency characteristics of the Neumann KU 100 internal microphones were investigated by attaching two reference microphones Brüel & Kjær 4193 modified with UC 0211 capsules. These microphones had a flat frequency response from 0.07 *Hz* to 20 *kHz* and were calibrated to their sensitivity at 1 *kHz*. The head with the attached microphones was put inside a sealed loudspeaker cabinet. A loudspeaker SEAS 33 F-WKA mounted on a 40 *cm* x 40 *cm* x 40 *cm* cabinet was used. The sound pressure inside a sealed cabinet is proportional to the displacement of the cone at very low frequencies -i.e. until the resonance of the system- and high sound pressure levels are reproduced. Outside the cabinet, the pressure is proportional to volume acceleration at those very low frequencies -it increases 12 *dB* per octave and very low sound pressure levels are reproduced. Therefore, all the measurements described in this section were done with the dummy-head inside the sealed cabinet.

The frequency characteristics of the Neumann KU 100 internal microphones were determined from 2 Hz to 20 Hz with a 1 Hz resolution and from 20 Hz to 315 Hz at the center frequencies of standard $\frac{1}{3}$ octave bands (ISO [1997]). A sine wave generator Brüel & Kjær 1027 was used to generate signals at each frequency of interest. For each of these sine waves, the voltage registered by the measuring amplifiers at the output of the microphones was recorded. This was done for both reference microphones and Neumann KU 100 internal microphones consecutively, without any change in the setup. These measurements were repeated in different days.

Since the reference microphones presented a flat frequency response down to 0.07 Hz, the difference in dB between the reference and internal microphones is equivalent to the frequency characteristics of the Neumann KU 100 internal microphones.

Results and analysis

The frequency characteristics of the Neumann KU 100 internal microphones are shown in Figure 2.11, normalized to their sensitivity at 1 kHz. It can be seen that the frequency characteristics of both left and right microphones are very different at very low frequencies. The left side frequency response complies with the specifications provided by the manufacturer (high-pass filter with cut-off frequency at 8 Hz), even though there is a small gain above 10 Hz. The right side frequency response, however, is far from the specifications.

The Neumann KU 100 dummy-head has options for high-pass filtering with cut-off frequencies at 40 Hz and 140 Hz. Measurements were also conducted with these settings and the results are shown in Figure 2.12. By inspection of Figures 2.11 and 2.12, a gain in the right internal microphone can be seen with respect to the left internal microphone. This explains the differences seen between left and right signals in previous figures -for example, see Fig.2.8.

Low frequency responses such as those seen in 2.12 have, as a counterpart, long impulse responses. However, those low frequencies should only be excited if a loudspeaker that reproduces sound at them is used as a source in the measurements. This is not the case of the VIFA loudspeakers used for M_{P1} and M_{P2} measurements presented in previous sections (a typical frequency response can be found in Møller *et al.* [1995a]) and yet low frequency differences in left and right signals can be seen. It is hypothesized that these differences are inherent to the diffuse field built-in circuits of the Neumann KU 100 dummy-head. In human HRTFs, on the other hand, measuring at low frequencies is not necessary since the responses should present asymptotic behavior towards DC. In

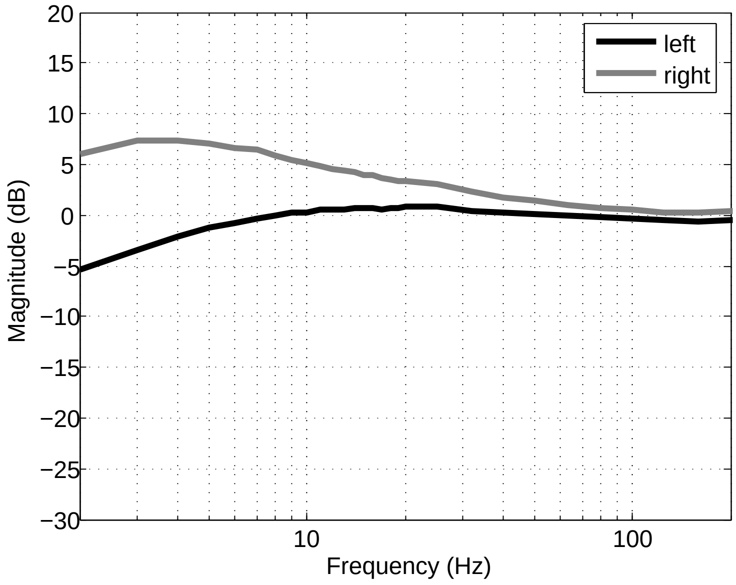

Figure 2.11: Low frequency response of the Neumann KU 100 internal microphones normalized to their 1 kHz sensitivity.

those cases, small loudspeakers that do not reproduce very low frequencies are preferred so that the impulse responses can be kept short and setup reflections can be windowed out. That was the approach reported by, for example, Han [1994]: it was mentioned that measurements were conducted with the woofer of a 2-way loudspeaker disconnected, so that low frequency ripples that had no significance were eliminated.

2.7.2 Low frequency range control

After investigating the low frequency characteristics of the Neumann KU 100 internal microphones, it was decided to control the low frequency range of all M_{P2} measurements by filtering. From the frequency responses shown in Fig. 2.11, inverse filters were constructed. The responses in frequency were completed in 1 Hz steps. In the range from 20 Hz to 1 kHz, the values were linearly interpolated between actual measured ones. Above 1 kHz, the responses were set to unity gain. From the obtained frequency response, linear phase FIR filters were computed by the windowing method. The minimum phase representation of these filters was obtained by the Hilbert Transform[2] (Hawksford [1997], Oppenheim & Schafer [1989]). The minimum phase filters

[2] Hilbert transform as a method for minimum phase decomposition is covered in Chapter 3.

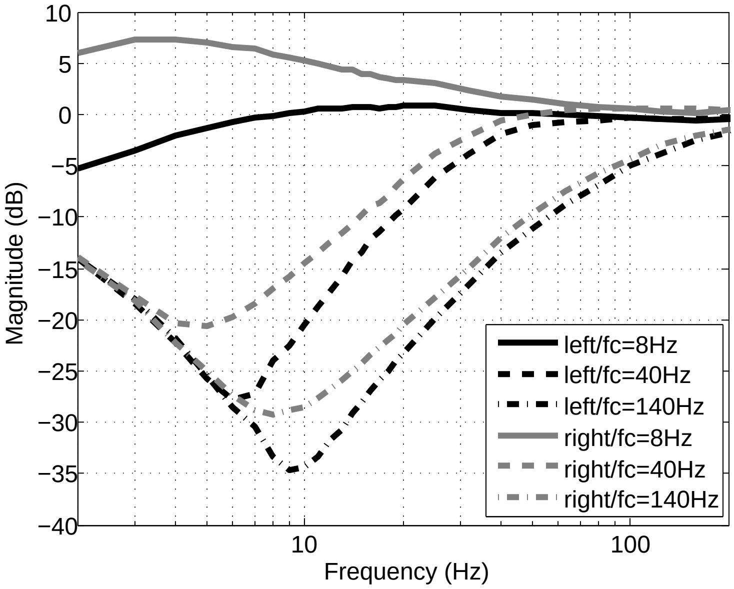

Figure 2.12: Low frequency response of the Neumann KU 100 internal microphones normalized to their 1 *kHz* sensitivity, with built-in high-pass filters applied. (Note change of scale with respect to Fig. 2.11)

were truncated to 256 taps and the inverse was computed. The results were implemented as FIR filters and applied to all M_{P2} measurements.

Results

After filtering all M_{P2} measurements with the aforementioned FIR inverse filters, the HRTFs were computed again. The results for direction $(0^\circ, 0^\circ)$ are shown in Figure 2.13. If compared with Figure 2.4 of Figure 2.9, it can be concluded that the low frequencies of the HRTFs are as expected for a direction in the MSP.

2.8 Discussion

The three issues examined in previous sections have been presented alone even though they are related among each other.

Calibration of the microphones is a requirement in any acoustical measurement. In the context of HRTFs, calibration ensures a correct gain. It is not straightforward to assess

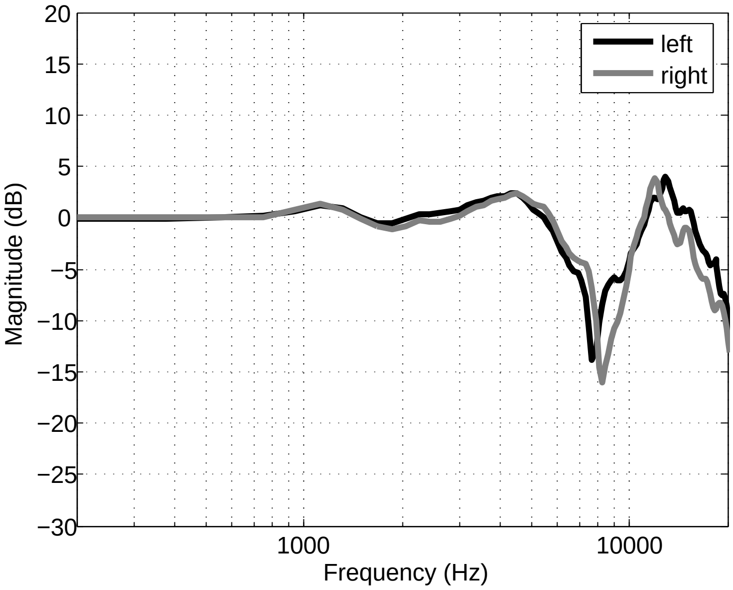

Figure 2.13: Same as Fig. 2.9, but low frequencies have been equalized to account for the characteristics shown in Fig. 2.11

calibration errors at high frequencies in HRTFs: in this range, they present much inter-subject variability that is furthermore direction-dependent. However, errors become self evident by inspection of the low frequency range: HRTFs are expected to reach 0 *dB* at 0 *Hz*, asymptotically. Deviations from this behavior -for example, in Figures 2.2 and 2.3- can be hypothesized to be closely related to poor or inexistent calibration. In the ideal case, a proper calibration accounts for the whole frequency response of the microphones if they are non-flat. This is possible if microphones are calibrated by methods such as comparison or substitution. An alternative is to investigate particular frequency ranges that deviate from a flat response, as in the reported low frequency investigation in this Chapter. The case of the Neumann KU 100 is unlike others, however, as measurement microphones usually present a flat frequency response in the whole range of audio frequencies.

Once HRTFs present the expected gain and an asymptotic decrease towards DC, it is only the value that DC takes which is meaningless. Therefore, proper calibration is also required for a valid DC correction. For example, controlling DC for the responses in Figure 2.4 gives correct results (after controlling the low frequency range, it gives Fig. 2.13), but controlling DC for the responses in Figure 2.2 would create a wrong jump of 25 *dB* between DC and the next frequency component.

As mentioned earlier in this work, HRTFs are often implemented as short FIR filters which convey all the necessary localization cues. However, short FIR filters define low frequencies with too few frequency components. The frequencies in-between those that are defined, are not controlled. Ripples appear in those frequencies in-between, as shown in Figure 2.10. One possibility of controlling the ripples would be to make longer measurements - then, more frequency points would be controlled. However, longer impulse responses require more demanding reflection-free setups and loudspeakers that can reproduce sound at those low frequencies. The procedure becomes troublesome, particularly when considering that the ripples are meaningless since HRTFs should decrease asymptotically. DC correction is a much more convenient way of minimizing those ripples which allows using loudspeakers with shorter impulse responses.

2.8.1 Measurement of HRTFs from human subjects

The concepts discussed previously were also applied to the measurements of HRTFs from the subjects that participated in the experiments reported in Chapters 5, 6 and 7. Since part of those experiments consisted of obtaining the evoked localization of virtual sound sources synthesized with individual HRTFs, the quality of the measured HRTFs was crucial. In this Section, the measurement procedure is described and the results are shown.

Subjects

HRTFs were obtained from 25 subjects: 13 female and 12 male. Most of them proceeded to different experiments according to their localization ability. Few subjects, however, only participated in pilot and/or side experiments which are not reported in this Thesis.

Measurement setup

The measurement setup has already been described in 2.4.1, but there were small variations which are explained in the following. The microphones used for M_{P2} measurements were a pair of miniature Sennheiser KE 4-211-2 which were calibrated to their sensitivity at 1 kHz before every set of measurements. The complete transfer function of each microphone had been obtained before the investigation by the method of comparison, where the reference was a $\frac{1}{4}$-inch Brüel & Kjær 4136 microphone. During the HRTFs measurement situation, the Sennheiser microphones were placed at the entrance of the blocked ear canal of the subjects and were connected to an in-house developed power supply which fed the signals to the Brüel & Kjær 2607 measuring amplifiers.

As only directions in the MSP were measured, the subjects needed not to be rotated and the turntable used for dummy-head measurements was not present. The elevations measured went from -67.5° (front hemisphere) to 247.5° (back hemisphere), in 22.5° steps[3]. The position of the subjects was checked before the measurements by a thin thread holding a miniature ball of 5 *mm* of diameter, whose position was calibrated with a laser passing through the ideal interaural axis position. Subjects were placed so that their interaural axis was coincident with the line crossing the miniature ball, which was not an obstacle from the point of view of the audio frequencies. Subjects stood on a small platform of variable height which was adjusted according to the height of the subjects.

The signal-to-noise ratio was measured to be around 60 *dB* in the mid- and high-frequencies. In the lower range of frequencies, the signal-to-noise ratio decreased to around 50 *dB*.

The measured HRTFs were processed according to 2.4.3, and the issues of DC correction and calibration were taken into consideration. The complete set of measurements is shown in Fig. 2.14. For the sake of synthesis, HRTFs from all subjects are graphically superimposed for each direction. The results can be compared to those of Fig. 13 in Møller *et al.* [1995a], which presents HRTFs measured for the same directions.

Fig. 2.14 is useful to understand the concept of HRTFs decreasing asymptotically towards DC and shows how proper calibration translates into a proper gain, which ensures that DC correction is properly applied. Therefore, all the previously covered concepts can be seen in context. This figure also shows that little happens in the lower frequency range of human HRTFs, which opens the possibility of implementing them as very short filters. The measurements on human subjects shown in Fig. 2.14 can be thought of as a generalization of the Neumann KU 100 case study.

2.8.2 Recommendations for correct HRTFs measurements

One of the issues that was considered while the dummy-head measurements were being conducted was whether it would be possible to establish a standard protocol, or at least a check-list, to ensure that HRTFs are correctly measured. This seems not to be easy due to the variety of purposes for which HRTFs can be measured. For instance, the issues covered in this Chapter are critical if HRTFs are used for binaural synthesis but might be

[3] As explained in 1.5, Chapter 1, a particular coordinate nomenclature was given to the measured HRTFs in the MSP for Chapters 5, 6 and 7.

irrelevant if they are used to analyze their spectral structure neglecting magnitude values, even though a strictly correct approach would take into consideration the technical validity of HRTFs as measured electroacoustical transfer functions. Begault [2000], for example, reviewed several laboratory procedures for HRTFs measurement and stated that the laboratory environment provided the more controlled situation for optimized repeatability and scientific accuracy. But he also acknowledged that, depending on the use of the HRTFs, other methods were also valid. For example, if a precise environmental context was to be included, or if HRTFs were to be used for certain artistic purposes where ITD and level relationships were more relevant than the actual HRTFs magnitude response. Therefore, it seems more plausible to make a list of recommendations for correct HRTFs measurements to be used in binaural synthesis. Some issues have already been pointed by Riederer [1998], Blauert *et al.* [1998][4], Wightman & Kistler [2005] and Hammershøi & Møller [2005]. As an extension of their contribution, and summarizing the discussions from this Chapter, the following points are considered:

- Measurements should be done in environments as anechoic as possible with the setup covered in absorbent material to avoid reflections and optimize the signal-to-noise ratio.
- Small loudspeakers with short impulse responses should be preferred.
- Calibration of the microphones should be performed by the methods of comparison or substitution to ensure that the whole frequency range is calibrated. If the frequency response of the microphones are known to be flat, a sound level calibrator or a pistonphone can be used.
- Calibration of the whole measurement chain should be performed.
- If the interaural difference in DC value is large, and the DC value is far from zero, DC correction procedures should be followed either in time domain or frequency domain.

2.9 Conclusion

The investigation presented in this Chapter addressed key technical aspects associated to measuring HRTFs which can compromise the validity of the HRTFs and their use in binaural synthesis. Emphasis was given to answering the question of whether it would be necessary or possible to work towards a suggested standard protocol to ensure good quality HRTFs measurements. In this context, a case study of dummy-head HRTFs

[4] The recommendations are included in the AUDIS CD but not in the conference publication.

measurements was presented to discuss the requirements of a proper calibration, DC correction and low frequency control. These are necessary conditions to ensure correct HRTFs: they should decrease asymptotically until they reach 0 dB at DC, preserving the audio quality. In the investigation presented here, it was seen that the three issues studied were connected: a proper calibration ensured a correct measurement at low frequencies and made DC control a valid procedure, and DC control ensured a meaningful value at 0 Hz apart from minimizing low frequency ripples. These principles were subsequently applied to measured HRTFs from human subjects, which were presented here as a generalization of the case study. Regarding the possibility of a consensual protocol for measuring HRTFs, it was concluded that it would very much depend on their use. A few recommendations were given from the perspective of binaural synthesis, which concentrated on the topics investigated.

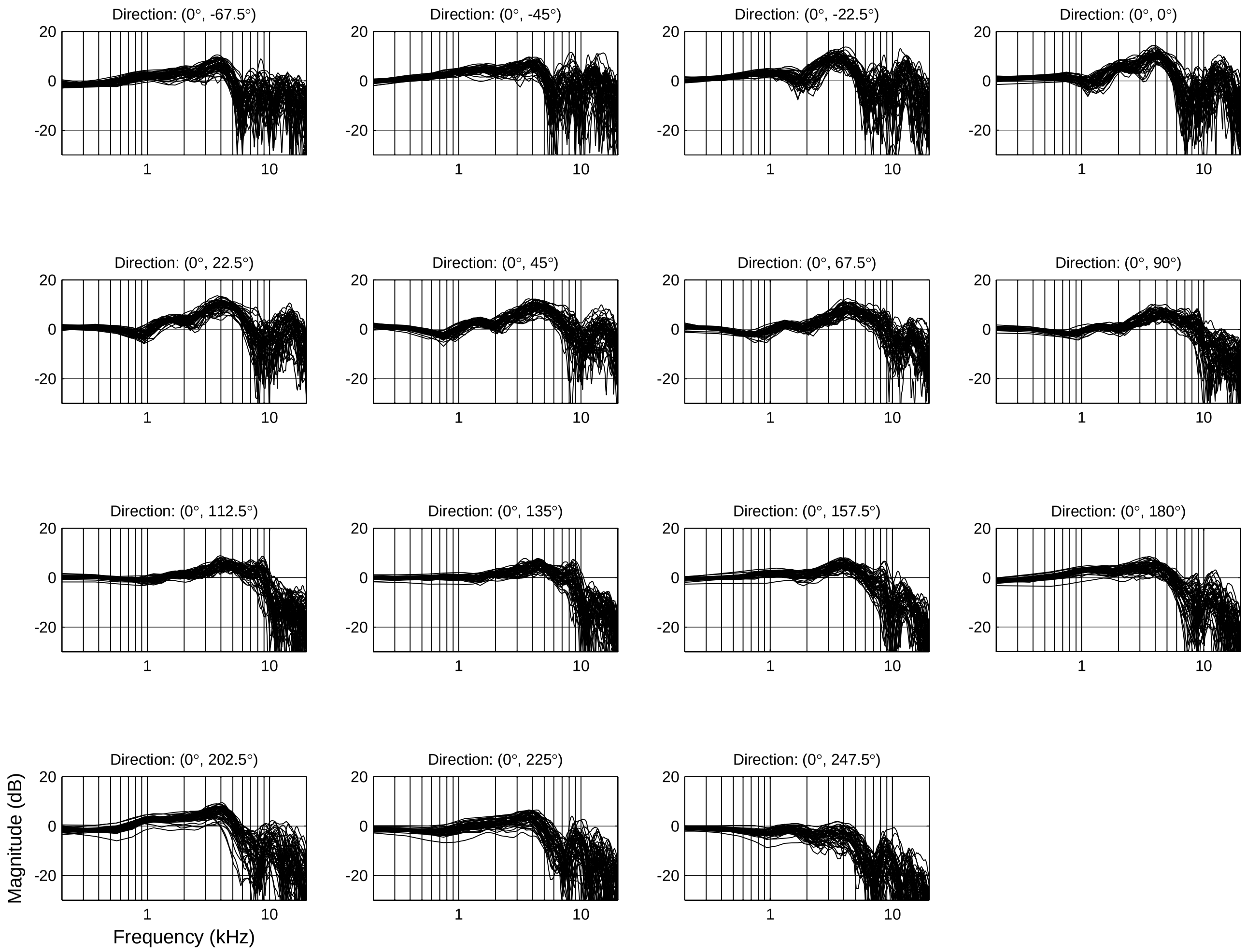

Figure 2.14: Measurements of human HRTFs for 15 directions in the MSP. Some of these measurements were used in Chapters 5, 6 and 7.

Chapter 3

Minimum phase decomposition of measured HRTFs

3.1 Introduction

Chapter 2 covered how to ensure that HRTFs are measured correctly. Obtaining the HRTFs filters to be used in binaural synthesis requires more post-processing steps, which will be discussed in this Chapter. More specifically, how to decompose measured HRTFs into minimum phase and excess phase components will be presented. The practicality of implementing HRTFs as minimum phase filters has already been mentioned in 2.6.2 and models that approximate HRTFs as minimum phase filters have been reported by many authors (Mehrgardt & Mellert [1977], Wightman & Kistler [1989a], Kistler & Wightman [1992], Sandvad & Hammershøi [1994], Kulkarni *et al.* [1995], Møller *et al.* [1995a], Jot *et al.* [1995], Minnaar *et al.* [1999], Minnaar *et al.* [2000], among others). In general terms, HRTFs are decomposed into minimum phase and excess phase components: the former contains all the spectral and frequency dependent information that provides cues for sound localization, while the latter encodes the time information that is relevant for sound localization. The properties of minimum phase filters have been well described by Oppenheim & Schafer [1989]. Briefly, these filters are causal and stable and therefore in the z-plane they have all their poles and zeros inside the unit circle. This characteristic ensures that the filter has, among other properties, the smallest phase possible of all filters with the same magnitude and that the energy is concentrated at the beginning of its impulse response. Even though the theoretical procedure to obtain minimum phase representations of a signal seems to be well understood, there are practical issues that are not always reported. During the first stages of this Ph.D. project, it was seen that algorithm implementations for minimum phase computation had restrictions. Therefore, two methods were compared: homomorphic filtering and Hilbert transform.

These are actually related to each other: Hilbert transform relationships hold in the cepstrum analysis, which is the core of homomorphic filtering, and the cepstrum is required in some implementations of the Hilbert transform (Oppenheim & Schafer [1989]). As a result of the present study, it was concluded that the methods were comparable in terms of success rate -i.e. percentage of signals converted to minimum phase in relation to the length of FFT used- but that the implementation of Hilbert transform chosen was more computationally demanding. A third method was also considered: reflection of zeros in the z-plane. This method was not part of the comparison since the implementation used was restricted to very short filters. The method was, however, used in Chapter 4 so as to keep consistency with previous experiments in the topic of all-pass sections audibility conducted at Aalborg University.

This Chapter is organized as follows. Firstly, relevant literature is reviewed and the phase characteristics of measured HRTFs is analyzed in terms of zeros outside the unit circle. Secondly, a review of the two procedures tested is presented, along with a description of the methods used to compare them. The z-plane method is also introduced. Thirdly, the results of the comparison are shown and discussed. Finally, conclusions are drawn.

3.2 Previous works

The perceptual validity of minimum phase HRTFs as conveyors of the spectral features required for sound localization has been shown by several investigations, and the topic will be discussed in Chapter 4. In the following, this Chapter will focus on the actual procedures to decompose a signal into minimum phase and excess phase parts. The literature shows that it is well established to use HRTFs minimum phase filters in binaural synthesis as reported by, for example, Wenzel & Foster [1993], Sandvad & Hammershøi [1994], Jot *et al.* [1995], Langendijk & Bronkhorst [2002], Minnaar *et al.* [2005] and Wightman & Kistler [2005]. Møller *et al.* [1995a] showed examples of minimum phase HRTFs and their associated all-pass sections. Huopaniemi *et al.* [1999] mentioned that HRTFs were *almost* minimum phase and that many dynamic virtual acoustic environments chose minimum phase FIR approximations for HRTF implementation due to straightforward interpolation, relatively good spectral performance and simplicity of implementation. Begault [2000] discussed the implementation of HRTFs as FIR filters and different methods to make the filter shorter. He showed minimum phase representations of HRTFs.

Despite the agreement regarding the practical use and perceptual validity of minimum phase HRTFs, issues around the actual procedure to decompose the signals into minimum phase and excess phase parts have received very little attention. For example, there are very few investigations that report the method used for decomposing the measured HRTFs. Mehrgardt & Mellert [1977] and Jot *et al.* [1995] mentioned Hilbert transform relationships as the basis for their procedure. Kistler & Wightman [1992] reported computing minimum phase HRTFs from their log magnitude. Sandvad & Hammershøi [1994] reported the use of the complex cepstrum as described by Oppenheim & Schafer [1989]. Kulkarni *et al.* [1995] and Kulkarni *et al.* [1999] also mentioned Hilbert transform relationships and gave the citation of Oppenheim & Schafer [1989] for the procedure employed. Brown & Duda [1998] measured HRTFs with a Snapshot^{TM} system built by Crystal River Engineering which provided both raw and minimum phase impulse responses, Huopaniemi *et al.* [1999] reported obtaining minimum phase HRTFs by means of the MATLAB built-in function *rceps.m*, which implements the real cepstrum. None of these investigations gave further details about the procedures, which is interesting since the literature also reveals that designing minimum phase FIR filters is not always straightforward. A review of different algorithms for minimum phase decomposition can be found in Damera-Venkata *et al.* [2000], along with a proposed novel algorithm. Typical constrains in the methods are numerical errors, inaccuracies due to truncation and excessive computational time. Karam [2006], on the other hand, discussed the difficulties in computing the unwrapped phase of signals and proposed a hybrid method which was compared to existing ones. Phase unwrapping is an important issue in minimum phase decomposition, as also mentioned by Wightman & Kistler [1989a], since Hilbert transform and cepstrum analysis assume a continuous phase.

As a preliminary investigation with measured HRTFs showed that methods for minimum phase decomposition had their limitations, it was decided to conduct a systematic study. The comparison presented here is not intended to be exhaustive and not all possible minimum phase algorithms and/or optimization methods are included. Two implementations already in use in audio signal processing were compared. In this Chapter, the results are shown and a third method is presented.

3.3 On the position of zeros in measured HRTFs

Measured HRTFs are mixed phase systems resulting from the pressure division $\frac{P_{ear}}{P_{ref}}$ or, as defined in Eq. 2.4 in the previous Chapter, $\frac{H_{P2}}{H_{P1}}$. They can be implemented as filters, which can be represented as rational polynomials in the z-domain in the form:

$$H(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + \cdots + b_{m-1} z^{1-m}}{1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_n z^{-n}} \tag{3.1}$$

The roots (points at which a polynomial evaluates to zero) of the numerator determine the location of the zeros in the z-plane, while the roots of the denominator determine the location of the poles. FIR filters have zeros determined by the *b* coefficients, but poles in the origin of coordinates of the z-plane since all *a* coefficients are zero. Therefore, HRIRs can be directly implemented as the *b* coefficients of FIR filters. This implementation can be optimized if the filters are minimum phase since those filters are the shortest possible. It was reported by Damera-Venkata *et al.* [2000] that given optimal minimum phase and linear phase digital FIR filters which have the same magnitude response, the minimum phase filter would have a reduced length which amounts from $\frac{1}{2}$ to $\frac{3}{4}$ of the linear phase filter length. Moreover, minimum phase filters require fewer computations and less memory in their implementation.

Minimum phase systems meet conditions of causalty and stability, as discussed by Oppenheim & Schafer [1989], which determine that all the zeros of the filter are inside the unit circle in the z-plane. The zeros outside the unit circle correspond to the excess phase part of the filter. These concepts will be discussed more in depth later in this Chapter (see 3.4.3) but are introduced here because an analysis of poles and zeros in measured HRTFs helps clarifying their phase characteristics. Mehrgardt & Mellert [1977], Huopaniemi *et al.* [1999] and Brown & Duda [1998] have stated that HRTFs were close to minimum phase but there were still some excess phase components. Kulkarni *et al.* [1995] showed that the excess phase part was not only composed by a linear phase which suggested the presence of all-pass sections, as also reported by Møller *et al.* [1995a].

In order to characterize measured HRIRs as mixed phase systems, the position of zeros in the z-plane was computed for an existing database of HRTFs (Møller *et al.* [1995a]). The procedure involved finding the roots of the polynomial[1]. The database used contained measurements from 40 subjects at 97 directions in the sphere, which are indicated in Table 3.1. Before the computation, the following post-processing was conducted on the measured signals:

- The raw data was low pass filtered and checked for causalty.
- Signals were DC corrected.
- The initial linear delay was computed with the 5% leading edge criterion (Sandvad & Hammershøi [1994]) and removed from the signals.
- Signals were truncated at 128 samples.
- Signals were DC corrected again.

[1] This was done with the MATLAB function *'roots.m'*, which will be discussed later in this Chapter.

Figure 3.1 shows how many zeros outside the unit circle each HRTFs had, which is a rough and simplistic measure of how far from minimum phase the filters were. The figures are color coded so that deep blue corresponds to no zeros outside the unit circle and deep red corresponds to 41 zeros outside the unit circle -the maximum found for a single signal. The ordinate of each figure shows the *Subject index* -all 40 subjects from the database were included. The abscissa of each figure indicates the *Direction index*, which is explained in Table 3.1. The coordinate system used has already been mentioned in Chapter 1, 1.5.

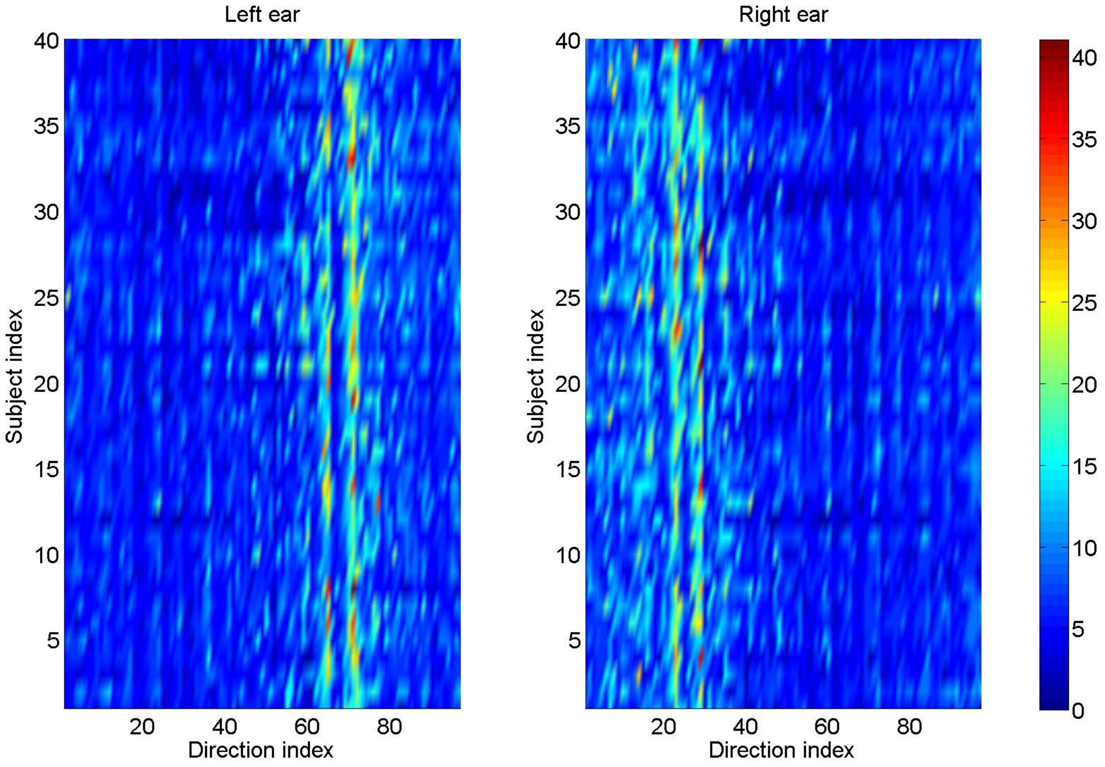

Figure 3.1: Number of zeros outside the unit circle for each signal in a database of 3880 pairs of HRTFs which were available from previous measurements (Møller *et al.* [1995a]). Left and right panels correspond to left and right ear signals of each HRTFs pair, respectively. The ordinate indicates the Subject index (40 subjects in total) and the abscissa indicates the Direction index (97 directions in the sphere around subjects were available, the indexes are explained in Table 3.1).

Table 3.1: Direction indexes used in Fig. 3.1. Each row corresponds to a different elevation angle ϕ, and each column corresponds to a different azimuth angle θ. Both elevation and azimuth were sampled in 22.5° steps, yielding a total of 97 directions in a sphere around the listener. The coordinate system is according to Chapter 1, 1.5.

ϕ \ θ	22.5°	45°	67.5°	90°	112.5°	135°	157.5°	±180°	−157.5°	−135°	−112.5°	−90°	−67.5°	−45°	−22.5°	0°
90°																90
67.5°		6		18		30		42		54		66		78		91
45°	1	7	13	19	25	31	37	43	49	55	61	67	73	79	85	92
22.5°	2	8	14	20	26	32	38	44	50	56	62	68	74	80	86	93
0°	3	9	15	21	27	33	39	45	51	57	63	69	75	81	87	94
−22.5°	4	10	16	22	28	34	40	46	52	58	64	70	76	82	88	95
−45°	5	11	17	23	29	35	41	47	53	59	65	71	77	83	89	96
−67.5°		12		24		36		48		69		72		84		97

Figure 3.1 clearly shows that for directions to the sides, the contralateral signals present more zeros outside the unit circle -i.e. the left ear for directions in the -90° azimuth vertical plane (*Direction indexes* from 66 to 72), and the right ear for directions in the 90° azimuth vertical plane (*Direction indexes* from 18 to 24). This is expected if the pressure division $\frac{P_{ear}}{P_{ref}}$ is analyzed for these directions: sound arrives first to the reference point in the center of the head (P_{ref} in the denominator) than to the contralateral ear (P_{ear} in the numerator), therefore determining non-causal sequences.

Figure 3.1 is in line with some of the findings reported by Kulkarni *et al.* [1995]: they found that the contralateral signals for sources to the sides ($\pm 90^\circ$ in azimuth) presented the excess phase parts that deviated more from linear phase. They suggested that this was due to scattering and diffraction components that contributed to the sound in the shadowed, contralateral ear. It has to be noted that Kulkarni *et al.* [1995] referred to the actual excess phase as all-pass[2]. However, the important observation is that in the reported case, the excess phase components deviated from a linear phase, suggesting that all-pass sections were present. Brown & Duda [1998] also reported that the contralateral side of HRTFs for directions in the ranges between $\pm 80^\circ$ and $\pm 100^\circ$ in azimuth were not minimum phase.

One widely acknowledged problem with minimum phase decomposition is phase unwrapping, as analyzed by Karam [2006]. Methods like Hilbert transform or homomorphic filtering require that the phase is unambiguous (Oppenheim & Schafer [1975]) -i.e. the unwrapped phase, or the instance where the phase jumps are defined so that the phase is continuous. On one side, in a mixed phase system the linear phase component leads to discontinuities in the unwrapped phase at $\omega = 2\pi$. Therefore, it is recommended to remove the linear phase (for example, by removing the initial time delay of the signal) prior to the computation. HRTFs do contain, by nature, a linear phase component. On the other side, Karam [2006] reported that methods for phase unwrapping fail due to zeros which are very closed to the unit circle, which cause very sharp phase changes. This difficulty is carried also to the minimum phase computation methods analyzed in this Chapter. It can be said, then, that the success of minimum phase decomposition methods is not related to how many zeros outside the unit circle the filter has, but to how close the zeros are to the unit circle. Figure 3.2 shows the histogram (relative to the pooled number of zeros from all HRTFs in the database) of the distance of zeros to the origin of coordinates in the z-plane, as computed from the available HRTFs database. It can be seen that in the context of HRTFs filters, the majority of zeros are very close to the unit circle.

[2] More about their work will be covered in Chapter 4 in the context of all-pass audibility.

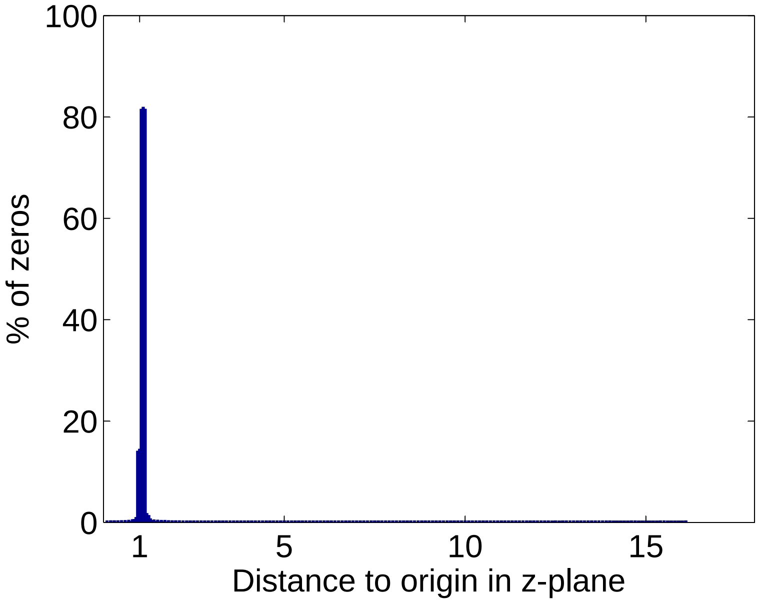

Figure 3.2: Histogram relative to the pooled number of zeros from all HRTFs in the database (127 zeros from each of 7760 filters). Bin size=0.1. The histogram shows that most of the zeros are very close to the unit circle, with only a few far outside the unit circle.

3.4 Methods for Minimum-Phase Decomposition

3.4.1 Hilbert transform

Theory

In mathematical terms, any real sequence $x(n)$ can be represented by a sum of even an odd parts:

$$x(n) = x_e(n) + x_o(n) \tag{3.2}$$

The even part is conjugate symmetric and the odd part is conjugate antisymmetric:

$$\begin{aligned} x_e(n) &= \frac{1}{2}[x(n) + x*(-n)] \\ x_o(n) &= \frac{1}{2}[x(n) - x*(-n)] \end{aligned} \tag{3.3}$$

This means that even and odd part are obtained from the original sequence $x(n)$ and its conjugate $x*(-n)$, and that the original sequence can be completely recovered by the even part or by the odd part -but the latter only for $n \neq 0$. Due to the overlapping characteristics of $x(n)$ and $x*(-n)$, the original sequence can by completely recovered from its even part by:

$$x(n) = x_e(n)u_+(n) \tag{3.4}$$

Where

$$u_+(n) = \begin{cases} 0, & n < 0 \\ 1, & n = 0 \\ 2, & n > 0 \end{cases} \tag{3.5}$$

This implies that $x(n)$ is causal, since $x(n) = 0$ for $n < 0$. By Fourier transform properties,

$$X(e^{j\omega}) = X_R(e^{j\omega}) + jX_I(e^{j\omega}) \tag{3.6}$$

Since $X_R(e^{j\omega})$ is the Fourier transform of $x_e(n)$ and $X_I(e^{j\omega})$ is the Fourier transform of $x_o(n)$, the original sequence $x(n)$ can also be completely recovered from the real part of its Fourier transform.

Hilbert transform relationships are integral relationships that express $X_R(e^{j\omega})$ in terms of $X_I(e^{j\omega})$ and $X_I(e^{j\omega})$ in terms of $X_R(e^{j\omega})$, so that the recovery of $x(n)$ is possible either from the real or imaginary part of its Fourier transform. In other words, for any real, causal and stable sequence, real and imaginary parts are uniquely related when they are Hilbert transforms of each other. It is a requirement, however, that both real and imaginary parts are continuous functions, as already mentioned. A sequence that complies with all these characteristics is a minimum phase sequence.

Minimum phase signals can also be recovered from their magnitude or phase. In such cases, the complex logarithm of the signal is computed:

$$\hat{H}(e^{j\omega}) = log[H(e^{j\omega})] = log|H(e^{j\omega})| + j\arg[H(e^{j\omega})] \tag{3.7}$$

Where $\hat{H}(e^{j\omega})$ is the Fourier transform of the minimum phase sequence $\hat{x}_{min}(n)$, and $log|H(e^{j\omega})|$ and $\arg[H(e^{j\omega})]$ are Hilbert transforms of each other.

If these concepts are applied to a discrete-time signal, it is found that the requirement of causalty imposes the following over a periodic signal of length N:

$$\begin{aligned} \hat{h}(n) &= 0, \qquad \frac{N}{2} < n < N \\ \hat{h}(n) &= 0, \qquad -\frac{N}{2} < n < 0 \end{aligned} \tag{3.8}$$

This means that, for a sequence of length N, the second half will be zero. The original sequence can also be completely recovered from its even part. Furthermore, as Hilbert transform relationships also hold for the discrete Fourier transform $\hat{H}(k)$, the imaginary part $\hat{H}_I(k)$ can be constructed from the real part $\hat{H}_R(k)$ and viceversa.

In the case of discrete-time signals, the log magnitude and phase cannot be related such as it was explained around Eq. 3.7, but still a minimum phase filter can be approximated with a similar process: given the log magnitude of the discrete Fourier transform, a phase can be constructed so that the magnitude response is preserved while the imaginary part is an approximation to the minimum phase. This is done by computing the inverse discrete Fourier transform of $log|H(k)|$, which is time aliased[3]:

$$\hat{h}(n) = \sum_{r=-\infty}^{\infty} \hat{h}(n+rN) \tag{3.9}$$

Then, the product $\hat{h}(n)u_N(n)$ has to be computed, where $u_N(n)$ is defined in a similar way as Eq. 3.5

$$u_N(n) = \begin{cases} 0, & n = \frac{N}{2}+1, ..., N-1 \\ 1, & n = 0, \frac{N}{2} \\ 2, & n = 1, 2, ..., \frac{N}{2}-1 \end{cases} \tag{3.10}$$

It has to be noted that the longer N is chosen, the better the results obtained for $\hat{h}(n)u_N(n)$. Finally, the discrete Fourier transform of $\hat{h}(n)u_N(n)$ is computed. The result of these operations is a signal $H_{min}(k)$ with a real part equal to $log|H(k)|$ and an imaginary part that approximates to the minimum phase.

The concepts behind Hilbert transforms can also be applied to complex sequences so that real and imaginary parts are related in ways similar to those explained before. Given a complex sequence $s(n) = s_r(n) + js_i(n)$, causalty is ensured if the negative frequencies are forced to zero. $S_r(e^{j\omega})$ and $S_i(e^{j\omega})$ are defined as the Fourier transforms of $s_r(n)$ and $s_i(n)$, respectively, and they are analogous to even and odd parts as explained

[3] The inverse Fourier transform of $log|H(k)|$ is equivalent to the complex cepstrum of $h(n)$, as it will be explained later in the Chapter.

around Eq. 3.2. Therefore, similar relationships to those in Eq. 3.3 can be established:

$$
\begin{aligned}
S_r(e^{j\omega}) &= \frac{1}{2}[S(e^{j\omega}) + S*(e^{-j\omega})] \\
jS_i(e^{j\omega}) &= \frac{1}{2}[S(e^{j\omega}) - S*(e^{-j\omega})]
\end{aligned}
\tag{3.11}
$$

Real and imaginary parts can be created keeping Hilbert transform relationships between each other, for instance, by computing:

$$
S(e^{j\omega}) = \begin{cases} 2S_r(e^{j\omega}), & 0 \leq \omega < \pi \\ 0, & -\pi \leq \omega < 0 \end{cases} \tag{3.12}
$$

And

$$
S(e^{j\omega}) = \begin{cases} 2jS_i(e^{j\omega}), & 0 \leq \omega < \pi \\ 0, & -\pi \leq \omega < 0 \end{cases} \tag{3.13}
$$

In this construction of the signal, the phase is obtained by a 90° phase shift of $S_r(e^{j\omega})$.

Implementation

The algorithm used in the present investigation was first proposed by Hawksford [1997]. It uses the MATLAB built in function *'hilbert.m'*, which applies the theory of Hilbert transforms for complex signals already explained.

The full procedure for obtaining the minimum phase component of HRIRs of length N can be described as follows:

- Compute the FFT of the signal.
- Compute the logarithm of the magnitude of the signal obtained in the previous step.
- Pass the result of the previous step through the function *'hilbert.m'*, which does the following:
  - Computes the FFT.
  - Obtains the even part of the signal.
  - Constructs a complex signal: the real part is the signal obtained in the previous step, the imaginary part is the same signal but shifted 90° in phase.
  - Computes the inverse FFT of the complex signal.

- Compute the conjugate to correct the time reversal of the signal.
- Compute the inverse logarithm.
- Compute the inverse FFT to go back to the time domain.
- Truncate the result to obtain a filter with the original length N.

3.4.2 Homomorphic Filtering

Theory

Homomorphic systems are a class of systems that obey a generalized principle of superposition and are represented by algebraically linear transformations between input and output. Considering the principle of superposition, where T is the system transformation, c is a scalar and $x_1(n)$ and $x_2(n)$ are two inputs:

$$\begin{aligned} T[x_1(n)+x_2(n)] &= T[x_1(n)]+T[x_2(n)] \\ T[cx_1(n)] &= cT[x_1(n)] \end{aligned} \tag{3.14}$$

This principle can be generalized so that $\Box$ is a rule of input combination and $\bigcirc$ is a rule of output combination:

$$H[x_1(n)\Box x_2(n)] = H[x_1(n)\bigcirc x_2(n)] \tag{3.15}$$

A similar generalization holds for combining scalars with inputs and outputs. The rules for combination can be addition, multiplication, convolution, etc., and there exists one characteristic system D for each of them, so that:

$$\begin{aligned} D_{\Box}[x_1(n)\Box x_2(n)] &= D_{\Box}[x_1(n)]+D_{\Box}[x_2(n)] \\ &= \hat{x}_1(n)+\hat{x}_2(n) \end{aligned} \tag{3.16}$$

And the corresponding generalization for the scalar exists. In this class of systems, perfect separation of $x_1(n)$ and $x_2(n)$ is possible as long as $\hat{x}_1(n)$ and $\hat{x}_2(n)$ can be perfectly separated by linear filtering. A mixed phase signal $x(n)$ can be considered as a convolution of minimum phase and maximum phase sequences:

$$x(n) = x_{min}(n) * x_{max}(n) \tag{3.17}$$

In such a case, homomorphic systems for convolution can be used to perfectly separate the minimum phase from the maximum phase components of the signal. These systems obey a generalized principle of superposition for convolution, which expressed in terms of Eq. 3.16 is:

$$\begin{aligned} D_*[x_1(n) * x_2(n)] &= D_*[x_1(n)] + D_*[x_2(n)] \\ &= \hat{x}_1(n) + \hat{x}_2(n) \end{aligned} \tag{3.18}$$

In order to make use of homomorphic filtering, the cepstrum $\hat{x}(n)$ and its properties have to be introduced. Considering the finite-length sequence $x(n)$, by Fourier transform:

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j(\frac{2\pi}{N})kn} \tag{3.19}$$

And defining

$$\hat{X}(k) = log[X(k)] \tag{3.20}$$

The complex cepstrum is given by the inverse Fourier transform of Eq. 3.20:

$$\hat{x}(n) = \frac{1}{N} \sum_{k=0}^{N-1} \hat{X}(k) e^{j(\frac{2\pi}{N})kn} \tag{3.21}$$

The properties of $\hat{x}(n)$ are well explained in Oppenheim & Schafer [1975] and Oppenheim & Schafer [1989]. For a minimum phase sequence $x(n)$,

$$x(n) = \hat{x}(n) = 0, \qquad n < 0 \tag{3.22}$$

Which means that both the sequence and its cepstrum are causal. As explained in the context of the Hilbert transform method, causal signals can be completely defined by their even part or the real part of their Fourier transform. This means that, in order to obtain the cepstrum, Eq. 3.20 can be replaced by:

$$\hat{X}_R(k) = log|X_R(k)| \tag{3.23}$$

And the even part of $\hat{x}(n)$ is the cepstrum:

$$c(n) = \frac{1}{N} \sum_{k=0}^{N-1} \hat{X}_R(k) e^{j(\frac{2\pi}{N})kn} \tag{3.24}$$

Which, following the discussion for Hilbert transforms, can be expressed as:

$$c(n) = \frac{1}{2}[\hat{x}(n) + \hat{x}(-n)] \tag{3.25}$$

Or, the familiar product:

$$\hat{x}(n) = c(n)u_+(n) \tag{3.26}$$

With

$$u_+(n) = \begin{cases} 0, & n < 0 \\ 1, & n = 0 \\ 2, & n > 0 \end{cases} \tag{3.27}$$

In the complex cepstrum, each minimum phase zero in the spectrum rises a causal exponential, and non-minimum phase zeros rises non-causal exponentials. The procedure to separate them make use of the relationships already mentioned for the Hilbert transform method. Is it easy to see that the implementation of homomorphic filtering to obtain a minimum phase signal follows very much the explanation given in the context of discrete-time signals, and that Eq. 3.9, which corresponds to the inverse discrete Fourier transform of $log|H(k)|$, is equivalent to the cepstrum of Eq. 3.24 which is, as well, time-aliased for time-discrete signals:

$$c(n) = \sum_{k=-\infty}^{\infty} c(n+kN) \tag{3.28}$$

It can further be seen that the application of $u_+(n)$ as defined in Eq. 3.27 follows that of $u_N(n)$ as defined in Eq. 3.10 for discrete-time signals. In this case, as well, the longer N is chosen, the better the results obtained for Eq. 3.26. Furthermore, care has to be taken when computing the correct phase unwrapping of the original signal.

Implementation

In the context of HRIRs, the cepstrum is used instead of the complex cepstrum (it has to be noticed that, for real sequences, the cepstrum and complex cepstrum give the same result). This is implemented with the MATLAB function *'rceps.m'*, also used by Huopaniemi *et al.* [1999], which includes the following steps:

- Compute the FFT of the input signal.
- Compute the log magnitude of the previous step.
- Compute the cepstrum by applying the inverse FFT of the previous setp and discarding the imaginary part.
- Apply a *lifter* (the equivalent to a frequency domain filter, but in the cepstral domain), as defined by Eq. 3.27, which in discrete-time follows Eq. 3.10.

- Compute the FFT and inverse logarithm of the previous result.
- Compute the inverse FFT to yield a result which is a real sequence.
- Truncate the result to obtain a filter with the original length N.

3.4.3 Z-plane Method

The z-plane method consists, basically, of finding the location of the zeros that lie outside the unit circe so that they can be mirrored inside the unit circle - poles have to be added, too, to cancel the added zeros. The implementation of the method presented here makes use of MATLAB numerical tools for finding roots in a polynomial. These functions are, more specifically, *'roots.m'* and its inverse *'poly.m'*, which are the base for a family of MATLAB functions that deal with zeros and poles of a filter (*'tf2zpk.m'*, *'zplane.m'*, *'isminphase.m'*, just to name a few). Therefore, in this particular case the problem of accuracy is partly given by round off errors in the computation of the two basic functions *'roots.m'* and *'poly.m'*, and which are inherent to the floating-point system in MATLAB.

This method was not compared to Hilbert transform nor homomorphic filtering since it was seen that it failed to obtain minimum phase sequences for filters longer than $N = 72$, approximately. However, the method is included here because it will be used in Chapter 4.

Theory

As explained before, minimum phase filters are causal and stable and therefore they must have all their poles and zeros inside the unit circle in the z-plane. If the filter is mixed phase, it can be decomposed into minimum phase and excess phase components, which expressed in the z-domain is:

$$H(z) = H(z)_{min.} \cdot H(z)_{exc.} \tag{3.29}$$

The excess phase component can be further decomposed into linear phase and all-pass components:

$$H(z)_{exc.} = H(z)_{lin.} \cdot H(z)_{all-pass} \tag{3.30}$$

A linear phase component is a delay or shift of a signal in time. In HRTFs filters, it is represented by the initial time before the arrival of the sound to each ear. A linear phase system has unity magnitude and its phase is linear with a negative slope. Hence, it can be accounted for without spectral consequences by removing the initial zeros in each signal of a HRTFs pair - the difference in arrival time to each ear is important, though, as it is necessary for constructing the ITD, but that is out of the scope of this Chapter. Without the linear phase component, the transfer function in Eq. 3.29 is replaced by:

$$H(z) = H(z)_{min.} \cdot H(z)_{all-pass} \tag{3.31}$$

The all-pass component, on the other hand, has also unity magnitude. In the z-plane, a first-order all-pass section consists of a single pole inside the unit circle and a zero at a conjugate reciprocal location. Second order all-pass sections are complex conjugated pairs of poles inside the unit circle and zeros at mirrored positions. General all-pass sections can be expressed as the product of first and second-order all-pass sections. A thorough theoretical background on all-pass sections can be found in Oppenheim & Schafer [1989] and Møller *et al.* [2007] and will not be repeated here. Furthermore, the audible consequences of removing all-pass sections from HRTFs will be discussed in the next Chapter.

Once the linear phase component has been removed from the filter, the knowledge on all-pass sections can be used to isolate them from the minimum phase filter. This is, each zero outside the unit circle can be mirrored inside the unit circle, and together with the original zeros (those originally inside the unit circle) they form the minimum phase filter. On the other hand, the original zeros outside the unit circle and the poles that cancel the added zeros (which are at conjugate reciprocal locations) conform the all-pass sections.

Summarizing, working in the z-domain allows to easily find zeros outside the unit circle (i.e. their distance from the origin in the z-plane is larger than 1) and accounting for them. This method has been outlined, for example, by Kulkarni *et al.* [1999] and Møller *et al.* [2007].

Implementation

The first step in the implementation consists of identifying and removing the initial linear delay from the HRTFs, which is the linear phase component in Eq. 3.30. From the methods available, the 5% leading edge criterion reported by Sandvad & Hammershøi [1994] was used. Removing the linear delay from the HRTFs yields:

$$HRTF(z) = HRTF(z)_{min.} \cdot HRTF(z)_{all-pass} \tag{3.32}$$

The next step consists of obtaining the location of poles and zeros. This is equivalent to finding the roots of the polynomial in Eq. 3.1. There are different ways to find the location of poles and zeros in MATLAB, but all of them rely on the function *'roots.m'* which uses the eigenvalue method to find the roots of a polynomial. Given a vector of length N, the algorithm of this function creates a N-by-N companion matrix A and computes its eigenvalues, which are the N roots of the characteristic polynomial. The results are not the exact roots of the polynomial, but the eigenvalues of the matrix A. More about the eigenvalue method for factoring polynomials and a comparison with other methods can be found in Sitton *et al.* [2003]. The function *'roots.m'* and its inverse *'poly.m'* are related by a scale or gain factor k and are affected by round-off errors. The latter can be problematic, since the mentioned functions are sensitive to the locations (and small changes in the locations) of poles and zeros, particularly when they are very close to the unit circle.

The issue of errors in poles and zeros placement was seen to increase around the Nyquist frequency. For a short vector of length $N < 72$, approximately, the error could be easily observed in the taps around $\frac{N}{2}$. This could be determined by either numerical noise caused by round-off errors in computation, or by truncation of the HRIRs themselves. HRTFs as 72 taps filters (at a sampling frequency of 48 kHz) have been shown to work well for localization purposes with binaural synthesis, as reported by Sandvad & Hammershøi [1994]. However, if the measured impulse responses have not decreased completely within 72 taps then truncation would determine, in some extent, errors in the location of poles and zeros. It is not clear whether the cause of the errors seen here was purely based on round-off or truncation issues, or a combination of both, but a preliminary inspection showed that windowing (forcing the last coefficients to decrease asymptotically to zero) decreased the error. With a proper window and for short polynomials, it was observed that there was negligible error in going back and forth from *'roots.m'* to *'poly.m'*. In the z-plane, the change affected the zeros around Nyquist, and those high frequencies were the ones affected in the frequency response of the filter. For filters longer than 72 coefficients, approximately, the error of re-computing the polynomial from its roots was seen to increase considerably. The results were meaningless, for which the method could not be implemented with MATLAB built-in functions. It was concluded, then, that the most important restriction with the use of the functions *'roots.m'* and *'poly.m'* was that they could not handle very long polynomials due to errors associated to the aforementioned floating-point system. This does not mean that factoring long polynomials with the function *'roots.m'* necessarily gives meaningless results. A review of the function and examples of successful factoring of high-degree polynomials has been reported by Sitton *et al.* [2003].

Considering the points mentioned before, the precise implementation of the method can be summarized in the following steps:

- Truncate the measured HRIRs to $N < 72$ coefficients, to ensure minimization of errors when reconstructing the polynomial from the modified poles and zeros.

- Create a vector of a coefficients (as in Eq. 3.1) with all of them equal to zero. The HRIRs are the b coefficients in Eq. 3.1.
- Find the location of poles and zeros, together with a gain factor, using the MATLAB function *'tf2zpk.m'*.
- Assess the location of each zero by computing the distance from the origin of coordinates, identify the zeros outside the unit circle.
- For those zeros placed outside the unit circle, add a pole and a zero at the conjugate reciprocal position.
- Separate the original zero outside the unit circle and the added pole as an all-pass section.
- In the original vector of zeros, those outside the unit circle are replaced by the added zeros inside the unit circle.
- Reconstruct the a and b coefficients by passing the vector of poles and modified vector of zeros (together with the gain factor) through the function *'zp2tf.m'*.

3.4.4 Assessment of the different methods

The methods of Hilbert transform and homomorphic filtering, implemented as described above, were evaluated by decomposing the database of 7760 signals already used in 3.3. Before the minimum phase computation, the signals were subject to the post-processing listed in 3.3.

As explained before, the theory behind minimum phase decomposition methods assumes that the Fourier transforms computed are continuous, for which the phase has to be correctly unwrapped. Zeros which are very close to the unit circle can be potentially problematic, as they cause sharp phase changes and make it difficult to successfully unwrap the phase (Karam [2006]). It was seen that most of the zeros in HRTFs are close to the unit circle (see Fig. 3.2). Furthermore, as the signals are assumed periodic but they have to be discrete and finite in order to implement the decomposition methods, the results are sensitive to the length of the FFT algorithm used which ultimately determines how well the product equivalent to $\hat{h}(n)u_N(n)$ can be implemented. The evaluation looked into:

- Percentage of the 7760 HRIRs which could be effectively decomposed into a minimum phase filter.

- Length N used in the FFT, with $N = 2^7$ being the original length -i.e. no zero padding.
- Run time of the MATLAB implementation on a PC with a 1.7 GHz Pentium M processor and 1 GB of RAM.

The assessment was done by computing the position of the zeros after the methods have been applied. This was done by means of the MATLAB function *'isminphase.m'*, which looks for zeros outside the unit circle. This function computes the distance from the origin of coordinates in the z-plane, considering a tolerance that can be modified by the user. The predefined tolerance in MATLAB is $eps^{(2/3)}$, where *eps* is the relative precision of the double-precision floating-point system, measured as the distance between 1 to the next larger double-precision floating point. Such a tolerance accounts for possible rounding errors in assessing whether a zero would be greater than one or not.

In the case of the Hilbert transform implementation, the variation of length N was introduced when the function *'hilbert.m'* was called -i.e. the second FFT computation. In the case of homomorphic filtering, the function *'rceps.m'* was directly called with different values of N -i.e. it was included in the first FFT computation.

3.5 Results

Table 3.2 shows the result of the minimum phase decompositions for both homomorphic filtering and Hilbert transform, for different lengths of zero padding.

3.6 Discussion

Table 3.2 shows that the methods were identical in terms of success rate as a function of N. This could be expected since they are both FFT-based methods.

With these methods, almost 64% of the HRIRs were converted into a minimum phase filter with their original length $N = 128$ (or $N = 2^7$). Most of the HRIRs that could not be converted with that length were for contralateral signals from direction to the sides, which were seen in Fig. 3.1 to possess the highest rates of zeros outside the unit circle. Not coincidentally, the signals which needed the longest N values were also from contralateral sides at those directions. For instance, there were 5 signals that could not be converted to minimum phase with the tested methods and using $N = 2^{17}$. These were left side signals for directions $(-90^\circ, -22.5^\circ)$ and $(-22.5^\circ, -22.5^\circ)$, and right

Method	Length N	% Correct decomposition	Run Time
Hilbert transform	2^7	63.28%	370 *sec*
	2^8	0.32%	613 *sec*
	2^9	21.12%	397 *sec*
	2^{10}	7.38%	360 *sec*
	2^{11}	3.9%	361 *sec*
	2^{12}	1.89%	408 *sec*
	2^{13}	0.96%	533 *sec*
	2^{14}	0.56%	785 *sec*
	2^{15}	0.27%	1374 *sec*
	2^{16}	0.16%	2686 *sec*
	2^{17}	0.05%	5431 *sec*
Homomorphic Filtering	2^7	63.28%	331 *sec*
	2^8	0.32%	587 *sec*
	2^9	21.12%	378 *sec*
	2^{10}	7.38%	329 *sec*
	2^{11}	3.9%	316 *sec*
	2^{12}	1.89%	332 *sec*
	2^{13}	0.96%	398 *sec*
	2^{14}	0.56%	509 *sec*
	2^{15}	0.27%	820 *sec*
	2^{16}	0.16%	1479 *sec*
	2^{17}	0.05%	2840 *sec*

Table 3.2: Comparison of Hilbert transform and homomorphic filtering as means to decompose a mixed phase signal into a minimum phase signal. *Length N* refers to the zero padding used in the FFT computation (2^7 was the original length, without zero padding implemented), *% Correct decomposition* refers at how many of the total 7760 signals could be decomposed, and *Run time* refers to how long it took to apply the method to the whole database.

side signals for directions (67.5°,−22.5°), (90°,−22.5°) and (135°,−45°) -all the signals corresponded to different subjects. In all these cases, the methods failed due to zeros that were very close to the unit circle.

The previous observations for contralateral signals from directions to the sides is in agreement with the findings reported by Kulkarni *et al.* [1995], as their minimum phase and linear delay representation failed to account for the excess phase of the HRTFs in the same directions.

In terms of run time, homomorphic filtering proved to be a much faster method, particularly for the longer zero padding. While both methods are based on computing the FFT two times and the inverse FFT another two times, which are the most computationally demanding steps of the algorithms, it is believed that the difference in run time is given mostly by a sub-optimal implementation of the Hilbert transform method.

3.7 Conclusion

This investigation focused on different methods to decompose HRTFs into minimum phase and all-pass components, as a step forward towards a reliable robust method that can be used with large HRTFs databases. Two methods were chosen for comparison: Hilbert transform and homomorphic filtering, both implemented in MATLAB. It was found that the methods were comparable in terms of success rate as a function of the FFT length used -i.e. the total number of HRTFs, from a database of 3880 pairs, that could be successfully decomposed into a minimum phase signal. The implementation of homomorphic filtering over-performed the implemented Hilbert transform in terms of run time, but possibly due to a sub-optimal implementation of the latter. It was also found that most HRTFs (63.28% out of 7760 signals) did not require zero padding for being successfully decomposed into a minimum phase filter. However, the rest of the signals showed that there was no zero padding length that could guarantee the success of minimum phase computation and each signal had to be checked individually. The most problematic cases were the contralateral signals for directions to the sides, particularly below the horizontal plane. These signals presented the highest rate of zeros outside the unit circle in a z-domain representation of the filters.

Chapter 4

Audibility of all-pass sections in measured HRTFs

4.1 Introduction

HRTFs, as other electroacoustical transfer functions, can be decomposed into minimum phase, linear phase and all-pass components (Møller *et al.* [1995a], Oppenheim & Schafer [1989]). The previous Chapter focused on methods for obtaining minimum phase HRTFs from mixed phase ones. That procedure eases the implementation of HRTFs as filters: the ITD can be computed from the linear phase part of the HRTFs and implemented as a delay, while the spectral characteristics given by the minimum phase components (Wightman & Kistler [2005], Kistler & Wightman [1992], Kulkarni *et al.* [1999]) can be implemented as short FIR filters. In such models, low Q-factor all-pass sections from HRTFs are included in the ITD computation when the interaural group delay of the all-pass components at low frequencies (IGD_0) is above 30 μs, since they are audible as lateral shifts above that threshold (Minnaar *et al.* [1999], Plogsties *et al.* [2000]). Current procedures, on the other hand, discard high Q-factor all-pass components even though they are audible as a ringing in general electroacoustical transfer functions, as reported by Møller *et al.* [2007]. In this latter investigation, it was suggested that all-pass sections in HRTFs would not be audible since they are centered at the same frequencies where notches occur in the magnitude response of their minimum phase components. The goal of the investigation reported in this Chapter was to test whether that assumption was valid -i.e. to test the audibility of removing high Q-factor all-pass sections from HRTFs. It was hypothesized that high Q-factor all-pass sections were audible when presented alone, but they would become inaudible when presented with their minimum phase HRTFs counterpart. A three-alternative forced choice experiment was conducted. Results suggested that the hypothesis held, and it was concluded that high Q-factor all-pass sections could be discarded from HRTFs when these were used for binaural synthesis.

This Chapter is organized as follows. Firstly, the relevant literature is reviewed and the theoretical background on all-pass sections is briefly revised. Secondly, the methods used in the experimental design and implementation are covered. The results are subsequently presented and discussed. The Chapter ends with some concluding remarks.

Part of the work presented in this Chapter has been already reported by Toledo & Møller [2008a].

4.2 Previous works

As reviewed in the previous Chapter (see 3.2), it is common practice to implement HRTFs as their minimum phase representation and a linear delay for ITD. Some of the studies that used such models for binaural synthesis were reported by Wenzel & Foster [1993], Sandvad & Hammershøi [1994], Langendijk & Bronkhorst [2002], Minnaar *et al.* [2005] and Wightman & Kistler [2005]. Minimum phase HRTFs are so widely used because several studies have confirmed the theoretical and perceptual validity of the model. The key points of those investigations will be reviewed in the following.

The seminal work on minimum phase HRTFs is that of Mehrgardt & Mellert [1977], who investigated the decomposition of HRTFs measured on human subjects. They found that the excess phase component was nearly linear up to 10 kHz and concluded that HRTFs were minimum phase up to that frequency. It has to be noted that Mehrgardt and Mellert used the term all-pass instead of excess phase. The difference between terms will be clarified later in this Chapter.

Wightman & Kistler [1989a] reported that the decomposition of their measured HRTFs into all-pass and minimum phase components supported the findings of Mehrgardt & Mellert [1977] that HRTFs could be modeled by a minimum phase system up to 10 kHz. Their psychophysical results showed that localization performance with minimum phase HRTFs was similar to that with measured (mixed phase) HRTFs -both conditions being played back through headphones. The authors discussed these findings and provided more evidence supporting the results in Wightman & Kistler [2005].

Møller *et al.* [1995a] showed graphical examples of decomposition of HRTFs into minimum phase and all-pass components, and reported that HRTFs, as a common trend, presented several second-order all-pass sections. In a work done at the same laboratory, Sandvad & Hammershøi [1994] compared different filter representations of HRTFs. They concluded that minimum phase HRTFs were good approximations, as the probability of perceiving differences between the minimum phase and the reference HRTF was very low.

Jot *et al.* [1995] reported that the excess phase of HRTFs was linear below 8 to 10 kHz, for which the all-pass associated to the HRTFs was approximately equivalent to a pure delay.

Brown & Duda [1998] compared measured dummy-head HRTFs with their minimum phase representations. They stated that, although HRTFs were generally minimum phase, the contralateral sides of HRTFs for directions to the sides were clearly not minimum phase. They acknowledged the need of listening tests to validate the minimum phase model.

Huopaniemi *et al.* [1999] reported that HRTFs were *nearly* of minimum phase, since the excess phase had been found to be approximately linear -which corresponded to a pure delay in time domain. Therefore, the excess phase part could be implemented as an all-pass filter or, as a special simplified case, a pure delay. They did not provide any psychoacoustical validation.

Kulkarni *et al.* [1999] studied the human sensitivity to phase structure in HRTFs. They psychoacoustically tested discriminability of three different HRTFs models: minimum phase HRTFs and a linear phase component as ITD, zero phase HRTFs and a linear phase component as ITD, and reversed phase HRTFs (equivalent to a time reversed signal) and a linear phase component as ITD. Of interest to the present study are the findings of their first experiment: they concluded that the minimum phase plus linear phase model was perceptually valid as long as the low frequency ITD was correctly computed. It was implicit in their report that they discarded the all-pass components from HRTFs, and they actually found that the interaural phase difference at low frequencies had to be included in the ITD computation in some cases, as they introduced a localization cue. This is in line with the findings of Minnaar *et al.* [1999] and Plogsties *et al.* [2000], who concluded that the interaural group delay of the all-pass components at low frequencies was audible as a lateral shift when it was above 30 μs. The evidence that the all-pass sections were discarded by Kulkarni *et al.* [1999] is implicit in their objective measure: they compared the computed excess phase part with a linear phase model and found large deviations in directions at the sides for contralateral signals. This is expected if Figure 3.1 from Chapter 3 is seen again: those directions present more all-pass sections than other ones. However, and in spite of a clarification of concepts such as all-pass and linear phase components, Kulkarni *et al.* [1999] failed to acknowledge the presence of all-pass sections and considered the excess phase of HRTFs as purely linear phase. It has to be noticed, however, that they referred to the excess phase as all-pass.

In a more recent study, Møller *et al.* [2007] tested the audibility of second-order all-pass sections in electracoustical transfer functions, centered at different frequencies. They found that low Q-factor all-pass sections could be audible as a shift of the auditory image when presented in one ear only, and that high Q-factor all-pass sections were audible as a ringing if the Q-factor was high enough. They presented thresholds of audibility for both cases. For the high Q-factor case, they found that the thresholds related to a constant $\frac{Q-factor}{f_0}$ ratio. They also mentioned that, if the Q-factor took a very high value, the presence of the all-pass section became inaudible again. Møller *et.al.* suggested that the found thresholds for ringing did not applied for HRTFs: when HRTFs were decomposed in the z-plane, zeros were added inside the unit circle, which gave rise to notches in the magnitude. At the same frequencies, the all-pass section were centered giving place to possible ringing. As the Q-factor got higher and was more likely to be audible, the more the ringing was minimized by the minimum phase counterpart. At the same laboratory, the experiments reported by Plogsties *et al.* [2000] were conducted. As mentioned previously, they studied the audibility of low Q-factor all-pass sections in HRTFs and concluded that they were audible as lateral shifts when the IGD_0 was above 30 μs.

From the reviewed literature, it can be seen that Kulkarni *et al.* [1999] and Mehrgardt & Mellert [1977] provided graphical evidence that HRTFs were not minimum phase functions and contained a linear phase and all-pass components. This has also been explicitly shown by Møller *et al.* [1995a], Minnaar *et al.* [1999] and Plogsties *et al.* [2000]. According to Plogsties *et al.* [2000], it is important that the ITD accounts for all-pass components that become audible. Even though high Q-factor all-pass sections are also present in HRTFs, minimum phase plus ITD models proved to be perceptually valid for binaural synthesis, as supported by Wightman & Kistler [1989a], Wightman & Kistler [2005], Kulkarni *et al.* [1999] and Sandvad & Hammershøi [1994]. This suggests that high Q-factor all-pass sections are not audible in HRTFs -i.e. they were discarded in all previous testing- even though they are audible in more general electroacoustical transfer functions, which has been already suggested by Møller *et al.* [2007].

4.3 Background concepts

The theory behind HRTFs as mixed phase systems has been already presented in Chapter 3. Summarizing the concepts presented there, it can be said that HRTFs are mixed phase systems that can be decomposed into minimum phase and excess phase components. The latter can be further decomposed into linear phase and all-pass components. Following the notation used in 3.4.3, these concepts can be expressed in the z-domain as:

$$HRTF(z) = HRTF(z)_{min.} \cdot HRTF(z)_{exc.} \tag{4.1}$$

With

$$HRTF(z)_{exc.} = HRTF(z)_{lin.} \cdot HRTF(z)_{all-pass} \tag{4.2}$$

And therefore,

$$HRTF(z) = HRTF(z)_{min.} \cdot HRTF(z)_{lin.} \cdot HRTF(z)_{all-pass} \tag{4.3}$$

Characteristics of minimum phase and linear phase components have already been outlined and will not be repeated here. A thorough theoretical background on systems with different phase characteristics can be found in Oppenheim & Schafer [1989] and Møller *et al.* [2007]. The following paragraphs will focus on all-pass sections.

All-pass components have unity magnitude. In the z-plane, a first-order all-pass section consists of a single pole inside the unit circle and a zero at a conjugate reciprocal location. Second order all-pass sections are complex conjugated pairs of poles inside the unit circle and zeros at mirrored positions. All-pass sections of any higher order can be obtained from the product of first and second-order all-pass sections. All-pass transfer functions can be expressed in terms of their center frequency (f_c) and quality factor (Q-factor). Considering the all-pass impulse response as an impulse followed by an exponentially decaying sinusoid, the center frequency of the all-pass section is close to the frequency of the exponential decay -i.e. ringing- and the Q-factor is associated to the peak in the phase and group delay of the all-pass. The Q-factor is related to the decay time of the impulse response: a low Q-factor implies that the all-pass impulse response dies out in a short time, while a high Q-factor implies that it remains ringing for a longer time.

4.4 Methods

The goal of the investigation presented here was to test the audibility of high Q-factor all-pass sections in HRTFs. All-pass sections that were more likely to be audible were selected from the large database of measured HRTFs (Møller *et al.* [1995a]) already used in Chapter 3. Signals with and without all-pass sections were presented to listeners in a psychoacoustical experiment.

4.4.1 Decomposition of HRTFs

Measured HRIRs at a sampling frequency (fs) of 48 kHz were used. In experiments reported by Sandvad & Hammershøi [1994], it was shown that HRIRs of 72 taps of length (with $fs = 48\ kHz$) were long enough to convey all the needed cues to sound localization. However, HRIRs of 64 samples were used in this investigation due to computational constrains of the algorithm implementing the z-domain method for HRTFs decomposition (see 3.4.3). The z-domain method was selected to keep consistency with a previous study conducted at our laboratory (Møller *et al.* [2007]). The theory behind the method and its implementation have already been described in 3.4.3. The first step in the computation was to identify the initial linear delay with the 5% leading edge criterion (Sandvad & Hammershøi [1994]) and remove it, yielding:

$$HRTF(z) = HRTF(z)_{min.} \cdot HRTF(z)_{all-pass} \tag{4.4}$$

The all-pass sections were then computed from the roots of the polynomial by finding the zeros outside the unit circle in the z-plane representation of the HRTFs. In that way, $HRTF(z)_{min.}$ was conformed by all the zeros of $HRTF(z)$ lying inside the unit circle plus zeros added at the conjugate reciprocal positions of those outside the unit circle. $HRTF(z)_{all-pass}$ consisted of all the zeros of $HRTF(z)$ lying outside the unit circle plus poles canceling the added zeros in $HRTF(z)_{min.}$. The computed $HRTF(z)_{all-pass}$ were implemented as IIR filters and $HRTF(z)_{min.}$ as FIR filters. Since high Q-factor all-pass components have long impulse responses, 1024 taps were used for all filters.

4.4.2 Selection of HRTFs

The database of measured HRTFs that was available for this experiment consisted of 3880 pairs of HRTFs, and it was the same one used in Chapter 3. As it was impossible to include all of them in the listening experiment, it was decided to analyze the database in order to select a few representative HRTFs with all-pass sections most likely to be audible according to their Q-factors[1]. The conditions imposed to select the HRTFs for the experiment were: 1) at least one all-pass section had to have its Q-factor well above the high Q-factor threshold of audibility in the mid-frequency range and 2) none of the Q-factors had to be below the threshold of audibility for low Q-factor. The second condition was also checked by computing the IGD_0 and ensuring that it was well below

[1] The thresholds of audibility in the context of electroacoustic transfer functions are given by Fig. 9, Møller *et al.* [2007], and are reproduced later in this Chapter.

30 μs. It was found that the second condition was easy to fulfill. However, the first condition was mostly found in the contralateral signals of HRTFs from the sides. This was not surprising given the analysis shown in 3.3, where it was seen that contralateral signals of HRTFs from the sides had the highest number of zeros outside the unit circle -particularly for directions to the sides.

The panels of Fig. 4.1 show the six pairs of HRTFs selected. Thick lines correspond to the left side signals and thin lines to the right side signals. At the top of each box, the directions to which the HRTFs correspond are marked. Each pair of HRTFs belongs to a different subject. Fig. 4.2, Fig. 4.3 and Fig. 4.4 show the impulse responses of the selected HRTFs in their minimum phase, all-pass and combined (minimum phase plus all-pass) forms respectively -only the contralateral side is shown. Fig. 4.5 shows the Q-factor of the all-pass components from the six HRTFs selected, ordered in the same fashion as in Fig. 4.1. Filled circles indicate all-pass sections from right ears, blank circles correspond to left ears. The thresholds of audibility found by Møller *et al.* [2007] for all-pass sections in electroacoustic transfer functions are also shown.

4.4.3 Conditions

Each condition compared an impulse response to the same impulse response with its all-pass sections removed. The all-pass sections from the selected HRTFs were either presented alone or with their corresponding minimum phase HRTFs. Presentations were done in binaural and diotic (same sound at both ears) conditions. In the former, both sides of the HRTFs were used and played back. In the latter, only the most unfavorable case (contralateral side of the HRTFs) was reproduced at both ears, without ITD. The following conditions were determined:

A- Minimum phase HRTFs (with corresponding ITD) with and without their associated all-pass sections (binaural reproduction).

$$HRTF(z)_{min.} \cdot HRTF(z)_{lin.}$$
$$vs.$$
$$HRTF(z)_{min.} \cdot HRTF(z)_{lin.} \cdot HRTF(z)_{all-pass}$$

B- Impulses (with ITD) with and without all-pass sections from HRTFs (binaural reproduction).

$$H(z) \cdot HRTF(z)_{lin.}$$
$$vs.$$
$$H(z) \cdot HRTF(z)_{lin.} \cdot HRTF(z)_{all-pass}$$

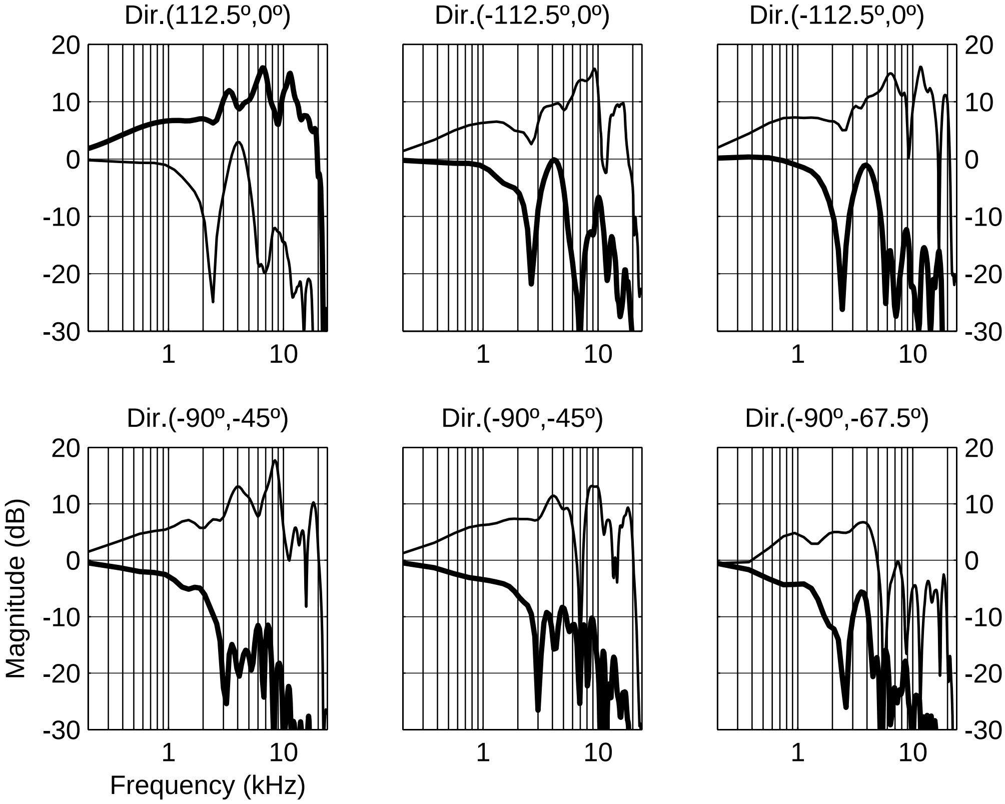

Figure 4.1: Minimum phase magnitude response of the HRTFs selected for the experiment. Thick lines correspond to the left side signals and thin lines to the right side signals. The HRTFs used belonged to different subjects. The direction for which the HRTFs were measured are shown at the top of each box.

C- Minimum phase components (without ITD) with and without their associated all-pass sections, same signal presented at both ears (diotic reproduction).

$$HRTF\ Diotic(z)_{min.}$$
$$vs.$$
$$HRTF\ Diotic(z)_{min.} \cdot HRTF(z)_{all-pass}$$

D- Impulses (without ITD) with and without all-pass sections from HRTFs, same signal presented at both ears (diotic reproduction).

$$H\ Diotic(z)$$
$$vs.$$
$$H\ Diotic(z) \cdot HRTF(z)_{all-pass}$$

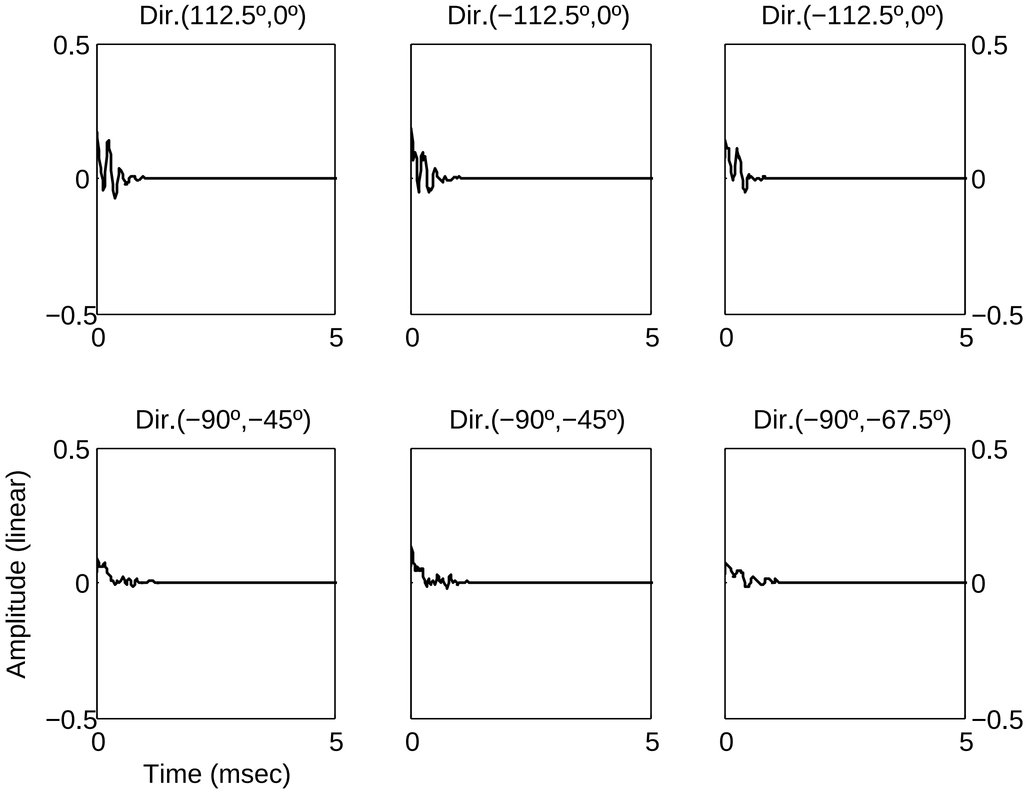

Figure 4.2: Minimum phase impulse responses of the HRTFs shown in Fig. 4.1. Only the contralateral signal of each HRIRs pair is shown.

Conditions C and D did not correspond to a natural situation. They were introduced to evaluate whether binaural interactions played a role in the audibility of high Q-factor all-pass sections: in the HRTFs selected there was mostly one side (contralateral) with potentially audible all-pass sections.

4.4.4 Procedure

A three alternative forced choice (3AFC) experiment was conducted. In a 3AFC experiment there are six possible sequences of ordering the presentations: AAB ABA BAA BBA BAB ABB. These six sequences were presented twice per each of the six HRTF/impulse pair. This gave a total of 72 trials per condition. Trials were randomized for every listener. Stimuli was presented in four blocks of 72 trials. The two first blocks tested conditions A and B, the two last blocks tested conditions C and D. The task of the subject was to report the sample that sounded different from the other two, regardless

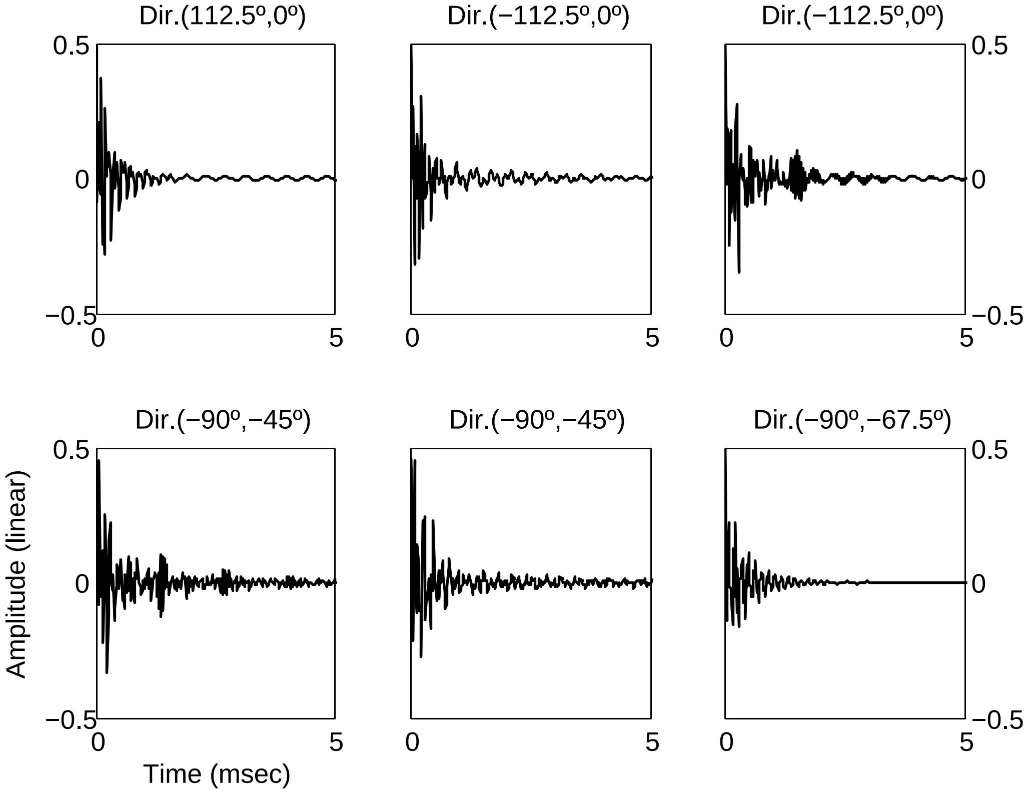

Figure 4.3: Impulse responses of the all-pass sections associated to the HRTFs shown in Fig. 4.1. Only the contralateral signal of each HRIRs pair is shown.

the nature of the difference.

Before the experiment, subjects were given written instructions about the task. They were taken to an anechoic chamber, where they conducted a training session consisting of 24 trials arbitrarily selected and without headphone equalization. After the training session, subjects proceeded with the experiment proper. During both training session and experiment, subjects interacted with the screen showed in Fig. 4.6. Each number in the screen was highlighted synchronously with a sound sample. In order to report the sample that sounded different, subjects had to touch the corresponding number in the screen. After their response to one trial, they had to press the NEXT button to hear the following trial. Subjects were given breaks between blocks, in which they were required to leave the anechoic chamber. Subjects were not given feedback during the training session nor the experiment. All subjects completed the experiment within two hours in a single day.

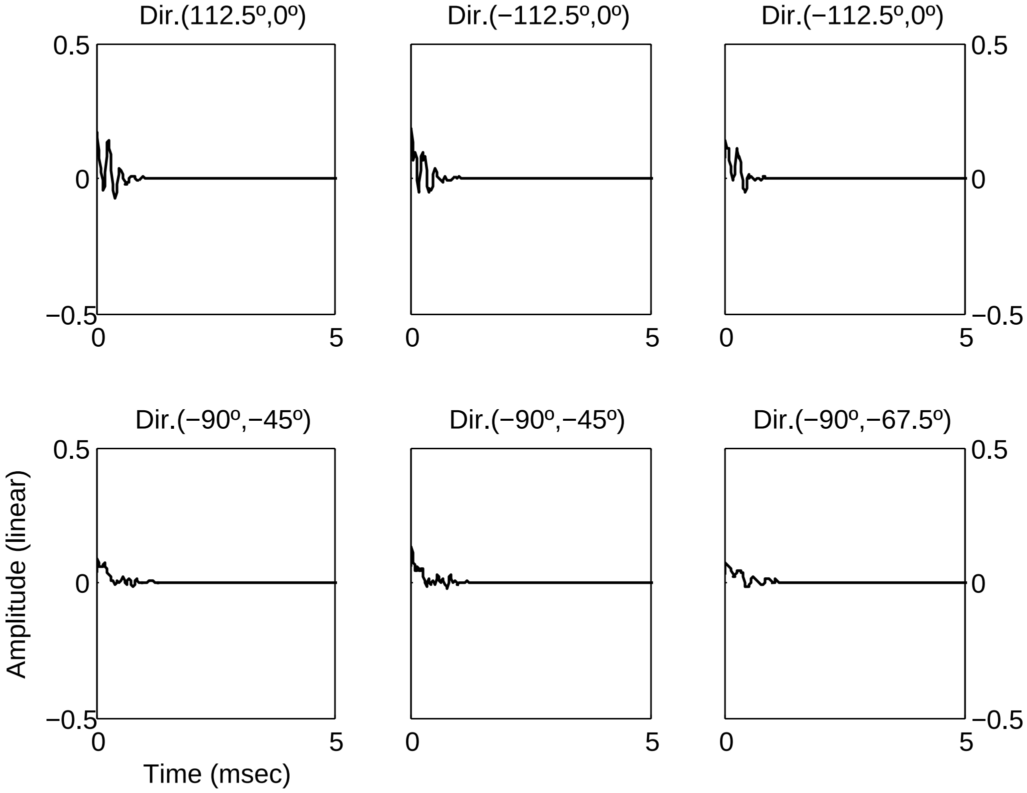

Figure 4.4: Impulse responses of the combined form $HRTF(z)_{min.} \cdot HRTF(z)_{all-pass}$ -i.e. the convolution of the impulse responses shown in Fig. 4.2 and Fig. 4.3.

4.4.5 Subjects

Twelve paid subjects with normal hearing participated in the experiment. They were six females and six males, with ages ranging from 20 to 30 years old. Their hearing thresholds were determined by a standard pure-tone audiometry in the frequency range from 250 *Hz* to 8 *kHz*. None of the subjects had hearing thresholds above 15 *dB HL*. Some of the listeners had participated in listening tests before, but all of them were unfamiliar with the procedure and the differences presented. Therefore, they were considered naïve for the purpose of this experiment.

4.4.6 Stimuli

The stimulus consisted of a single impulse (perceived as a click) of 21.3 milliseconds (1024 taps sampled at 48 *kHz*). This impulse was filtered with the appropriate filters

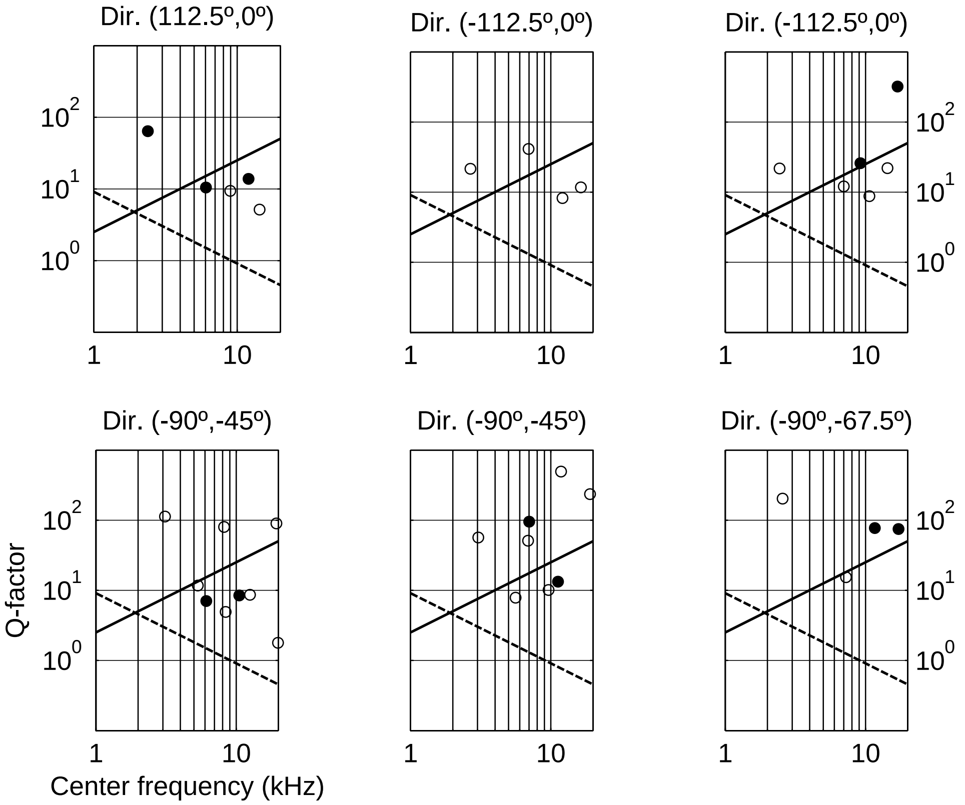

Figure 4.5: Q-factors of the selected all-pass sections from the HRTFs in Fig. 4.1. Filled circles indicate all-pass sections from right ears, blank circles correspond to left ears. The high Q-factor and low Q-factor thresholds of audibility for all-pass sections found by Møller *et al.* [2007] are indicated with the solid and dashed lines, respectively.

obtained from the decomposition in order to produce the signals required to test the conditions of interest. All processing was done with MATLAB. In each trial, subjects listened to three intervals of sound. The silence between clicks was 500 milliseconds. The time between trials was controlled by the subjects.

4.4.7 Signal generation and reproduction

The equipment was placed in a control room next to the anechoic chamber where the subject sat. Signals were played back through a PC with a digital sound card RME HDSP 9632 connected to an external AD/DA converter RME ADI-8 DS. The signals fed a power amplifier Pioneer A-616. The level of the amplifier was raised and a passive

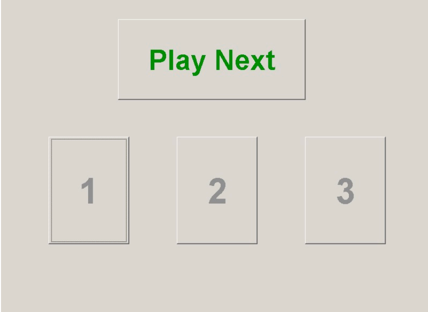

Figure 4.6: Interface presented to the subjects in a touchscreen for judgments report.

attenuator used to reduce the overall noise. The output of the attenuator was delivered to the subjects through headphones Beyerdynamic DT990, individually equalized. Typical transfer functions for these headphones have been shown by Møller *et al.* [1995b] and Wightman & Kistler [2005].

4.4.8 Headphone equalization

Headphones were equalized individually for each subject. An in-house developed dual channel *MLS* system (Olesen *et al.* [2000]) was used to measure the headphones transfer functions (PTFs) in each subject. The signals were collected by two Sennheiser KE 4-211-2 miniature microphones placed at the blocked entrance of the ear canals of the subjects. Microphones were calibrated and connected to a power supply that provided a gain of 20 *dB*. Two measuring amplifier Bruel & Kjær 2607 were used which fed the signals to the AD/DA converter and back to the PC. All measurements were done at a sampling frequency of 48 *kHz*. Appropriate post-processing to construct the inverse filters was implemented with MATLAB. Equalization filters were constructed from the average of 5 repeated measurements, and they only accounted for the minimum phase response of the PTFs.

4.5 Results

Fig. 4.7 shows results for conditions A (white bars) and B (black bars). Fig. 4.8 shows results for conditions C (white bars) and D (black bars). The bars indicate the percentage of correct answers (ordinate) for each subject (arbitrarily numbered in the abscissa). Results from all the HRTFs in a given condition were pooled. In a 3AFC experiment, the number of correct answers is binomially distributed. The probability of guessing

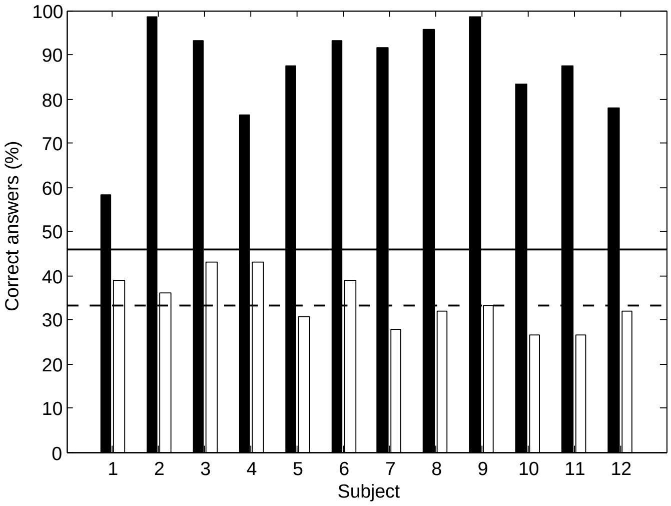

Figure 4.7: Results for conditions A (white bars) and B (black bars). The bars indicate the percentage of correct answers (ordinate) for each subject (arbitrarily numbered in the abscissa). Results from all the HRTFs in a given condition were pooled. The probability of guessing is $1/3$ (dashed line) and the null hypothesis can be rejected at the 1% of significance level if the percentage of correct answers is greater than 46% (solid line).

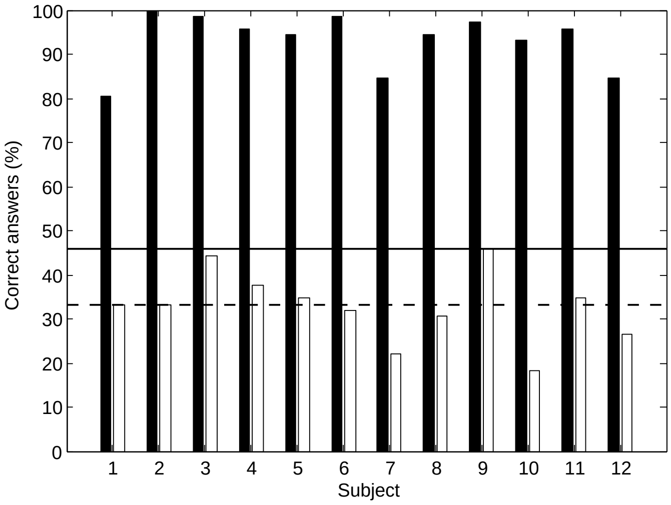

Figure 4.8: Same as Figure 4.7, but with results for conditions C (white bars) and D (black bars).

is $1/3$ and the null hypothesis can be rejected at the 1% of significance level if the percentage of correct answers is greater than 46%. In Fig. 4.7 and Fig. 4.8, the 1% significance level boundary is shown by the solid line and the chance level is shown by the dashed line.

4.6 Discussion

The null hypothesis -i.e. the subjects are guessing- can be rejected if the percentage of correct answers is above 46%, represented by the solid lines in Fig. 4.7 and Fig. 4.8. The results obtained here show that the null hypothesis was rejected for all subjects in conditions B and D, and for none in conditions A and C.

The results of condition A (Fig. 4.7) indicate that removing the all-pass component to HRTFs was inaudible for all subjects. When the same all-pass components were presented alone, the differences became audible as it is seen in the results of condition B in the same figure. The research hypothesis was then confirmed by these results. Therefore, it is concluded that high Q-factor all-pass sections from HRTFs can be discarded without audible consequences.

The results of condition C (Fig. 4.8) show that when the same all-pass section was added at both ears with a minimum phase counterpart, the differences were inaudible to all subjects. On the other hand, the differences became audible to all subjects when the all-pass sections were presented alone (condition D). These results would suggest that binaural interaction does not play a role in the lack of audibility of high Q-factor all-pass sections, as the same trend seen in Fig. 4.7 is followed.

Analysis of HRTFs from the used database has shown that most HRTFs contain all-pass sections -see Fig. 3.1 in Chapter 3. Furthermore, the center frequencies of high Q-factor all-pass sections correspond to deep notches in the magnitude response (see Figures 4.1 and 4.5). These all-pass sections would not be expected to produce the perception of ringing: if a deep notch is present in the magnitude, the amplitude of the ringing becomes smaller. This can also be understood from the impulse responses: the all-pass impulse responses are long (Fig. 4.3), but minimum phase impulse responses convolved with all-pass impulse responses are rather short (Fig. 4.4). It would seem plausible to hypothesize that it is the particular high-frequency magnitude response associated to each all-pass section which is responsible for the inaudibility of the latter. This was also suggested by Møller *et al.* [2007] in terms of the z-domain representations: in the reported process to decompose the HRTFs, zeros are added inside the unit circle that give rise to notches in the magnitude response. At the same frequency, the all-pass section is centered giving place to possible ringing. As the Q-factor gets higher and is more likely to be audible, the more the ringing is minimized by the minimum phase counterpart.

4.7 Conclusion

This Chapter investigated the perceptual consequences of removing high Q-factor all-pass sections from HRTFs, in order to verify the hypothesis proposed by Møller *et al.* [2007] that there are no perceptual consequences in doing so, since the all-pass sections are centered at frequencies with deep notches in the HRTFs frequency response. A listening experiment to assess the audibility of high Q-factor all-pass sections was conducted. The results showed that high Q-factor all-pass sections from HRTFs were audible if tested alone, but were inaudible when they were combined with their associated minimum phase HRTFs. Therefore, it is concluded that high Q-factor all-pass sections from HRTFs can be discarded without audible consequences in binaural synthesis. In other words, it is perceptually valid to represent HRTFs by minimum phase functions and a linear delay as ITD.

Chapter 5

The role of HRTFs' spectral features in sound localization

5.1 Introduction

It has been shown in previous Chapters that HRTFs can be represented by their minimum phase components (Chapter 3) and an ITD, which is a perceptually valid implementation of HRTFs (Chapter 4). In such a model, the ITD includes all the relevant temporal information while the minimum phase component comprises all the spectral details required for sound source localization. While ITDs and low frequency spectra of HRTFs are known to change little across subjects, high frequency spectral components of HRTFs are highly dependent on the anthropometric characteristics of the individual subject, particularly on his or her pinnae shape. These very individual characteristics are responsible for localization errors in binaural synthesis with non-individual HRTFs: localization performance is degraded if the spectral characteristics of the directional filters used do not match the individual characteristic of the listener's HRTFs (Butler & Belendiuk [1977], Wenzel *et al.* [1993], Wightman & Kistler [1993], Begault *et al.* [2001]). How similar the HRTFs should be to avoid degradation in the performance, or whether the similarity covers single features (as proposed by Blauert [1969/70], Hebrank & Wright [1974b], Bloom [1977], Han [1994]) or broad ranges of frequencies (as proposed by Macpherson [1994], Wightman & Kistler [1993], Langendijk & Bronkhorst [2002]), is still unknown. The investigation presented in this Chapter focused on the spectral characteristics of HRTFs that are relevant as localization cues. The research hypothesis was that non-individual HRTFs would evoke the direction for which the relevant spectral features that cue localization were present, regardless the direction for which they were measured, and that they could therefore be identified and parameterized. The hypothesis was tested by matching individual and non-individual

HRTFs from different subjects according to the results of localization experiments, and comparing simple parameters of peaks and notches such as Q-factor, bandwidth, center frequency and slopes. It was found that the relevant cues covered a broad range of frequencies in which the first peak, first notch and second peak in a windowed version of the HRTFs were relevant. For some subjects, the Q-factor of the first peak seemed to disambiguate front-high, back-high and above directions. The global Q-factor of the first notch seemed to disambiguate front, back and back-low directions, while maintaining a trend for high and above regions as well. The second peak seemed to convey redundant information that was also relevant for disambiguating back and back-low directions. Parameters were seen to span ranges of values and frequencies, and still evoke the same angle span of directions. These ranges of values and frequencies appeared to be individual for each subject.

This Chapter is organized as follows. Firstly, a literature review is presented. Secondly, the experimental design and implementation is explained, along with the statistical methods used for analyzing the behavioral results. Two groups can be defined according to the experimental design: Group A and B. The parameterization methods used in the HRTFs spectral analysis are also given. The Chapter follows with the presentation of the results for both Group A and B. Subsequently, the results are discussed covering topics such as correlation analysis, externalization of binaurally reproduced sound and spectral features analysis. Finally, conclusions are presented.

Part of the work presented in this chapter has already been reported by Toledo & Møller [2008b].

5.2 Previous works

A general review on the topic of spectral cues to sound localization was published by Carlile *et al.* [2005]. Their approach was broad and covered a wide range of issues within the field, for which relatively little emphasis was given to the particular discussion of individual spectral features of HRTFs and pinnae transfer functions that are relevant for localization in the MSP, at least as it is needed for this Chapter. Of particular interest to the experiment presented here are those works based on localization and sound source identification in the MSP. There is a general understanding that localization off the MSP make use of interaural cues and that findings cannot be directly extended to that plane since there is a theoretical lack of interaural differences in HRTFs from it[1]. As a direct consequence, it is localization in the elevation dimension

[1] The lack of interaural differences is theoretical in the case of human subjects, since the ITD is not always equal to zero due to inaccuracies in the measurement procedure and there are interaural spectral

that becomes relevant for this Chapter, and the literature review that follows focuses on it. Many authors have assessed localization performance for real and virtual sound sources in order to understand its relationship to spectral features. As an overview, it can be stated that some authors have been advocates of spectral peaks as important cues (Blauert [1969/70], Humanski & Butler [1988], Middlebrooks [1992], Carlile & Pralong [1994]) and others have argued in favor of notches in the spectra (Bloom [1977], Hebrank & Wright [1974b]). Later studies suggested that several spectral features, covering a somewhat larger frequency range, were relevant (Asano *et al.* [1990], Wightman & Kistler [1993], Han [1994], Langendijk & Bronkhorst [2002]). The interest around spectral cues to sound localization increased since binaural synthesis started to be implemented in virtual auditory systems to emulate realistic virtual environments, even though the topic is of broader relevance. Some studies showed that the use of non-individual HRTFs degraded the localization performance in binaural synthesis (Wenzel *et al.* [1993], Begault [2000]), particularly if the HRTFs used did not match those of the listener (Wightman & Kistler [1993]), but the impossibility of measuring individual HRTFs for each user of a virtual auditory system was acknowledged. In the following, the literature that is relevant to the hypothesis of this Chapter will be covered. The works cannot, in general, be directly compared since the differences in the methods used are substantial. Therefore, a more detailed description of the experiments is given than in literature reviews from other Chapters of this Thesis. Studies conducted off the MSP are included only if relevant.

One of the seminal works on spectral features that cue sound source localization was that of Blauert [1969/70], who analyzed the localization of $\frac{1}{3}$ -octave band noise samples presented through loudspeakers. Stimuli was located in the MSP directly in front, above, and in the back. Consistently, he divided the hemisphere above the horizontal plane into sectors *front*, *above* and *behind*. Since the judgements of direction clustered depending on the center frequency of the noise presented, Blauert formulated the concept of directional bands: each direction in space would be associated to particular frequency bands. A sound with a center frequency (f_c) in a given directional band would evoke a sound image in the associated direction to that frequency band. Blauert also measured HRTFs and subtracted the mean sound pressure levels from the front and rear hemispheres, yielding his proposed concept of boosted bands: depending on the band where the peak excitation was located, the sound would be perceived either to the front or to the rear. In a different experiment, Hebrank & Wright [1974b] presented filtered noise through a loudspeaker that moved in elevation from 30° to 210° in the MSP. Filters were pass-band, stop-band, high-pass and low-pass. They identified

differences between right and left HRTFs signals as a consequence of assymetries of the human anthropometry. While some authors have argued that the small spectral differences could be cues to localization, or that elevation in the MSP is a binaural process (Ivarsson *et al.* [1980]), Hebrank & Wright [1974a] showed that localization in the MSP is fundamentally a monaural process.

those directions for which subjects judgements were biased towards a particular direction and they measured artificial pinnae to correlate the filtering needed to the actual one provided by the pinnae transfer functions. The findings reported by Blauert [1969/70] and Hebrank & Wright [1974b] were consistent with each other. Hebrank and Wright found that perception of frontal direction was determined by a 1-octave notch with low cut-off frequency between 4 *kHz* and 8 *kHz*, with peaks below and above those frequencies and increased energy above 13 *kHz*. They identified that the high and low cut-off frequencies of the notch were responsible for elevation discrimination in the frontal directions. This is in agreement with both directional and boosted bands reported by Blauert for the frontal direction. The cue for directions above was found by Hebrank and Wright to be a $\frac{1}{4}$ -octave peak between 7 *kHz* and 9 *kHz*. This is also in accordance to Blauert's findings of a peak centered at 8 *kHz* as directional band for above. The cue for behind reported by Hebrank and Wright was a small peak between 10 *kHz* and 12 *kHz*, with decreased energy above and below the peak, and which was consistent with Blauert's directional and boosted bands. Blauert also found a likely band around 1 *kHz*, but this frequency was not tested by Hebrank and Wright neither as cut-off nor center frequency. Lastly, Hebrank and Wright showed that the pinnae transfer functions provided the filtering needed in each direction. They hypothesized that the main cues were given by reflections from the back wall of the concha and formulated a model in time domain that adjusted well to their data. A later study reported by Itoh *et al.* [2007] showed that individual differences existed for the directional bands proposed by Blauert, but that they were strong parameters since directional bands occurred even for subjects who had naturally degraded free field localization performance with broadband noise. Furthermore, Moore *et al.* [1989] measured detection thresholds for spectral peaks and notches and showed that those reported by Hebrank & Wright [1974b] were detectable cues. Moore *et al.* [1989] reported determining the detection thresholds for spectral peaks and notches, thresholds for notch-depth and peak-height discrimination and thresholds for detecting changes in f_c of peaks and notches. In general, it appeared that spectral notches were perceptually less salient than peaks, particularly at high frequencies. Thresholds for peaks decreased for increasing bandwidth. The 1 *kHz* thresholds for bandwidths of $0.125f_c$, $0.25f_c$ and $0.5f_c$ were 2.9, 2.4, and 2.1 *dB*, respectively. Corresponding thresholds at 8 *kHz* were 5.3, 3.9, and 2.5 *dB*. For notches at 1 *kHz*, thresholds were similar to those for peaks, but performance decreased with decreasing bandwidth. Thresholds for notches at 8 *kHz* could not be measured, possibly because narrow notches are represented by only a small dip in the excitation pattern evoked by the noise. Regarding the height of a peak and depth of a notch, thresholds increased with decreasing bandwidth in all cases. They also decreased with increasing noise level. Averaged across bandwidths, the thresholds at 1 *kHz* were 2.3, 1.5, and 1.5 *dB* for three different spectrum levels. Performance was poorer at 8 *kHz*. The authors concluded that spectral peaks were more salient than notches as pinna cues, even though changes in the f_c of notches were detectable. They suggested that the thresholds to detect changes in f_c of notches from the anterior MSP could explain localization.

Other studies confirmed the relevance of HRTFs' high frequency range in sound source localization. For example, Roffler & Butler [1968a] reported a real sound sources identification experiment where sources were located in the frontal MSP at elevations -13°, -2°, 9° and 20°. Subjects did not see the loudspeakers, which were covered by a cloth panel with numbers from 1 to 13. Subjects had to identify the number behind which the sound was coming from. Stimuli comprised broadband, low-pass (with cut-off at 2 kHz) and high-pass (with cut-off at 2 kHz and 8 kHz) filtered noise and tonal bursts ($f_c = 0.6$ kHz and 4.8 kHz). Subjects could not identify tonal stimuli nor the low-pass filtered stimulus. They found that the 0.6 kHz tonal stimuli was localized systematically around -2°, while the same happened for the 4.8 kHz tonal stimuli around 11°. The low-pass filtered noise tended to be located between -10° and -3°. These results are somewhat consistent with Blauert's result, but they cannot be directly compared since the elevation resolution imposed by Roffler & Butler [1968a] can be entirely generalized as belonging to the *frontal* directions of Blauert [1969/70]. Roffler & Butler [1968b] reported, in a similar experiment as that of Roffler & Butler [1968a], the localization of nine tone bursts ranging from frequencies 250 Hz to 7.2 kHz. Perhaps as an antecedent to Blauert [1969/70], they also found that tones were not located according to their actual position but to their spectral content.

The role of the high frequency range in elevation perception was also investigated by Gardner & Gardner [1973], who tested sound source identification performance in various planes around subjects. Of interest to this Chapter are their results in the MSP, with the pinnae cavities with and without occlusion. Their setup consisted of 9 loudspeakers of similarly flat frequency response, placed from elevations -18° to 18°, in 4.5° steps. Stimuli were samples of filtered broadband noise: full bandwidth and $\frac{1}{2}$ -octave narrow bands centered at 2, 3, 4, 6, 8 and 10 kHz. For 2 participant subjects, they reported that the more the pinnae cavities were occluded, the more the identification ability decreased. They also found that identification ability increased if the stimuli presented high frequency content. The best condition was always for broadband stimuli, regardless the degree of pinnae cavity occlusion. Without occlusion, sound source identification of narrow band noise samples centered at 8 and 10 kHz was equivalent to that of broadband noise. The trends were maintained for a similar experiment that tested the rear hemisphere, with the difference that identification was, in general terms, poorer than for the front hemisphere. The findings of Gardner & Gardner [1973] agree in some extent with those of Blauert [1969/70] and Hebrank & Wright [1974b], since the frequency range from 4 kHz to 12 kHz would be sufficient to let the pinnae filter and cue directions to the front and to the back. Furthermore, a $\frac{1}{2}$ -octave narrow band signal could be broad enough to allow the pinnae filter the slopes and center frequency of the frontal notch mentioned by Hebrank & Wright [1974b], even though the latter stated that the notch found was 1-octave broad. A later experiment reported by Gardner [1973] extended the results of Gardner & Gardner [1973]. He conducted a real sound sources identification listening test. He used the same setup of 9 loudspeakers in the frontal hemisphere of the MSP and the same test signals as in Gardner & Gardner [1973]. Different levels of pinnae occlusion were tested: no occlusion of pinnae cavities, occlusion of only one pinna cavity, occlusion of both pinnae cavities. The general trend was that performance was best without pinnae occlusion and worst with both pinnae occluded. As reported by Gardner & Gardner [1973], sound source identification was best with the full broadband stimulus and, for the narrow band stimuli, performance was more degraded as the center frequency was lower. Identification was severely impaired in all conditions for the narrow band signal centered at 2 kHz. They concluded that the cues provided by the cavities of the pinnae in the MSP were largely monaural, even though a binaural interaction was needed for optimum results. Regarding the spectral content required for sound source identification in the MSP, the results reported are similar to the findings reported by Gardner & Gardner [1973].

In a similar line of work, Butler & Planert [1976] reported a sound source identification experiment with real sound sources in the MSP and a signal of varying bandwidth as stimulus. Five loudspeakers were positioned from -30° to 30° in 15° steps. The baseline condition was broadband noise, and the band limited signals were centered at 8 kHz, with a bandwidth changing from 1 kHz to 6 kHz. A similar experiment was done where the left ear of the listeners was occluded. Pooled results showed good identification performance for the binaural condition for signals of bandwidth 4, 5 and 6 kHz. The best performance was for the broadband stimulus. The monaural condition was impared in comparison to the binaural condition, and good results were achieved only with the broadband stimulus. Individual analysis of the results showed strong differences across subjects, for example 1 subject out of the 7 participants was able to identify sound sources with good performance when the signals had a bandwidth of 1 kHz or broader, and another subject was able to identify sounds with a bandwidth of 2 kHz or more. It is difficult to frame these results in terms of the findings reported by Blauert [1969/70] and Hebrank & Wright [1974b], since frontal directions would not be cued by a band of noise centered at 8 kHz according to those results. It is believed that the paradigm of sound source identification strongly affected the results, according to a comparative literature review done in the topic of response paradigms in sound perception studies and not included in this Thesis[2]. Moreover, there is no report on loudspeaker equalization in Butler & Planert [1976], for which it is possible that subjects had other cues to identify sound sources. For a later study reported by Butler & Belendiuk [1977] an experiment

[2] See, for example, Perrett & Noble [1995].

was conducted where sound source identification performance in the MSP under different conditions was tested: real life, individual and non-individual binaural recording conditions, being the latter reproduced through headphones. They also analyzed $\frac{1}{3}$ -octave band measurements from the HRTFs of the subjects, in search of spectral cues to elevation. They tested 5 sources, from -30° to 30° in 15° steps. Test signals were trains of broadband noise bursts. Performance in free field and with individual recordings were equivalent for 6 out of 8 subjects. From the $\frac{1}{3}$ -octave band -centered at 4 *kHz*, 5 *kHz*, 6.3 *kHz*, 8 *kHz* and 10 *kHz*- spectral analysis, they concluded that a notch that moved up in frequency and became narrower in bandwidth as the sound source was moving up in elevation, coded localization in the frontal region of the MSP. The notch, however, was not evident above the horizontal plane in their figures, suggesting that it was a low-frontal cue. By testing 4 subjects with non-individual binaural recordings, they found that subjects did not always performed best with their own recordings. Furthermore, there were two particular sets of recordings with which all subjects performed best, and 1 set of recordings with which all subjects performed at chance level. Their results supporting the hypothesis of a notch with moving fc as elevation moves agrees with the results of Hebrank & Wright [1974b]. It has to be noted, however, that their broadband stimuli was restricted by the microphone frequency response -which started to roll-off around 9 *kHz*-, and that the authors did not report on equalization procedures for the headphones. It is unclear whether this could have affected the results. Unlike in Butler & Planert [1976], the authors reported choosing the loudspeakers so that they all had similar frequency responses.

The relevance of notches as cues for elevation has also been explored and reported by Bloom [1977]. Even though his experiment tested the perception of sounds only in the frontal plane (90° azimuth), his findings are important since they suggested that a certain elevation could be evoked by manipulating spectral cues. He concentrated on the dips of simplified versions of HRTFs and tried to emulate them by filtering a 1 -octave random noise sample centered at 8 *kHz* with a notch filter with varying f_c. A preliminar experiment showed that, as the center frequency of the notch was increased from 6.3 *kHz* to 10 *kHz*, the perceived elevation went from -30° to $+45^\circ$.

The candidacy of notches as the primary cue to elevation is, however, controversial. Humanski & Butler [1988] studied localization performance with real sources at three different azimuth angles outside the MSP, for elevations ranging from $+30^\circ$ to -30°. The study is interesting despite not testing the MSP since they applied the concept of overt and covert characteristics. Overt features were those that could be identified in a single HRTFs signal as its maximum or minimum. Covert features were obtained from comparison of HRTFs across directions. They analyzed the quality of covert and overt peaks and dips (their uniqueness to be contained in a single direction) and their relationship to localization of high-pass filtered noise with cut-off frequency at 4.5 *kHz*. They found that, for the ipsilateral ear, covert peaks and dips and overt dips accounted well for localization. However, they suggested that covert peaks were a more robust cue since it was the only one that contributed to sound localization at the contralateral ear when the ipsilateral ear was occluded. Even though the works cannot be directly compared due to the azimuth tested, these results would agree with the findings of Blauert [1969/70] as the latter was advocate for increased energy in certain frequency bands as cue for elevation. Another study that does not support the idea of notches as primary cue to elevation is that reported by Macpherson [1994]. He analyzed the primary high frequency notch patterns of 6 subjects and tried to predict perception of elevation on the basis of their f_c, in what he called a single-notch model. Several planes were tested, including the MSP. The model could not account for the behavioral data of the subjects in free field condition, i.e. the scatter in elevation judgements and biases in elevation introduced by front-back reversals. It was concluded that more spectral features would act as cues, than notches alone.

The importance of broader ranges of frequencies, covering more than one spectral feature, as cue to localization in the elevation dimension was studied by several authors. Asano *et al.* [1990] tested sound source localization with individual and non-individual HRTFs in binaural synthesis, for directions in the MSP. The HRTFs were simplified in different ranges of their spectrum by changing the order of the ARMA model used to construct the filters. The stimuli was broadband noise. They found that when the frequencies above 5 *kHz* were modeled with the lowest order tested, the elevation component of localization was equivalent to that with the original measured HRTFs, i.e. elevation was cued by macroscopic spectral features above 5 *kHz*. They hypothesized that the power in the range from 5 *kHz* to 10 *kHz*, compared to the power in neighboring ranges, was a likely cue. A second experiment showed that the critical boundary for modeling HRTFs without compromising front-back confusion rate lied around 2 *kHz*, meaning that the microscopic characteristics of the lower frequencies helped discerning front from back directions -along with the macroscopic features above 5 *kHz*. These findings are interesting in the light of the results presented by Carlile & Pralong [1994]. They measured HRTFs from human subjects in order to identify spectral characteristics that systematically changed with direction and provided perceptual cues for localization. They filtered the HRTFs with an auditory filter model in order to account for the audiometric sensitivity[3] and the spectral smoothing effects of the cochlea, which eliminated sharp notches in the mid and high frequency ranges. They worked with several locations in the sphere, being their findings on the anterior (frontal) MSP relevant for this Chapter. It was reported that the steep roll-off of the mid frequency gain in the range

[3] They used the minimum audible field (MAF), which is determined by the minimum detectable pressure level for a free field stimulus at the position of the subject's head.

from 2 to 5 kHz was in general smoothed by the auditory filters. Variation in elevation from -45° to 60° resulted in a systematic increase of frequency in the high frequency roll-off of the principal gain feature from around 3 to 6 kHz. The notch that otherwise characterized the range around 8 kHz was smoothed for elevations below the horizontal plane.

The idea of different cues used for decoding different directions, as suggested by Asano *et al.* [1990] and already mentioned, was also proposed by Han [1994]. He reported measurements on KEMAR and a characterization of how HRIRs and HRTFs changed with azimuth and elevation. Several planes were measured, including the MSP for which the following features were proposed: from -60° to -30°, the low frequency slope of a notch changed from 6 kHz to 7 kHz, respectively; from -30° to 0° the low frequency slope of a double notch feature changed from 7 kHz to 8 kHz; from 10° to 80° the notch became less deep and its f_c remained around 9.5 kHz; directions above were characterized by a level increase. It was concluded that the low frequency slope of the notch centered between 6 kHz and 13 kHz was the main cue to elevation, and that secondary cues existed which resolved ambiguities. A degree of correspondence with the findings of Blauert [1969/70] and Hebrank & Wright [1974b] was reported.

Another study covering a larger range of spectral cues was that reported by Langendijk & Bronkhorst [2002]. A binaural synthesis localization performance experiment was conducted where spectral cues were removed from the individual directional transfer functions[4] from the HRTFs and the signals were presented through headphones. Gaussian noise samples (200 Hz to 16 kHz) were filtered with HRTFs from which cues in specific frequency ranges had been removed -i.e. the spectrum was set to its average within the band in question-. The bandwidth of the removed frequency ranges were from 2 -octave to $\frac{1}{2}$ -octave bands, spanning from 4 kHz to 16 kHz. Of interest to this Chapter are their results for the MSP. It was suggested that the frequencies that more prominently cued front-back directions in the MSP lied in the range from 8 kHz to 16 kHz -however, removing cues below 8 kHz also generated front-back reversals. Up-down cues seemed to be located in the range from 5.7 kHz to 11.3 kHz. Other observation made by the authors was that frontal directions contained a peak in the range from 8 kHz to 16 kHz that was not present in rear directions. The idea that elevation was cued by a notch in the range 5.7 kHz to 11.3 kHz (with increasing f_c as elevation increased) was not supported by their experimental results, since removing that frequency band did not affect localization -results that agreed with those of Macpherson [1994]. It was suggested that, in general, spectral features that cued sound localization were broader than $\frac{1}{2}$ -octave. The results are in agreement with Humanski & Butler [1988], that stated that

[4] Directional transfer functions are defined as HRTFs without the non-directional components such as ear canal resonance.

notches were not as robust cues as peaks. The results also agree, partially, with Hebrank & Wright [1974b] since they reported that increased energy in the high frequency range (above 8 *kHz*) was needed to cue frontal directions. They are also in agreement with Middlebrooks [1992] in identifying the high frequency range as important for elevation perception. In any case, according to Carlile *et al.* [1999] the spectral cues included in the full-range HRTFs would be redundant. Carlile *et al.* [1999] reported a real sound sources localization experiment where the test signals covered three alternatives: broadband noise, low-pass filtered noise (with cut-off at 2 *kHz*) and high-pass filtered noise (with cut-off at 2 *kHz*). They pooled the results across subjects from 76 locations in the sphere surrounding them, for which no separate conclusions for the MSP were made. Even though their main goal was to test the use of low frequency differences as cues, it is relevant to report their finding that localization with the high-pass filtered noise was comparable to that with broadband noise, suggesting that the cues over the whole range of frequencies were redundant -i.e. when the low frequency range was not present, high frequency cues could be used to supplement ITD cues. With the low-pass filtered noise, however, the rate of front-back confusions increased substantially. The importance of the high frequency range was also supported by Algazi *et al.* [2001b]. They tested sound source localization with binaural signals synthesized with individual HRTFs and reproduced through headphones. Test signals were either low-pass filtered (with cut-off at 3 *kHz*) or full bandwidth (incorporating frequencies up to 22 *kHz*) Gaussian noise modulated bursts. They tested several planes, including the MSP which is of interest to this Chapter. They found that localization with the low-pass filtered signal was severely degraded in the MSP, and that subjects were guessing their answers -evidence of low frequency cues for elevation were seen but in planes far from the MSP.

Regarding the likely bandwidth or scale of the relevant cues to sound localization (Langendijk & Bronkhorst [2002] suggested broader than $\frac{1}{2}$ -octave), it is relevant to mention the study reported by Macpherson & Middlebrooks [2003]. They conducted a free field localization experiment where stimuli were noise bursts with log-ripple spectra of varying ripple density. They focused on the density of ripples that would obscure the spectral cues inherent to the directional transfer functions causing localization errors. They tested several vertical planes, including the MSP. In the latter, they tested elevation angles from -60° to 60°. Even though there were differences among subjects regarding error rate and ripple parameters, the highest and most consistent increase in elevation errors was seen for ripple densities from 0.5 to 2 ripples/octave (pooled results across directions), being the worst case that of 1 ripple/octave. For this latter case, it was also found that ripple depths larger than 20 *dB* also caused increased elevation errors, suggesting that the depth had to exceed the magnitude level differences seen in spectral features from directional transfer functions.

The need of similarity between the signal reaching the ears and the particular spectrum of the listener's own HRTFs in order to evoke a direction, which is implicit in the concept of directional bands, was further tested by Middlebrooks [1992]. He studied the localization of $\frac{1}{6}$ octave narrow band noise with f_c at 6 kHz, 8 kHz, 10 kHz and 12 kHz and the relationship with the spectral features of the external ear. He presented the sound from a loudspeaker that could be positioned in one of 66 locations around the subjects. Of interest to the present work was his comparison of the directional filtering provided by the ear to the signal reaching the ear when sound came from actual and reported locations. It was shown that a given narrow band spectrum was localized in elevation according to the listener's directional filter that could better account for the combination of stimulus and actual directional filter being used -being the latter the one corresponding to the actual source location. In other words, subjects localized largely independently from the actual source location, reporting elevation answers constricted to particular ranges. These results can be considered as an extension of those of Blauert [1969/70], but they did not provide evidence that peaks were the relevant cues to sound localization: they rather suggested that the tested bands and how they were filtered by the HRTFs were relevant in sound localization. On the other hand, it is important to note that the results of the experiments presented by Middlebrooks [1992] were largely individual for every subject, as Figure 4 from that study showed for the elevation component[5]. The author hypothesized, following some previous work, that the trends found related to the height of the subjects. Another relevant study that explored the idea of spectral similarities was reported by Wightman & Kistler [1993]. They conducted a feature analysis on 15 subjects with high proficiency localizing virtual sound sources. Through multidimensional scaling, they identified HRTFs sets that were similar and used them in binaural synthesis localization experiments to test how subjects performed with individual, similar non-individual, and dissimilar non-indiviudal HRTFs. Each set used in the analysis consisted of HRTFs from 265 directions, which were simplified by means of principal component analysis. Even though the results were pooled for several directions, the findings are interesting since it was reported that listeners performed equally with individual and similar non-individual HRTFs. Elevation judgements were seen to be less accurate and the rate of front-back confusion increased when dissimilar non-individual HRTFs were used.

Despite the extensive work towards identifying specific frequency ranges and spectral cues to sound localization, little research has been done towards parameterizing the relevant spectral features. An important work in that direction is that of Iida *et al.* [2007], who parameterized peaks and notches of individually measured HRTFs from human subjects, in terms of center frequency, level and sharpness. By modeling simplified

[5] It has to be noted that Figure 4 in Middlebrooks [1992] corresponds to the pooled data of the experiment across directions, not only for sound sources in the MSP.

HRTFs versions from combinations of these features and presenting them in a binaural synthesis localization experiments, they tested their contribution as localization cues in the upper MSP, from 0° to 180° in 30° steps. They extracted peaks and notches above 4 *kHz*, from smoothed representations of the measured HRTFs. Outside the range of the spectral features, the magnitude response of the modeled HRTFs was flat. They concluded that the first two notches and the first peak of the HRTFs could contribute to elevation perception. The findings are in agreement with those reported by Hebrank & Wright [1974b]: the first and second notch would correspond to those that evoked front and back directions. According to Iida *et al.* [2007], the second peak would correspond to the above cue reported by Hebrank & Wright [1974b]. They also hypothesized that, as the first peak did not change with direction, it would be used as a reference to analyze further spectral cues at higher frequencies -this is in the line of thought proposed by Asano *et al.* [1990], of a power comparison between the ranges above and below 2 *kHz*, approximately, for front-back disambiguation.

Summarizing and crossing the findings of the works reviewed above, the following can be suggested:

- Candidate cues for localization in the elevation dimension in the MSP would be coded among the first peak and first two notches of the HRTFs.
- In general, these candidate relevant spectral features would be broader than $\frac{1}{2}$ -octave, they would possibly be 1-octave wide.
- The macroscopic features in the frequency range from 8 *kHz* to 12 *kHz* would be necessary for localization in the frontal elevation. An important feature could be a 1-octave notch (candidate parameters that would act as cues are f_c, low frequency slopes, bandwidth), but it is unlikely that it would be the main and only relevant feature. Peaks would be more salient, and the increased energy in the higher end of the range would be relevant.
- The general features of a $\frac{1}{4}$ -octave peak (possibly of wider bandwidth, according to the second point of this list) between 7 *kHz* and 9 *kHz*, could act as cue for the above directions.
- A small peak between 10 *kHz* and 12 *kHz*, with decreased energy above and below the peak, is suggested as a cue for behind.
- Front-back perception would be disambiguated by the band around 1 *kHz* - 2 *kHz* and the macroscopic features in the range from 8 *kHz* to 16 *kHz*.

These points were used to narrow down the spectral features parameterization conducted as part of the study presented in this Chapter, as will be covered in 5.3.2.

5.3 Methods

The study presented here comprised two main work packages. The first one consisted of conducting psychoacoustic experiments with stimuli binaurally synthesized with individual and non-individual HRTFs, and correlating the behavioral results in order to obtain groups of HRTFs that evoked similar sound source directions. The methods for this work package are described in *Experimental methods for psychoacoustic testing*. The second work package consisted of analyzing the previously mentioned groups in search of relevant spectral cues and the parameters describing them. The methods for this work package are described in *Parameterization methods*.

5.3.1 Experimental methods for psychoacoustic testing

An experiment was conducted where individual and non-individual HRTFs that evoked similar sound source directions were obtained, for a small number of subjects and individually for each of them. This means that the original directions for which non-individual HRTFs were measured were not relevant, and the HRTFs were only associated with the direction they evoked. Evoked directions were obtained by applying spherical statistical techniques to the results of the localization experiments.

The nature of the study required subjects that could perform equivalently well with real sound sources under anechoic conditions and with virtual sound sources synthesized with individual HRTFs -more about the assumption of *correct* localization performance within binaural synthesis will be discussed in Chapter 6. Due to the performance requirement, the listening experiments tested localization under three conditions: real sound sources, binaural synthesis with individual HRTFs and binaural synthesis with non-individual HRTFs. The testing environment was an anechoic chamber in the case of real sound sources presentations and a sound insulated booth in the case of binaural synthesis presentations.

Subjects

Ten paid subjects with normal hearing participated in the experiment. Their hearing thresholds were determined by means of a standard pure-tone audiometry in the frequency range from 250 *Hz* to 8 *kHz*. None of the subjects had hearing thresholds above 15 *dB HL*. The 10 participant subjects were allocated to two groups: 5 subjects in Group A and the remaining 5 subjects in Group B. These groups differed slightly in the answer procedure, awareness of sound externalization and number of repetitions, as it will be explained and discussed later in this Chapter.

Some of the listeners had participated in listening tests before, and one of them participated also in the pilot test for this experiment. Most of the listeners were unfamiliar with the experimental procedure and were considered naïve for the purpose of this experiment. Participation of naïve and experienced subjects was allowed due to the difficulty in finding subjects that could pass the audiometric test and meet the localization performance requirement.

HRTFs measurement

HRTFs from all the participant subjects were measured in an anechoic chamber as previously reported in Chapter 2, Sec. 2.8.1. Measured HRTFs from the participants are among those ones seen in Fig. 2.14, Chapter 2. The measurement procedure was described in that Chapter and will not be repeated here. Measurements were done for 15 directions in the MSP with elevations from -67.5° to 247.5° in 22.5° steps. See Chapter 1, 1.5 for more on the coordinate system used in this Chapter.

Selection of non-individual HRTFs

Non-individual HRTFs used in the experiments presented here were selected from the large database of measured HRTFs (Møller *et al.* [1995a]) already used in Chapters 3 and 4. Møller *et al.* [1996b] reported experiments where these HRTFs were used for non-individual HRTFs binaural testing with a group of 20 listeners (different from the participants in the experiment reported here). Figure 2 in Møller *et al.* [1996b] shows the rate of median plane localization errors associated to each set of HRTFs from the database. The 10 sets that produced less median plane localization errors were selected, and arbitrarily re-named from *Non-individual 1* to *Non-individual 10*. Each HRTFs set consisted of 15 pairs of signals measured from elevation -67.5° to 247.5° in 22.5° steps. These 10 sets of HRTFs were tested with all subjects under the non-individually synthesized condition.

Real sound sources condition

This condition was used as part of the control, which ensured that the localization performance with real sound sources was equivalent to that with individual HRTFs binaurally synthesized sounds. It has to be noted that the results from the real sound sources condition were not used for the parameterization procedures that will be explained later in this Chapter. In the following, the methods used in the real sound sources condition are explained.

Loudspeaker setup. The localization experiment with real sound sources was conducted in the same anechoic chamber where the HRTFs measurements took place, with the same loudspeaker setup as described for the HRTFs measurements in Chapter 2.

Signal generation and control. The equipment was placed in a control room next to the anechoic chamber. Signals were played back through a PC equipped with a digital sound card RME HDSP 9632 connected to an external AD/DA converter RME ADI-8 DS. The signals fed a power amplifier Pioneer A-616 modified to provide 0 *dB* gain. The output of the amplifier was connected to a custom made switch, controlled through the PC, which fed the signal to the corresponding loudspeaker.

Stimuli. The stimulus was broadband noise of 1 *s* length, equalized to account for the frequency response of the corresponding loudspeaker from which it was being played back. The transfer characteristic of each loudspeaker was measured in order to build minimum phase inverse filters which were applied to the signals offline. In this way, a specific signal was delivered to each loudspeaker ensuring that the sound being reproduced was spectrally flat and only the filtering imposed by the anthropometry of the listener was taking place during the experiment. Onset and offset ramps of 50 *ms* length were also applied to the signals. The overall gain was calibrated so that the free field sound pressure at the position of the center of the head, with the listener absent and for every loudspeaker, was 57 *dB SPL*. The main restriction encountered to set the reproduction level was the voltage to be fed to the loudspeakers, which could not exceed 0.75 $Vrms$.

Procedure - Group A. Signals coming from each of the 15 loudspeakers were repeated 15 times. This gave a total of 225 presentations which were randomized and presented in 3 blocks of 75 presentations each.

Subjects were familiar with the setup and the anechoic chamber, since the experiment took place after the HRTFs measurements. Besides, neither during the measurements nor the experiment were the subjects blindfolded or the setup hidden. This could have biased the subjects in their responses, even though they were clearly instructed not to identify sound sources but to make absolute localization judgements. Since localization of real sound sources was a control condition to assess the ability of subjects to localize sound, and the statistical procedures comprised responses to binaurally synthesized stimuli only, it was considered that eventual biases would be minimal.

Before the experiment, subjects were given written instructions about the task and how to enter their answers in a touchscreen. An absolute localization response paradigm was used. The task given to the subjects was to report the direction in space from which they perceived the sound as coming from. After each presentation, subjects were instructed to answer the question *'Did you perceive the sound outside your head?'*, with YES and NO as possible answers. This question was included to explicitly assess externalization. Subjects were further instructed to point inside the head in the touchscreen if they perceived the sound inside the head. After reading the instructions, subjects were taken to the anechoic chamber and proceeded with a training session to familiarize themselves with the task. The length of the training session varied from subject to subject, but in general one session with 15 signals randomly presented was sufficient. Once the experimenter was sure that the task had been understood, subjects could keep still in the same position to listen to the sounds and they could master the use of the touchscreen, the experiment proper began. The position of the subjects, who were standing up, was monitored at all times by two cameras: the first one capturing the subjects from the side and the second one capturing the subjects and the loudspeaker arc from behind. This control ensured that the subjects kept still in the same position during the sounds being played back. Subjects interacted with the screen showed in Figure 5.1 to enter their localization answers, where they had to report the elevation angle (left side of the screen) and the azimuth angle (right side of the screen). Subjects could correct their answers as many times as they wanted before proceeding to the next presentation. Even though all sounds came from directions in the MSP, a lateral effect was perceived in some cases. The graphics in which the subjects had to give their angular answers presented parallels and meridians, as seen in Figure 5.1. The intersection of these did not correspond to the actual location of the loudspeakers, but were provided to ease the construction of reference points by the subjects. Subjects were given breaks between blocks, in which they were required to leave the anechoic chamber. Feedback was not provided at any stage of the experiment. Typically, subjects completed a block within 15-17 minutes.

Some aspects of the methodology, like the question about externalization or reporting the perceived azimuth of the sounds, were not relevant in the real sound sources condition -with the exception of some situations which are reported in Chapter 6. However, these aspects of the methodology were included in this condition so as to keep consistency with the later parts of the experiment, namely the individual and non-individual HRTFs binaural conditions. In these latter cases, issues about externalization and perception to the sides were relevant.

Procedure - Group B. Signals coming from each of the 15 loudspeakers were repeated 16 times. This gave a total of 240 presentations in 4 blocks of 60 each.

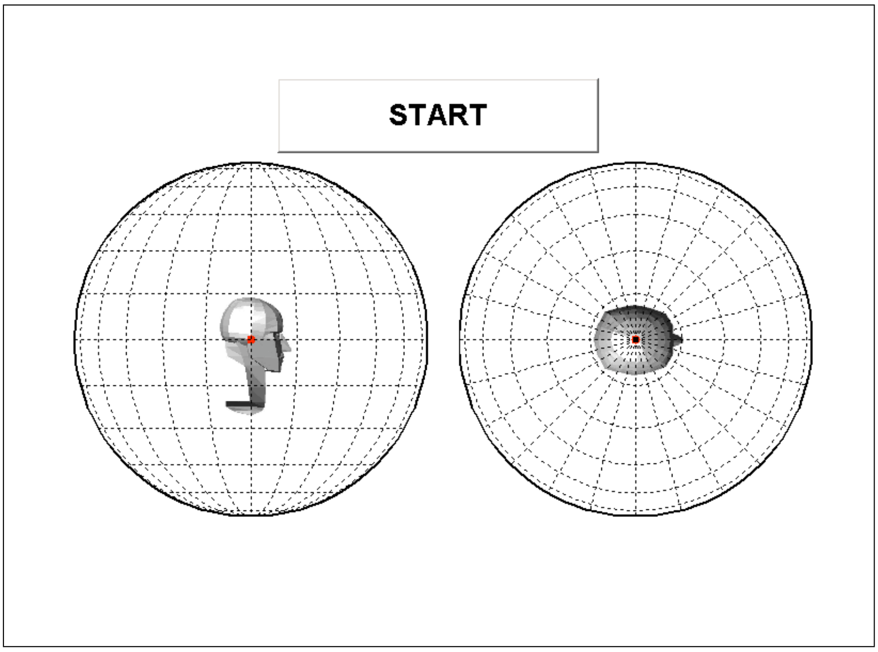

Figure 5.1: Graphical interface of the touchscreen presented to subjects in Group A.

The procedure for this group was similar to that already described for Group A, but there were some variations which will be explained in the following. Firstly, subjects did not have to answer the question about externalization. On the basis of the screening that was done to potential subjects and the results from Group A, it was hypothesized that a lack of externalization correlated with a poor localization performance and the question was no longer required. More on this will be discussed in 5.5. Furthermore, subjects were presented with the screen showed in Figure 5.2 to enter their localization answers, in which neither parallels nor meridians were given, only a pair of orthogonal axis as a guideline. This change was introduced to avoid any possible bias determined by the parallels and meridians -more about the different graphical interfaces used for Groups A and B will be covered in the Discussion. Subjects were instructed to report their answers inside the head if they could not localize the sound source and they had to guess their answer. As subjects in Group B were presented with shorter blocks than Group A, they typically completed a block within 10-12 minutes.

Binaural synthesis conditions

Experiments under binaural synthesis conditions are the core of the study presented in this Chapter, as the spectral features identification procedures were based on the results

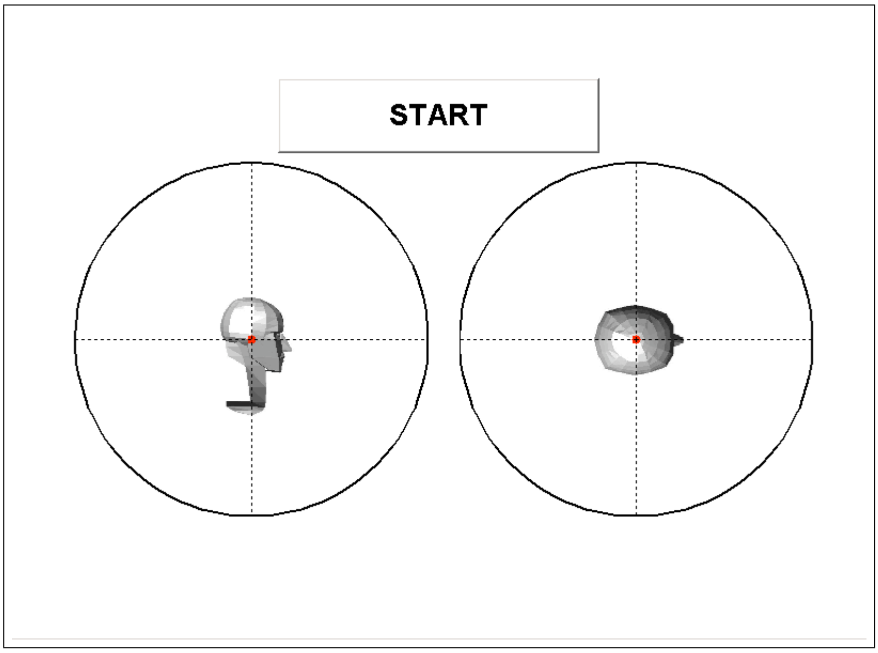

Figure 5.2: Graphical interface of the touchscreen presented to subjects in Group B

of these experiments. In the following, the methods under this condition are explained.

Headphones setup. The experiments under binaural synthesis conditions, both with individual and non-individual HRTFs, were conducted in a sound insulated listening cabin. Inside the cabin, subjects were seated comfortably in front of a touchscreen that had the same features and graphical interface as in the real sound sources condition. The only other device inside the cabin was a pair of individually equalized headphones Beyerdynamic DT990. Typical transfer functions for these headphones have been shown by Møller *et al.* [1995b] and Wightman & Kistler [2005]. The equalization filters were designed as the inverse of the average of five repeated frequency response measurements done for each particular subject. More about the equalization procedure can be found in Chapter 4, 4.4.8.

Signal generation and control. The equipment was placed in a control room close to the sound insulated cabin. Signals were played back through a PC with a digital sound card RME HDSP 9632 connected to an external AD/DA converter RME ADI-8 DS. The signals fed a power amplifier Pioneer A-616 modified to provide 0 *dB* gain, which fed the headphones.

Stimuli. The stimulus was broadband noise of 1 *s* length with onset and offset ramps of 50 *ms*. The stimulus was filtered offline with the appropriate HRTFs -individual or non-individual, according to the condition- and further processed with individual filters to equalize the headphones response. The overall gain of the system was calibrated so that the unprocessed broadband noise was reproduced at a level equivalent to, approximately, 76 *dB SPL* in free field.

Procedure - Group A. The procedure was similar to that reported for the real sound sources localization condition. Subjects were not familiar with the sound insulated cabin, and they ran an individual HRTFs binaural synthesis practice block before the experiment proper began. They were not given written instructions, as the task and procedure were the same as in the real sound sources condition. For each virtual source, 15 repetitions were presented. After the practice session, the 3 blocks of 75 presentations each which corresponded to the individual HRTFs condition were tested. Once the performance had been verified as equivalent to real sound sources localization (the procedure is explained later in this Chapter), subjects proceeded with the non-individual HRTFs condition. To avoid familiarization with a given set of HRTFs -i.e. belonging to a particular subject- the 30 blocks from all non-individual HRTFs presentations (3 blocks of 75 presentations each, for each of the 10 non-individual HRTFs sets) were presented in a randomized order. Each block contained stimuli which had been filtered with non-individual HRTFs belonging to only 1 set. This means that the level relationships among HRTFs from every single non-individual set were kept.

Procedure - Group B. The procedure was similar to that used for real sound sources for Group B, with the same considerations mentioned for Group A regarding differences between real sound sources and binaural synthesis conditions.

Statistical methods

Due to the nature of the responses, which were obtained through an absolute localization response paradigm, results were analyzed with spherical statistics in order to assess whether they evoked similar directions. The theory behind these methods has already been described by Fisher *et al.* [1987], Wightman & Kistler [1989b] and Leong & Carlile [1998], among others. This theory will be briefly presented here for the sake of completeness.

Since reporting distance perception was allowed but neither encouraged nor required in the experiment, only the angles were taken into account for the analysis. As already explained, subjects were asked to give their answers inside the heads if they felt so (Group A) or if they had to guess their answer (Group B). These responses were also included in the statistical analysis - no answer was discarded. All responses were considered with their actual angles, even if they showed evidence of front-back confusion.

As an overview, it can be said that localization results were firstly analyzed with descriptive statistics for vectors in 3 dimensional space. Descriptives in spherical statistics are quantitative -such as centroid, dispersion and rotation matrix, among others- and qualitative -unimodal vs. multimodal distribution, or isotropic vs. anisotropic distribution, among others. Secondly, parameters were estimated according to the distribution model to which the data adjusted. Available analysis procedures vary according to the model chosen and the type of test to be performed. The classical concept of localization error[6] was not relevant in the context of this experiment. The characteristics of the distribution for the evoked localization responses were of interest, and how they matched other distributions. All these mentioned procedures are explained in detail in the following.

In order to perform a descriptive statistical analysis for vectors in 3 dimensional space, the following analysis protocol was followed for the results of each virtual source:

Computation of cosine angles. The cosine angles were computed for each reported answer. Given an answer in the form of (colatitude θ, longitude ϕ), which for convenience is corresponded to (azimuth θ, elevation ϕ), the cosine angles (x, y, z) were given by:

$$\begin{aligned} \sin\theta\cos\phi &= x \\ \sin\theta\sin\phi &= y \\ \cos\theta &= z \end{aligned} \tag{5.1}$$

Computation of the resultant length for each sound source. From the summation of the cosine angles of all n answers for a given sound source, the resultant length R could be obtained:

$$S_x = \sum_{i=1}^{n}(\hat{x}_i), S_y = \sum_{i=1}^{n}(\hat{y}_i), S_x = \sum_{i=1}^{n}(\hat{z}_i) \tag{5.2}$$

$$R = \sqrt{S_x^2 + S_y^2 + S_z^2} \tag{5.3}$$

The resultant had direction cosines $(\hat{x}, \hat{y}, \hat{z})$ that were obtained from the resultant length:

$$(\hat{x}, \hat{y}, \hat{z}) = (S_x/R, S_y/R, S_z/R) \tag{5.4}$$

[6] This is described at the angular deviation of the perceived direction from the target or original direction -i.e. the direction for which the HRTFs were measured.

Computation of the centroid for each sound source. Once $(\hat{x}, \hat{y}, \hat{z})$ were known, the mean direction or centroid $(\hat{\theta}, \hat{\phi})$ of the distribution was obtained by:

$$\hat{\theta} = \arccos(\hat{z})$$
$$\hat{\phi} = \arctan(\hat{y}/\hat{x}) \qquad (5.5)$$

Computation of the mean resultant length for each sound source. The mean resultant length $\overline{R} = R/n$ gave an impression of how clustered around the mean direction the localization answers were. In the case of unimodal distributions, small values of $\overline{R}$ indicated scattered answers, while large $\overline{R}$ values indicated concentrated answers.

Testing for the hypothesis of a uniform distribution. It was relevant to know whether a given distribution was uniform or unimodal. In the first case, answers were uniformly distributed in the spherical space surrounding the subjects - they were distributed in an isotropic fashion. In other words, uniform results showed that the subjects were guessing their answers for that particular sound source. In turn, unimodal distributions were non-isotropic. They indicated clustering of the answers around a mean direction. To test for uniformity against a unimodal alternative, the critical values of R tabulated in Fisher *et al.* [1987] were used. If the value of R was equal or less than that of the tables, it was reasonable to assume that the judgements were isotropic, i.e. they were drawn from a uniform distribution. Uniform or isotropic distributions were not of interest, since they did not evoke a particular direction. Therefore, they were discarded without further analysis.

Estimation of shape and concentration parameters according to a Kent distribution model. Even though the correlation could be made with the already obtained data, it was interesting to use parametric models to estimate different values. From the exploratory analysis, it was decided to apply a Kent (rotational assymetrical or elliptical) distribution model. The procedures for estimating the parameters are given by Fisher *et al.* [1987][7]. The estimated shape parameters obtained were $\hat{\kappa}$ and $\hat{\beta}$. $\hat{\kappa}$ is a concentration parameter that gives an estimation of how clustered the data is about the centroid. Usually, it is used in its inverse form κ^{-1}. β, on the other hand, is an ovalness parameter that results from the ratio of the two density axis of the distribution and indicates if the distribution departs from circular symmetry. Whether a given distribution is unimodal or bimodal, it can be defined from the ratio of $\frac{\kappa}{\beta}$.

Computation of correlation coefficients. Once the main descriptors were obtained,

[7] It has to be noted an errata in later editions of the book.

the correlation among distributions was computed as:

$$\hat{\rho}v = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} \tag{5.6}$$

where

$$\begin{aligned} S_{xy} &= \det\left\{ \sum_{i=1}^{n} (\mathbf{X_i Y_i}) \right\} \\ S_{xx} &= \det\left\{ \sum_{i=1}^{n} (\mathbf{X_i X_i}') \right\} \\ S_{yy} &= \det\left\{ \sum_{i=1}^{n} (\mathbf{Y_i Y_i}') \right\} \end{aligned} \tag{5.7}$$

Test hypothesis. $\hat{\rho}v$ in Eq. 5.6 ranged from -1 to 1, corresponding to complete negative and complete positive correlation, respectively. The null hypothesis was that $\hat{\rho}v = 0$ and there was no correlation between the vectors. Alternative hypotheses that could be tested were: $\hat{\rho}v \neq 0$, $\hat{\rho}v > 0$ and $\hat{\rho}v < 0$. In this work, the last two hypotheses were tested.

To test at $100\alpha\%$ significance level, the null hypothesis was rejected in favor of $\hat{\rho}v > 0$ if $\hat{\rho}v > \hat{\rho}_{\alpha}$. Similarly, the null hypothesis was rejected in favor of $\hat{\rho}v < 0$ if $\hat{\rho}v < \hat{\rho}_{1-\alpha}$.

The critical values $\hat{\rho}_{\alpha}$ and $\hat{\rho}_{1-\alpha}$ were obtained using a permutation test described by Fisher *et al.* [1987]. Briefly, given two sequences $\mathbf{X}$ and $\mathbf{Y}$, the correlation $\hat{\rho}v$ was obtained by computing Eq. 5.6. Afterwards, different pairs of $\mathbf{X}$ and $\mathbf{Y}$ were obtained by permuting the values of $\mathbf{Y}$. For each of the new combinations, a new correlation value $\hat{\rho}v^*$ was obtained. This procedure was to be repeated for all possible samples or, at least, a large number of samples. Once all the correlation values were obtained, they were arranged in increasing order and the critical values $\hat{\rho}_{\alpha}$ and $\hat{\rho}_{1-\alpha}$ were computed. That is, to test at the $100\alpha\%$ significance level, the null hypothesis was rejected in favor of $\hat{\rho}v > \hat{\rho}_{\alpha}$ if the value of $\hat{\rho}v$ was among the $100\alpha\%$ largest values of $\hat{\rho}v^*$ -i.e. $\hat{\rho}v$ was significantly large compared to the other $\hat{\rho}v^*$ values. A similar procedure was conducted to test $\hat{\rho}v < \hat{\rho}_{1-\alpha}$.

In this work, the hypothesis that performance under real sound sources condition was equivalent to that under individual HRTFs binaural synthesis condition was tested by:

- Computation of the centroids of localization judgements with Eq. 5.5, for both conditions.
- Computation of correlation between centroid vectors with Eq. 5.6.
- Computation of critical regions for a 5% significance level -two sided test-, with 10.000 permutations.
- Testing of the alternative hypothesis $\hat{\rho}v > \hat{\rho}_{\alpha}$ for positive $\hat{\rho}v$ values, and testing $\hat{\rho}v < \hat{\rho}_{1-\alpha}$ for negative $\hat{\rho}v$ values.

There is agreement in the literature regarding evidence of high sensitivity to ITD differences in the MSP. That means that even differences in ITD of 1 sample (at $fs = 48\ kHz$) are noticeable, and therefore small errors carried from measurements can have audible consequences[8]: they could be responsible for the perception of a sound source to the left or to the right hemispheres instead of lying precisely on the MSP. The perception of a sound source to the right or to the left are partly responsible for defining whether Eq. 5.6 takes a positive or negative result. On the other hand, the elevation perception is partly responsible for lower or higher correlation values. For instance, the correlation of two vectors of localization responses with positive azimuth values (i.e. to the left of the MSP) would lead to a positive result of Eq. 5.6. However, if the azimuth values of only one vector were forced to take a negative value, the result of Eq. 5.6 would take a negative sign.

Computation of standard deviations along the axes of the distribution. One of the parameters of a Kent distribution is G or the rotation matrix, which is composed by three column vectors that contain the cosine angles of those points which define the mayor and minor axes of the distribution. In this work, the approach of Leong & Carlile [1998] was followed and the standard deviation along these axes was computed. As a result, an ellipse centered on the centroid of the distribution could be drawn.

Matching of HRTFs according to the localization performance associated to them. There were no available spherical statistics methods to compare small samples of Kent distributions. Therefore, the following approach was used: once the standard deviations along the axes of a distribution associated to a pair of HRTFs were computed, the centroids of all other distributions were compared to it. If the centroid of the distribution associated to another pair of HRTFs lied within the standard deviation ellipse of the initial distribution, then they were registered

[8] Jot *et al.* [1995] reported that, at $fs = 50\ kHz$, rounding the ITD to the nearest sample led to a worst case error of around 2.7°, which was smaller than the 3.6° localization blur for frontal directions reported by Blauert [1983].

as matching HRTFs pairs. As mentioned before, isotropic distributions were not considered in the matching procedure and only those non-individual centroids derived from non-isotropic distributions were checked with the standard deviations of the individual distributions. It has to be noted that distributions resulting from non-individual HRTFs were not matched among themselves.

5.3.2 Parameterization methods

It is well known that HRTFs can be modeled as filters with very short impulse responses, as discussed in Chapter 3. After direct sound, the reflections and the diffraction caused by the pinnae take place. Later, reflections and scattering from the head and torso occur. As reported before in this Chapter, previous studies have shown that the reflections and diffraction from the pinnae are sufficient to encode elevation information (Hebrank & Wright [1974b], Gardner & Gardner [1973], Gardner [1973], among others). An extensive pilot investigation was conducted in the first period of this Ph.D. study, in which the pinnae transfer functions measured for 40 subjects (which were measured and reported by Christensen [2001]) were analyzed, together with their role in HRTFs. That study was in the framework of a possible hypothesis that cues to sound localization could be identified in the time domain, and which was not followed. However, that pilot investigation proved useful as it showed that the very first samples of minimum phase HRTFs already presented distinctive characteristics for directions generally defined as front-low, front-high, back-high and back-low, and which matched the characteristics of the pinnae transfer functions in time domain. Furthermore, pinnae transfer functions from directions close to the MSP were seen to contain all the relevant information in the first 20 samples, completely dying out before reaching the 30 samples (with $fs = 48\ kHz$). On the basis of those results, it was decided to use windowed HRTFs for the spectral features parameterization in this Chapter. A half-hanning window of 20 samples length ($fs = 48\ kHz$) was applied to all HRIRs, so that from sample 20 to sample 256 the impulse responses were set to zero.

The following parameters were obtained for all windowed HRTFs (in their log magnitude form) belonging to groups of matching evoked perception:

High frequency slope of the first spectral peak. This was computed as the rate of magnitude change between the frequency component of maximum value of the first peak and the frequency component where the value had decreased by 3 dB, with $f_c < f_{-3\ dB}$. Figure 5.3 shows the points used for the computation. Defining $f_{-3\ dB} = f_H$, this parameter can be expressed as:

$$HFS\_1P = \frac{HRTF(f_C) - HRTF(f_H)}{f_C - f_H}$$

Q-factor of the first spectral peak, computed as the ratio of the center frequency to the bandwidth (defined by those frequencies where the magnitude was -3 *dB* with respect to the peak value), or $Q = \frac{f_c}{bw}$. In some cases, the peak value of the first spectral peak was less than 3 *dB* above the value of the first spectral component. In those cases, the first spectral component was used for computation of the Q-factor. Figure 5.3 shows the points used for the computation. This parameter can be expressed as:

$$Qf\_1P = \frac{f_C}{f_H - f_L}$$

Q-factor of the first spectral peak at global level. This was computed with a modified concept of Q-factor, where the bandwidth was defined by the center frequency of the first notch and the frequency component where the magnitude of the peak had decreased by 3 *dB*, with $f_c > f_{-3\ dB}$. This modified value, which is called *global Q-factor* in the following, accounts for different slopes in the peak. Figure 5.3 shows the points used for the computation. The parameter can be expressed as:

$$GQf\_1P = \frac{f_C}{f_{H-global} - f_{L-global}}$$

Low frequency slope of the first notch. This was computed as the rate of magnitude change between the frequency component with magnitude 3 *dB* higher than the minimum value of the notch and the frequency component of the minimum value of the notch, where $f_{+3\ dB} < f_c$. Figure 5.4 shows the points used for the computation. Defining $f_{+3\ dB} = f_L$, this parameter can be expressed as:

$$LFS\_1N = \frac{HRTF(f_L) - HRTF(f_C)}{f_L - f_C}$$

High frequency slope of the first notch. This was analogous to the low frequency slope, but with $f_{+3\ dB} > f_c$. Figure 5.4 shows the points used for the computation. Defining $f_{+3\ dB} = f_H$, this parameter can be expressed as:

$$HFS\_1N = \frac{HRTF(f_C) - HRTF(f_H)}{f_C - f_H}$$

Q-factor of the first notch at $+3$ *dB* level. This was computed with the same concept of Q-factor as in the analogous parameter for the first peak, but with the $+3$ *dB* values instead of the -3 *dB* ones. Figure 5.4 shows the points used for the computation. This parameter can be expressed as:

$$Qf\_1N = \frac{f_C}{f_H - f_L}$$

Q-factor of the first notch at global level. In this case, the bandwidth was defined between the frequency components of maximum values of the first and second peaks, and the center frequency was that of the first notch. Figure 5.4 shows the points used for the computation. The parameter can be expressed as:

$$GQf\_1N = \frac{f_C}{f_{H-global} - f_{L-global}}$$

Low frequency slope of the second peak. This was computed as the rate of magnitude change between the frequency component that was 3 dB lower than the maximum value of the peak, and the frequency component of the maximum value of the peak. Figure 5.5 shows the points used for the computation. Defining $f_{-3\ dB} = f_L$, this parameter can be expressed as:

$$LFS\_2P = \frac{HRTF(f_L) - HRTF(f_C)}{f_L - f_C}$$

High frequency slope of the second peak. This was analogous to the low frequency slope, but with $f_{-3\ dB} > f_c$. Figure 5.5 shows the points used for the computation. Defining $f_{-3\ dB} = f_H$, this parameter can be expressed as:

$$HFS\_2P = \frac{HRTF(f_C) - HRTF(f_H)}{f_C - f_H}$$

Q-factor of the second peak. This was computed in the same way explained for the first peak. Figure 5.5 shows the points used for the computation. This parameter can be expressed as:

$$Qf\_2P = \frac{f_C}{f_H - f_L}$$

It has to be noted that the frequency resolution of the HRTFs did not allow to identify an exact sample with a value $-3\ dB$ or $+3\ dB$ with respect to another one. Therefore, the sample chosen was always $>= -3\ dB$ or $<= +3\ dB$. Furthermore, in some cases there were minor peaks and or dips. These were considered as actual peaks and dips only if it was possible to identify neighboring ±3 dB points. This approach had a basis on the results shown by Moore *et al.* [1989] -see the literature review for further details.

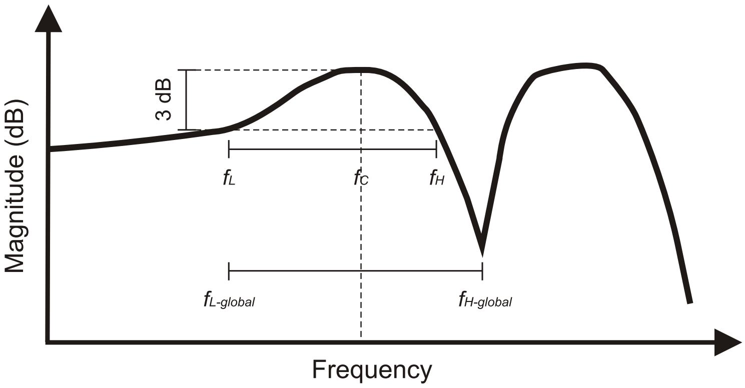

Figure 5.3: Conceptualization of the dimensions used for computing the parameters for the first peak, see text for details.

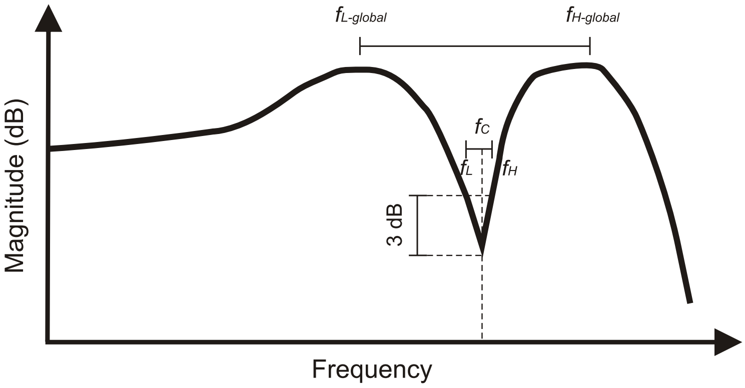

Figure 5.4: Conceptualization of the dimensions used for computing the parameters for the first notch, see text for details.

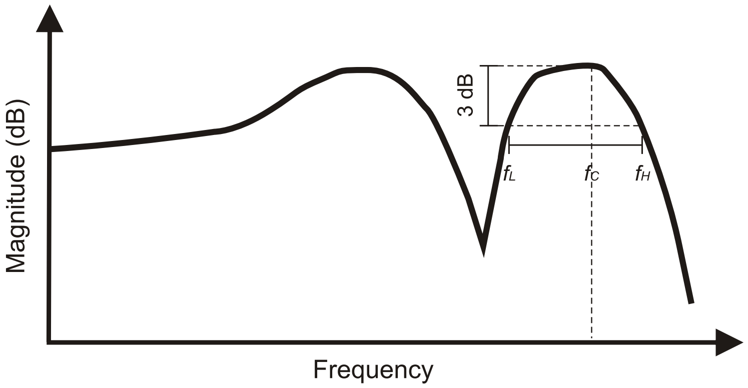

Figure 5.5: Conceptualization of the dimensions used for computing the parameters for the second peak, see text for details.

Another important remark is that some of the HRTFs did not present a notch and second peak, but they followed the shape of a low-pass filter. This could be seen, for example, in those HRTFs measured from the above directions. More about this will be covered in the Discussion.

5.4 Results

5.4.1 Experiment Group A

Extensive results were obtained from the localization experiments, HRTFs matching and parameterization procedures. Therefore, not all of them will be shown in the main text of this Chapter. Examples of results from two subjects will be given for the cases of localization experiments and HRTFs matching procedures, results for the other three subjects can be found in Appendix B. Results from all subjects will be given in the following for the parameterization procedures.

Figures 5.6 and 5.7 show localization judgements for two of the subjects categorized in Group A, for the three conditions tested. Subplots in each panel are labeled according to the corresponding condition: *Real Life*, *Individual* and from *Non-individual 1* to *Non-individual 10*. While the big subplots correspond to the elevation dimension, the small plots contained in them correspond to the azimuth dimension. Subject codes are given in the legend of each Figure panel: MT (Fig. 5.6) and EMS (Fig. 5.7). The localization judgements shown correspond to the raw data, being the original direction (that representing the actual position of the sound sources) in the abscissa and the perceived direction in the ordinate -this applies to both elevation and azimuth plots. The original direction is only given to organize the data and ease its visualization, but it was of no importance to the statistical analysis. The answers plotted in red correspond to those cases in which subjects considered that they perceived the sound inside their heads. Those situations corresponded to either reporting the answer inside the head in the touchscreen -i.e. clicking inside the head- or answering NO to the question *'Did you perceive the sound outside your head?'*. These answers did not receive any special treatment in the statistical analysis or before it, and their rates of occurrence will be covered in the Discussion. The legend of each figure also shows the correlation between real sound sources localization performance and binaural synthesis with individual HRTFs localization performance. Despite the apparent low values of $\hat{\rho}v$, the null hypothesis was rejected in all cases. This topic will also be covered in the Discussion.

Figures 5.8 and 5.9 show matching HRTFs, with the subject codes in the legends, which were obtained as explained in 5.3.1. It has to be noted that some individual HRTFs determined isotropic distributions. In those cases, the individual HRTFs pair was discarded and the subplot in the corresponding figure was left empty -i.e. they were not used in the matching procedures. These cases are marked with the title *'Evoked centroid: Isotropic'*. All the other subplots correspond to one individual HRTFs pair measured for the direction indicated in the title of the subplot, and which evoked a direction that is also given in the title. Graphically superimposed in each subplot are all the non-individual HRTFs that were found as matching according to the localization results.

The results of the parameterization are given in Figures 5.10 to 5.14, with the subject code in the legends. Results for the 5 subjects in Group A are given. Each figure panel is organized so that the first, second and third columns show parameters computed for the first peak, first notch and second peak, respectively. Moreover, the first row of each figure panel shows Q-factors, the second one shows high frequency slopes, the third one shows global Q-factors and the fourth one shows low frequency slopes. The values of these parameters are given in the ordinate of each subplot, while the abscissa always shows the center frequency of the spectral feature. The marker and color codes in Figs. 5.10 to 5.14 follow a division into different regions of the MSP. All circle markers correspond to parameters computed from HRTFs that evoked a direction in the frontal hemisphere, while square markers were computed from HRTFs that evoked a direction in the back hemisphere. The precise elevation span assigned to each marker is shown in the legend of the figures. In general, the colors of the markers in these figures are mirrored with respect to the frontal plane. For example, red circles were computed from HRTFs that evoked directions from -22.5° to 22.5°, while red squares were computed from HRTFs that evoked directions from 157.5° to 202.5°. This latter approach was chosen so as to ease the visual evaluation of front-back relationships.

5.4.2 Experiment Group B

The results for two of the five subjects categorized in Group B are given in the following. Results for the other 3 subjects are found in Appendix B. The figures follow the same directives explained in 5.4.1 for subjects in Group A. Figures 5.15 and 5.16 show localization judgements for the three conditions tested. Subject codes are given in the legend of each Figure panel: ME (Fig. 5.15) and YB (Fig. 5.16). As before, the title of each subplot indicates which condition was being tested and the ten non-individual sets of HRTFs are termed from *Non-individual 1* to *Non-individual 10*. The answers plotted in red correspond to those cases in which subjects reported that they did not know where the sound was coming from -i.e. by clicking inside the head in the touchscreen. Moreover, and as for Group A, correlation values (given in the legends) between real sound sources localization performance and binaural synthesis with individual HRTFs localization performance also show that the null hypothesis of no correlation was rejected in all cases. The reader is referred to 5.4.1 for more details regarding the organization of the figures.

The results of the HRTFs matching procedure for the two subjects are given in Figures 5.17 and 5.18, with the subject codes in the legends. Those results were obtained as explained in 5.4.1 on the basis of the answers shown in Figs. 5.15 and 5.16. Results for the other three subjects of Group B are given in Appendix B. The same explanation given for similar results from Group A are valid, and more information regarding the organization of Figures 5.17 and 5.18 can be found in 5.4.1.

The results of the parameter computations for all subjects in Group B are given in Figures 5.19 to 5.23. These Figures follow the same directives given in 5.4.1 for parameters computed from subjects in Group A.

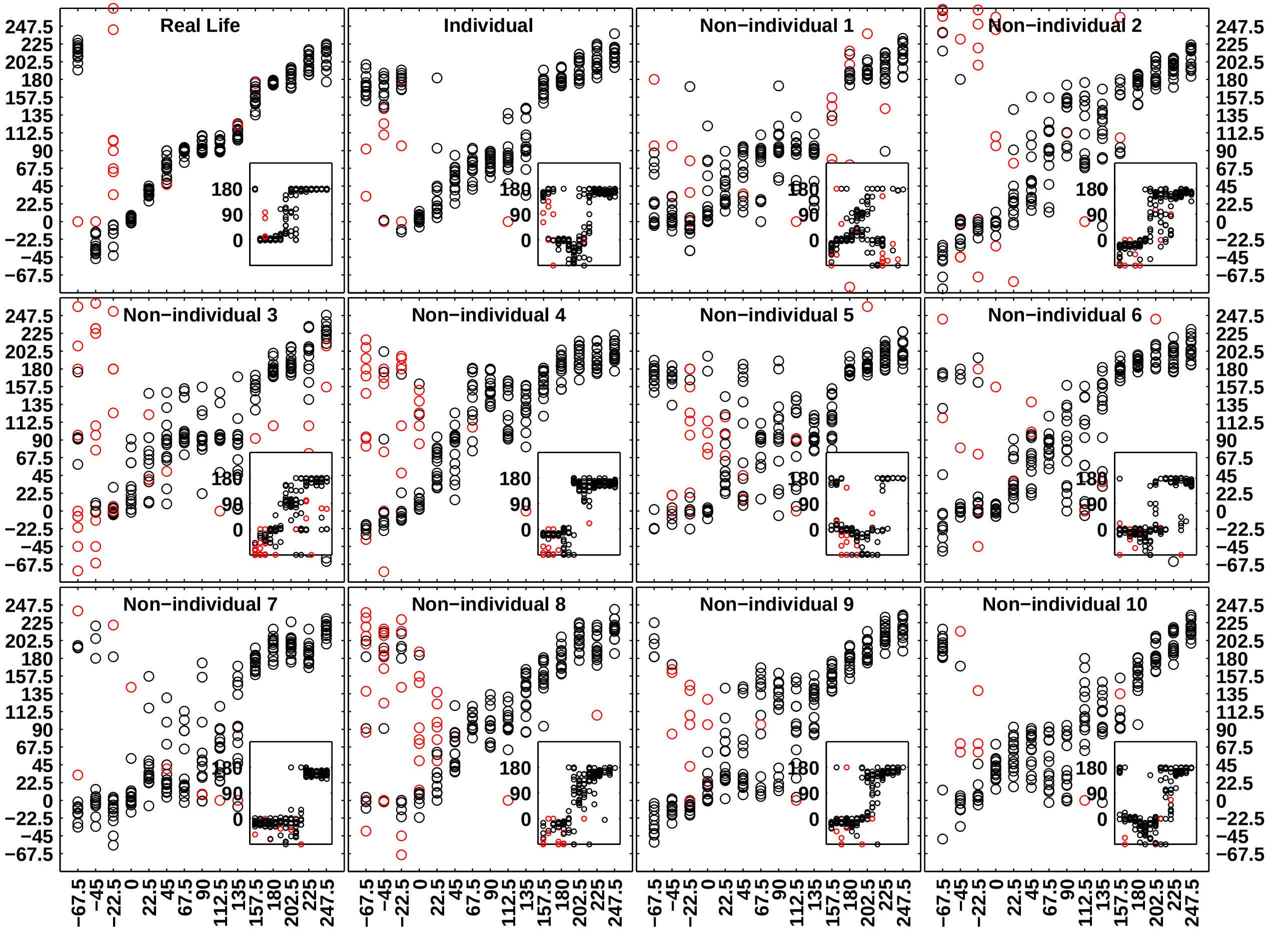

Figure 5.6: Localization performance results for the different experimental conditions tested, for subject MT. Big subplots correspond to perceived elevation, while small plots contained in them correspond to perceived azimuth. The computed correlation value between real sound sources and individual HRTFs binaural synthesis centroids is $\hat{\rho}v = 0.3$ and the null hypothesis $\hat{\rho}v = 0$ is rejected in favor of $\hat{\rho}v > \hat{\rho}_{\alpha}$.

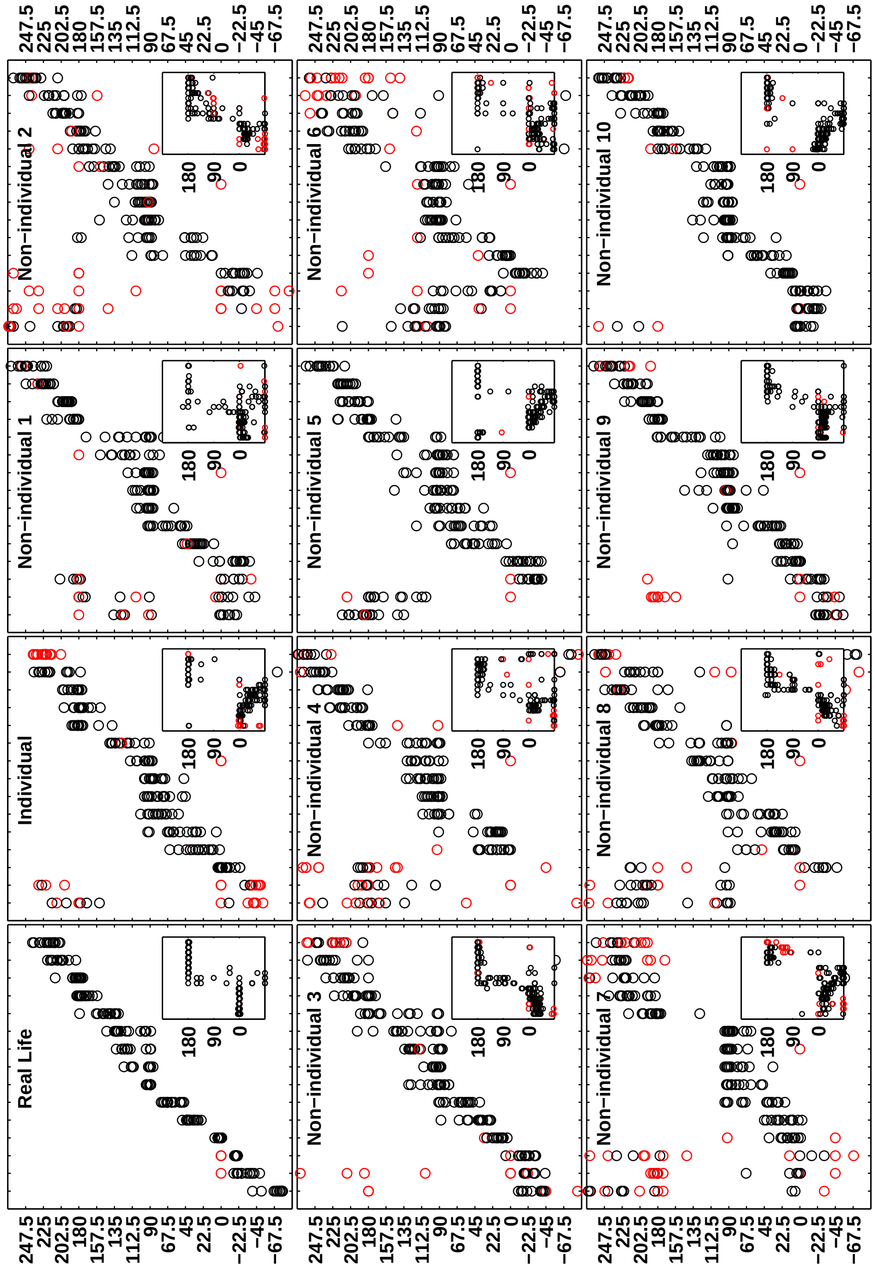

Figure 5.7: Same as Fig. 5.6, but for subject EMS. The computed correlation value between centroids is $\hat{\rho}v = 0.4$ and the null hypothesis $\hat{\rho}v = 0$ is rejected in favor of $\hat{\rho}v > \hat{\rho}_\alpha$.

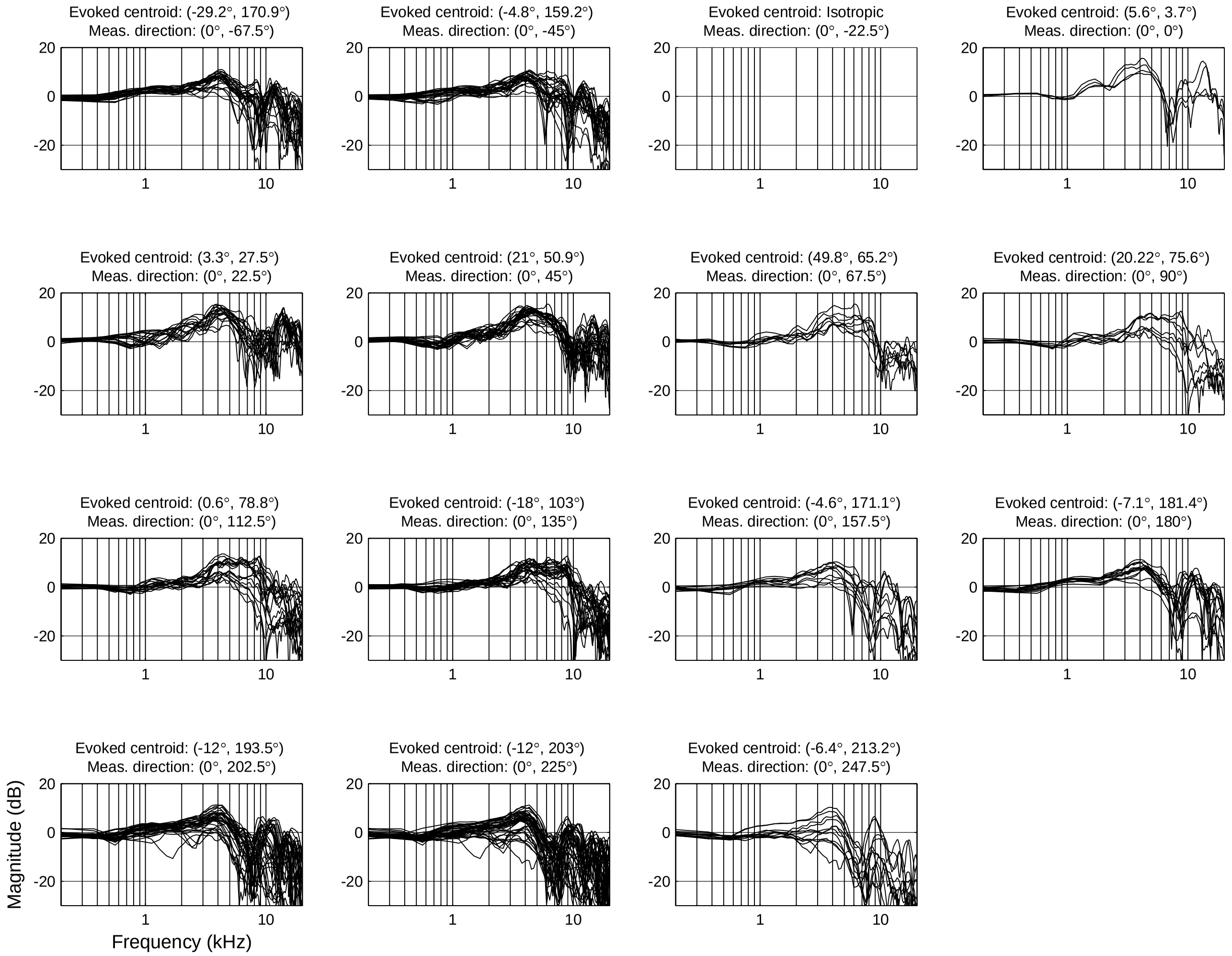

Figure 5.8: Results from the HRTFs matching procedure, for subject MT. The figure shows the non-individual HRTFs that evoked a localization that matched those of the individual HRTFs, as obtained from the behavioral data presented in Fig.5.6. The empty subplot correspond to a individual HRTFs pair that evoked isotropic performance. See text for more details.

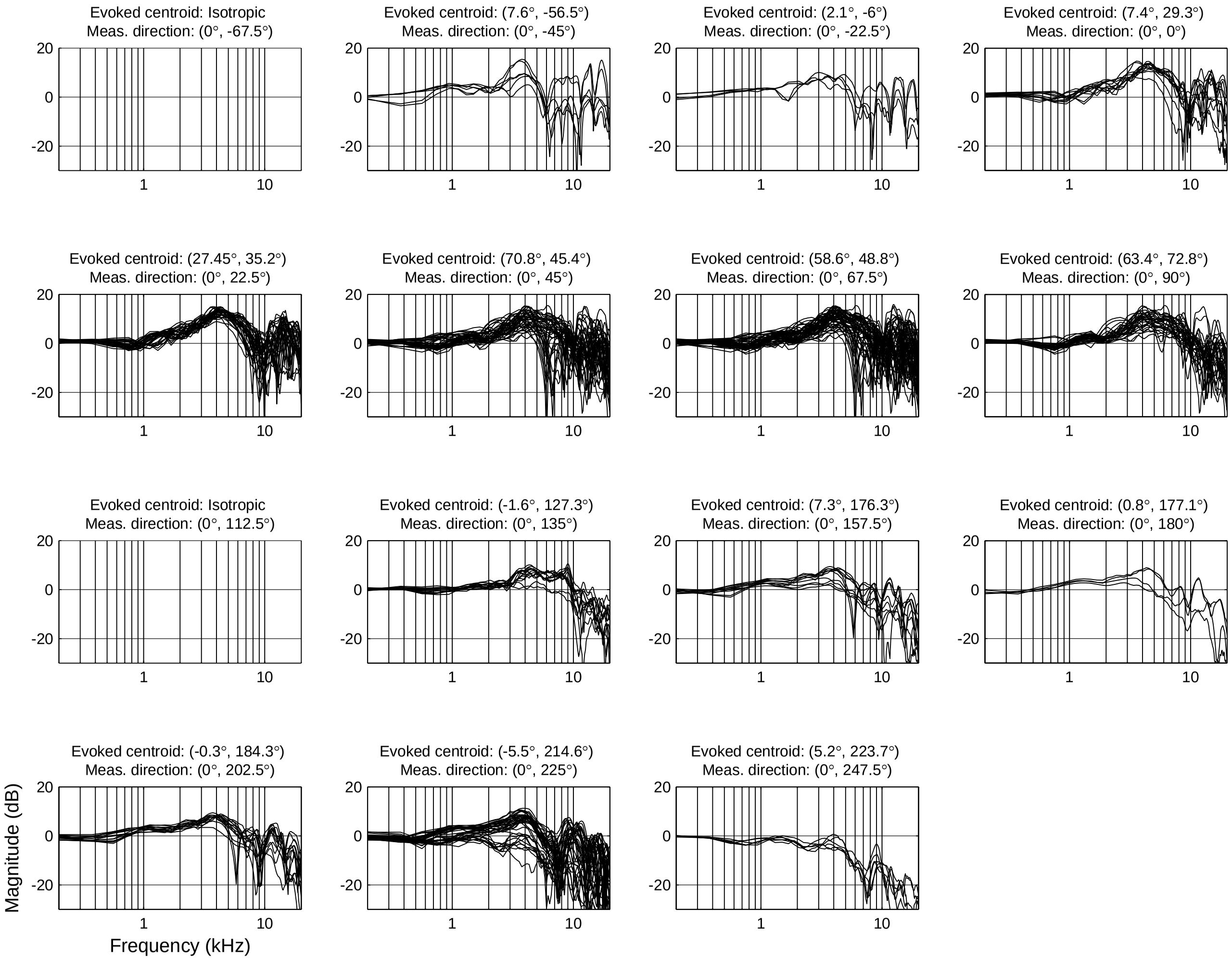

Figure 5.9: Same as Fig. 5.8, but for subject EMS. The HRTFs matching procedure was based on the behavioral data presented in Fig.5.7. See text for more details.

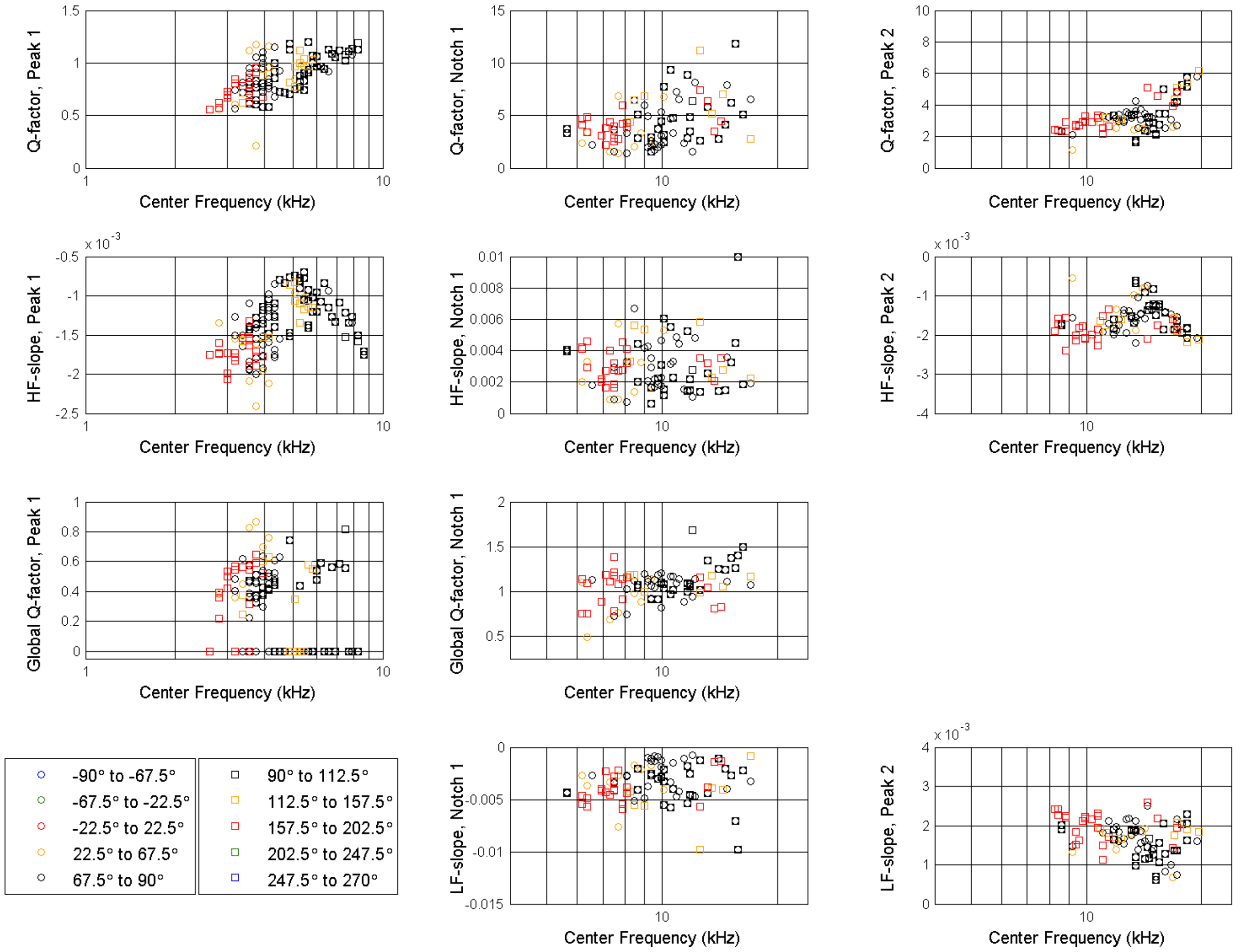

Figure 5.10: Spectral features parameters computed for subject JC, from the HRTFs shown in Fig. B.4. See text for further details.

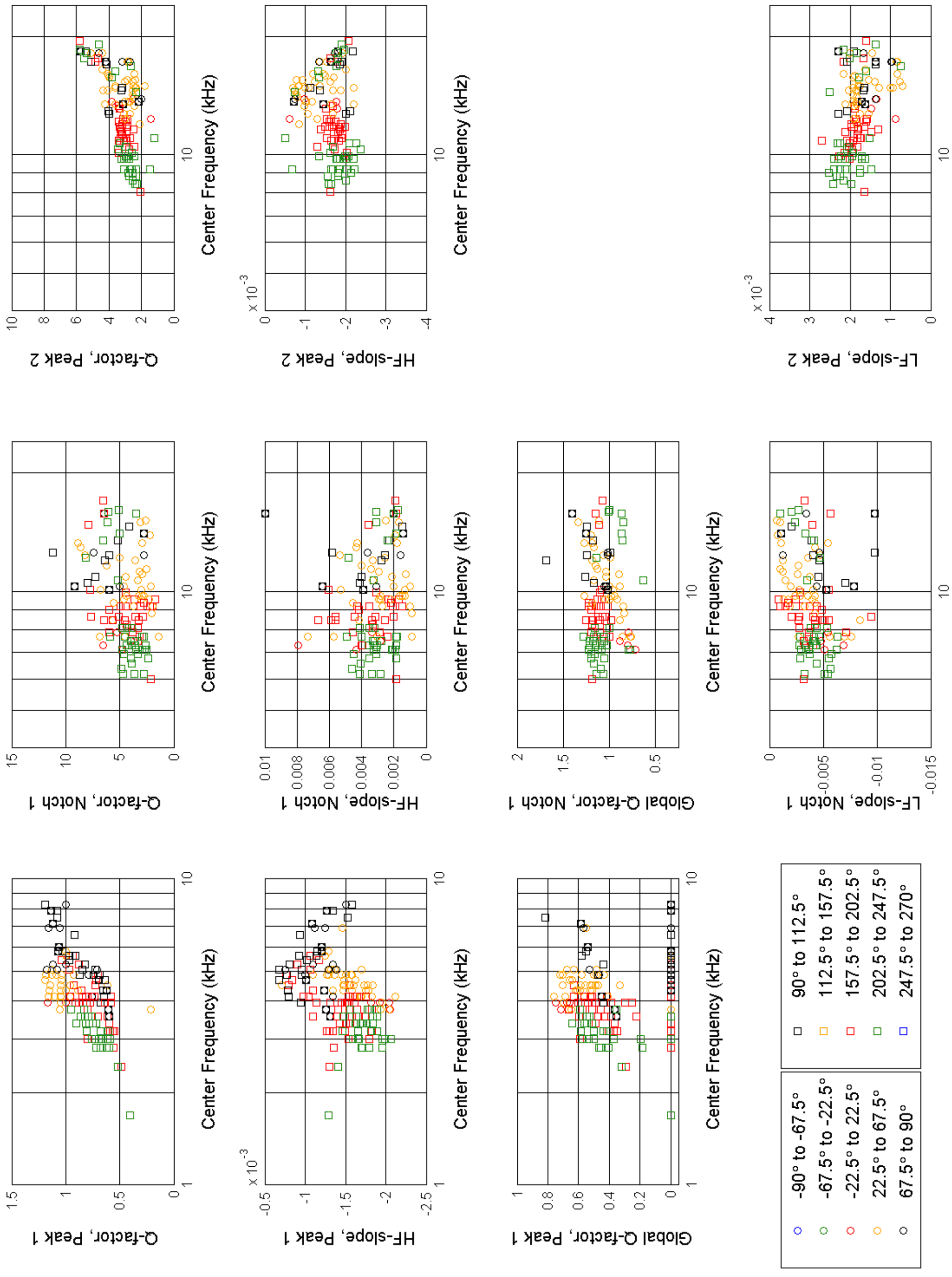

Figure 5.11: Same as Fig. 5.10, but for subject MT. Parameters computed from the HRTFs shown in Fig. 5.8.

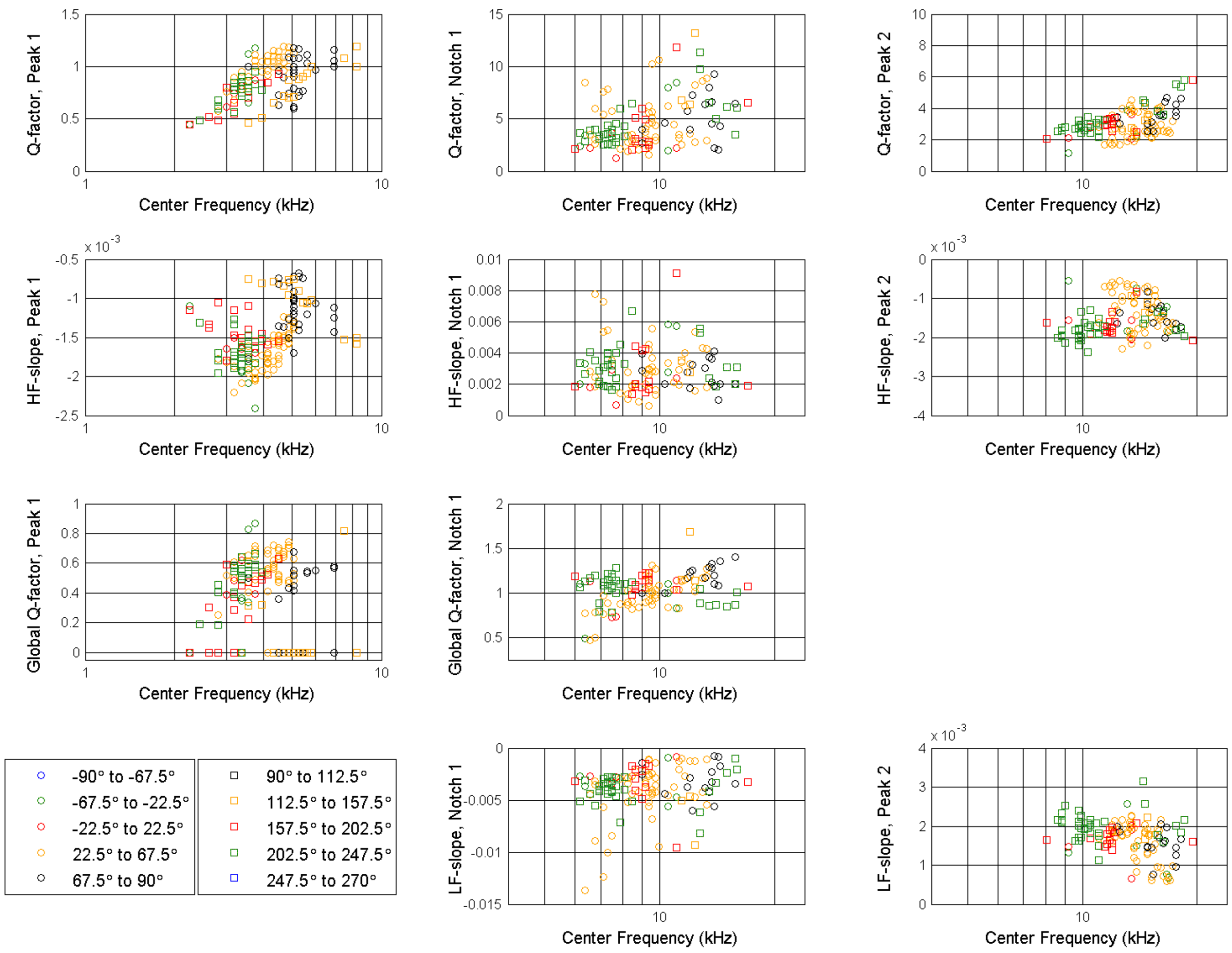

Figure 5.12: Same as Fig. 5.10, but for subject EMS. Parameters computed from the HRTFs shown in Fig. 5.9.

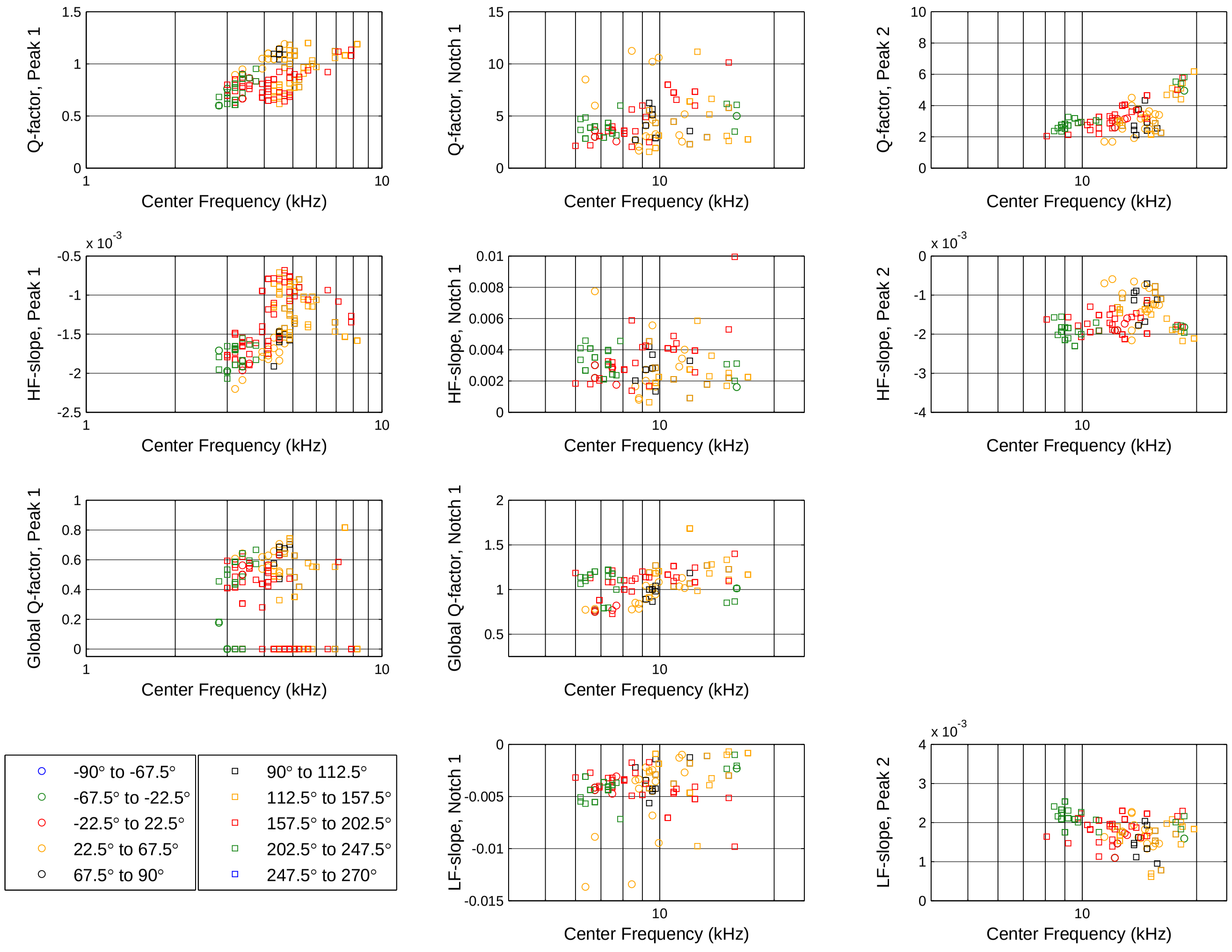

Figure 5.13: Same as Fig. 5.10, but for subject BG. Parameters computed from the HRTFs shown in Fig. B.5.

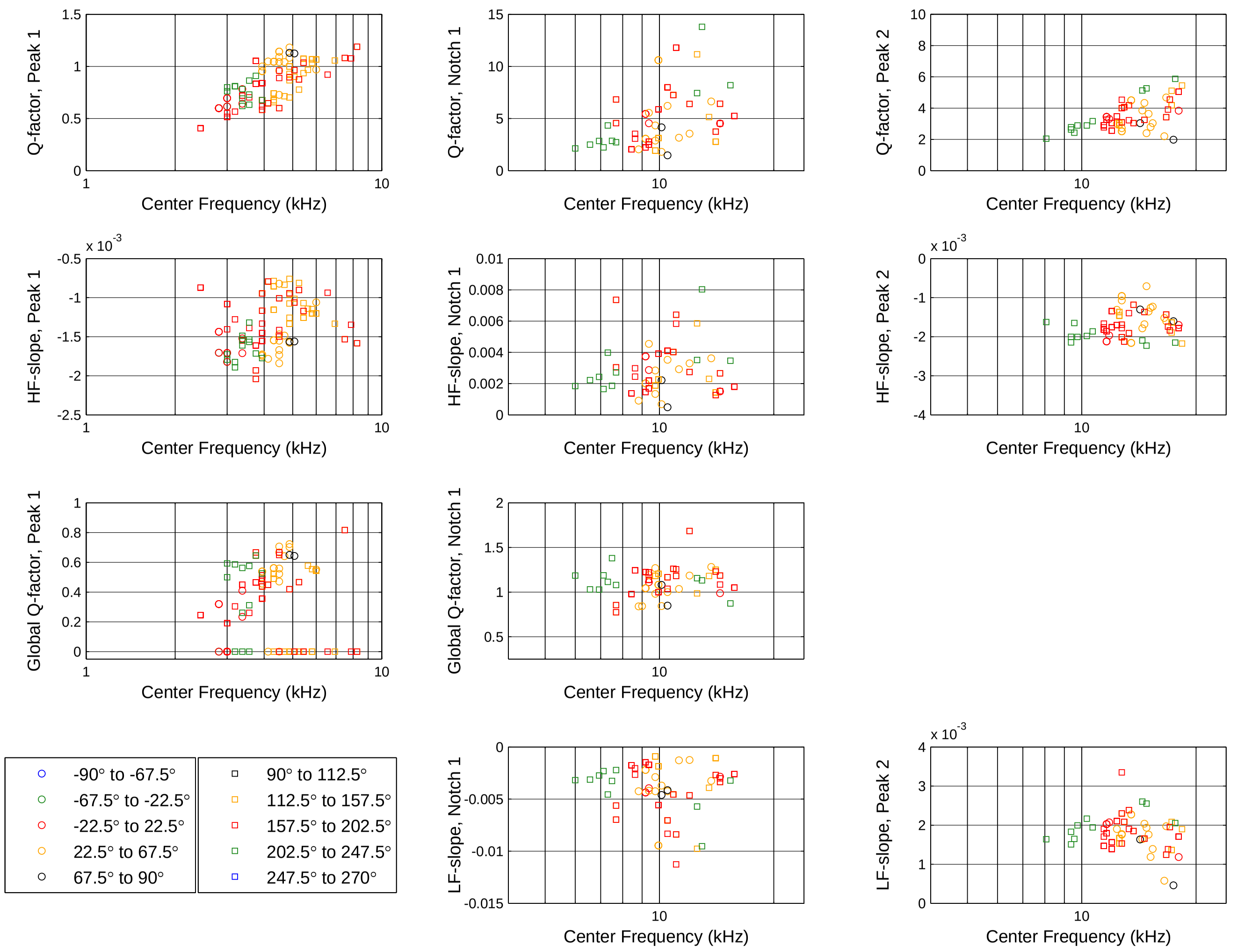

Figure 5.14: Same as Fig. 5.10, but for subject AM. Parameters computed from the HRTFs shown in Fig. B.6.

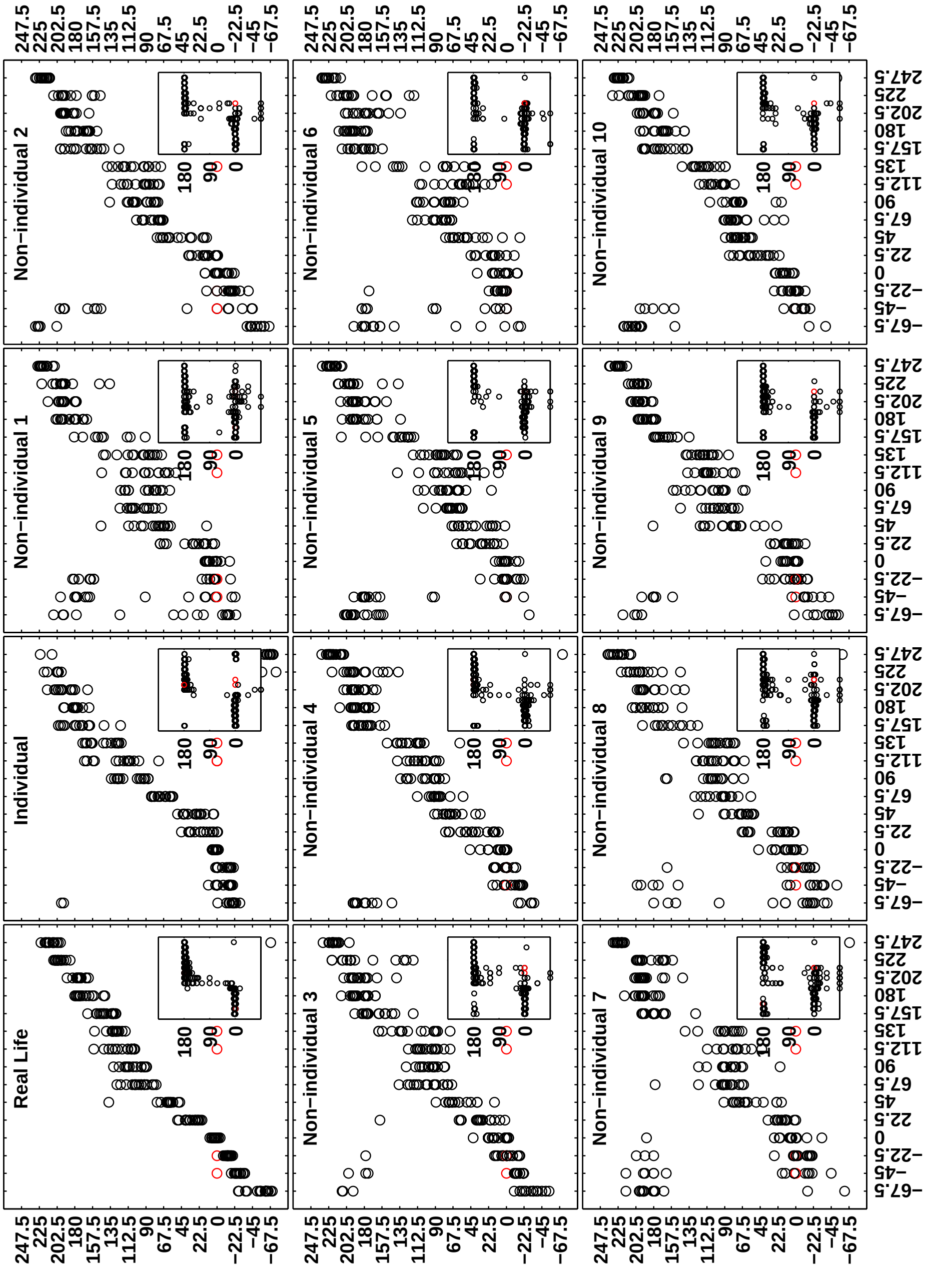

Figure 5.15: Localization performance results for the different experimental conditions tested, for subject ME. The computed correlation value between real sound sources and individual HRTFs binaural synthesis centroids is $\hat{\rho}\nu = -0.7$ and the null hypothesis $\hat{\rho}\nu = 0$ is rejected in favor of $\hat{\rho}\nu < \hat{\rho}_{1-\alpha}$.

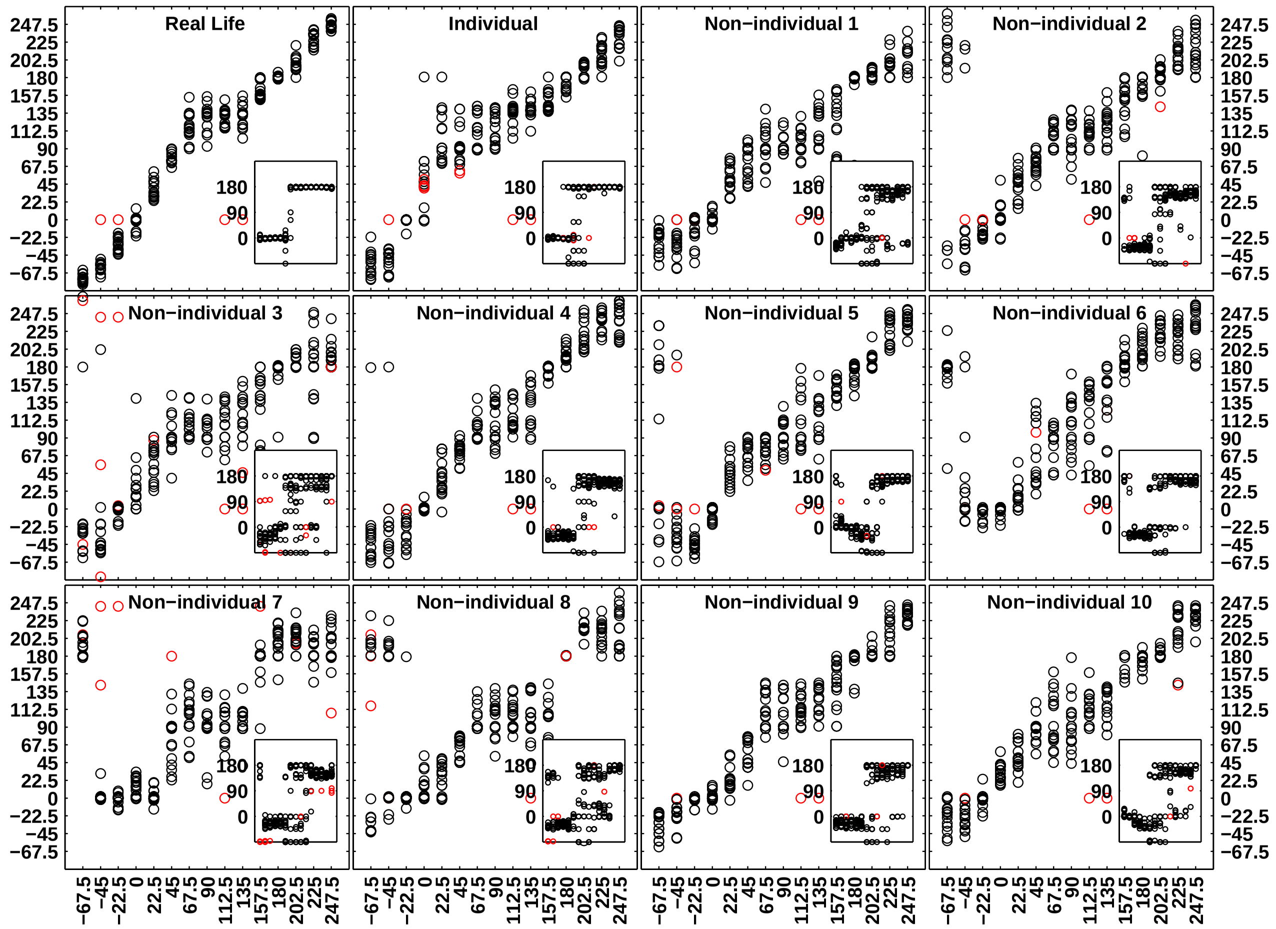

Figure 5.16: Same as Fig. 5.15, but for subject YB. The computed correlation value between centroids is $\hat{\rho}v = -0.2$ and the null hypothesis $\hat{\rho}v = 0$ is rejected in favor of $\hat{\rho}v < \hat{\rho}_{1-\alpha}$.

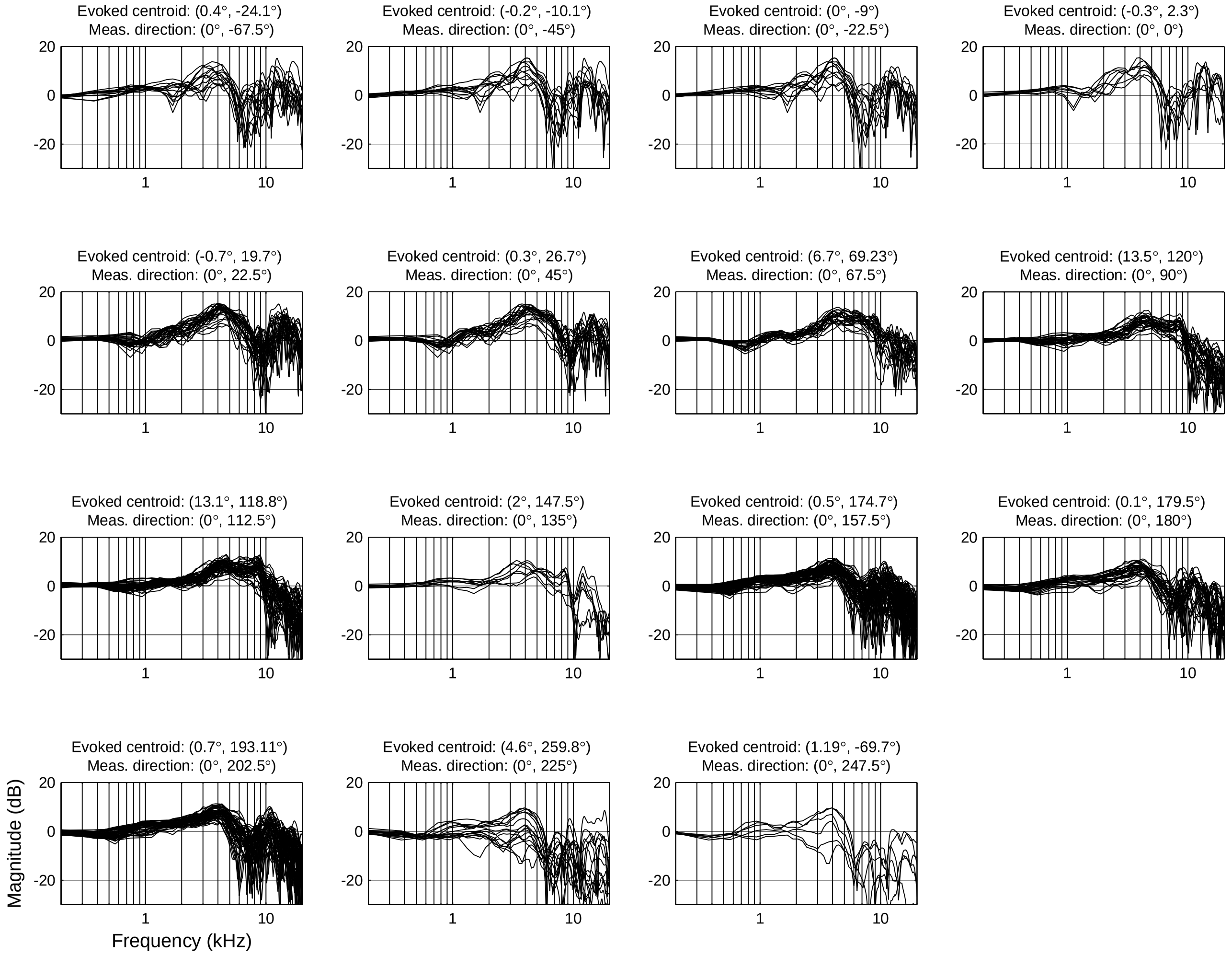

Figure 5.17: Results from the HRTFs matching procedure, for subject ME. The figure shows the non-individual HRTFs that evoked a localization that matched those of the individual HRTFs, as obtained from the behavioral data presented in Fig.5.15. See text for more details.

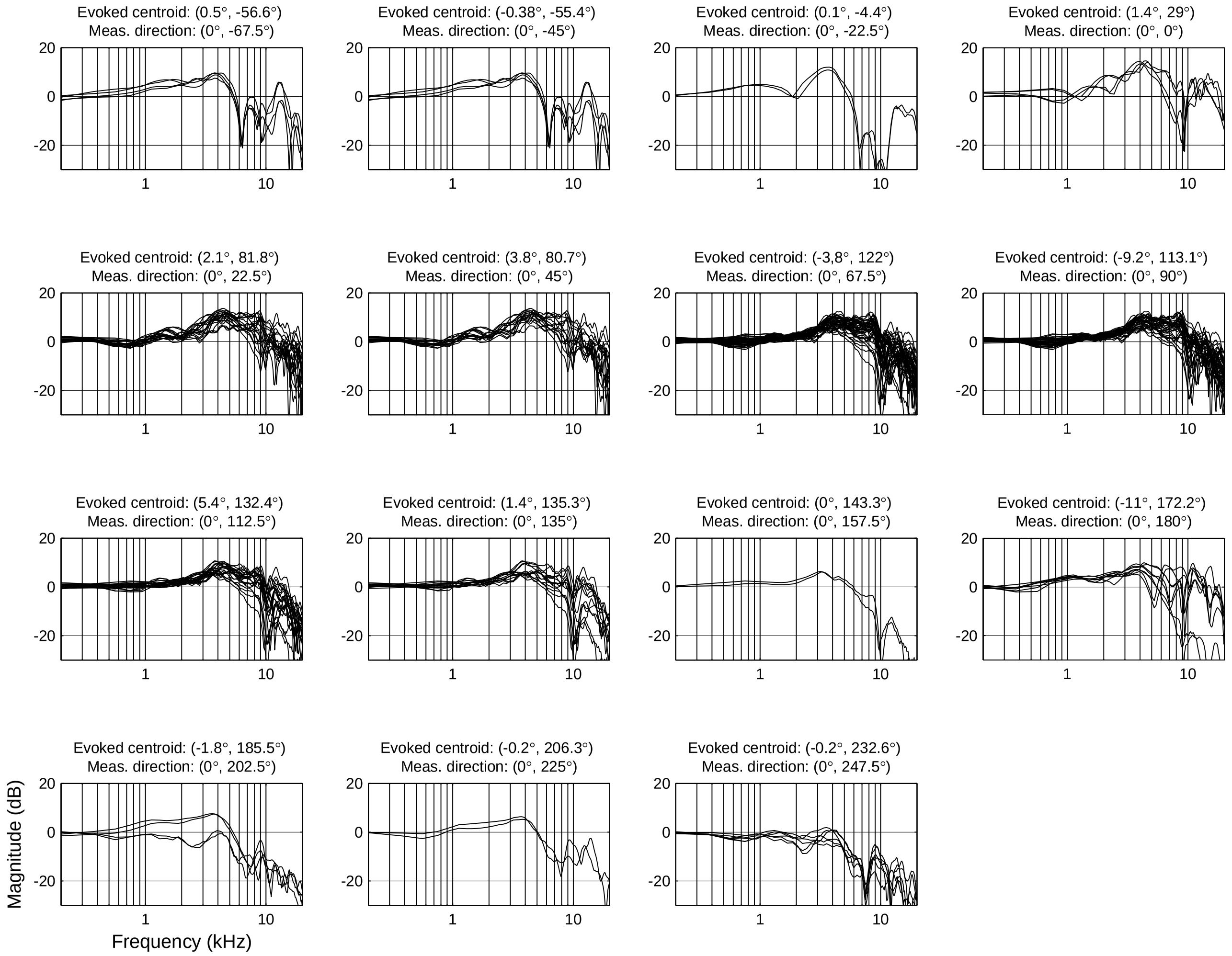

Figure 5.18: Same as Fig. 5.17, but for subject YB. The HRTFs matching procedure was based on the behavioral data presented in Fig.5.16. See text for more details.

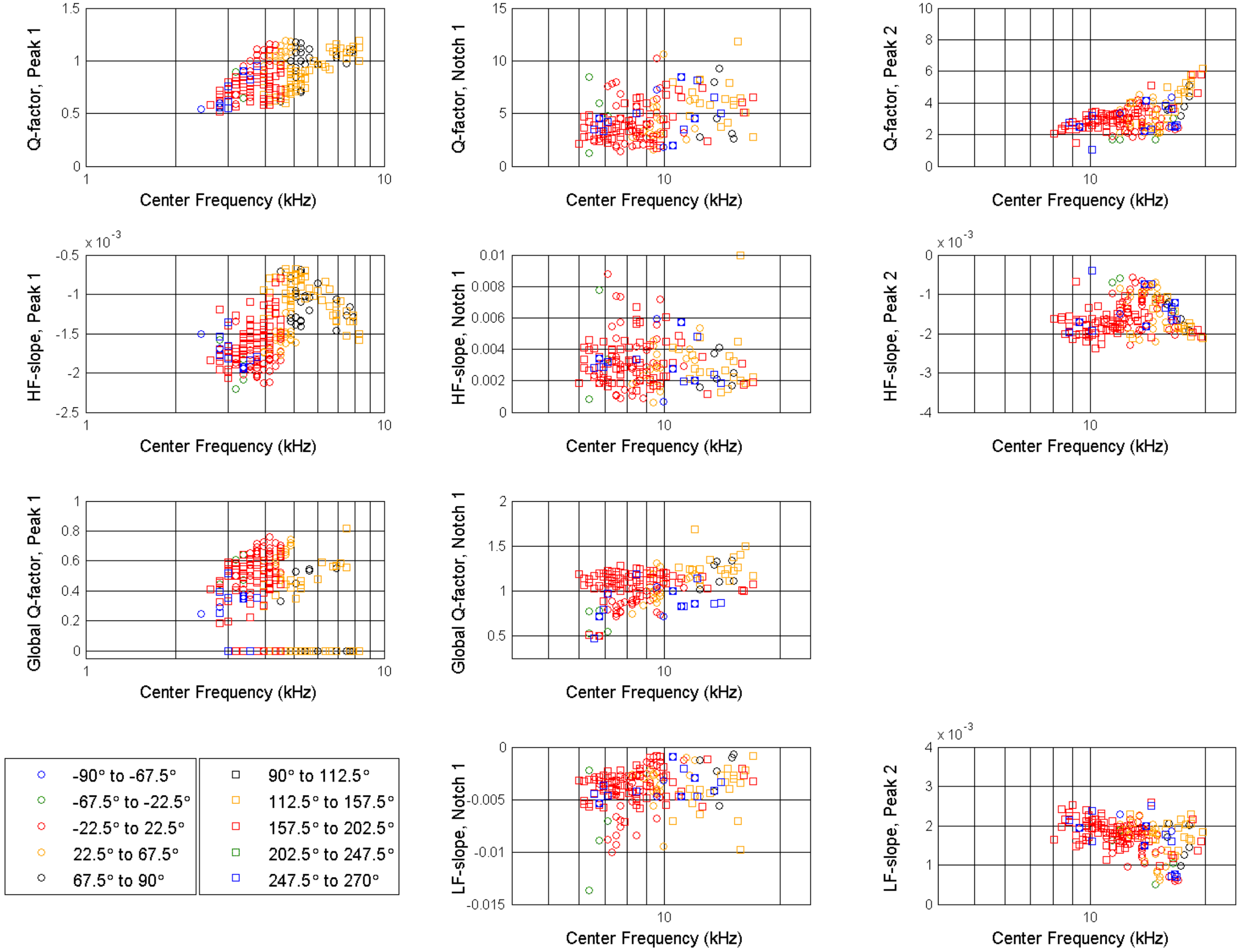

Figure 5.19: Spectral features parameters computed for subject ME, from the HRTFs shown in Fig. 5.17. See text for further details.

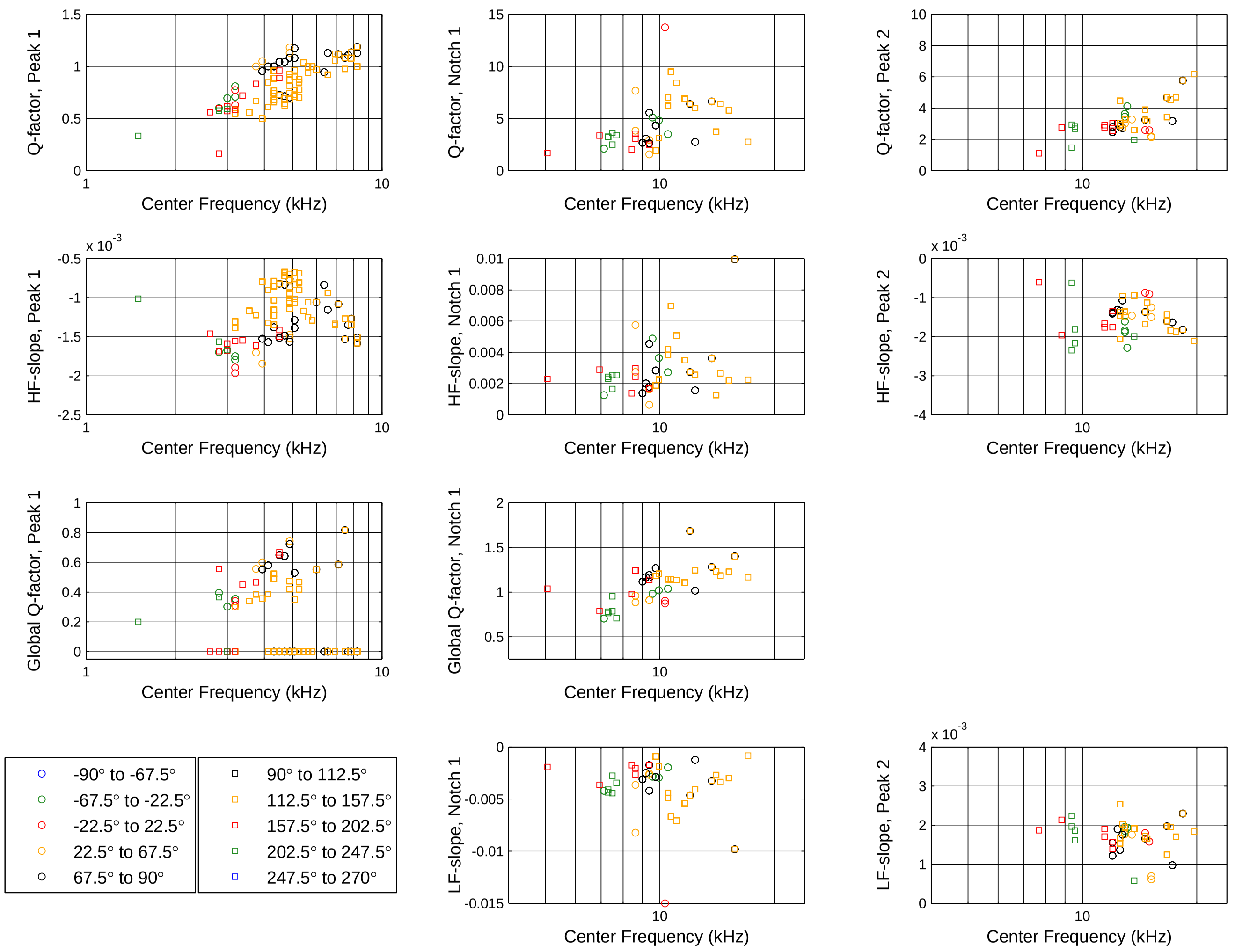

Figure 5.20: Same as Fig. 5.19, but for subject YB. Parameters computed from the HRTFs shown in Fig. 5.18.

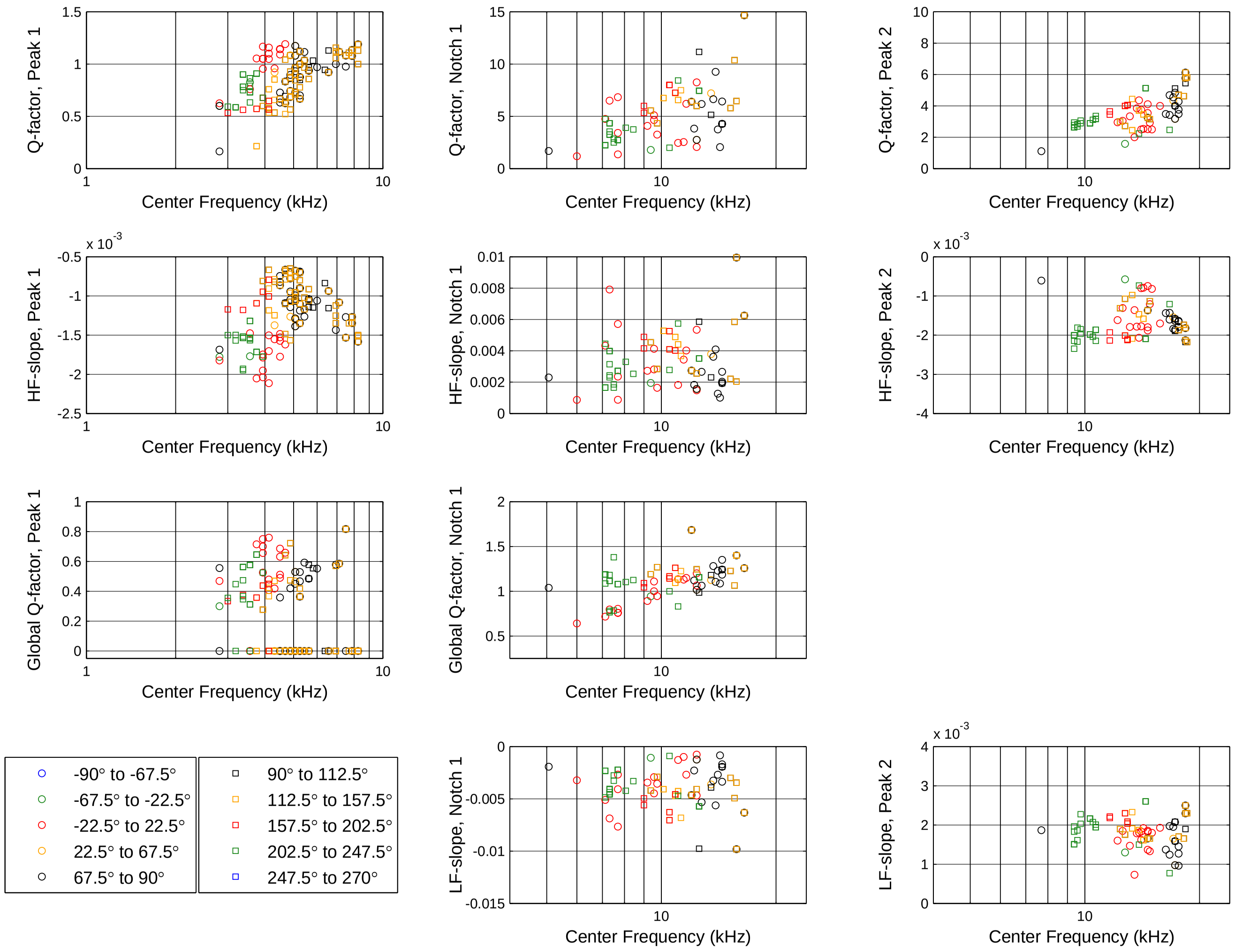

Figure 5.21: Same as Fig. 5.19, but for subject AW. Parameters computed from the HRTFs shown in Fig. B.10.

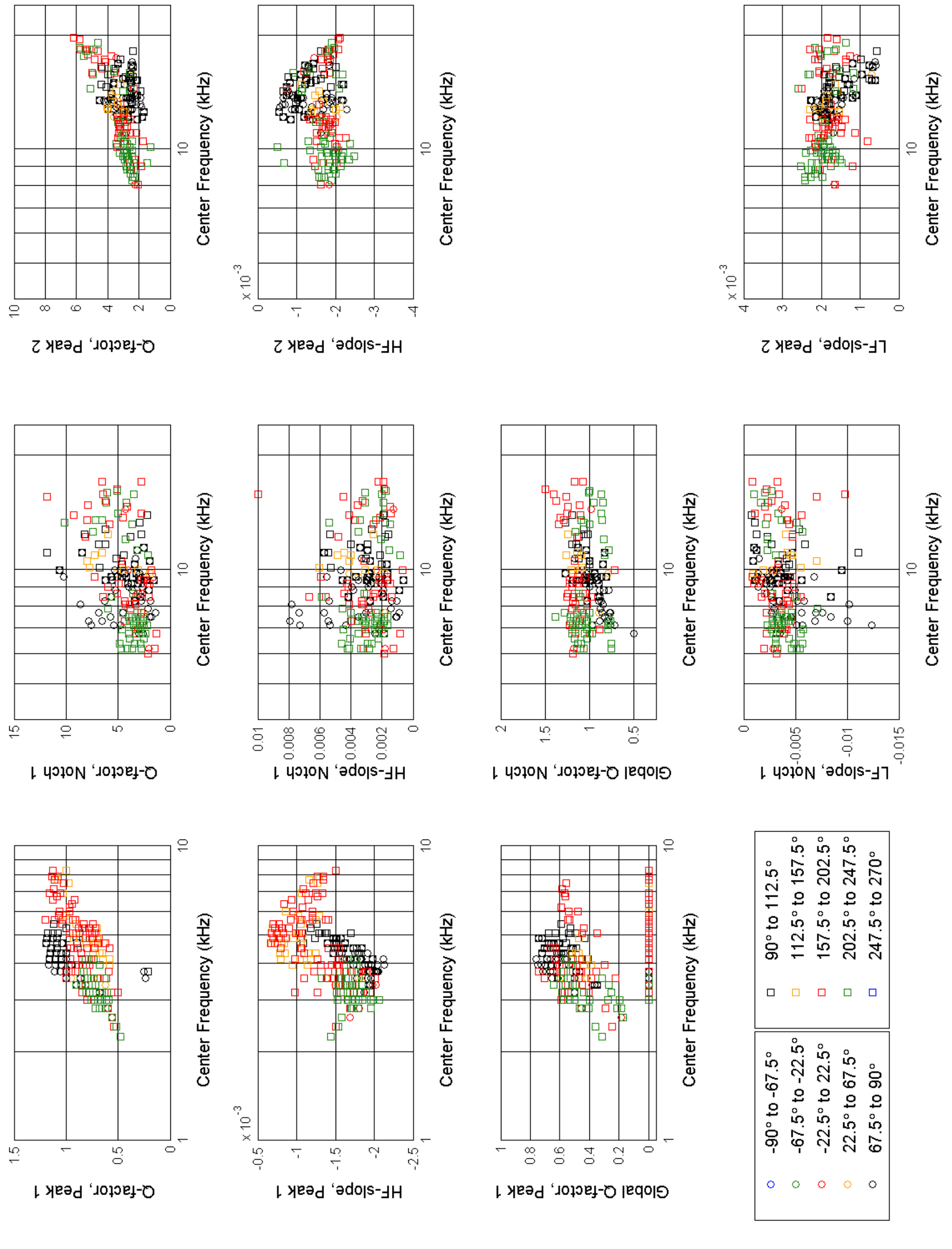

Figure 5.22: Same as Fig. 5.19, but for subject SR. Parameters computed from the HRTFs shown in Fig. B.11.

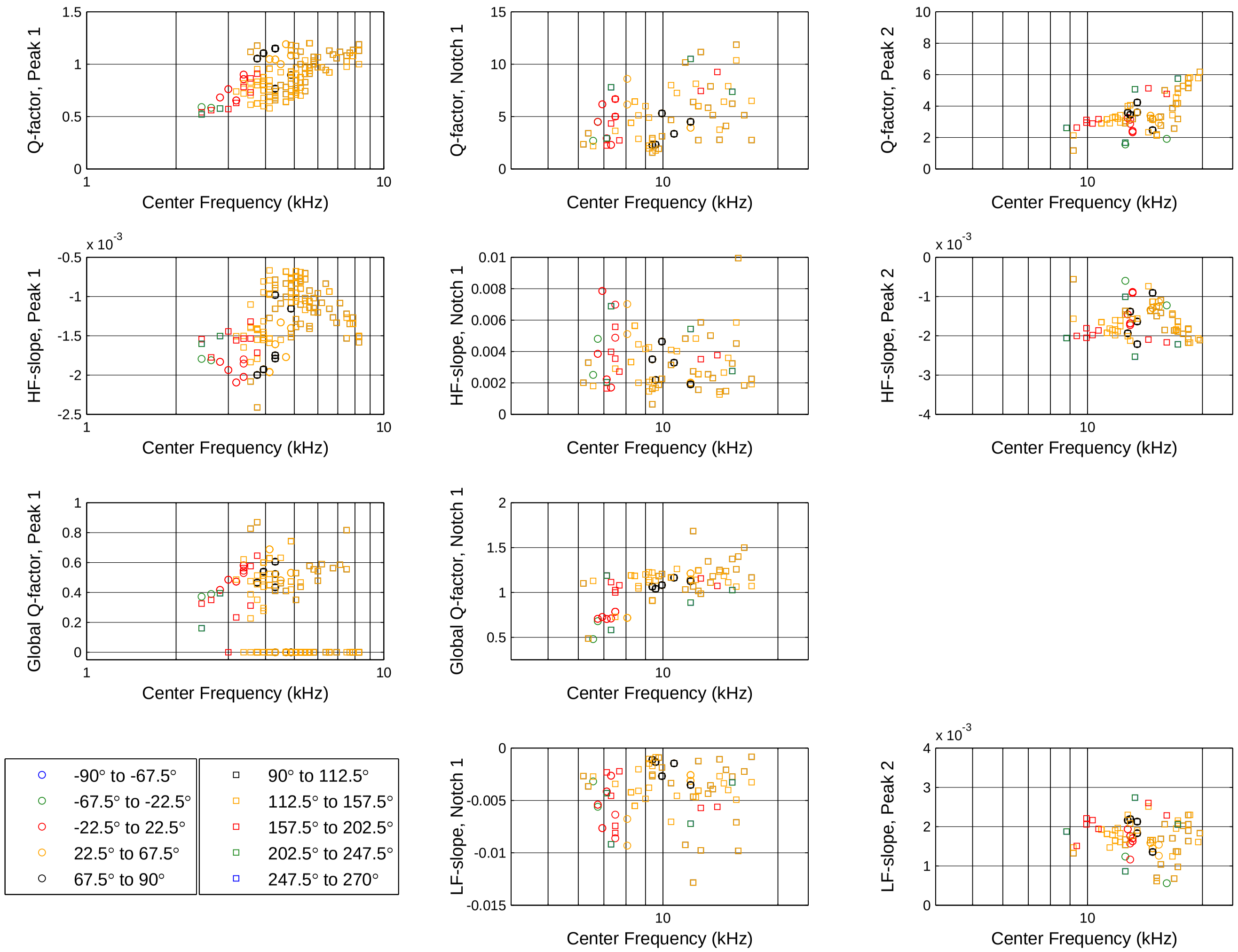

Figure 5.23: Same as Fig. 5.19, but for subject LA. Parameters computed from the HRTFs shown in Fig. B.12.

5.5 Discussion

5.5.1 Correlation analysis

The results of the correlation analysis between real sound sources localization performance and individual binaural synthesis localization performance for both Group A and B are condensed in Table 5.1. The null hypothesis that there was no correlation between centroids was tested with permutation tests as described in 5.3.1, and in all cases the null hypothesis was rejected in favour of the alternative hypothesis -which in some cases was $\hat{\rho}v > \hat{\rho}_{\alpha}$ and in others $\hat{\rho}v < \hat{\rho}_{1-\alpha}$ - at the 5% significance level.

Group	Subject	Correlation Coefficient
A	JC	-0.2
	MT	-0.3
	EMS	0.4
	GB	-0.4
	AM	-0.4
B	ME	-0.7
	YB	-0.2
	AW	-0.2
	SR	-0.3
	LA	0.3

Table 5.1: Correlation coefficients between centroids obtained from localizing real sound sources and virtual sound sources synthesized with individual HRTFs. The null hypothesis (there is no correlation between centroids) is rejected in all cases at the 5% significance level.

The correlation values presented are, however, much lower than those reported in the literature. Wightman & Kistler [1989b] showed ’goodness of fit’ values between answers and target directions of above 0.89. Carlile *et al.* [1997] reported individual spherical correlation coefficients of above 0.9, and a pooled correlation coefficient of 0.98. The explanation for these higher values is based on two differences between those studies and the one presented here. Firstly, they correlated perceived and target sources for a variety of directions around the listeners and computed an average measure, while in this study perceived sources under two conditions, for a restricted number of sources from the MSP only, were correlated. Secondly, in those studies the data was subject to some pre-statistical processing: Wightman & Kistler [1989b] resolved front-back confusions and Carlile *et al.* [1997] extracted answers with evidence of front-back confusion from the data prior to the statistical analysis. Both approaches bias the results by reducing the actual scattering of answers and therefore overestimating the concentration parameter. The computed centroids, along with other parametes, are also changed. This was acknowledged by Wenzel *et al.* [1993] who reported goodness of fit values (analogous measure to that presented by Wightman & Kistler [1989b], correlating perceived and target directions) for answers where front-back confusions were not resolved nor discarded. Not surprisingly, the individual values ranged from 0.52 to 0.76 for the binaural synthesis condition and from 0.55 to 0.91 for the free field condition. It is hypothesized that the values reported by Wenzel *et al.* [1993] are higher than those of Table 5.1 because in the former study the measures were computed as averages of goodness of fit across different directions.

Both resolving and removing front-back confusions indiscriminately -i.e. by applying an algorithm that treats all judgments equally and resolves or removes all front-back confusions found- can be more or less risky depending on the context: in the case of a distribution like that of the fifth source in the individual condition for subject MT, Fig. 5.6, with original coordinates $(0^\circ, 22.5^\circ)$, resolving the few isolated cases of front-back confusion would result in a more concentrated distribution with a smaller standard deviation cone. The centroid would be closer to the original coordinates $(0^\circ, 22.5^\circ)$ and it could be argued that it better represents the perception of the subject. If the answers to remove were too many, the risk would be that too few ones would be left for the statistical computation. However, in a distribution like that of the first source in the real sound sources condition for the same subject, Fig. 5.6, almost all the answers lie in the opposite hemisphere. Resolving all these answers would result in a centroid that nothing has to do with the perception of the subject, who consistently heard the sound in a very restricted area - the answers are indeed concentrated. Removing them would lead to no perception for that source. Therefore, removing and/or resolving front-back conditions should be assessed case by case. On the other hand, it has been argued that cases of front-back confusion lead to bimodal distributions, which are undesirable. The statistical analysis for the experiment presented here showed no evidence of bimodal distributions -which was tested with the procedures described in 5.3.1.

Table 5.2 compares the correlation results between centroids from localizing real sound sources and virtual sound sources synthesized with individual HRTFs, for subjects in Group A, with and without front-back and up-down reversals resolved. In the line of what was exposed, an improvement can be seen for those subjects who presented some extent of front-back and up-down confusion -i.e. JC, MT and EMS. No change is seen for AM and a decrease is seen for BG.

Subject	Corr. Coeff. raw data	Corr. Coeff. confusions resolved
JC	-0.2	0.5
MT	-0.3	-0.5
EMS	0.4	0.5
GB	-0.4	-0.3
AM	-0.4	-0.4

Table 5.2: Correlation coefficients between centroids obtained from localizing real sound sources and virtual sound sources synthesized with individual HRTFs, for the raw data and the data with front-back and up-down confusions resolved.

Processing the data to account for front-back confusions would seem appealing as results appear closer to what is expected from binaural technology. However, this procedure seems to neglect the cognitive processes that work together in order to construct a sound source image. These processes include the ability of the subject to disambiguate his or her own HRTFs ambiguous cues. In order to understand binaural hearing and how to use it, it has to be considered that localization answers are more scattered and farther from the target than theoretically expected. More about this topic will be presented in Chapter 7.

5.5.2 Externalization of binaurally reproduced sound

One of the fundamentals of binaural technology is that virtual sound sources are externalized -i.e. perceived as outside the head-, as opposed to stereo technology where the sound source image is located inside the head. The number of perceived sound sources plotted in red in Figs. 5.6 and 5.7 (and Figs. B.1 to B.3 in Appendix B for the other subjects), i.e. subjects from Group A, shows that this is not always the case. When actively asked about whether they externalized the sound, there were instances in which subjects answered NO even under real sound sources conditions. It would be expected that inside-the-head perception connects to a scattered or isotropic performance -i.e. guessing the answers-, see for example Fig. 5.6, third real sound source with original coordinates $(0^\circ, -22.5^\circ)$. However, Fig. 5.7 (see also Fig. B.2) shows, in its fifteenth sound source under individual HRTFs binaural condition with original coordinates $(0^\circ, 247.5^\circ)$, that the direction perceived can be consistent and *'correct'*, determining a non-isotropic distribution, even though virtual sources are still perceived as not externalized. Table 5.3 condenses the rates of inside-the-head perception (as percentage over the total number of answers for the condition, for each subject). The rates shown in Table 5.3 suggest that, when the subjects are actively asked about sound externalization, the issue is rather minor with real sound sources or under individual binaural synthesis conditions but it becomes more important under non-individual binaural synthesis condition. Interestingly, subjects from Group B, which were not asked about externalization at all, did not report so many answers inside the head as those from Group A. It does not imply that their localization performance was better, but rather that they did not have to use cognitive resources in determining externalization issues.

The problem of inside-the-head perception under anechoic conditions has long been acknowledged: daily environments are not anechoic and subjects are too used to using reflections from their surroundings as cues to sound localization which enhance or potentiate their own. Besides, the non-anechoic environment provides the baseline for the cognitive assessment of how real sounds (as opposed to imaginary sounds or sounds which are thought of) sound. For example, Toole [1970] conducted a dedicated experiment to assess inside-the-head perception with real sound sources, where the stimuli were wideband and bandpass filtered noise samples. He tested different combinations of sound source locations, and showed that subjects reported inside-the-head perception at different rates but in all conditions. Begault *et al.* [2001] showed that the addition of reverberation enhanced externalization from a mean rate of 40% (anechoic binaural reproduction) to 79% (full auralization of a highly reverberant space), and that including early reflection up to 80 *ms* was sufficient to aid externalization of speech signals. It is interesting to note, however, that the rate of inside-the-head perception in Table 5.3 is higher with binaural synthesis reproduction than with real sound sources even if both environments (virtual and real, respectively) are anechoic. This suggests that there are other processes involved in the externalization of sound apart from the non-anechoic character of the environment and uncorrelated signals at the ears, and which would be intrinsically related to binaural technology. These could range from potential errors in the measuring and/or reproduction chains to cognitive paradoxes: how is it possible to perceive the sound from any other place that is not the headphones provided? With real sound sources, whether the subjects are blindfolded or not, they get the message that sound is being generated from devices far from their bodies. In binaural reproduction through headphones, on the other hand, it is cognitively demanding to *neglect* that sound is being generated by devices very close to the ears and concentrate to relate the sound to a direction in space. The cognitive task starts being rather confusing, even if it is not consciously accepted as such by the subjects.

5.5.3 Differences between the graphical interfaces used by Group A and B

As mentioned before in 5.3.1, Group A used a graphical interface where meridians and parallels were given (see Fig. 5.1), while Group B used another one were only a pair of

HRTFs set	Group A					Group B				
	JC	MT	EMS	BG	AM	ME	YB	AW	SR	LA
Real life	4%	6%	1%	15%	24%	2%	2%	2%	1%	1%
Individual	6%	4%	19%	20%	29%	1%	4%	1%	4%	2%
Non-Ind.1	8%	13%	8%	78%	18%	2%	2%	2%	3%	3%
Non-Ind.2	25%	11%	20%	65%	10%	1%	2%	2%	3%	2%
Non-Ind.3	19%	20%	10%	84%	16%	2%	5%	2%	3%	3%
Non-Ind.4	23%	13%	16%	74%	5%	2%	2%	2%	3%	2%
Non-Ind.5	19%	10%	2%	88%	30%	1%	3%	2%	4%	3%
Non-Ind.6	21%	7%	18%	94%	17%	2%	3%	2%	3%	2%
Non-Ind.7	28%	5%	19%	86%	16%	2%	5%	2%	3%	3%
Non-Ind.8	29%	18%	12%	82%	16%	2%	3%	2%	3%	2%
Non-Ind.9	21%	7%	9%	47%	9%	2%	2%	2%	3%	4%
Non-Ind.10	8%	5%	4%	80%	23%	2%	2%	2%	3%	4%

Table 5.3: Percentage of sound sources that were not externalized, discriminated by subject and by HRTFs set.

orthogonal axis was marked as guideline (see Fig. 5.2). It was believed that the parallels and meridians of the former case could introduce biases in how subjects would sample space around them. In order to check whether biases could happen, a side pilot test was conducted with three subjects -which did not participate in the experiment for spectral features analysis and parameterization.

The pilot test consisted of running the individual HRTFs binaural synthesis localization experiment with the procedure and graphical interface depicted for Group B -i.e. without parallels and meridians- at a first stage, and running the same condition but with the graphical interface depicted for Group A at a second stage. Prior to the experiment, the 3 subjects participated of localization experiments with real sound sources in the anechoic chamber, and their HRTFs were measured.

The centroids of the distributions obtained by using both graphical interfaces with individually synthesized sound sources were correlated. The individual values for each of the three subjects were $\hat{\rho}v = 0.63$, $\hat{\rho}v = 0.94$ and $\hat{\rho}v = 0.42$. In all cases, the null hypothesis $\hat{\rho}v = 0$ was rejected in favor of $\hat{\rho}v > \hat{\rho}_{\alpha}$. It was concluded that there would not be significant biases introduced by the graphical interfaces, and no further analysis was attempted regarding potential differences between the results obtained from Group A and B -at least concerning that issue.

It is interesting to note that the correlation values are much higher than those reported in Table 5.1. This confirms the hypothesis that the low values shown in that Table were due to, partly, small differences in azimuth and differences in concentration parameters inherent to comparing performance with real and virtual sound sources.

5.5.4 Spectral features analysis - Group A

The markers in Figures 5.10 to 5.14 are coded in color and shape. For simplicity, the ranges of directions that are depicted in the legends of those figures will be referred to as explained in Table 5.4.

Angular Region (Elevation)	Name
From -90° to -67.5°	Below-front
From -67.5° to -22.5°	Front-low
From -22.5° to 22.5°	Front
From 22.5° to 67.5°	Front-high
From 67.5° to 90°	Above-front
From 90° to 112.5°	Above-back
From 112.5° to 157.5°	Back-high
From 157.5° to 202.5°	Back
From 202.5° to 247.5°	Back-low
From 247.5° to 270°	Below-back

Table 5.4: Names given to the regions arbitrarily determined for the color coding in Figs. 5.10 to 5.23. These names are used in the text to analyze the distribution of the computed parameters from spectral features.

Some trends can be seen in the results from Group A which can be discussed in general. As for the Q-factor of the first peak:

- The relationship between Q-factor of the first peak and center frequency draws a tendency with a positive slope: the higher the center frequency, the higher the Q-factor tends to be. This is not surprising, since it is a known feature that the first peak of the HRTFs *'moves'* from the lower to the upper frequencies as elevation in the frontal hemisphere gets higher, while the peak is narrowed until reaching the above region. As elevation gets lower in the back hemisphere, the peaks decrease in both Q-factor and center frequency.

- HRTFs for the above-front (black circle markers) and above-back (black square markers) directions seem to have a fist peak with superimposing Q-factor and center frequency, which is expected since there is little spectral change in the HRTFs for those directions. In other words, the same HRTFs pairs were evoking both regions. In general, they seem to either maintain a low Q-factor profile -see Figs. 5.10 and 5.11- extending over a broad range of frequencies but disposed at the bottom of the group distribution while maintaining the positive slope, or to concentrate in a very specific and narrow range of frequencies -see Fig. 5.12, and also 5.13 and 5.14 even though fewer HRTFs were perceived in the above region by these last three subjects.

- In general terms, HRTFs from the back hemisphere (square markers) seem to present a first peak with lower Q-factor than HRTFs from the frontal hemisphere. This is not a strict trend, but it is explained by looking at the spectra of the HRTFs: those from the front have the energy concentrated in a narrower frequency range -i.e. a narrow first peak. Those from the back hemisphere present the energy spread in a broader peak that determines a lower Q-factor.

- HRTFs for back-high (yellow square markers) and front-high (yellow circle markers) directions seem to have a first peak with distinctive center frequencies. They seem to be in parallel distributions -except for subject MT in Fig. 5.11, which did not perceive sound in the back-high region and for which front-high spectral features are clustered at the top of the Q-factor values. For the rest, the Q-factor of the first peak in HRTFs from the front-high region seems to be centered at lower frequencies compared to those in the back-high region, while maintaining a high Q-factor profile. The distribution of both groups of markers tend, in general, towards a positive slope.

- Distributions for the front-high (yellow circle markers) Q-factors of the first peak superimpose in some extent with those in the above-front (black circle markers) region -see Figs. 5.10, 5.11, 5.12 and 5.14- or with the above-back (black square markers) region -see Fig. 5.13. Similarly, back-high (yellow square markers) is superimposed with above-back. This overlapping is expected, since in reality the transition between regions is smooth and the division in *'regions'* or *'directions'* is arbitrary.

- The Q-factor of the first peak of HRTFs in the back (red square markers) and back-low (green square markers) regions seem superimposed. They present the lowest center frequencies but occupy a large range of Q-factor values, therefore superimposing as well with other regions that are centered in the lower frequency range: front (red circle markers) and front-low (green circle markers).

The above points suggest that the first peak disambiguates front-high from back-high and the above region -i.e. directions above the horizontal plane. Directions below the horizontal plane seem to rely on other spectral characteristics for disambiguation. The case of subject EMS in Fig. 5.12 seems paradigmatic, since the relationships shown in Fig. 5.24 can be drawn from its distribution.

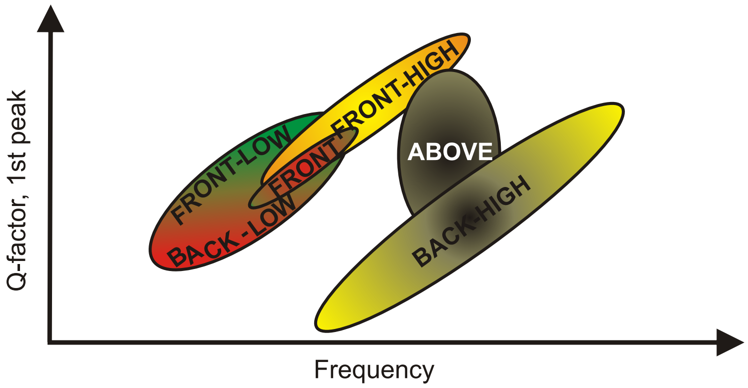

Figure 5.24: Diagram conceptualizing the characteristics observed for the Q-factor of the first peak, subject EMS.

Regarding the high frequency slope of the first peak, they present trends that relate in some extent to the Q-factor distributions discussed previously, possibly because of two reasons. Firstly, the values are also given in relation to the center frequency of the first peak just as in the Q-factor case. Secondly, the first peak is in many cases dominated by the high frequency slope, as the peak value does not reach $+3$ dB over DC.

As a general trend, it can be seen that the high frequency slope is steeper for the first peak of HRTFs in the front-high (yellow circle markers) region than for those in the back-high (yellow square markers) region. However, the high frequency slope of the first peak of HRTFs in the back (red square markers) and back-low (green square markers) regions seems to be as steep as those in the front-high region, but just centered at lower frequencies. As a particular trend seen for subject JC -Fig. 5.10-, the above region is characterized by a high frequency slope that can be less steep if the peak is centered around 5 kHz or steeper if the center frequency of the peak increases towards 9 kHz or decreases towards 3 kHz. Overall, this discussion suggests that the high frequency slope of the first peak could help disambiguating front-high and back-high directions, taking the leading parameter role for the first peak.

The computation of the other parameters involved taking into account other spectral features than the first peak. Therefore, before analyzing them it is worth noticing that not always more spectral features were present in the windowed HRTFs. This is more intuitive for the above region where, even before windowing, the HRTFs take the spectral shape of a low-pass filter. Fig. 5.25 shows the percentage of HRTFs that evoked a direction but for which only the Q-factor and high frequency slope of the first peak could be computed, since they did not present a first notch nor a second peak. In this figure, 100% correspond to the overall number of HRTFs signals (where each HRTFs pair counts for 2 signals) that evoked a direction in a particular region, pooling the results for all subjects in a particular Group (A or B). In some cases, it was only for one side in a HRTFs pair that the parameters could not be computed -those cases are labeled *'Monaural'*. In the case that both sides in the HRTFs pair were affected, the case was labeled *'Binaural'*. Care has to be taken with this classification, as it was not chosen to imply that perception of elevation in the MSP is a binaural process but to ease the understanding of whether the parameter computation was compromised for one or two sides in the HRTFs pair. Figure 5.25 shows that it is in the back-high and above directions, mainly, that the parameters could not be computed. It is suggested that these directions rely mainly on the first peak to be disambiguated. This is supported by the previous analysis, which suggested that the Q-factor and high frequency slope of the first peak were relevant cues.

Another interesting point that can be drawn from Fig. 5.25 is that there would seem to be a hierarchy in the cues: several spectral cues might be present but it would be one of them that would take the leading role to evoke a direction. Furthermore, cues would seem to be able to act in that hierarchy even if they were present in one side only of the HRTFs pair. For example, according to Fig. 5.25 it was only for a few HRTFs evoking the front-high directions that the first notch and second peak were not present. However, the previous analysis conducted on the parameters computed for the first peak suggested that the latter was enough to explain localization above the horizontal plane. Later in this Chapter, it will be shown that the parameters computed for the first notch also present clear tendencies for directions front-high. In other words, front-high directions could be characterized by a series of meaningful and consistent parameters computed from different spectral features. It is therefore hypothesized that there is a hierarchy that organizes them.

Going back to the analysis of the computed parameters for the first peak, it has to be noted that for those HRTFs that did not present a first notch nor a second peak, the global Q-factor for the first peak was set to zero -as plotted in the corresponding figures- since at least there were center frequencies from each peak.

Regarding the global Q-factor of the first peak, similar trends to those mentioned in the context of the Q-factor and high frequency slope are seen: front-high, above and back-high seem to separate themselves even though they are superimposed in a minor

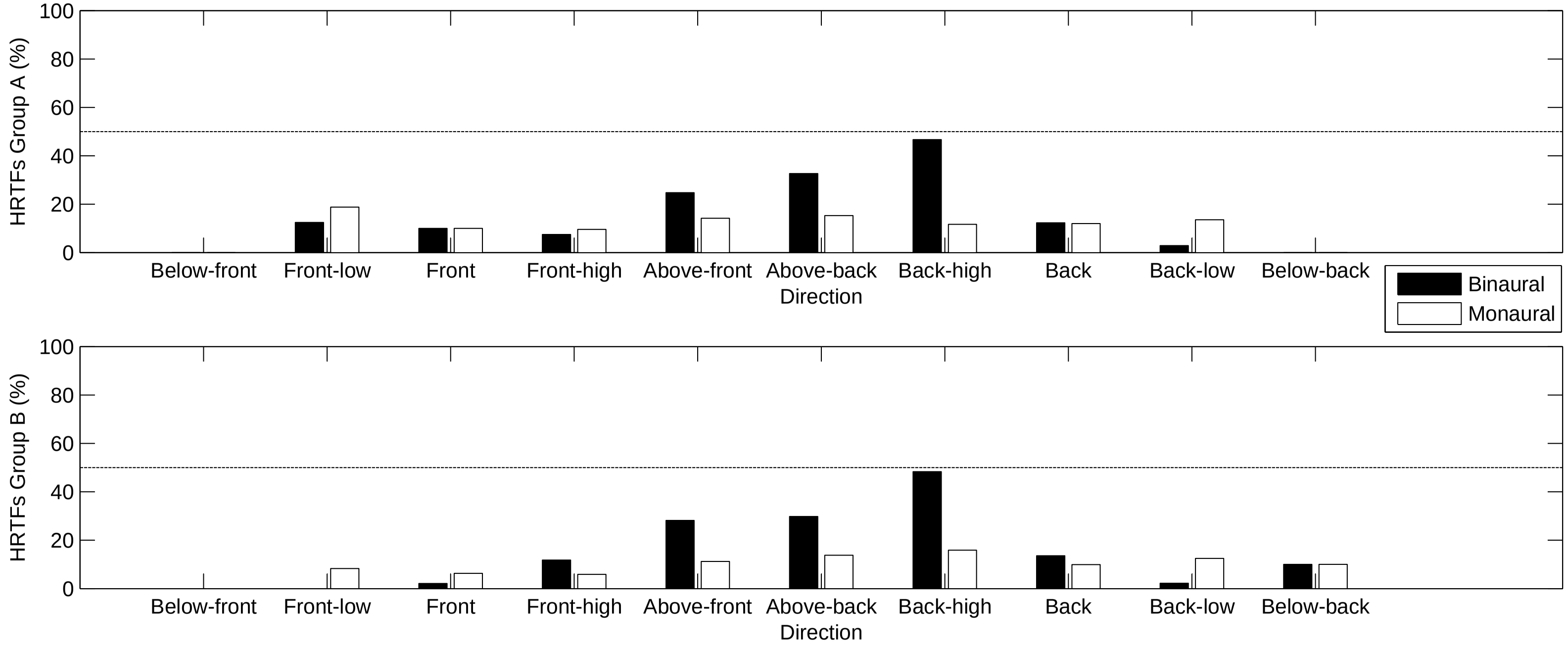

Figure 5.25: Percentage of HRTFs that evoked directions but did not present a first notch nor a second peak in their windowed spectral characteristics. For each evoked region (abscissa), 100% correspond to the total number of HRTFs signals that evoked a direction in it (pooled results for all subjects), where each HRTFs pair counts for 2 signals. ’Binaural’ refers to the percentage of HRTFs pairs that did not present a second peak nor a first notch. ’Monaural’ refers to the percentage of individual HRTFs sides that did not present those spectral features. The 50% level is marked for reference purposes.

degree with other regions. There are resemblances between the Q-factor and global Q-factor distributions for this spectral feature, which are expected.

As for the first notch, the most evident feature is that back (red square markers) and back-low (green square markers) regions present their notches centered at different frequencies -except for subject JC (Fig.5.10) who did not perceive sources in the low directions. Back-low notches seem to be centered at frequencies ranging from 6 *kHz* to up to 8 *kHz*, with a very few cases around 16 *kHz*. Back notches, on the other hand, are centered from 8 *kHz* to 10 *kHz*, with some cases ranging up to around 16 *kHz*. This difference is present in all the figures corresponding to first notch parameters. However, it is the global Q-factor parameter which seems to disambiguate those regions from the rest. Subjects MT and EMS -Figs. 5.11 and 5.12, respectively- show that the global Q-factor of the first notch in regions front-high (yellow circle markers), above (black markers) and back-high (yellow square markers) takes a positive slope in the whole range from 6 *kHz* to 16 *kHz*. In the lower frequencies, the global Q-factor is lower than for the back (red square markers) and back-low (green square markers) regions. In the higher frequencies, the global Q-factor is higher than for the back and back-low regions. The trend is less clear for subjects GB and AM -Figs. 5.13 and 5.14, respectively-, since there is more overlapping between back (red square markers) and front-high (yellow circle markers) regions. The trend is clear for subject JC -Fig. 5.10- with the peculiarity that, as already mentioned, this subject did not perceive sound sources below the horizontal plane.

Front (red circle markers) directions also seem to be disambiguated by the the first notch. A trend is evident for subjects MT, EMS and BG -Figs. 5.11, 5.12 and 5.13, respectively-, in which front directions present a global Q-factor for the first notch that seems to be separated from back (red square markers) directions. The global Q-factor is lower than for back-low (green square markers) and back directions, and the notch is centered at lower frequencies than for front-high (yellow circle markers) directions. This suggests that the disambiguation between front and back is given not by one single spectral feature, but a combination of features in a given range - it is the notch with its center frequency and how distant in frequency the peaks besides it are. Taking the case of subject MT -Fig. 5.11- as a paradigmatic case, the behavior of the global Q-factor of the notch can be described conceptually as in Figure 5.26.

Analysis of the second peak shows some shared trends with the global Q-factor of the first notch. More specifically, in the plots for Q-factor of the second peaks, the back (red square markers) and back-low (green square markers) directions seem to be well separated from the rest of the regions: they are centered at lower frequencies and have a low Q-factor. There are some exceptions, however, where the center frequency is at the higher end of the range (around 18 *kHz*) and the Q-factor is higher -in general, the

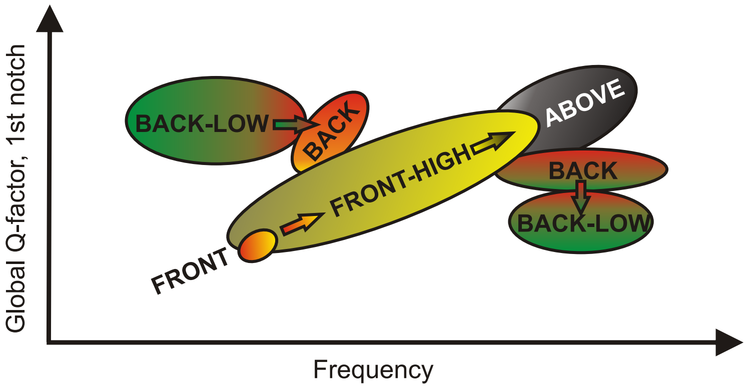

Figure 5.26: Diagram conceptualizing the characteristics observed for the global Q-factor of the first notch, subject MT.

distribution of the Q-factor is disposed with a positive slope tendency. Furthermore, the back-low (green square markers) direction is centered at lower frequencies than the back (red square markers) direction. While a similar trend is seen in the global Q-factor parameter of the first notch, it has to be noted that the difference in center frequency is inherent to the spectral features under consideration, regardless the parameter being computed. This suggests that, as the first notch moves in frequency with increased elevation, so does the second peak but only for directions back-low (green square markers), back (red square markers) and above (black markers). Frontal directions have a different behavior: while for the global Q-factor of the first notch parameters seem to dispose themselves in a diagonal fashion (see Fig. 5.26 for a conceptual diagram) where the value of the parameter disambiguates back and front, for the Q-factor of the second peak these frontal directions seem to be superimposed in a narrow frequency region delimited, approximately, between 12 *kHz* and 18 *kHz*. There is no clear tendency about how the frontal hemisphere directions could be disambiguated in this range. In this context, it can be suggested that the elevation in the back hemisphere could be disambiguated by the second peak, but front-back disambiguation would most likely occur according to the characteristics of the first notch.

5.5.5 Spectral features analysis - Group B

In the following, the results shown in Figures 5.19 to 5.23 are discussed using the same nomenclature for angular regions as mentioned before in Table 5.4.

Many of the trends that were observed in Group A are also present in Group B. Regarding the first peak:

- The tendency that shows a positive slope of the Q-factor with respect to the center frequency of the peak is maintained, as for Group A.

- The tendency that shows that HRTFs for the above-front (black circle markers) and above-back (black square markers) directions seem to have a first peak with superimposing Q-factor and center frequency is also maintained, except for those subjects that did not perceive sound sources in the above-back directions like ME and YB -Figs. 5.19 and 5.20, respectively. Once again, the center frequency of the first peak for these directions seems to either concentrate in a narrow range of frequencies -subjects ME, SR and LA with Figs. 5.19, 5.22 and 5.23, respectively- or to extend in a broader one -subjects YB and AW with Figs. 5.20 and 5.21, respectively-.

- The results also suggest, even though not conclusively, that HRTFs from the back hemisphere (square markers) present a first peak with lower Q-factors than HRTFs from the frontal hemisphere (circle markers).

- The trend in which HRTFs from back-high (yellow square markers) and front-high (yellow circle markers) directions seem to be disambiguated by the Q-factor of the first peak is also maintained. More specifically, both regions appear as parallel distributions where the Q-factor for the front-high directions is higher, for a given center frequency. However, the trend cannot be distinguished in the results from subject AW -5.21- nor subject SR -5.22-, possibly due to the very few HRTFs that evoked directions in the front-high region to them. In the case of subject AW, it is interesting to note that the front (red circle markers) region is the one being disambiguated by concentrating the center frequency of the peak in a very specific range of frequencies and taking a Q-factor value which is also very restricted.

- Distributions for the front-high (yellow circle markers) Q-factors of the first peak are overlapping in some extent with those in the above-front (black circle markers) region -see Figs. 5.19, 5.20 and 5.23- or with the front (red circle markers) region -see Fig. 5.19-.

- The Q-factor of the first peak of HRTFs from the back (red square markers) and back-low (green square markers) regions seem superimposed, just as for Group A. The case of subject SR -Fig. 5.22- presents particularities, since the back region is distributed all along the frequency range of interest superimposing itself with the back-high (yellow square markers) region in a great extent.

From the above description, and with the exception of subject SR -Fig. 5.22- it is suggested that the first peak could disambiguate front-high from back-high regions. It could also disambiguate the above region, except for subject AW (Fig. 5.21). Therefore, the Q-factor of the first peak does not appear to be such a strong cue as it was for Group A.

With respect to the high frequency slope of the first peak, there was a trend in Group A in which the slope seemed steeper in peaks of HRTFs from the front hemisphere, specially when comparing front-high (yellow circle markers) and back-high (yellow square markers) region. That trend is not so clear in Group B.

The other parameters do not seem to present trends that are intuitively recognized. However, there are some individual characteristics. For example, the global Q-factor of the first notch for subject ME -Fig. 5.19-, shows that front (red circle markers) and back (red square markers) regions are disambiguated by it. Even though the notch is centered at the same range of frequencies, the global Q-factor value is lower for the front region. Furthermore, the parameters suggest that while the global Q-factor value determines whether the perception is to the front (circle markers) or to the back (square markers) , the center frequency of the notch determines whether the perception is in the front-low (green circle markers), front (red circle markers) or front-high (yellow circle markers) regions. The parameter does not seem to explain perception in the below-front (blue circle markers) and below-back (blue square markers) regions. Similar trends, though less marked, are evident in results for subject AW -Fig. 5.21-.

Subjects YB and LA -Figs. 5.20 and 5.23, respectively- presented too few HRTFs from which parameters could be computed. It is, therefore, difficult to draw any tendency from them. A trace of the trend shown for Group A is seen for those two subjects, and it is also evident in the parameters computed for Subject SR -Fig. 5.22-: back (red square markers) and back-low (green square markers) regions seem to be disambiguated by the global Q-factor parameter of the first notch.

Similar observations to those mentioned for Group A can be made for this Group regarding the second peak. In general, the tendency shown for Group A in which back and back-low directions would be disambiguated by the center frequency of either the first notch or second peak, seems to be maintained.

5.5.6 Valid ranges for the parameters computed from spectral features

One of the underlying hypothesis in binaural synthesis with non-individual HRTFs is that the cues for sound localization are the same for the great majority of listeners.

Within the framework of the current investigation, it would mean that the valid ranges for the parameters of the spectral cues are the same for all subjects -or, at least, they have some common ground. Figures 5.27, 5.28, 5.29 and 5.30 show two of the computed parameters pooled over all subjects and separated in Groups A and B. The figures show that the ranges are highly individual. Some ranges might evoke a certain direction for one subject and a different one for another subject -for example, it might evoke the above region for one subject and the back region for another, as in Fig. 5.29. It is hypothesized that each subject holds an inner balance, i.e. they allow a certain distortion of their own features without degrading localization performance. This idea is supported by the fact that, in the individual analysis, the parameters clearly appears in the results as working in ranges. This supports the findings reported by Wightman & Kistler [1993], who suggested that several representative HRTFs sets would be needed to cover the spectral matching requirements for a large sample of listeners.

5.5.7 Thresholds for the parameters computed from spectral cues

Regarding whether thresholds for the parameters could be defined, the results shown in 5.4.1 and 5.4.2 suggest that it would be very unlikely -at least in the context of the methods used for this experiment: the ranges usually overlapped defining 'transition zones' in which the evoked perception associated to a parameter changed. For example, it could change from being perceived to the front region to being perceived to the front-high region. This is expected as changes in perception occur smoothly, without abrupt jumps. On the other hand, what could be seen in the figures as 'transition zones' are actually the effect of having defined angles arbitrarily for the analysis. It is hypothesized that a different spatial sampling approach could help determining thresholds, but these would not be *natural*.

5.5.8 On how independent the spectral cues are from each other

The results shown in 5.4.1 and 5.4.2 suggest that the spectral cues have to be consistent with each other. In other words, in order to evoke a certain direction, the HRTFs pair needs to present peaks and notches whose parameters lie within the range for that direction. The topic was further analyzed for subject ME from Group B, by inspecting those HRTFs pairs that did not evoke any direction or for which the spectral features could not be parameterized. The cases for which no parameters could be computed can be categorized in two groups. The first group gathers all those HRTFs that, in their original form, present a low-pass filter characteristic. These are typically HRTFs from the above

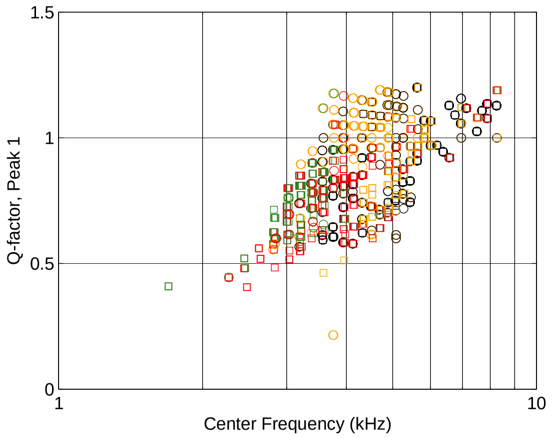

Figure 5.27: Q-factor of the first peak for subjects in Group A, pooled results.

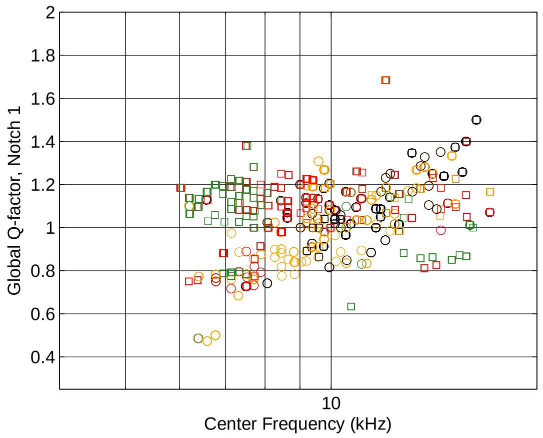

Figure 5.28: Global Q-factor of the first notch for subjects in Group A, pooled results.

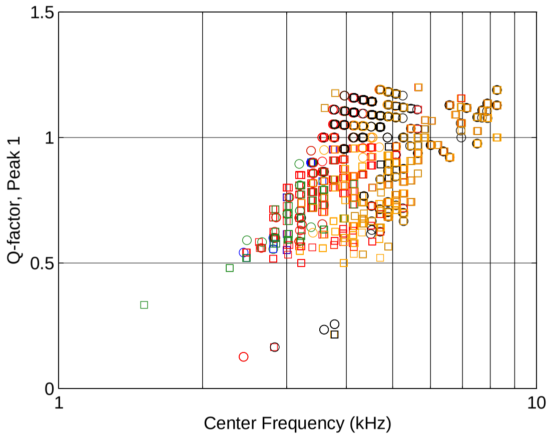

Figure 5.29: Q-factor of the first peak for subjects in Group B, pooled results.

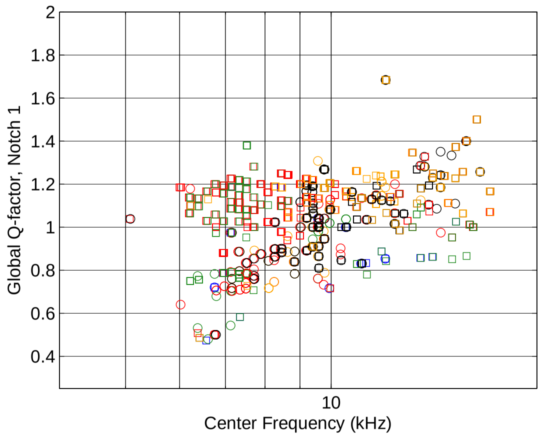

Figure 5.30: Global Q-factor of the first notch for subjects in Group B, pooled results.

directions. The second group gathers all those HRTFs that present a low-pass filter characteristic after windowing the first samples.

From the 165 HRTFs pairs tested in binaural synthesis condition (individual and non-individual), there were 31 pairs of HRTFs for which the first notch and second peak could not be parameterized -for either side of the HRTFs pair. 24 of those HRTFs pairs evoked a direction, while 7 did not. Analysis of the 24 pairs that evoked a direction showed that in many cases the Q-factors of the first peak for both sides in the HRTFs pair were close together in value and center frequency. In all cases, even if they were farther from each other, they were well within the range of values for a certain direction -for example, back-high. However, this also seemed to be the case for some of the 7 pairs of HRTFs that did not evoke any direction, suggesting that there were some other factors that played a role.

Furthermore, there were 33 pairs of HRTFs, out of the 165 total pairs, for which the parameterization could not be done for only one side in the HRTFs pair. From these 33 pairs, there were 24 pairs that evoked a direction and 7 pairs that did not. Analysis of the pairs that evoked directions showed that most of the directions lied in the back-high and back regions. In general, the parameters of the first peak for both sides in the HRTFs pair were very close together both in center frequency and Q-factor. The global Q-factor of the first notch for the HRTF side that could be parameterized lied in a well defined trend. This suggests that a strong first peak cue present in both sides of the HRTFs pair took the lead as spectral cue, while being supported by a consistent first notch global Q-factor cue. In the case of the HRTFs that did not evoke a direction, the Q-factors of the first peak were far from each other in some cases and far from the valid ranges of parameters for the subject in some other cases. On the other hand, the ranges in which the parameters for the first peak and first notch lied were not consistent with each other -for example, in one case the Q-factor of the first peak would lie in ranges for above directions (one side of the HRTFs pair) and back directions (other side of the HRTFs pair), while the global Q-factor for the first notch would lie in ranges for front-high directions.

Regarding the final 101 HRTFs pairs, for which all parameters could be computed for both sides, there were 75 pairs which evoked a direction and 26 which did not. Most of the parameters computed for those 75 pairs which evoked directions were close to each other, and parameters for peaks and notches were consistent with each other. This was also the case for some of the parameters computed for HRTFs that did not evoke a direction, but mostly the parameters for peaks and notches were not consistent (they lied in ranges for one direction for the peak, and in ranges for a different direction for the notch).

A similar analysis was conducted for subject EMS from Group A. The tendencies were similar, but the number of HRTFs that evoked a direction changed. From the 165 pairs of HRTFs used in binaural synthesis, there were 31 for which neither side of the pair could be parameterized beyond the first peak. From those, 10 pairs evoked a direction and 21 pairs did not. From the total 165 pairs, there were 36 pairs for which one side could not be parameterized. From those, 21 pairs evoked a direction and 15 did not. There were 98 pairs of HRTFs for which both sides could be parameterized. 52 of those evoked a direction and 46 did not. The differences with subject ME from Group B are based, mainly, on differences between the individual HRTFs characteristics.

This analysis suggests that cues within a HRTFs pair need to be consistent with each other -i.e. they do not work independently. In other words, it is a broad range of spectral features that needs to be in place to evoke a certain direction. This is not surprising, since Carlile *et al.* [1999] already mentioned a degree of *'redundancy of information'* in the spectral cues of HRTFs: they concluded that the high frequency range contained information also available in the low frequency range. The existence of redundancy in one aspect makes it plausible to exist in more aspects. In this context, it is not surprising that directions from the back hemisphere could be disambiguated by either the first notch or second peak, or that directions above the horizontal plane needed also notches that lied within the *'valid ranges'* to evoke certain directions which seemed to be disambiguated by the first peak. Redundancy could also explain why occluding single pinnae cavities do not disrupt completely localization: the contribution of the other cavities would also play a role (Gardner & Gardner [1973], Gardner [1973]). It would also seem that there is a hierarchy, as for some directions there is a strong first peak cue which leads sound localization. It is also interesting to discuss these assumptions in the light of the results reported by Wightman & Kistler [1997]. In their study, they concluded that even though monaural spectral cues were important for elevation perception, a true monaural localization paradigm eliminated localization. Localization impairment seemed to be related to the degree of attenuation at the occluded ear. This suggests that there could be an interaction between the monaural cues at each ear, which either establishes a hierarchy for each subject or works with a mechanism in which cues from one ear complement cues from the other one.

It would be interesting to extend these hypotheses with further work, which will also help explaining why some HRTFs that present a first peak and first notch with parameters within the *proper* ranges, do not evoke a direction. The answer could lie in the spectral features that were left out in the present analysis (Iida *et al.* [2007] suggested the second notch as one of the relevant spectral cues). It has to be remembered that the windowing imposed to HRIRs to analyze the first samples only, in great extent determined the values of the parameters computed. If the windowing changes, the values of the parameters change and some further relationships could be found. Possible interactions between the monaural cues at both ears could also be investigated.

5.5.9 Spectral features as cues for sound localization

As analyzed in 5.4.1 and 5.4.2, the results suggest that the first peak in the HRTFs, and particularly its high frequency slope, would disambiguate directions front-high, above and back-high, and that the global Q-factor of the first notch would disambiguate front-low, front, back-low and back directions -apart from presenting redundant information about above and high regions. The second peak seems to be also relevant in disambiguating directions back-low and back, but it is not clear how this cue interacts with the global Q-factor of the first notch and which of these two cues would have more weight. Summarizing, these three parameters would serve as cues to sound localization.

In general, the result that the first and second peaks, and the global Q-factor of the first notch (where the first and second peak play dominant roles) are candidate cues is in good agreement with previous works. For example, Humanski & Butler [1988] proposed spectral peaks from HRTFs as possible cues, which was later supported by the work of Moore *et al.* [1989], who suggested that peaks were more salient than notches as spectral cues as they were more easily perceived. It is also in agreement with Macpherson [1994], who stated that changing the f_c of the notch was not sufficient to evoke consistent localization in elevation in the MSP. It also agrees with Langendijk & Bronkhorst [2002], who showed that flattening the narrow frequency band where the notch would lie did not affect localization. The results presented here are also partly consistent with Iida *et al.* [2007], who proposed the first peak and first two notches of the HRTFs as cues for elevation in the upper MSP. A difference between their results and those from this Chapter is that the latter suggest that the first peak alone would be a candidate cue for directions above the horizontal plane -even though a trend for those directions can also be seen in the global Q-factor of the first notch. In turn, Iida *et al.* [2007] reported that the first peak did not change with direction, which was not the case for the HRTFs analyzed in the present study. Another difference with Iida *et al.* [2007] is that the second notch was not tested in the experiment presented here -but the second peak was included, which partly determines the second notch: in HRTFs, high frequency spectra is characterized by an alternation of peaks and notches. It is hypothesized that analyzing more spectral features would potentially explain the unanswered questions from the present study.

On the other hand, the frequency ranges in which the spectral features and parameters computed lie are broader than 1 -octave, which is consistent with the reports made by Macpherson & Middlebrooks [2003], and Langendijk & Bronkhorst [2002]. Those two works concluded that candidate spectral features were broader than $\frac{1}{2}$ -octave, they would possibly be 1-octave wide.

Regarding the spectral features identified here and how they would cue elevation perception, there are more interesting coincidences with previous works. For example, Hebrank & Wright [1974b] proposed a notch with lower cut-off frequency between 4 kHz and 8 kHz as cue for front directions, with increased energy above 13 kHz -this was in the line of the findings reported by Blauert [1969/70]. Their notch would correspond to the first notch and its global Q-factor identified in the present work, even though the range of frequencies for the f_c of the notch seen here goes from 5 kHz to around 16 kHz. It has to be noted that Asano *et al.* [1990] proposed the band around 5 kHz to 10 kHz as playing a central role in elevation perception, particularly due to a cross-frequency band power comparison. That finding is comparable, to a certain extent, to the global Q-factor defined here since the peaks at both sides of the notch take a central role, perhaps enabling such a power comparison. This could also be discussed in the light of Langendijk & Bronkhorst [2002] results, who showed that removing the narrow frequency band of the notch between 5.7 kHz to 11.3 kHz did not affect localization: that frequency band does not necessarily removes the first and second peak, still enabling a spectrum comparison. Furthermore, Asano *et al.* [1990] proposed that the microscopic characteristics around 2 kHz together with the macroscopic features above 5 kHz helped disambiguating front-back perception. This is also in agreement with the results presented here: HRTFs from different subjects do not change much below 2 kHz, and only the macroscopic details of HRTFs were used to compute the parameters - which in turn seemed to explain front-back perception.

The frequency band identified here for the f_c of the first notch (from 5 kHz to around 16 kHz) is also consistent with the findings of Langendijk & Bronkhorst [2002]: they suggested that the frequencies that more prominently cued front-back directions in the MSP lied in the range from 8 kHz to 16 kHz -but removing cues below 8 kHz also generated front-back reversals in their results. Up-down cues seemed to be located, according to Langendijk & Bronkhorst [2002], in the range from 5.7 kHz to 11.3 kHz. That range is relevant for the parameters computed in this Chapter. It has to be noted that Langendijk & Bronkhorst [2002] reported that frontal directions contained a peak in the range from 8 kHz to 16 kHz that was not present in rear directions and suggested that it could be a possible cue. This could correspond to the second peak parameterized in this Chapter -which however is present for other directions but with different center frequencies. A small peak in the same region (from 10 kHz to 12 kHz, with decreased energy above and below the peak) was suggested as a cue for behind by Hebrank & Wright [1974b] and Blauert [1969/70]. Therefore, the particular spectral feature that cues that range is controversial unless a feature like the parameters computed here are used for disambiguation. In that context, the frequency range proposed by both Langendijk & Bronkhorst [2002] and Hebrank & Wright [1974b] are in good agreement with the findings presented here - and the global Q-factor of the notch does take the peak in the range into consideration.

According to Hebrank & Wright [1974b], the general features of a $\frac{1}{4}$ -octave peak between 7 kHz and 9 kHz, could act as cue for the above directions - and Blauert [1969/70] also reported a directional band for above directions centered at 8 kHz. As already mentioned, this range is included in that reported by Langendijk & Bronkhorst [2002] for a peak as cue for frontal directions. In complex signals like the ones presented here, however, the first peak seems to encode enough information to distinguish the above directions from the high-front and high-back ones. This latter finding seems to be supported by Han [1994] who considered that the low frequency slope of a notch cued elevation in the frontal plane, which could be related to the high frequency slope of the first peak presented in this Chapter, depending on how it is computed.

5.5.10 Possible applications

Even though the ranges for the parameters of the spectral features are not the same for all subjects, it is interesting to note that the relationship between the parameters from different regions seem to be maintained. This is potentially useful for applications of binaural technology - and it is relevant here to describe alternative implementation strategies since the overall goal of the investigation is to work towards sets of non-individual HRTFs that produce better performance in sound localization tasks than the current ones. While it has been shown by Zahorik *et al.* [2006] that training seems to be an option for commercial implementations to reduce the rate of localization errors (front-back confusions, in particular), the results of the experiment presented here could be used for an alternative *'individualization'* procedure. It would be similar to a calibration procedure, where the system would need to set the first peak and its Q-factor to a center frequency and value, respectively, for representative HRTFs - for example, one direction in the front-high region and one direction in the back-high region- according to the user's input. The relationships found in the investigation here could be used to set the spectral features for the other HRTFs spanning from front-high to back-high. A similar procedure could be conducted for the first notch and its global Q-factor, which would individualize those HRTFs for directions front-low, front, back-low and back. If only directions in the back hemisphere are needed, tuning the second peak would seem enough to evoke different ranges of elevation.

The strategy presented above is related to one of the first ideas that were tested informally at the initial stages of this Ph.D. project: manipulating spectral features. Extensive listening testing with this author as only subject was conducted to preliminary evaluate whether spectral features manipulation could help disambiguating cues and resolving front-back confusions. The results were promising, even though they were excluded form this Thesis for lacking statistical design. Manipulation of spectral features is suggested as further work to extend the present experiment by using the results as it has been described before. More specifically, the relationship between center frequency and Q-factor of the first peak and the relationship between center frequency and global Q-factor of the first notch could be exploited to enhance existing cues or provide new ones.

The results presented in this Chapter are applicable to the MSP, only. As concluded by Carlile & Pralong [1994], different spectral features may be employed as cues to elevation, which depend on the vertical plane being considered.

5.6 Conclusion

This Chapter investigated whether spectral features in HRTFs that cue sound localization in the MSP could be identified and parameterized, how much spectral mismatch was allowed by the subjects without compromising localization, and whether it would be feasible to construct a non-individual set of HRTFs that provides the necessary spectral cues to sound localization for a large sample of users. Listening experiments were conducted with 10 subjects, in which the results were analyzed individually in search of spectral cues to sound source localization. Individual and non-individual HRTFs that evoked similar directions were windowed to 20 samples length ($fs = 48\ kHz$) and their spectral characteristics were parameterized. It was found that, for some subjects, the first spectral peak helped disambiguating directions above the horizontal plane: front-high, back-high and above. Front, back and back-low seemed disambiguated by the global Q-factor of the first notch in the magnitude response. The second peak was seen to convey redundant information to disambiguate directions in the back hemisphere. The described trends were not present for all subjects even though they were consistent with evidence from the literature. It was also seen that the parameters took ranges of values and frequencies without degradation in the perception of direction. These working ranges seemed to be individual, ruling out the possibility of a unique non-individual set of HRTFs that overperforms current ones. Parameters were also seen to change smoothly in value and frequency as different regions of directions were evoked, and establishing thresholds for the parameters seemed not natural. Finally, it was seen that the spectral features that act like cues belonged to broad ranges of frequencies (this was inherent to the definition of the relevant parameters) and that most cues within a HRTFs pair have to be consistent with the direction they are to evoke -in other words, redundancy would be necessary. In some cases, a single cue seemed to determine the evoked direction (for example, in those cases for which parameters could not be computed for one side in the HRTFs pair) and further studies are required to explain those cases. Similarly, it is still not understood why some HRTFs did not consistently evoked a direction even though their computed parameters lied in valid ranges of evoked perception. It is hypothesized that the answer to these questions will be found by analyzing more features of the HRTFs.

Chapter 6

Observations on human sound localization

6.1 Introduction

Binaural reproduction of sound synthesized with -or recorded through- individual and non-individual HRTFs has been traditionally assessed in terms of sound localization, where localization performance with real sound sources is taken as the reference. Relatively accurate sound localization performance with both real sources and individual HRTFs is a condition that binaural technology assumes *a-priori*, and which has been validated by studies such as that reported by Wightman & Kistler [1989b]. Localization with real sound sources and the associated errors have been reported by Oldfield & Parker [1984] and Carlile *et al.* [1997], among others. More emphasis has been given, however, to characterizing sound localization with non-individual HRTFs: partly due to the high rate of localization errors reported (Wenzel *et al.* [1993], Møller *et al.* [1996a], Chapter 5 of this Thesis, among others) and partly due to their significance in actual applications of the technology. In this scenario, little has been said regarding those subjects who fail localizing real sound sources and exhibit strong biases in their responses. These subjects present localization errors that do not classify within the definition of localization blur and do not correspond to the low rate of front-back confusions usually reported in the literature (Makous & Middlebrooks [1990], Carlile *et al.* [1997], Carlile *et al.* [1999], among others). Considering the definition of *good localizers* as those who perceive and report sound sources accurately where they actually are (real life condition) or in those directions where the individual HRTFs used for synthesis were actually measured (synthesis condition), and considering that *poor localizers* are those who fail to do this, the subjects which are the focus of this Chapter could be categorized as *biased localizers*. These subjects were found during the real sound sources condition screening performed to recruit participants for the experiment reported in Chapter 5. These subjects were considered *poor localizers*, but they interestingly presented characteristic biases in their responses: real sound sources were consistently perceived either in the frontal or rear hemisphere, or their localization was degraded in specific ranges of space. In some cases, for example, listeners had an equally biased perception with real sound sources and binaural synthesis with individual HRTFs, but still there was no correlation between the two distributions. Even though the classification between *good* and *poor localizers* seems to be well accepted (Begault [2000]), little has been reported on cases of *biased localizers*. It is the aim of this Chapter to report descriptively the localization performance from these subjects who presented strong biases, and discuss the topic in the light of possible applications of binaural technology. Five examples of these *biased localizers* are presented. It is not clear how these cases could be approached, for which no systematical statistical analysis is attempted. It seems important, however, to present these cases as observations on human sound localization so as to have a broader perspective of what binaural technique can realistically achieve. As a reference, two examples of subjects with degraded performance under individual HRTFs binaural synthesis condition are included.

This Chapter is organized as follows. Firstly, relevant literature where sound source localization is either quantitatively or qualitatively assessed is presented. Secondly, the methods used to obtain localization answers from *biased* and *poor localizers* are covered. Subsequently, results from the localization experiments are shown. A discussion follows where the importance of these cases in the context of binaural synthesis applications is analyzed. The Chapter ends with some concluding remarks.

6.2 Previous works

The existence of some degree of sound localization errors has been acknowledged throughout the literature. Blauert [1997] has defined the term *localization blur* as the spreading in space of localization answers derived from a single sound source. The span angles of the *localization blur* are direction and signal dependent. According to Blauert [1997], they have been reported as ranging from 4 ° to 17 ° for the elevation dimension in the front direction of the MSP. *Localization blur* -confounded with the inherent accuracy of subjects to report the sound source location with a given response paradigm-could be seen in Chapter 5, in those figures presenting localization answers with real sound sources under anechoic conditions. A different type of localization error is that of front-back confusions, which has been mentioned in the psychoacoustical validation studies that compared free field and individual HRTFs synthesis condition (Wightman & Kistler [1989b]). Front-back confusions are those in which the angle of the perception with respect to the MSP is correct, but the hemisphere of the target is confused. It is now well established that the occurrence of front-back confusion rises under binaural synthesis, at a minor degree with individual HRTFs and much more notoriously with non-individual HRTFs (Wenzel *et al.* [1993]).

The fact that there are individual differences in how accurately subjects can localize sound led to the classification between *good* and *poor localizers*, where *good localizers* are those subjects who can perceive a sound source at its correct position with reasonable accuracy. Begault [2000], for example, presented behavioral data from Wightman & Kistler [1989b] and regarded to it as typical of a *good localizer*. *Poor localizers*, on the other hand, are those subjects who have disrupted localization ability. The classification between *poor* and *good localizers* became important in the field of binaural synthesis since it was reported that HRTFs from *good localizers* could provide useful spectral cues for localization when used as non-individual HRTFs (Wenzel *et al.* [1993]). The idea that localization performance was related to spectral cues in the HRTFs of *good localizers* was suggested by Wightman & Kistler [1989b], who studied how the spectra changed in elevation for each of the subjects that participated in their studies. They showed that all the *good localizers* shared a trend, but that the one *poor localizer* in their study had a very particular and different tendency. The experiments presented in Chapter 5 of this Thesis are examples of localization performance from 10 subjects that were considered *good localizers*. As a requisite to participate in the experiments, subjects had to present similar localization performance with real sound sources and virtual sound sources synthesized with individual HRTFs. This was evaluated by computing the correlation between both distributions and testing the null hypothesis of no correlation between them by means of permutation tests.

The focus of this Chapter is, however, the localization errors made with real sound sources under anechoic conditions. In particular, errors which are beyond the *localization blur* are of interest. Gardner & Gardner [1973] reported a sound source identification experiment in free field where they mentioned encountering some *localization anomalies*, which they classified as belonging to three different types: image displacements of various magnitudes, apparent locations that tended to be independent of the actual location of the source in use, and nebulosity of source location where there was no definite sense of signal direction. They reported that these *anomalies* were more present as pinnae occlusion increased.

Some other studies have mentioned the occurrence of front-back and up-down confusions when localizing or identifying real sound sources. These confusions were usually quantified in percentage terms. For example, Carlile *et al.* [1999] reported front-back confusion errors under free field conditions when testing 76 locations in the ranges of ±180° in azimuth and ±40° in elevation. With broadband stimuli (from 0.4 *kHz* to 16 *kHz*), subjects presented front-back confusion rates that ranged from 0.3% to 4.7% with 2.3% as mean value. The percentages for a similar study reported by Carlile *et al.* [1997] ranged from from 0.7% to 7.9% with 3.2% as mean value. They reported that only a few front-back confusions were seen on the MSP. Makous & Middlebrooks [1990] reported a free field localization experiment where the average front-back confusion rate for six subjects was 6% and individual rates ranged from 2% to 10%. Stimuli had been bandpassed in the range from 1.8 *kHz* to 16 *kHz*. They also provided some interesting descriptive analysis: they reported that two of the subjects showed a significant majority of back-to-front confusions and other two subjects showed a majority of front-to-back confusions. Furthermore, one subject exhibited a large number of front-back confusions when sound sources were in the back-high region, while another had a high incidence of confusions for sources that were in the frontal hemisphere. Wenzel *et al.* [1993] tested 16 subjects in free field and non-individual binaural synthesis localization experiments. Free field front-back confusions rates ranged from 2% to 43% with a mean of 19%. Up-down confusions ranged from 1% to 26% with a mean of 6%. While two of the subjects presented poor performance in both conditions, another two presented degraded performance only under non-individual binaural conditions -trends which were obtained after resolving front-back confusions. In general, poor localization in the elevation dimension corresponded to compressing all the answers in a limited angle range.

Other studies reported generalities about *poor performance* of some of the participant subjects, in the line with the already mentioned description made by Makous & Middlebrooks [1990] for some of their subjects. For example, Butler & Belendiuk [1977] tested identification of real sound sources and virtual sources obtained through individual binaural recordings with 8 subjects. The experiments consisted of source identification studies where five sound sources were arranged in an arc ranging from $-30°$ to $30°$. Two of the subjects performed poorly -i.e. at chance level- in free field, and one of these two subjects performed poorly in the recording condition also. Oldfield & Parker [1984] reported an extensive localization study with real sound sources that could be located in positions with elevation in the ±40° range and with azimuth angles ranging from 0° to 180°. The eight participant subjects showed, in general, poorer sound localization performance in the back-high directions -which is in agreement with Makous & Middlebrooks [1990]. Bronkhorst [1995] tested localization of sound sources under real free field conditions and binaural synthesis with individual and non-individual HRTFs. Two paradigms of response were tested in order to evaluate head movements. Of interest to this Chapter are the reported results from a *confusion task* experiment where subjects had to report in which quadrant around them the sound was coming from. They were given 8 choices, and the closest to the MSP they tested were directions with azimuth at ±27.7° and ±152.3°. It was shown that under real free field conditions, the rate of confusions (particularly front-back) decreased as the spectral content of the signal included higher frequencies (they tested 8 kHz and 16 kHz as cut-off frequencies). Mean values of confusion rates for the free field condition were around 15%-20%. These results were in line with others that also reported an increase in front-back confusion with bandpass filtered signals (Carlile *et al.* [1999], Middlebrooks [1992]).

In order to avoid individual biases in behavioral results due to localization performance, some laboratories decided to provide training to prospective participants. For example, Middlebrooks [1992] trained the subjects that participated in his experiment by performing a localization task with broadband sound, with and without visual feedback, until the root mean square error was less than 12° in the horizontal and vertical dimensions. This suggests that the error of naïve subjects would be much larger.

One of the few cases where radically biased localization performance was presented is that of Itoh *et al.* [2007]. They tested localization of real sound sources in the MSP, in an anechoic environment. Sound sources ranged from 0° to 180° in 30° steps. For the wide-band noise stimuli condition, they showed that two of the seven subjects localized all sound sources to the rear hemisphere. One subject localized all the front-high and above sources to the front. The performance of these three subjects are very similar to those which will be reported here as belonging to *biased localizers*. Whether these degrees of bias have appeared previously in other localization experiments is not known: most of the studies either resolve the reversals or discard them before presenting their data, as discussed in Chapter 5, Section 5.5, under the analysis of the correlation results. The description made by Makous & Middlebrooks [1990] for two of their six subjects, however, fits the aforementioned concept of *biased localizers*.

It is not in the spirit of this Chapter to discuss the nature of front-back confusions nor possible ways to overcome them[1]. The original contribution of this Chapter is to present some observations made on biased human sound localization, which have received little attention in the literature. In the following, real sound sources localization performance from five naïve subjects under anechoic conditions will be presented. These subjects were considered *biased localizers*, since the errors they presented did not correspond to the *localization blur* that *good localizers* also presented, nor to the degree of front-back errors that has been reported for both *poor* and *good localizers*. Further examples of *poor localization* are the results from two participants who had degraded performance under individual HRTFs binaural synthesis condition -their results will also be shown here. The approach of this Chapter is purely descriptive. It builds on the current knowledge that states that the existence of reversals in the localization of real sound

[1] Zahorik *et al.* [2006] showed that front-back confusion rates could decrease by training, and Wightman & Kistler [1999] reported lower rates of occurrence when head movements were allowed.

sources under anechoic conditions provides evidence of a degree of ambiguity inherent to HRTFs spectral cues (Wenzel *et al.* [1993]). On this basis, the cases presented here are aimed to a broader understanding of how humans localize. It is not clear, for example, how common -or how rare- these cases are or how training would affect these subjects.

6.3 Methods

In this section, the methods used to obtain and describe the localization performance from *poor* and *biased localizers* will be presented. As these subjects were found through the screening for the experiments reported in Chapter 5, much of the methods is based on what has already been described there. Therefore, only a synthesis will be presented here and the reader will be referred to different parts of that Chapter for further details.

6.3.1 Subjects

Five *biased localizers* and two *poor localizers* -the latter, under individual HRTFs binaural synthesis condition- are included in this descriptive study. They all had normal hearing as tested by a standard pure-tone audiometry in the frequency range from 250 *Hz* to 8 *kHz*: none of the subjects had hearing thresholds above 15 *dB HL*. *Biased localizers* were those whose responses under real sound sources condition (and individual HRTFs binaural synthesis condition, if tested) were consistently biased towards a particular range of space. *Poor localizers* were those whose responses under real sound sources condition was equivalent to those from a *good localizer* but whose individual HRTFs binaural synthesis condition performance was degraded.

6.3.2 HRTFs measurements

HRTFs from 5 of the participants subjects (three *biased localizers* and the two *poor localizers*) were measured in the anechoic chamber as previously reported in Chapter 2, 2.8.1. The measurement protocol will not be repeated here and the reader is referred to that Chapter for a detailed description. Briefly, it can be said that HRTFs were measured at the blocked entrance of the ear canals from 15 directions in the MSP, ranging from -67.5° to 247.5° in 22.5° steps.

6.3.3 Localization of real sound sources

The loudspeaker setup, signal generation and control, and stimuli used have already been thoroughly described in Chapter 5, 5.3.1 and will not be repeated here for the sake of synthesis. The reader is referred to that Chapter for further details.

Procedure

During the recruiting process, subjects were assigned to either Group A or Group B, as already explained in Chapter 5. Therefore, some of the subjects presented here used the graphical interface of Fig. 5.1 while others used that of 5.2, both of which corresponded to two slightly different experimental procedures. Subjects IS and EM from the *biased localizers* group, for instance, proceeded with the experimental procedure described for Group A. That means, among other things, that they were asked to reflect on the externalization of the perceived sound and were presented with 15 repetitions per sound source. On the other hand, subjects PG, MK and KAK from the *biased localizers* group had been recruited with the objective of participating according to the experimental procedure described for Group B. Therefore, they were not asked to reflect on externalization, and were presented with 16 repetitions per sound source. This was also the case for the two subjects from the *poor localizers* group. The reader is referred to Chapter 5 for further details.

6.3.4 Localization of virtual sound sources - individual HRTFs

The headphones setup, signal generation and control, and stimuli have already been described in Chapter 5, 5.3.1.

Procedure

For each subject, the procedure was consistent with the corresponding one used under real sound sources condition. From the five *biased localizers* that participated, three were invited to test individual HRTFs binaural synthesis condition: subjects EM, MK and KAK.

6.4 Results

Figures 6.1 to 6.5 show localization judgements for the 5 *biased localizers*: they not only did not present a good localization performance with real sound sources, but they also showed evidence of strong biases. Subjects EM, MK and KAK (with Figures 6.3, 6.4 and 6.5, respectively) were invited to test the individual HRTFs binaural synthesis condition. Their measured HRTFs are among those shown in Figure 2.14, Chapter 2.

Figures 6.6 and 6.7 correspond to two other subjects whose real sound sources localization performance was acceptable according to the classification of *good localizers*, but their individual HRTFs binaural synthesis performance was either strongly biased (subject EK, Fig. 6.6) or with isotropic tendencies (Fig. 6.7).

Figures 6.1 to 6.7 follow the guidelines already used for figures in Chapter 5 and Appendix B. Figures 6.1 and 6.2 correspond to real sound sources localization from subjects IS and PG, respectively. Figures from 6.3 to 6.7, on the other hand, show panels for real sound sources and individual HRTFs binaural synthesis conditions. These are respectively titled *Real Life* and *Individual*. Subject codes are given in the legend of each figure. The localization judgements shown correspond to the raw data, being the original direction (that representing the actual position of the sound sources) in the abscissa and the perceived direction in the ordinate. The answers plotted in red correspond to those cases in which subjects considered that they perceived the sound inside their heads. Those situations corresponded to either reporting the answer inside the head in the touchscreen -i.e. clicking inside the head- or answering NO to the question *'Did you perceive the sound outside your head?'*. In the figures, the main plots correspond to the perceived elevation while the small plots inside them correspond to perceived azimuth.

6.5 Discussion

Figures 6.1 to 6.5 show three different and well defined trends of perception under anechoic conditions with real sound sources: subjects IS and KAK perceived all the sources in the rear hemisphere, subject MK perceived all the sources in the frontal hemisphere, and subjects PG and EM presented individual spatial ranges in which they had degraded perception -the above region for subject PG and the front-low region for subject EM. The cases seen here are very similar to those subjects with disrupted localization ability reported by Itoh *et al.* [2007].

For those subjects that tested individual HRTFs synthesis, it can be seen that the general trends from real sound sources localization were maintained: subject EM perceived all synthesized sources above the horizontal plane, subject MK had all the answers compressed in the front-high and above ranges, and subject KAK perceived most of the synthesized sources to the back hemisphere. Correlation values were computed between real life and virtual individual distributions, but in none of the cases was the null hypothesis $\hat{\rho}v = 0$ rejected. In other words, the trends were maintained but they were not strong enough as to reject the null hypothesis of no correlation between distributions. It is interesting, however, to see that the technology worked with a similar mechanism as that reported in Chapter 5: binaural synthesis evoked a similar, but somehow scattered, perception to that with real sound sources.

Following the convention adopted across this thesis, red circles in Figs. 6.1 to 6.5 indicate those answers that were reported as perceived inside the head. Subjects IS and EM had to actively relate to sound source externalization due to the paradigm they were being tested with. It can be seen that they presented a high rate of answers reported as inside-the-head even with real sound sources. It is hypothesized that these *biased localizers* have little awareness of their own perception and perhaps their own self. It would be interesting to explore the psychological aspects involved in localization, and whether difficulties to define one's own body boundaries can translate to difficulties assessing a sound source as external. It would also be interesting to analyze the recurrence of auditory hallucinations with respect to sound source externalization. Subjects PG, MK and KAK had a much lower rate of inside-the-head answers in the real sound sources localization test, which is in line with the difference already seen between Groups A and B - i.e. subjects perceived sound much more naturally as outside-the-head when they did not have to think about it, see Table 5.3 in Chapter 5. Regarding binaural synthesis condition, and for those subjects that were tested, the rates of answers that were not externalized was higher than for real sound sources. This was also expected as it agrees with the findings of the experiment in Chapter 5.

It is interesting to note that subjects could see the sound sources during the real sound sources localization experiment. They were asked to report where the sound came from, regardless of the position of the loudspeakers. Poor localization performance with real sound sources under anechoic conditions is nothing new, as already discussed in Chapter 5, Section 5.5. However, some subjects reported some sounds as coming from such unexpected directions as from the lateral angles (subject MK), even though they were being tested in the MSP. In a personal and informal communication with these subjects that had degraded localization with real sound sources, they reported that they were being cautious as they thought they were being *cheated* by the experimenter. It is hypothesized that their own high expectations to outperform in the test could have been a strong cognitive aspect affecting their perception. Others mentioned that the aesthetic of the anechoic environment was *too surreal*, probably playing a role in the cognitive processes of the subjects.

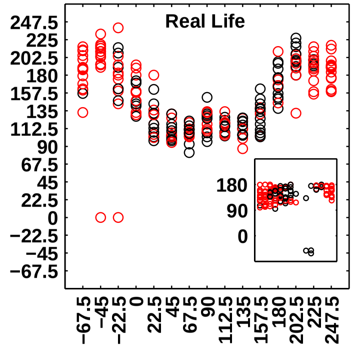

Figure 6.1: Subject IS, see text for further details.

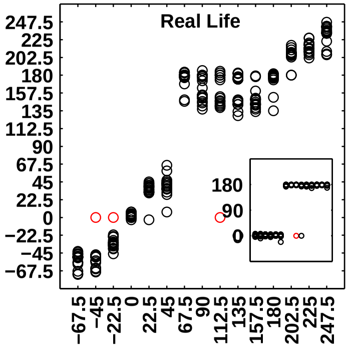

Figure 6.2: Same as Fig. 6.1, but for subject PG.

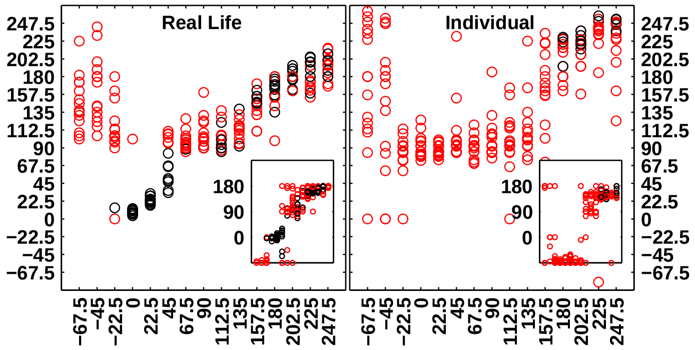

Figure 6.3: Subject EM, localization performance under real sound sources condition (left) and individual HRTFs binaural synthesis condition (right).

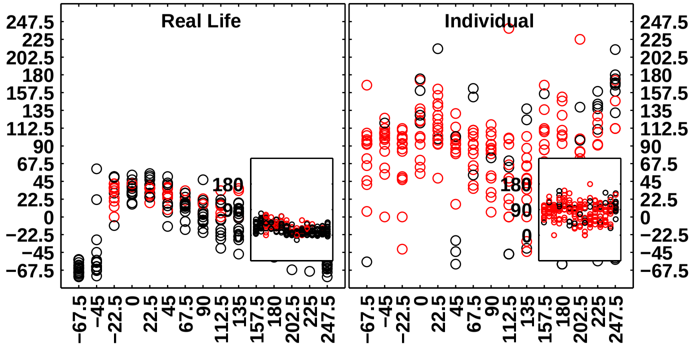

Figure 6.4: Same as Fig. 6.3, but for subject MK.

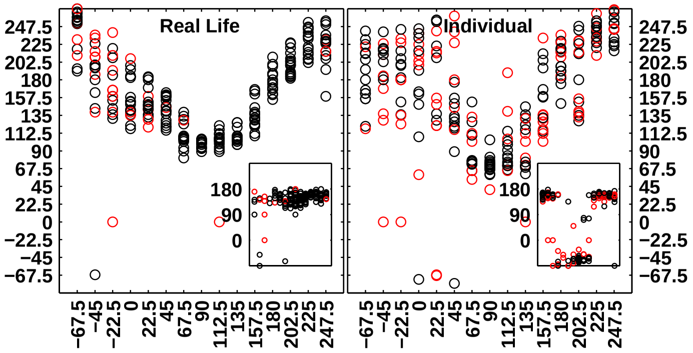

Figure 6.5: Same as Fig. 6.3, but for subject KAK.

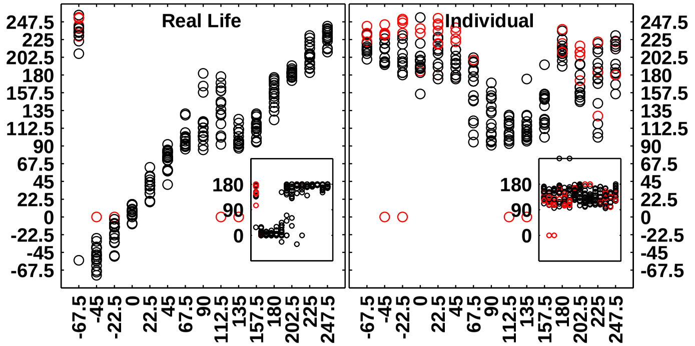

Figure 6.6: Same as Fig. 6.3, but for subject EK.

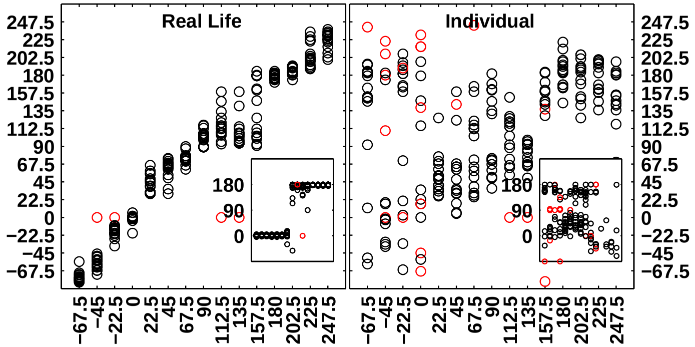

Figure 6.7: Same as Fig. 6.3, but for subject KK.

Two examples of *poor localizers* under individual HRTFs binaural synthesis condition are shown in Figures 6.6 and 6.7. These were subjects whose real life localization performance was considered relatively good, but the performance under individual HRTFs binaural synthesis condition was degraded. These subjects were not tested under non-individual HRTFs binaural synthesis condition. Figures 6.6 and 6.7 show localization judgements from these subjects, for both real sound sources and binaural synthesis with individual HRTFs conditions. Real sound sources performance was accurate except for one case of front-back confusion presented by subject EK. Subjects were not asked about externalization, for which the rate of inside-the-head answers under both real sound sources and binaural synthesis conditions was low. Regarding the latter condition, it can be seen that subject EK perceived all synthesized sources to the back hemisphere while subject KK presented a high occurrence of front-back confusion, tending towards isotropic distributions.

Results from *biased* and *poor localizers* are shown here to emphasize that *good localization* is not an inherent characteristic of human beings -partly due to the ambiguity of spectral cues in human HRTFs, as already mentioned. As suggested before, there might be not only cognitive circumstances but also emotional or psychological ones. It appears that, in general, the technology successfully recreates binaurally the events that subjects hear in real life -though it was not the case seen in Figs. 6.6 and 6.7. That means that the target of the technology should not be to evoke those directions for which the HRTFs are measured for, but to evoke those directions that the subjects can actually perceive in real life. This rises the natural question of why previous validation studies correlated perceived and target directions for both real and virtual sound sources separately, instead of correlating perceived real and perceived virtual sound sources -as reported in Chapter 5 of this Thesis. In this line of thought, the current classification between *good* and *bad localizers* seems debatable. A subject is considered a *good localizer* if he/she can identify accurately the actual position of a sound source, and it is assumed that his/her spectral cues will provide the tools so that his/her localization performance with individual binaural synthesis is comparable to that in real life. Whether *good localization* in real life is the standard for the average listener, is still not clear. Whether the reasons for *good localization* actually have an acoustical basis -i.e. in the HRTFs spectra- is not clear, neither. Furthermore, it would appear sensible to use binaural synthesis to emulate real life localization. In such a context, trying to provide fully *accurate* localization performance with binaural synthesis to a subject that has severely *degraded* localization performance with real sound sources is, at least, unnatural. This would call for a redefinition of the goals of binaural technology and what a true realistic virtual experience is.

Another perspective, which has not been mentioned before in this Thesis, refers to the physiological and cultural aspects that affect hearing and which happen to be closely related. Binaural technology emerged and became more or less popular in an era where the use of personal music players was not broadly extended. However, nowadays the use of MP3 players is the norm across different age groups and awareness of the physiological consequences -i.e. hearing damage- of their abuse is raising. Together with an increased rate of sound pollution and unregulated exposure to high sound pressure levels during leisure activities, fast changes in the hearing sensitivity of people and the general way to perceive and respond to sound would be expected. Whether these changes could impair particular functions like sound source localization, even though having normal hearing, is not clear and should be further studied. This could be, for example, in the form of an updated topography of auditory space as investigated by Oldfield & Parker [1984].

6.6 Conclusion

This Chapter investigated the localization ability of subjects who presented strong biases when localizing real sound sources, focusing on whether localization errors could be explained, under which circumstances sound direction was being *correctly* evoked, and whether the conceptual duality *poor localizers* vs. *good localizers* was appropriate in the context of binaural technology. Behavioral results from localization experiments with real sound sources were shown for five *biased localizers*, from which three were invited to further test individual HRTFs binaural synthesis conditions. Two subjects perceived all the real sound sources in the rear hemisphere, one subject perceived all the real sound sources in the frontal hemisphere, and the remaining two subjects presented individual spatial ranges in which they had degraded perception -for one of them, it was the above region while for the other, it was the front-low region. The general trends seen with real sound sources were maintained under binaural synthesis conditions with individual HRTFs, even though no correlation was seen between corresponding distributions. Results from two other subjects which presented *good* localization performance with real sound sources but degraded performance with individual HRTFs binaural synthesis were also shown. The cases were presented descriptively, as it is not clear how representative they are or how to approach them analytically. Antecedents are three subjects reported by Itoh *et al.* [2007], and perhaps two subjects from the study reported by Makous & Middlebrooks [1990]. It is hypothesized that the basis for the biased responses could be either acoustical, cognitive, emotional or cultural/physiological, and further studies should be made to bring more light into the topic. It was also concluded that the technology seemed to evoke a real life situation for these subjects and that the dichotomy *poor* vs. *good localizers* did not allow them a place within binaural technology.

Chapter 7

On the perceived quality of sound synthesized with non-individual HRTFs

7.1 Introduction

Spatial sound synthesized with non-individual HRTFs has been mainly assessed through localization errors across the literature (Wenzel *et al.* [1993], Møller *et al.* [1996a]). In Chapter 5, it was shown that subjects could allow a certain range of distortion to their own HRTFs without compromising localization performance. It remains unclear whether the non-individual HRTFs that successfully provided localization cues would compromise, however, the perceived quality of sound: the study of the potentially degraded sound quality due to the use of non-matching spectra has been given little importance in the literature. This Chapter focuses on perceptual aspects beyond sound source localization that are relevant in the implementation of binaural synthesis. These aspects are of varied nature. On the one side, there exist subjective characteristics that are related to what the individual listener considers as *good quality*. In this context, the listener has his/her inner reference. On the other side, there are quality characteristics that are measurable: fidelity of music and intelligibility of speech, for example. These are also related to the subjective perception of sound quality. Due to time constrains inherent to the Ph.D. study, this Chapter presents a theoretical approach to the problem. This approach is based on the experiences obtained from the listening experiments conducted for Chapters 5 and 6. Informal communication with the participant subjects of those investigations revealed that perceived sound quality issues were relevant when using non-individual HRTFs, which was in line with the *a-priori* hypothesis that those issues could exist. Furthermore, the filtering that non-individual HRTFs imposed to the own HRTFs of the listeners was computed for all the resulting HRTFs from the matching procedures reported in Chapter 5. It was seen that, according to the evidence in the literature (Moore & Tan [2003]), the filtering imposed would be audible in terms of perceived quality of sound.

This Chapter is organized as follows. Firstly, a few relevant studies on the topic of perceived quality of sound are reviewed. Secondly, the theoretical framework and the development of a research hypothesis are presented. The Chapter ends with an outline of a possible course to investigate the proposed research hypothesis.

7.2 Previous works

While much emphasis has been given to the degradation of sound localization due to the use of non-individual HRTFs (Wenzel *et al.* [1993], Møller *et al.* [1996a]), little has been said about how the perceived quality of sound would be affected under such conditions. Sound quality in binaural synthesis was marginally mentioned in Chapter 2 of this Thesis and by Hammershøi & Møller [2005], where it was stated that sound quality could be affected by not controlling the low frequency spectral characteristics of HRTFs. This, however, is inherent to the HRTFs measurement procedure and affects both individual and non-individual HRTFs. The lack of studies on the perceived sound quality in the context of binaural synthesis is interesting if compared to the attention given to the perceived quality of sound with loudspeaker reproduction. From a structural point of view, the distortion that subjects allowed to their own HRTFs spectra without compromising localization, as already reported in Chapter 5, can be compared to the non-flat frequency response of a loudspeaker in a reproduction chain. In that analogy, whether the distortion would be audible as coloration or would introduce some other perceptually relevant characteristic, is yet to be studied. In the following, a few relevant previous investigations are presented which will help developing a research hypothesis for further tests in the topic.

Begault *et al.* [2001] evaluated the perceived *realism* in binaural synthesis with individual and non-individual HRTFs, by testing three different scenarios: anechoic environment, virtual environment with early reflections simulated, and virtual environment with full reverberant simulation. The *realism* could be evaluated in a scale from 1 (least realistic) to 4 (most realistic), with bad, poor, fair, good and excellent as anchor points. They used speech as test signal, and they did not provide an interpretation of *realism* to the subjects. They tested the effect of reverberation type simulated in the synthesized scenario, the effect of individual and non-individual HRTFs and the effect of head tracking. Their results showed a lack of variability that meant that participants could either not differentiate among conditions in order to evaluate perceived *realism*, or they did not have a common understanding of the concept of *realism*. It was concluded that *realism* would be a more important cue with other types of signals, like music.

Moore & Tan [2003] conducted a study to quantify how the perceived *naturalness* of music and speech signals were affected by different forms of linear filtering applied to their amplitude response. They evaluated spectral tilts, bandpass filtering and inclusion of spectral ripples, and the modifications affected different frequency ranges up to 7 kHz. Signals were speech and jazz music segments reproduced through headphones. They tested 10 subjects who had to report the perceived *naturalness* in a scale from 1 to 10, with 1 being *'very unnatural - highly colored'* and 10 being *'very natural - uncolored'*. They found that the perceived *naturalness* decreased as ripple depth increased from 5 dB to 15 dB, and that ripples were more quality disrupting when they covered broad ranges of frequency. Speech was also affected when the ripples were in the mid or low ranges. For tilts, the perceived *naturalness* decreased with increasing tilt magnitude, independently of the sign of the slope. Tilts were also more disrupting when they occurred over wide frequency ranges. Mid and low frequency ranges seemed to have a more central role in *naturalness* perception than high frequency ranges. When tilts and ripples were combined, *naturalness* was most affected when both alterations occurred in wide frequency ranges. Regarding bandpass filtering the signals, it was seen that keeping the range from 55 Hz to 16.8 kHz gave the higher rates of *naturalness* for music - narrower bands introduced decreased *naturalness*. A similar effect was found for speech signals, but reducing the band to the 123 $Hz - 10.8\ kHz$ range had little effect in the perceived *naturalness*.

Pedersen & Zacharov [2008] reported an overview of perceptual attributes used in different areas of acoustics, for a variety of purposes. Since sound characterization is very much domain specific, the authors covered how new attributes could be developed and evaluated for each particular interest by means of sensory descriptor development and sensory evaluation methods. New attributes would supplement existing ones, some of which have a correlation with physical metrics: for example, the psychoacoustical attributes loudness and loudness level have the sone and phon as their respective metrics. Pedersen & Zacharov [2008] described the semantic space of sounds -i.e. the vocabulary used to describe the perception of sounds. Of particular interest to this Chapter is their comparison of sensory descriptors used in different studies evaluating sound reproduction systems -including spatial systems. It could be seen that, throughout the literature, spatial sound reproduction has not only been evaluated in terms of sound source direction -which relates to the descriptors *localization* and *sense of direction*. While evaluating such descriptors has been the norm in the field of binaural synthesis with non-individual HRTFs (with the exception of the previously mentioned work reported by Begault *et al.* [2001]), the overview provided by Pedersen & Zacharov [2008] shows that spatial sound reproduction allows for a much richer evaluation in terms of the perceived quality of sound.

Lorho [2010] conducted an extensive research project on perceived quality of spatial sound reproduced through headphones. The main focus of his research was to look for suitable test methodologies for the reliable measurement of the perceived characteristics of audio systems and part of his work was included in the review reported by Pedersen & Zacharov [2008]. The quality evaluations conducted by Lorho [2010] focused on a variety of algorithms for headphones spatial enhancement, some of which made use of non-individual HRTFs, but the specific role of the spectral mismatch was not evaluated. In any case, if these mismatches played a role, it was confounded with the general results of overall perceived quality. Of relevance to this Chapter, Lorho [2010] reported a series of attributes to be used in quality assessment of spatial sound reproduced through headphones, and which were grouped as follows: tone color, timbral aspects, localization aspects, room perception, externalization, broadness, artifact aspects, temporal aspects, and other aspects such as realism or naturalness. It has to be noted that the descriptors in each of these groups were chosen so that they could be related to different reproduction techniques and algorithms, and therefore it could be expected that not all of them would be relevant when comparing perceived quality of sound synthesized with individual and non-individual HRTFs. On the contrary, it could be the case that new parameters would be required for this latter case. Another relevant finding reported by Lorho [2010] was that, when comparing overall quality of spatial reproduction algorithms for sound reproduced over headphones, timbral degradation was seen to play a more important role in overall quality judgement than spatial characteristics. The author hypothesized that individual differences would also be important when evaluating perceived quality.

7.3 Informal observations on sound quality

Some of the subjects which participated in the experiment described in Chapter 5 informally reported some attributes of the binaurally synthesized sound they were presented with, particularly when non-individual HRTFs had been used to synthesize the sound samples. Subjects tended to associate the perceived qualities to sound localization, but possibly because the latter was the only parameter they had to assess. Subjects informally reported the following:

- Some sounds were perceived as *’uncomfortable’* or as *’not sounding good’*. Some subjects said that these sounds were more difficult to localize.

- Some subjects reported that there were sounds characterized by a *'hiss'* among the non-individual HRTFs presentations.
- Some subjects reported loudness as a cue associated to direction and distance perception: some non-individually synthesized sound sources were perceived louder and very close to the surface of the face, being these sounds easier to localize.
- Other subjects reported using the loudness cue in the way that less loud sounds were localized in the rear hemisphere.

It was interesting to receive these comments, which were consistent for several subjects and which came spontaneously from them -i.e. without being directly asked about sound quality issues, but about their general feeling on the task. It is hypothesized that the *'uncomfortable'* sound samples belonged to those HRTFs that did not evoke a direction consistent with the perception under individual HRTFs binaural synthesis conditions. The basis for this assumption is that subjects reported that those samples were difficult to localize. This suggests that the spectral shape conferred to the white noise by means of the non-individual HRTFs was so foreign that it not only challenged their sense of localization (the could not say where the sound came from), but also their sense of sound quality (it did not sound good). However, this cannot be concluded until further research is conducted. In the context of the application of the technology, every system that implements binaural synthesis with non-individual HRTFs is under the potential risk of degraded sound quality. The extent of the problem is believed to be minor, since the potential quality degradations would not be as disruptive as, for example, to affect speech intelligibility. This assumption is supported by previous binaural synthesis studies conducted with speech as a source, where degraded intelligibility was not reported as an issue (Begault *et al.* [2001], Møller *et al.* [1996a]).

Regarding the use of loudness as a cue, it is not clear whether subjects related more on loudness or on spectral content as relevant cues for certain directions. It is suggested that loudness as a cue with non-individually synthesized sounds was caused by the experimental protocol used in Chapter 5, in which repetitions from only 1 set of non-individual HRTFs were used per block and hence the inherent balance between the HRTFs of each subject was kept.

7.4 Research hypothesis

As it has been mentioned above, the effect of using a non-individual HRTFs pair for sound synthesis can be thought of as using a loudspeaker with a non-flat frequency response for sound reproduction. In such a model, the sound filtered by the HRTFs of the individual subject would undergo some linear filtering with non-flat frequency response, yielding the non-individual HRTFs. Defining that non-flat linear filtering as $X(f)$, this can be expressed in the frequency domain as:

$$HRTF_{individual}(f) \cdot X(f) = HRTF_{non-individual}(f) \tag{7.1}$$

The role of the non-individual HRTFs as conveyors of distortion which affect the individual HRTFs spectra -i.e. $X(f)$ in Eq. 7.1- can be then computed by the ratio of non-individual to individual HRTFs in the frequency domain. This approach is valid as long as the reproduction of sound is performed through individually equalized headphones. Otherwise, an additional factor of spectral filtering would have to be accounted for, and isolating the filtering effect $X(f)$ introduced by the non-individual HRTFs would not be possible. On the other hand, the approach is useful only if the non-individual HRTFs pair provides the necessary spectral cues for sound localization, which is the case of the perceptually matching HRTFs identified in Chapter 5.

The steps mentioned above are relevant considering the work reported by Moore & Tan [2003], where different ways of filtering were imposed. The ratio between non-individual and individual HRTFs was performed for all the subjects participating in the experiments of Chapter 5, for those non-individual HRTFs that perceptually matched individual HRTFs. The results for the four subjects for which HRTFs matching results were presented in Chapter 5 (MT and EMS from Group A, and ME and YB from Group B) are shown in Figures 7.1 to 7.4. Results for the rest of the subjects are included in Appendix C. It has to be noted that some of the subplots in each figure panel showing the results are empty. Those are either cases in which there were not non-individual HRTFs pairs perceptually matching the individual HRTFs pair, or in which the perception with the individual HRTFs pair led to an isotropic distribution.

There are several interesting points that can be drawn from Figures 7.1 to 7.4. In general, it can be seen that the shape of $X(f)$ is approximately flat until 2 kHz, frequency from which a slope of, generally, positive sign defines the envelope at mid and high frequencies. At higher frequencies, $X(f)$ seems to take rapidly fluctuating shapes, and the magnitude of the ratio increases. However, there are other cases where a ripple shape is exhibited in the mid and low ranges, and which starts at frequencies well below 1 kHz. The magnitude of the ripples is, in general, relatively modest. The ripples are mostly present in broad ranges of frequencies and they are combined with a slope in some cases. In the light of the findings reported by Moore & Tan [2003], it is hypothesized that the filtering imposed by $X(f)$ would be audible in terms of perceived quality, particularly in those cases where the frequency response of $X(f)$ is non-flat in the lower range and exhibits a general tilted envelope which takes a broad frequency range. Moore & Tan [2003] reported results to filtering imposed up to 7 kHz, for which it is unclear whether

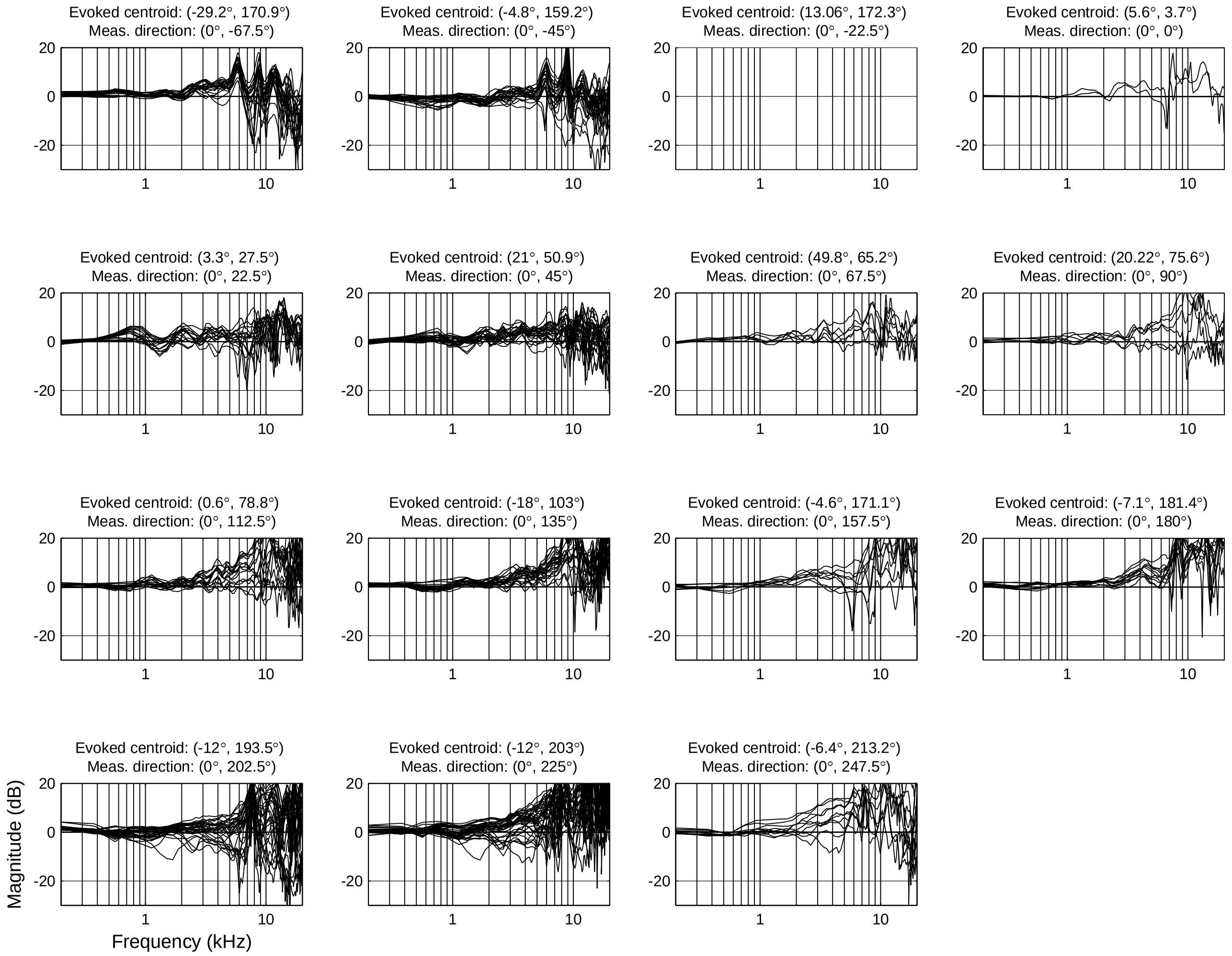

Figure 7.1: Subject MT. Computed ratio $HRTF_{non-individual}/HRTF_{individual}$, based on those matching HRTFs already shown in Figure 5.8.

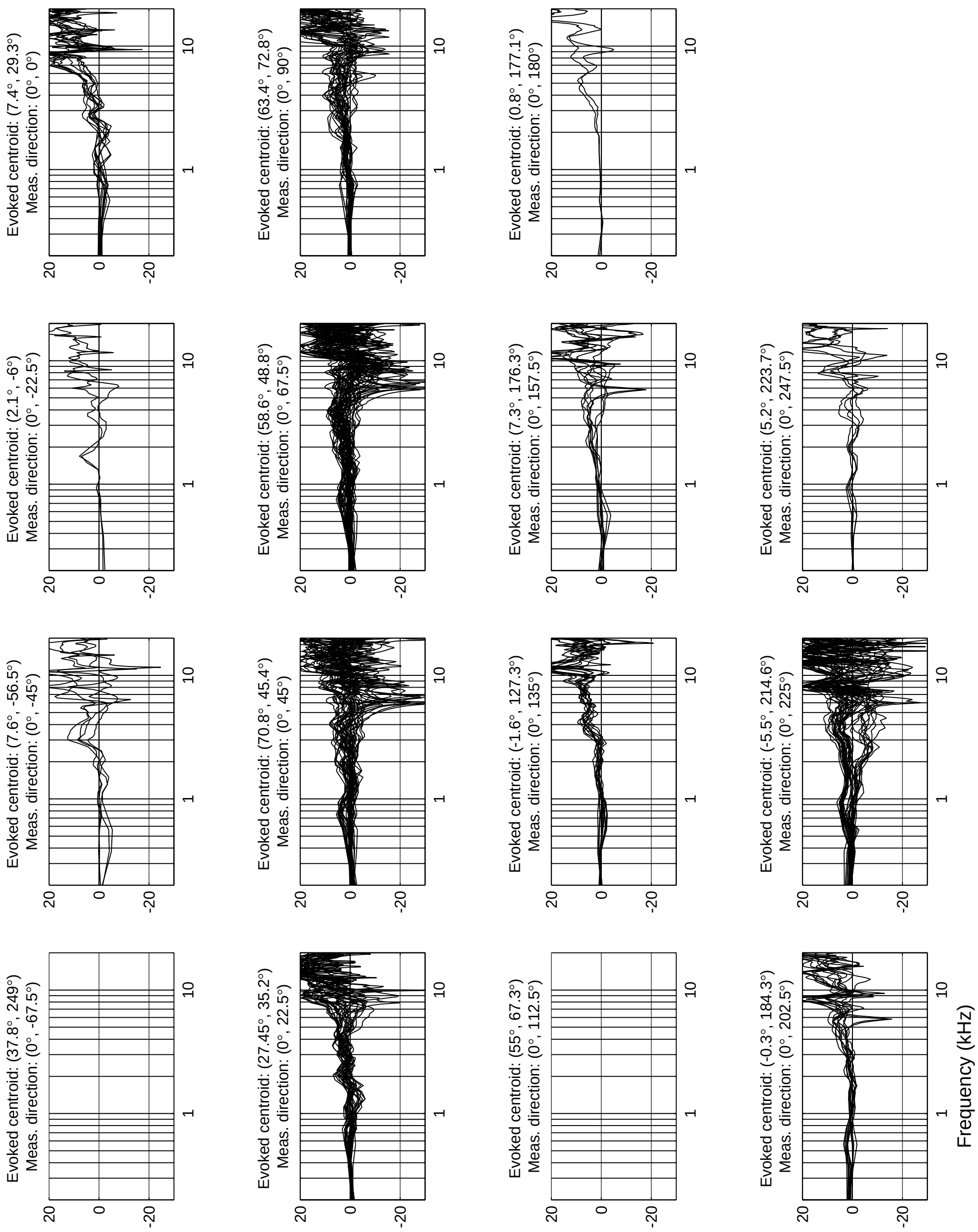

Figure 7.2: Subject EMS. Computed ratio $HRTF_{non-individual}/HRTF_{individual}$, based on those matching HRTFs already shown in Figure 5.9.

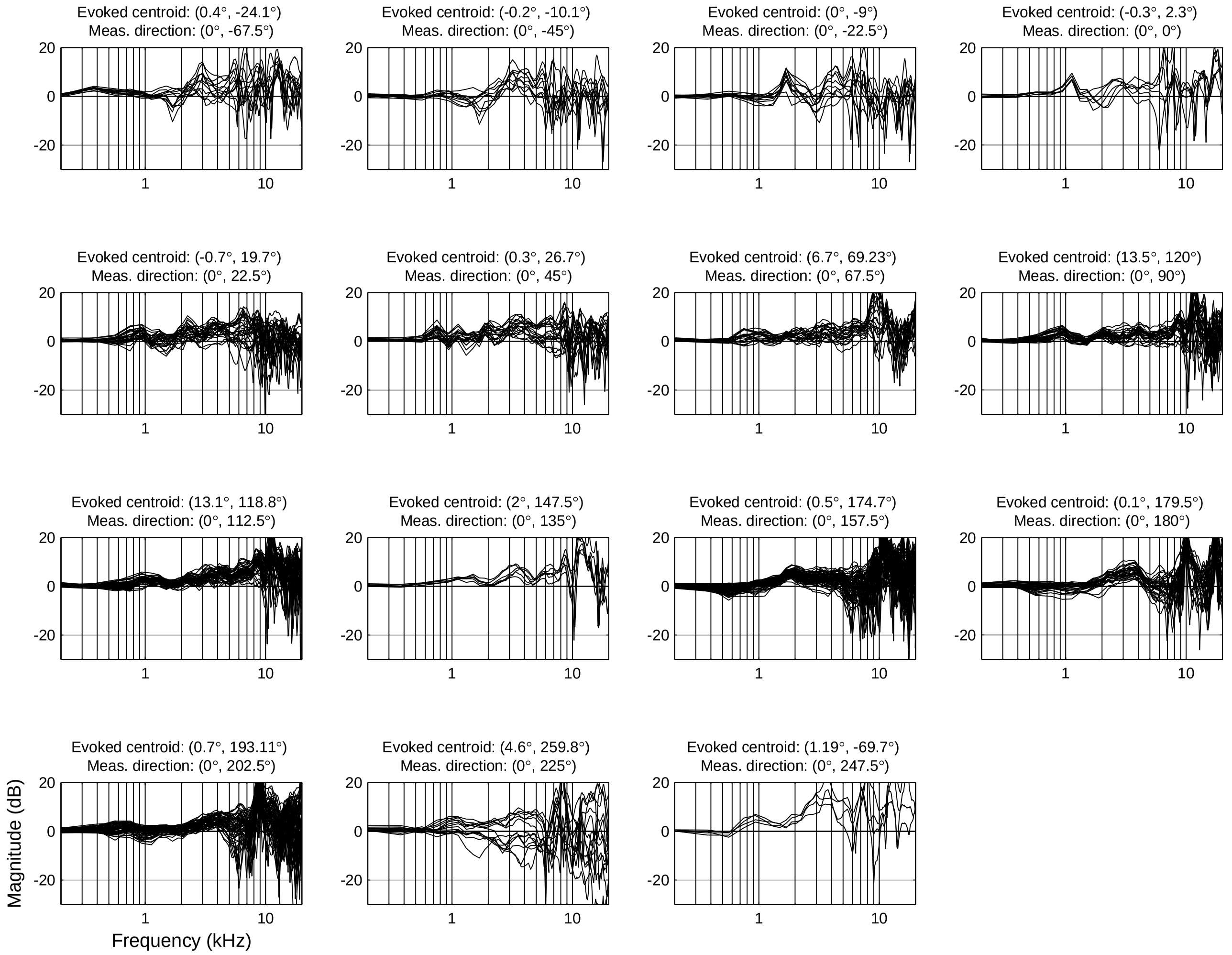

Figure 7.3: Subject ME. Computed ratio $HRTF_{non-individual}/HRTF_{individual}$, based on those matching HRTFs already shown in Figure 5.17.

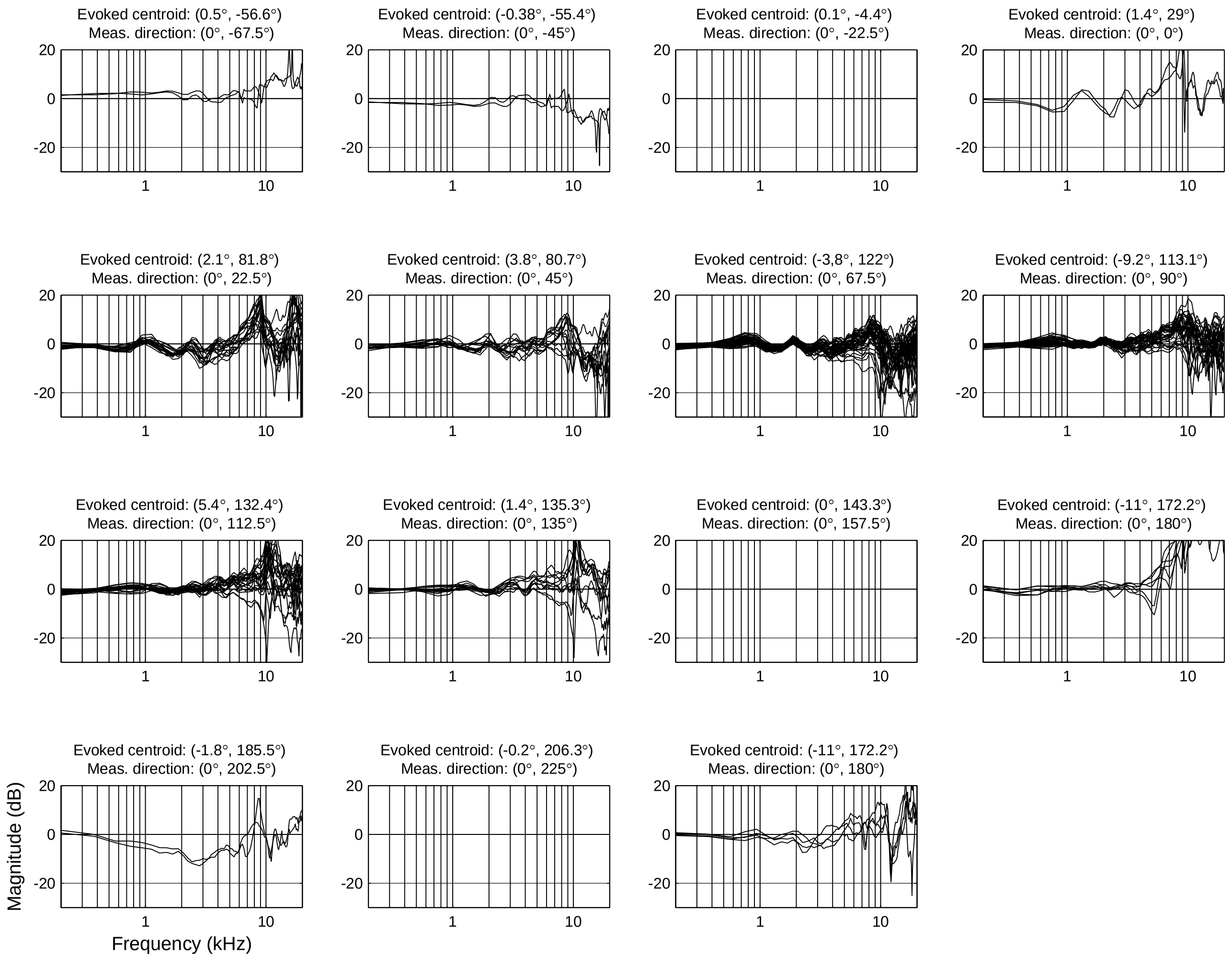

Figure 7.4: Subject YB. Computed ratio $HRTF_{non-individual}/HRTF_{individual}$, based on those matching HRTFs already shown in Figure 5.18.

the high frequency magnitude of $X(f)$ seen in Figures 7.1 to 7.4 would affect the perceived sound quality. Another interesting observation is that there seem to be strong individual differences: the degree of deviation of $X(f)$ from a flat frequency response seems to be different from subject to subject. Just as localization seems to work while the listener's own HRTFs are distorted within a (subject dependent) range, as shown in Chapter 5, it is hypothesized that the allowed degree of non-flatness to $X(f)$ would also relate to an inner reference. The shape of $X(f)$ also seems to be direction dependent. It is hypothesized that subjects would be more sensitive to quality degradation at ecological directions like those at eye level and to the front. It is interesting to note that the ripples and tilted behavior are also seen for $X(f)$ at those directions.

The research hypothesis for the study of the perceived quality of sound when using spectrally non-matching non-individual HRTFs in binaural synthesis can be formulated as follows:

Provided that the non-individual HRTFs used in binaural synthesis perceptually match the individual HRTFs of the listener in sound localization terms, it is hypothesized that the perception of sound quality would be degraded when the ratio of non-individual to individual HRTFs spectra deviates from a flat response in the low, mid and mid-high ranges. It is further hypothesized that some direction dependent effect in the perception of sound quality degradation will be found.

It has to be noted that emphasis is given to the perceived quality of sound once the localization cues are secured. It seems irrelevant to analyze sound quality issues if the HRTFs used for sound synthesis do not comply with the basal requirement of providing the necessary cues for sound source localization.

7.5 Proposed methodology

If perceptual attributes relevant to perceived quality of sound with HRTFs were to be studied systematically, several descriptors could be used to characterize them. For instance, Lorho [2010] reported a series of attributes that could be assigned to reproduced sound over headphones, as condensed in Table 7.1. One of the most important points when conducting sound quality evaluation is the definition of the attributes, so that all subjects use them in the same way. There are different approaches for selecting the definitions of the attributes, such as the consensus vocabulary development and individual vocabulary methods reviewed by Lorho [2010]. It is believed that the *realism* evaluation reported by Begault *et al.* [2001] did not yield significant results due to the lack of common understanding of the concept of *realism*, which was also suggested by the authors of that study. If subjects are not provided with a reference situation (as reported by Moore & Tan [2003]) or clear definitions, it is intuitive to expect some of the following situations: a) subjects will use the same attribute to represent the same perception, b) subjects will use different attributes to represent the same perception, c) subjects will use the same attribute to represent different perceptions, and d) subjects will use different attributes to represent different perceptions. For example, when some of the participant subjects from the experiment in Chapter 5 reported that some sounds *'did not sound good'*, it was not clear which attribute they were evaluating, how they were evaluating it, and whether a different subject would choose the same attribute for representing the same perception and would evaluate it in the same way.

Localization parameters	Space parameters	Timbre parameters
Sense of distance	Quality of echo	Separability
Sense of direction	Amount of echo	Tone color
Sense of movement	Sense of space	Richness
Ratio of localizability	Balance of space	Distortion
	Broadness	Disruption
		Clarity
		Balance of sounds

Table 7.1: Attributes proposed by Lorho [2010] to evaluate perceived quality of spatial sound over headphones, as obtained through a consensus vocabulary development procedure. The definition of each attribute can be found in the cited source.

A series of different listening tests are proposed to systematically study the perceptual attributes that are relevant in the perception of quality of sound synthesized with non-individual HRTFs. The goals with the proposed experiments would be:

- To define a set of relevant attributes with their respective scales.
- To quantify the perception of quality of sound in relation to each attribute.
- To quantify the overall perception of quality of sound.
- To study whether the individual attributes are linked to sound localization.
- To study whether the overall perception of quality of sound is direction dependent.

The experiments would require, as a first step, to conduct a similar study to that reported in Chapter 5, so that individual and non-individual HRTFs that are perceptually equivalent in terms of sound source localization are available. It can be stated *a priori* that the experiments to study perceived quality of sound would concentrate on attributes that are not directly classified under *Localization Parameters* in Table 7.1. However, localization would be studied indirectly in the sense that the relationship between attributes and localization would be quantified. The experiments would not be intended to integrate distance perception, which is a topic that has also received a fair amount of attention before (Zahorik *et al.* [2005]). The proposed experiments are:

- ***1.- Localization experiment with individual and non-indidvidual HRTFs.*** This is a pre-requisite experiment, as mentioned before, by which: a) a panel of *good localizers* would be selected, b) individual and non-individual HRTFs which evoke similar directions would be obtained -for the panel of listeners-, and c) individual and non-individual HRTFs which do not evoke similar directions would be obtained -for the panel of listeners. In other words, this experiment would be similar to that reported in Chapter 5, but discarding the spectral parameterization.

- ***2.- Identification of the perceptual attributes that are relevant for the assessment of the perceived quality of sound, and which relate to a) the use of binaural synthesis as reproduction technique, and b) to the use of different HRTFs filters.*** This experiment can make use of the findings reported by Lorho [2010] regarding how to select the relevant attributes. Methods like consensus vocabulary development could be used, for example, and the following steps would be included: a) panel selection with extensive pilot testing to identify whether *good localizers* possess the required skills for participating in a descriptive analysis experiment, b) consensus vocabulary generation to define the actual attributes to be evaluated, and c) panel training to ensure that all the participants use the attributes and their respective scales in the same way.

- ***3.- Perceptual attributes that relate to the use of individual and non-individual HRTFs in binaural synthesis, with real sound sources as a baseline condition.*** Based on the HRTFs matching procedures of 1.- and the perceptual attributes identified in 2.-, the quality of sound obtained through binaural synthesis and reproduction could be evaluated. Signals would be speech and music samples, and the simulation of non-anchoic environments would be preferred. It would be interesting to divide the study so that a) the quality of sound as defined by each attribute and as a global parameter could be quantified, b) the link of the attributes to sound localization could be defined (non-individual HRTFs that did not evoke a consistent direction could be also used in this part of the listening tests), and c) differences in the perceived quality across directions could be quantified.

7.6 Conclusion

This Chapter investigated issues that relate to the perceived quality of sound when spectrally non-matching non-individual HRTFs are used in binaural synthesis, more specifically, whether the spectral mismatch would be audible as quality degradation. An *a priori* assumption that quality issues would play a role in binaural synthesis with non-individual HRTFs was informally confirmed by the subjects participating in the experiment of Chapter 5. In this Chapter, the ratio between non-individual to individual HRTFs resulting from the matching procedures reported in Chapter 5 was computed. Results showed that, according to the evidence in the literature, the filtering imposed by the non-individual HRTFs to the individual HRTFs would be potentially audible in terms of the perceived quality of sound -at least from a theoretical approach. The filtering imposed by the non-individual HRTFs showed some possible subject and direction dependences. Even though generally flat in the low and mid ranges, there were ripples and slopes in many of the computed ratios. The following hypothesis was defined: *Provided that the non-individual HRTFs used in binaural synthesis perceptually match the individual HRTFs of the listener in sound localization terms, it is hypothesized that the perception of sound quality would be degraded when the ratio of non-individual to individual HRTFs spectra deviates from a flat response in the low, mid and mid-high ranges. It is further hypothesized that some direction dependent effect in the perception of sound quality degradation will be found.* Due to time constrains, this Chapter ended with a proposed methodology for investigating the hypothesis.

Chapter 8

Discussion and Conclusion

8.1 Discussion

The work presented in this Thesis investigated technical and perceptual aspects of broad importance in binaural synthesis. The findings reported are relevant if optimizations on how the technology is implemented are attempted. This Chapter discusses the different studies from this Thesis in a more comprehensive context, which reveals also how the different topics are interconnected.

One of the concepts on which all the Chapters of this Thesis are based, is that technical and perceptual aspects are not completely separated since overlooking the former can potentially have perceptual consequences. That is the case of DC correction and low frequency control, as reported in Chapter 2. The perceptual consequences of overlooking them were seen, however, to be so obvious that a dedicated listening test was not necessary: an uncomfortable feeling of *boominess* due to the low frequency ripples provided a perception of degraded sound quality (also reported by Hammershøi & Møller [2005]). The issue of the perceived quality is a key factor when HRTFs are used in binaural synthesis, and yet it has received little attention: most studies concentrate on localization performance. Some comments on low frequency control and DC correction can be found in the literature (Algazi [1998], Brown & Duda [1998], Riederer [1998], Zotkin *et al.* [2006], among others), but mainly with a purely technical approach. The consequences in the perceptual side of binaural synthesis applications, on the contrary, has not been given much importance. The same occurred with other sound quality aspects, like the perception of degraded sound quality caused by the use of non-individual HRTFs with spectra that does not match that of the listener. Even though Chapter 7 took a theoretical approach due to time constrains, the computations of the ratio between non-individual to individual HRTFs taken from groups of HRTFs that evoked the same direction for a given subject (Chapter 5), showed that the filtering imposed by the non-individual HRTFs in many cases took the form of ripples and slopes that spread over a broad range of frequencies. According to a previous work (Moore & Tan [2003]), this imposed filtering would be audible in terms of sound quality, which demonstrates that more work is needed in the field of perceived quality of sound within the context of binaural synthesis.

The technical aspects discussed in Chapter 2 have also relevance in a purely non-perceptual context. Chapter 2 originated from a contribution to a current round robin of HRTFs measuring systems. From the differences seen in previously published results (Katz & Begault [2007]), it was concluded that a common understanding on how to ensure the validity of measured HRTFs was needed. This was supported by the difficulties faced during an otherwise trivial activity such as calibrating the internal microphones used -trivial in terms of technical protocol, which does not mean that calibration has a minor importance. The work of Chapter 2 is a step forward towards the goal of a consensual protocol on HRTFs measurements. One interesting aspect on which Chapter 2 is built, is the fact that measuring valid HRTFs does not finish with the acquisition of the signals at the ears of the subject and at the center of the head. There are post-processing issues that have to be controlled. It would have been interesting to measure HRTFs from the Neumann KU-100 dummy-head with an external pair of microphones such as the Sennheiser KE-4-211-2 used at the blocked entrance of the ear canals when measuring human HRTFs. That procedure would have allowed constructing a more complete picture of the spectral features introduced by the built-in diffuse field equalization of the dummy-head.

As said, obtaining valid HRTFs for binaural synthesis does not end with the acquisition of the signals and post-processing is requried. For example, the signals have to be conditioned so that they can be used as filters. Chapter 3 touched the topic of minimum phase decomposition, which is relevant for filter design but of relatively restricted importance in binaural synthesis, as not necessarily are HRTFs required as minimum phase filters and an ITD. It is agreed that the implementation is eased if the filters are minimum phase FIR, but there is no perceptual nor technical invalidity in representing them as mixed-phase FIR or IIR, topic that partially relates to the findings in Chapter 4. In that context, the applicability of the findings from Chapter 3 is limited. Besides, the differences in computational cost between applying the Hilbert transform or homomorphic filtering algorithms tested in Chapter 3 were not found to be drastic -this is because HRTFs are by nature short filters, a different scenario would have been obtained if signals had been from different nature. Then again, if filters were much longer probably FIR implementations would not be preferred. As for the computation of the minimum phase representation, it was interesting to see that long zero paddings were needed for more than 30% of the signals -particularly for the contralateral signals in measured HRTFs from directions at the sides and below the horizontal plane. This is relevant, since MATLAB provides built-in methods to compute minimum phase signals. Chapter 3 showed that they have to be used with care, as 80% of the pooled zeros from the database used lied very close to the unit circle, which can make the methods fail. In those cases, appropriate zero padding has to be implemented. It is specially important to highlight this in a cross-disciplinary field like that of binaural technique.

One interesting contribution within Chapter 3 is the quantitative analysis of zeros outside the unit circle as a measure of non-causality, and the direction dependency exhibited. The latter has been suggested before by Kulkarni *et al.* [1995], but none of the studies on phase characteristics of HRTFs showed it systematically as in Chapter 3. The fact that contralateral signals in HRTFs from directions to the sides tend to present more all-pass sections due to their non-causal nature was one of the underlying concepts of Chapter 4, which constitutes another example of how technical and perceptual issues are closely related. The benefit of representing HRTFs as a minimum phase filter and an ITD is purely technical, but the approach proved to have audible consequences in the special case of some low Q-factor all-pass sections, as reported by Plogsties *et al.* [2000]. The audibility of removing high Q-factor all-pass sections had not been tested before. Therefore, and from a strictly methodological point of view, the experiment reported in Chapter 4 was the last one required to fully accept and embrace the concept of HRTFs as a minimum phase filter and an ITD. The hypothesis had been already formulated by Møller *et al.* [2007], who stated that there would not be perceptual consequences from removing high Q-factor all-pass sections from HRTFs since they were centered at frequencies where deep notches occurred in the HRTFs magnitude response. In that context, the contribution of the Chapter lies in the testing of the hypothesis itself for 12 subjects, both under binaural and diotic reproduction conditions.

Representing HRTFs as a minimum phase filter and ITD is an elegant approach as it presents a clear conceptualization of the separation of temporal and spectral elements -which is relevant in binaural hearing. The conceptualization is useful in a study like the one reported in Chapter 5 which required concentrating on the spectral characteristics of HRTFs that cue localization. The basic question that Chapter 5 tried to answer was whether the spectral cues to elevation in the MSP could be identified and parameterized. This research question is not new, but the approach to look for the answer is: to the knowledge of this author, it was not attempted to obtain individual and non-individual HRTFs that provided equivalent evoked directions before, and it was this methodology that allowed finding the similarities among HRTFs. Wightman & Kistler [1993] reported a study based on similarities, but the individual relevant spectral cues were not identified. There has also been an attempt to identify those cues: Iida *et al.* [2007]

concluded that the first peak and two first notches would provide the necessary cues, but that study did not establish which individual parameters of the peaks and notches needed to be preserved -they mentioned center frequency, level and sharpness as a generalization. The parameterization, on the other hand, is important. Otherwise, the single-notch model proposed by Macpherson [1994] would have worked in some extent. Chapter 5 proposes the Q-factor of the first peak, with a main role of the high frequency slope, to disambiguate directions above the horizontal plane (front-high, back-high and above), and the global Q-factor of the first notch to disambiguate front, back and back-low - together with providing redundant cues for the high and above directions. The second peak, specially the frequency at which it is centered, would have some redundant information shared with the first notch to disambiguate back and back-low directions. These findings agree with the evidence from the literature, and their importance is broad. On one side, the knowledge of the underlying mechanisms in binaural hearing could be extended -for example, by investigating the detailed role of the pinnae in the different features. On the other side, it has a direct impact on binaural synthesis implementations -both with individual and non-individual HRTFs. For the former case of individual HRTFs, knowledge of the relevant spectral features and the parameters that have to be preserved would help discarding redundant information or those spectral features that do not aid localization: HRTFs could be simplified. In the non-individual HRTFs case, the findings of Chapter 5 would help *individualization* processes in which spectral features could be manipulated to better match the HRTFs spectra parameters of the listener. This would be aided by the findings of Chapter 5 that relate to the ranges in frequency and value that the parameters take without compromising localization. The results from Chapter 7 are also relevant in this context, as they showed that listeners allowed large variations in the high frequency range, with a low-, mid-, and mid-high ranges very similar to their own HRTFs. There were also cases were more important deviations were allowed in those ranges, including some deviations that started well below 1 *kHz*, and yet localization was not affected. Both in the individual and non-individual HRTFs cases, however, quality issues have to be considered, for which the experiments proposed in Chapter 7 would have to be followed. It is implicit, then, in the results from Chapter 5 that a single set of non-individual HRTFs that provide localization cues for a large sample of listeners would most likely not be feasible. The results, showing that the ranges that the parameters take are highly individual and that the parameters from different features have to be consistent with each other, support the approach of, as said, *individualizing* a single set. As an alternative, and as proposed by Wightman & Kistler [1993], several representative sets could be used, which would be targeted to different groups of subjects.

Chapter 6 has a different character than the rest of the Thesis, as it relies more on real sound sources localization than on synthesized sound. It was shown in Chapter 6 that there are subjects which perceived all the sound sources either in the rear or frontal hemisphere, or had degraded localization in specific ranges of frequencies. The findings described in that Chapter challenge the current notion of how binaural synthesis is validated and what is to be expected in terms of localization. The approach towards the listener is more integrative, as the Chapter proposes to consider those subjects who localize in a different way also as normal, breaking the dichotomy between *good* and *poor localizers*, and integrating other possible aspects that play a role in localization processes. On the other hand, the term *poor localizer* seems to convey a certain negative connotation, when in reality it should not. *Biased localizer* seems to be a more neutral term (despite the paradox). Chapter 6 advocates, also, for a more realistic approach: not everyone localize with the same proficiency and the technology has to grow from that basis, without expecting to one day provide *correct* localization performance for all. It is believed that wrong expectations were the cause for overlooking *biased localizers* in previous studies, it is interesting that there is not much in the literature describing these localizers -apart from Itoh *et al.* [2007]. The results from Chapter 6, moreover, showed that those *biased localizers* that participated in the individual HRTFs binaural synthesis condition had a performance that kept the general trends of the real sound sources condition, suggesting that the technology was actually providing an experience close to real life. Overall, the message underlying Chapter 6 is that more ground discussion is needed regarding the conceptualization of binaural hearing and the expectation on localization performance from an integrative point of view. The issue about expectations that the technology imposes over itself is also related to some of the findings reported in Chapter 5. More precisely, to the discussion on sound externalization of virtual sound sources reproduced binaurally through headphones. It was seen that subjects were actually confused when asked about sound sources inside or outside the head, probably because they had to use too many cognitive resources. It was interesting to see that the rate of inside-the-head answers was related to the awareness of the subjects about the concept of sound externalization.

8.2 General conclusion

This Thesis reported several investigations on technical and perceptual aspects of HRTFs to be used in binaural synthesis, all of which can be directly used for optimizing how the technology is implemented. The investigation reported in Chapter 2 on calibration, DC correction and low frequency control can be used as a step forward towards a consensual protocol for HRTFs measurements among the community of HRTFs users. The comparison between minimum phase decomposition methods reported in Chapter 3 showed that care has to be taken when decomposing the contralateral signals of HRTFs from the sides and below the horizontal plane, which are non-causal by nature, and for which decomposition methods can require longer zero padding. Chapter 4 confirmed that minimum phase representations of HRTFs are valid as long as audible low Q-factor all-pass sections are accounted for in the conformation of the ITD: high Q-factor all-pass sections can be discarded without audible consequences. Once the technical aspects are under control, the perceptual issues take a major role, particularly regarding the relevant spectral information that non-individual HRTFs have to convey in order to provide the necessary localization cues to the listeners. Chapter 5 showed that the first peak, particularly its Q-factor and its high frequency slope, would cue directions front-high, above and back-high; the first notch and its global Q-factor would cue front, back and back-low directions, and would provide redundant information for the above and high directions; and the second peak, particularly its center frequency, would provide some redundant information to also disambiguate back and back-low directions. These findings are of broad interest, as they would help simplifying HRTFs so that they only keep relevant information. The findings would also help *individualization* processes of non-individual HRTFs. Chapter 7 showed that non-individual HRTFs that provide the necessary localization cues to a subject could potentially introduce spectral components that would theoretically be audible in terms of degraded perceived quality of sound, calling for further testing. Finally, Chapter 6 raised the question of revising the dichotomy between *good* and *poor localizers* so that a better understanding of how humans localize is obtained from the study of *biased localizers*. It is believed that these latter localizers should be better reported in the literature, so that their performance can be studied more thoroughly.

8.3 Future work

This Thesis is a step forward in many directions, and some of the Chapters have already made explicit the need of further work to use, extend or validate the results and/or hypotheses raised. Some of the opportunities for further work will be outlined in the following.

There seems to be a need for a consensual protocol for HRTFs measurement, which would have to be agreed among the relevant stakeholders. That work is to be done, and it calls for a collaborative project which could be an extension of the current round robin for which the dummy-head measurements in Chapter 2 were done. Moreover, the work presented in Chapter 2 could be used as a step forward towards that goal.

It was shown that there is a degree of direction dependency on whether minimum phase decomposition methods can succeed or fail. Hilbert transform and homomorphic filtering were also seen to be dependent on the choice of zero padding to the signals in order to successfully decompose them. In this framework, optimized algorithms for minimum phase decomposition could be tested. For example, it has been reported that using exponential weighting can prevent the methods from failing when the zeros are very close to the unit circle. This procedure, and others, could be investigated in the context of binaural synthesis so that one method that works in a standard form, and regardless the direction at which the HRTFs were measured, is recommended.

In this Thesis, potential spectral features that cue localization in the MSP, and their relevant parameters, were identified. However, a closer inspection showed that some non-individual HRTFs presented those cues and the parameters within the working range of the subjects, and yet they did not provide localization cues. Further work should be done to explain that phenomenon and validate the results presented in Chapter 5. Further testing is also needed to understand the suggested hierarchy among cues, where a side of the HRTFs pair did not present the spectral features identified and yet the cues to localization were still conveyed. The methodology used, on the other side, could be extended to the study of planes off the MSP.

The understanding on human sound localization also needs to be extended with more localization results of real sound sources, including anechoic and non-anechoic conditions, as well as synthesized sound. More information is needed about *biased localizers*. As it was mentioned in the discussion of Chapter 6, an updated topography of auditory space as investigated by Oldfield & Parker [1984] could be attempted, so as to integrate the changes in perception due to cultural factors - particularly the explosion in the use of portable music players and the exposure to high sound pressure levels during leisure activities.

Finally, this Thesis has shown that there is evidence to hypothesize that, provided that the non-individual HRTFs used in binaural synthesis perceptually matches the individual HRTFs of the listener in sound localization terms, the perception of sound quality would be degraded when the ratio of non-individual to individual HRTFs spectra deviates from a flat response in the low, mid and mid-high frequency ranges. Furthermore, a direction dependency and strong individual results would be expected. Further work is needed to test the proposed hypothesis, and a possible course of action has already been outlined in Chapter 7.

Bibliography

Algazi, V. R. 1998. *Documentation for the UCD HRIR files*. Technical Report. CIPIC Interface Laboratory, University of California at Davis.

Algazi, V. R., Duda, R. O., Thompson, D. M., & Avendano, C. 2001a. The CIPIC HRTF database. *Pages 99–102 of: Proceedings of 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics*. IEEEE.

Algazi, V. R., Avendano, C., & Duda, R. O. 2001b. Elevation localization and head-related transfer function analysis at low frequencies. *Journal of the Acoustical Society of America*, **109**(3).

Algazi, V. R., Duda, R. O., Duraiswami, R., Gumerov, N. A., & Tang, Z. H. 2002a. Approximating the head-related transfer function using simple geometric models of the head and torso. *Journal of the Acoustical Society of America*, **112**(5), 2053–2064.

Algazi, V. R., Duda, R. O., & Thompson, D. M. 2002b. The use of head-and-torso models for improved spatial sound synthesis. *In: Proceedings of 113th Audio Engineering Society Convention*. Audio Engineering Society.

Asano, F., Suzuki, Y., & Sone, T. 1990. Role of spectral cues in median plane localization. *Journal of the Acoustical Society of America*, **88**(1), 159–168.

Begault, D. R. 2000. *3-D sound for virtual reality and multimedia*. National Aeronautics and Space Administration, Ames Research Center ; Available from NASA Center for AeroSpace Information, Moffett Field, Calif., Hanover, MD.

Begault, D. R., Wenzel, E. M., & Anderson, M. R. 2001. Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source. *Journal of the Audio Engineering Society*, **49**(10), 904–916.

Best, V., Carlile, S., Jin, C., & van Schaik, A. 2005. The role of high frequencies in speech localization. *Journal of the Acoustical Society of America*, **118**(1), 353–363.

Blauert, J. 1969/70. Sound Localization In The Median Plane. *Acustica*, **22**(4), 205–213.

Blauert, J. 1983. *Spatial Hearing*. MIT Press.

Blauert, J. 1997. *Spatial Hearing. Revised Edition*. MIT.

Blauert, J., Brüggen, M., Hartung, K., Bronkhorst, A.W., Drullman, R., Reynaud, G., Pellieux, L., Krebber, W., & Sottek, R. 1998 (June). The AUDIS catalog of human HRTFs. *Pages 2901–2902 of: Proceedings of the 16th ICA*. ICA.

Bloom, P. J. 1977. Creating source elevation illusions by spectral manipulation. *Journal of the Audio Engineering Society*, **25**(9), 560–565.

Bronkhorst, A. W. 1995. Localization of real and virtual sound sources. *Journal of the Acoustical Society of America*, **98**(5), 2542–2553.

Brown, C. P., & Duda, R. O. 1998. A structural model for binaural sound synthesis. *Speech and Audio Processing, IEEE Transactions on*, **6**(5), 476–488.

Brüel&Kjær. 1996. *Microphone Handbook Vol.1: Theory*. Tech. rept. Brüel & Kjær.

Brungart, D. S., & Rabinowitz, W. M. 1999. Auditory localization of nearby sources. Head-related transfer functions. *Journal of the Acoustical Society of America*, **106**(3), 1465–1479.

Brungart, D. S., & Scott, K. R. 2001. The effects of production and presentation level on the auditory distance perception of speech. *Journal of the Acoustical Society of America*, **110**(1), 425–440.

Burandt, U., Posselt, C., Ambrozus, S., Hosenfeld, M., & Knauff, V. 1991. Anthropometric contribution to standardizing mannequins for artificial-head microphones and to measuring headphones and ear protectors. *Applied Ergonomics*, **22**(6), 373–378.

Burkhard, M. D., & Sachs, R. M. 1975. Anthropometric manikin for acoustic research. *Journal of the Acoustical Society of America*, **58**(1), 214–222.

Butler, R. A., & Belendiuk, K. 1977. Spectral cues utilized in the localization of sound in the median sagittal plane. *Journal of the Acoustical Society of America*, **61**(5), 1264–1269.

Butler, R. A., & Planert, N. 1976. Influence of stimulus bandwidth on localization of sound in space. *Perception & Psychophysics*, **19**(1), 103–108.

Carlile, S., & Pralong, D. 1994. The location-dependent nature of perceptually salient features of the human head-related transfer-functions. *Journal of the Acoustical Society of America*, **95**(6), 3445–3459.

Carlile, S., Leong, P., & Hyams, S. 1997. The nature and distribution of errors in sound localization by human listeners. *Hearing Research*, **114**(1-2), 179–196.

Carlile, S., Delaney, S., & Corderoy, A. 1999. The localisation of spectrally restricted sounds by human listeners. *Hearing Research*, **128**(1-2), 175–189.

Carlile, S., Martin, R., & McAnally, K. 2005. Spectral information in sound localization. *Auditory Spectral Processing*, **70**, 399–434.

Chen, J. S., Vanveen, B. D., & Hecox, K. E. 1995. A spatial feature-extraction and regularization model for the head-related transfer-function. *Journal of the Acoustical Society of America*, **97**(1), 439–452.

Christensen, F. 2001. *Binaural technique with special emphasis on recording and playback*. Ph.D. thesis, Aalborg University.

Chung, W., Carlile, S., & Leong, P. 2000. A performance adequate computational model for auditory localization. *The Journal of the Acoustical Society of America*, **107**(1), 432–445.

Damera-Venkata, N., Evans, B. L., & McCaslin, S. R. 2000. Design of optimal minimum-phase digital FIR filters using discrete Hilbert transforms. *IEEE Transactions on Signal Processing*, **48**(5).

Fels, J., Buthmann, P., & Vorlander, M. 2004. Head-related transfer functions of children. *Acta Acustica united with Acustica*, **90**(5), 918–927.

Firestone, F.A. 1930. The phase difference and amplitude ratio at the ears due to a source of pure tone. *Journal of the Acoustical Society of America*, **2**(2), 260–270.

Fisher, N. I., Lewis, T., & Embleton, B. J. J. 1987. *Statistical analysis of spherical data*. Cambridge University Press.

Gardner, B., & Martin, K. 1994. *HRTF measurements of a KEMAR dummy-head microphone*. http://sound.media.mit.edu/resources/KEMAR.html.

Gardner, M. B. 1973. Some monaural and binaural facets of median plane localization. *Journal of the Acoustical Society of America*, **54**(6), 1489–1495.

Gardner, M. B., & Gardner, R. S. 1973. Problem of localization in median plane: effect of pinnae cavity occlusion. *Journal of the Acoustical Society of America*, **53**(2), 400–408.

Grantham, D. W., Willhite, J. A., Frampton, K. D., & Ashmead, D. H. 2005. Reduced order modeling of head related impulse responses for virtual acoustic displays. *Journal of the Acoustical Society of America*, **117**(5), 3116–3125.

Hammershøi, D., & Møller, H. 1996. Sound transmission to and within the human ear canal. *Journal of the Acoustical Society of America*, **100**(1), 408–427.

Hammershøi, D., & Møller, H. 2005. *Binaural technique - Basic Methods for Recording, Synthesis and Reproduction. Chapter in book Communication Acoustics. Edited by J. Blauert*. Springer-Verlag, ISBN 3-540-22162-x. Pages 223–254.

Han, H. L. 1994. Measuring a dummy head in search of pinna cues. *Journal of the Audio Engineering Society*, **42**(1-2), 15–37.

Hawksford, M. O. 1997. Digital signal processing tools for loudspeaker evaluation and discrete-time crossover design. *Journal of the Audio Engineering Society*, **45**(1-2), 37–62.

Hebrank, J., & Wright, D. 1974a. Are 2 ears necessary for localization of sound sources on median plane. *Journal of the Acoustical Society of America*, **56**(3), 935–938.

Hebrank, J., & Wright, D. 1974b. Spectral cues used in the localization of sound sources on the median plane. *Journal of the Acoustical Society of America*, **56**(6), 1829–1834.

Hiranaka, Y., & Yamasaki, H. 1983. Envelope representations of pinna impulse responses relating to 3-dimensional localization of sound sources. *Journal of the Acoustical Society of America*, **73**(1), 291–296.

Humanski, R. A., & Butler, R. A. 1988. The contribution of the near and far ear toward localization of sound in the sagittal plane. *Journal of the Acoustical Society of America*, **83**(6), 2300–2310.

Huopaniemi, J., Zacharov, N., & Karjalainen, M. 1999. Objective and subjective evaluation of head-related transfer function filter design. *Journal of the Audio Engineering Society*, **47**(4), 218–239.

Iida, K., Itoh, M., Itagaki, A., & Morimoto, M. 2007. Median plane localization using a parametric model of the head-related transfer function based on spectral cues. *Applied Acoustics*, **68**(8), 835–850.

IRCAM. 2002. *Listen Project - http://recherche.ircam.fr/equipes/salles/listen/index.html*.

ISO. 1997. *266:1997 Acoustics - Preferred Frequencies*.

Itoh, M., Iida, K., & Morimoto, M. 2007. Individual differences in directional bands in median plane localization. *Applied Acoustics*, **68**(8), 909–915.

Ivarsson, C., Deribaupierre, Y., & Deribaupierre, F. 1980. Functional ear asymmetry in vertical localization. *Hearing Research*, **3**(3), 241–247.

Jin, C., Schenkel, M., & Carlile, S. 2000. Neural system identification model of human sound localization. *The Journal of the Acoustical Society of America*, **108**(3), 1215–1235.

Jin, C., Corderoy, A., Carlile, S., & van Schaik, A. 2004. Contrasting monaural and interaural spectral cues for human sound localization. *Journal of the Acoustical Society of America*, **115**(6), 3124–3141.

Jot, J. M., Larcher, V., & O., Warusfel. 1995 (February). Digital Signal processing issues in the context of binaural and transaural stereophony. *In: Proceedings of 98th Audio Engineering Society Convention.* Audio Engineering Society.

Kahana, Y. 2000. *Numerical modelling of the head-related transfer function.* Ph.D. thesis, University of Southampton, UK.

Karam, Z. N. 2006 (January). *Computation of the One-Dimensional Unwrapped Phase.* M.Phil. thesis, Department of Electrical Engineering and Computer Science, M.I.T.

Katz, B. F. G. 2001. Boundary element method calculation of individual head-related transfer function. II. Impedance effects and comparisons to real measurements. *Journal of the Acoustical Society of America*, **110**(5), 2449–2455.

Katz, B. F. G., & Begault, D. R. 2007. Round robin comparison of HRTF measurement systems: preliminary results. *In: Proceedings of the 19th International Congress on Acoustics.*

Kistler, D., & Wightman, F. 1992. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *Journal of the Acoustical Society of America*, **91**(3), 1637–1647.

Kulkarni, A., Isabelle, S.K., & Colburn, H.S. 1995. On the minimum-phase approximation of head-related transfer functions. *Pages 84–87 of: Applications of Signal Processing to Audio and Acoustics, 1995., IEEE ASSP Workshop on.*

Kulkarni, A., Isabelle, S. K., & Colburn, H. S. 1999. Sensitivity of human subjects to head-related transfer-function phase spectra. *Journal of the Acoustical Society of America*, **105**(5), 2821–2840.

Langendijk, E. H. A., & Bronkhorst, A. W. 2000. Fidelity of three-dimensional-sound reproduction using a virtual auditory display. *Journal of the Acoustical Society of America*, **107**(1), 528–537.

Langendijk, E. H. A., & Bronkhorst, A. W. 2002. Contribution of spectral cues to human sound localization. *Journal of the Acoustical Society of America*, **112**(4), 1583–1596.

Larcher, V., Vandernoot, G., & Jot, J. M. 1998. Equalization methods in binaural technology. *In: Proceedings of the 105th Convention of the Audio Engineering Society*. Audio Engineering Society.

Leong, P., & Carlile, S. 1998. Methods for spherical data analysis and visualization. *Journal Of Neuroscience Methods*, **80**(2), 191–200.

Lopez Poveda, E. A., & Meddis, R. 1996. Physical model of sound diffraction and reflections in the human concha. *Journal of the Acoustical Society of America*, **100**(5), 3248–3259.

Lorho, G. 2010. *Perceived quality evaluation: An application to sound reproduction over headphones*. Ph.D. thesis, Aalto University, School of Science and Technology, Department of Signal Processing and Acoustics.

Macpherson, E. A. 1994 (November). On the role of head-related transfer function spectral notches in the judgement of sound source elevation. *In:* Kramer, G., & Smith, S. (eds), *ICAD '94*. ICAD.

Macpherson, E. A., & Middlebrooks, J. C. 2003. Vertical-plane sound localization probed with ripple-spectrum noise. *The Journal of the Acoustical Society of America*, **114**(1), 430–445.

Majdak, P., Balazs, P., & Laback, B. 2007. Multiple exponential sweep method for fast measurement of head-related transfer functions. *Journal of the Audio Engineering Society*, **55**(7/8), 623–637.

Makous, J. C., & Middlebrooks, J. C. 1990. Two-dimensional sound localization by human listeners. *Journal of the Acoustical Society of America*, **87**(5), 2188–2200.

Mehrgardt, S., & Mellert, V. 1977. Transformation characteristics of the external human ear. *Journal of the Acoustical Society of America*, **61**(6), 1567–1576.

Mellert, V., Siebrasse, K. F., & Mehrgardt, S. 1974. Determination of transfer-function of external ear by an impulse response measurement. *Journal of the Acoustical Society of America*, **56**(6), 1913–1915.

Middlebrooks, J. C. 1992. Narrow-band sound localization related to external ear acoustics. *Journal of the Acoustical Society of America*, **92**(5), 2607–2624.

Middlebrooks, J. C. 1999. Individual differences in external-ear transfer functions reduced by scaling in frequency. *Journal of the Acoustical Society of America*, **106**(3), 1480–1492.

Middlebrooks, J. C., & Green, D. M. 1990. Directional dependence of interaural envelope delays. *Journal of the Acoustical Society of America*, **87**(5), 2149–2162.

Middlebrooks, J. C., & Green, D. M. 1992. Observations on a principal components analysis of head-related transfer functions. *Journal of the Acoustical Society of America*, **92**(1), 597–599.

Middlebrooks, J. C., Makous, J. C., & Green, D. M. 1989. Directional sensitivity of sound-pressure levels in the human ear canal. *The Journal of the Acoustical Society of America*, **86**(1), 89–108.

Mills, A. W. 1958. On the minimum audible angle. *Journal of the Acoustical Society of America*, **30**(4), 237–246.

Minnaar, P., Christensen, F., Møller, H., Olesen, S. K., & Plogsties, J. 1999. Audibility of all-pass components in binaural synthesis. *In: Proceedings of the 106th Audio Engineering Society Convention*.

Minnaar, P., Plogsties, J., Olesen, S. K., Christensen, F., & Møller, H. 2000 (February 19-22). The interaural time difference in binaural synthesis. *In: Proceedings of the 108th Audio Engineering Society Convention*. Audio Engineering Society, Paris, France.

Minnaar, P., Olesen, S. K., Christensen, F., & Møller, H. 2001. Localization with binaural recordings from artificial and human heads. *Journal of the Audio Engineering Society*, **49**(5), 323–336.

Minnaar, P., Plogsties, J., & Christensen, F. 2005. Directional resolution of head-related transfer functions required in binaural synthesis. *Journal of the Audio Engineering Society*, **53**(10), 919–929.

Møller, H. 1992. Fundamentals of binaural technology. *Applied Acoustics*, **36**(3/4), 171–218.

Møller, H., Friis Sørensen, M., Hammershøi, D., & Boje Jensen, C. 1995a. Head-related transfer functions of human subjects. *Journal of the Audio Engineering Society*, **43**(5), 300–321.

Møller, H., Hammershøi, D., Jensen, C. B., & Sørensen, M. F. 1995b. Transfer characteristics of headphones measured on human ears. *Journal of the Audio Engineering Society*, **43**(4), 203–217.

Møller, H., Friis Sørensen, M., Boje Jensen, C., & Hammershøi, D. 1996a. Binaural technique: Do we need individual recordings? *Journal of the Audio Engineering Society*, **44**(6), 451–469.

Møller, H., Boje Jensen, C., Hammershøi, D., & Friis Sørensen, M. 1996b (May 11-14). Using a typical human subject for binaural recording. *Pages 1–18 of: Proceedings of the 100th Audio Engineering Society Convention*. Audio Engineering Society, Copenhagen.

Møller, H., Hammershøi, D., Boje Jensen, C., & Friis Sørensen, M. 1999. Evaluation of artificial heads in listening tests. *Journal of the Audio Engineering Society*, **47**(3), 83–100.

Møller, H., Minnaar, P., Olesen, S. K., Christensen, F., & Plogsties, J. 2007. On the audibility of all-pass phase in electroacoustical transfer functions. *Journal of the Audio Engineering Society*, **55**(3), 115–134.

Moore, B. C. J., & Tan, C. T. 2003. Perceived naturalness of spectrally distorted speech and music. *Journal of the Acoustical Society of America*, **114**(1), 408–419.

Moore, B. C. J., Oldfield, S. R., & Dooley, G. J. 1989. Detection and discrimination of spectral peaks and notches at 1 and 8khz. *Journal of the Acoustical Society of America*, **85**(2), 820–836.

Neumann. 2009. *http://www.neumann.com/?lang=en&id=current_microphones&cid=ku100_publications*. Tech. rept. Neumann GmbH.

Nicol, R., Lemaire, V., Bondu, A., & Busson, S. 2006. Looking for a relevant similarity criterion for HRTF clustering: a comparative study. *In: Proceedings of the 120th Audio Engineering Society Convention*.

Oldfield, S. R., & Parker, S. P. A. 1984. Acuity of sound localization: a topography of auditory space. I. Normal hearing condition. *Perception*, **13**.

Olesen, S. K., Plogsties, J., Minnaar, P., Christensen, F., & Møller, H. 2000. An improved MLS measurement system for acquiring room impulse responses. *Pages 117–120 of: Proceedings of NORSIG 2000*. IEEE Nordic Signal Processing Symposium, Kolmården, Sweden.

Oppenheim, A. V., & Schafer, R. W. 1975. *Digital signal processing*. Prentice-Hall, Englewood Cliffs, N.J.

Oppenheim, A. V., & Schafer, R. W. 1989. *Discrete-Time Signal Processing*. Prentice Hall.

Paul, S. 2009. Binaural recording technology: a historical review and possible future developments. *Acta Acustica united with Acustica*, **95**(5), 767–788.

Pedersen, T. H., & Zacharov, N. 2008. How many pshycho-acoustic attributes are needed? *In: Acoustics '08 Paris: 155th Meeting of the Acoustical Society of America, 5th Froum Acusticum (EAA), 9th Congrés Français d'Acoustique (SFA)*.

Pernaux, J. M., Emerit, M., Daniel, J., & Nicol, R. 2002. Perceptual evaluation of static binaural sound synthesis. *In: Proceedings of the 22nd Audio Engineering Society International Conference on Virtual, Synthetic and Entertainment Audio*. Audio Engineering Society.

Perrett, S., & Noble, W. 1995. Available response choices affect localization of sound. *Perception & Psychophysics*, **57**(2), 150–158.

Plogsties, J., Olesen, S. K., Minnaar, P., Christensen, F., & Møller, H. 2000 (February 19-22). Audibility of all-pass components in head-related transfer functions. *Pages 1–14 of: Proceedings of the 108th Audio Engineering Society Convention*. Audio Engineering Society, Paris.

Pralong, D., & Carlile, S. 1994. Measuring the human head-related transfer-functions - a novel method for the construction and calibration of a miniature in-ear recording-system. *Journal of the Acoustical Society of America*, **95**(6), 3435–3444.

Raykar, V. C., Duraiswami, R., & Yegnanarayana, B. 2005. Extracting the frequencies of the pinna spectral notches in measured head related impulse responses. *Journal of the Acoustical Society of America*, **118**(1), 364–374.

Riederer, K. A. J. 1998. Repeatability analysis of head-related transfer function measurements. *In: Proceedings of the 105th Convention of the Audio Engineering Society*.

Roffler, S. K., & Butler, R. A. 1968a. Factors that influence localization of sound in vertical plane. *Journal of the Acoustical Society of America*, **43**(6), 1255–&.

Roffler, S. K., & Butler, R. A. 1968b. Localization of tonal stimuli in vertical plane. *Journal of the Acoustical Society of America*, **43**(6), 1260–&.

Sandvad, J., & Hammershøi, D. 1994. Binaural auralization. comparison of FIR and IIR filter representation of HIRS. *In: Proceedings of the 96th Audio Engineering Society Convention*.

Searle, C. L., Braida, L. D., Cuddy, D. R., & Davis, M. F. 1975. Binaural pinna disparity - another auditory localization cue. *Journal of the Acoustical Society of America*, **57**(2), 448–455.

Shaw, E. A. G. 1974. Transformation of sound pressure level from free field to eardrum in horizontal plane. *Journal of the Acoustical Society of America*, **56**(6), 1848–1861.

Shaw, E. A. G., & Teranishi, R. 1968. Sound pressure generated in an external-ear replica and real human ears by a nearby point source. *Journal of the Acoustical Society of America*, **44**(1), 240–&.

Shimada, S., Hayashi, N., & Hayashi, S. 1994. A clustering method for sound localization transfer-functions. *Journal of the Audio Engineering Society*, **42**(7-8), 577–584.

Sitton, G. A., Burrus, C. S., Fox, J. W., & Treitel, S. 2003. Factoring very-high-degree polynomials. *Signal Processing Magazine, IEEE*, **20**(6).

Toledo, D., & Møller, H. 2008a. Audibility of high Q-factor all-pass components in head-related transfer functions. *In: Proceedings of the 125th Audio Engineering Society Convention*.

Toledo, D., & Møller, H. 2008b. The role of spectral features in sound localization. *In: Proceedings of the 124th Audio Engineering Society Convention*.

Toledo, D., & Møller, H. 2009. Issues on dummy-head HRTFs measurements. *In: Proceedings of 126th Audio Engineering Society Convention*.

Toole, F. E. 1970. In-Head Localization Of Acoustic Images. *Journal of the Acoustical Society of America*, **48**(4), 943–&.

Watkins, A. J. 1978. Psychoacoustical aspects of synthesized vertical locale cues. *The Journal of the Acoustical Society of America*, **63**(4), 1152–1165.

Wenzel, E. M., & Foster, S. H. 1993 (Oct). Perceptual consequences of interpolating head-related transfer functions during spatial synthesis. *Pages 102–105 of: Applications of Signal Processing to Audio and Acoustics, 1993. Final Program and Paper Summaries., 1993 IEEE Workshop on*.

Wenzel, E. M., Arruda, M., Kistler, D., & Wightman, F. 1993. Localization using nonindividualized head-related transfer functions. *Journal of the Acoustical Society of America*, **94**(1), 111–123.

Wightman, F., & Kistler, D. 1989a. Headphone simulation of free-field listening .1. stimulus synthesis. *Journal of the Acoustical Society of America*, **85**(2), 858–867.

Wightman, F., & Kistler, D. 1989b. Headphone simulation of free-field listening. 2. Psychophysical validation. *Journal of the Acoustical Society of America*, **85**(2), 868–878.

Wightman, F., & Kistler, D. 1993. Multidimensional scaling analysis of head-related transfer functions. *In: 1993 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. IEEE.

Wightman, F., & Kistler, D. 1997. *Binaural and Spatial Hearing in Real and Virtual Environments. Edited by R.H. Gilkey and T.R. Anderson.* Lawrence Erlbaum Associates, Mahwah, NJ. Chap. Factors Affecting the Relative Salience of Sound Localization.

Wightman, F., & Kistler, D. 1999. Resolution of front–back ambiguity in spatial hearing by listener and source movement. *Journal of the Acoustical Society of America*, **105**(5), 2841–2853.

Wightman, F., & Kistler, D. 2005. Measurement and validation of human HRTFs for use in hearing research. *Acta Acustica united with Acustica*, **91**(3), 429–439.

Xu, S., Li, Z., & Salvendy, G. 2009. Identification of anthropometric measurements for individualization of head-related transfer functions. *Acta Acustica united with Acustica*, **95**, 168–177.

Zahorik, P., Brungart, D. S., & Bronkhorst, A. W. 2005. Auditory distance perception in humans: A summary of past and present research. *Acta Acustica united with Acustica*, **91**(3), 409–420.

Zahorik, P., Bangayan, P., Sundareswaran, V., Wang, K., & Tam, C. 2006. Perceptual recalibration in human sound localization: Learning to remediate front-back reversals. *Journal of the Acoustical Society of America*, **120**(1), 343–359.

Zotkin, D. N., Duraiswami, R., & Davis, L. S. 2004. Rendering localized spatial audio in a virtual auditory space. *IEEE Transactions on Multimedia*, **6**(4), 553–564.

Zotkin, D. N., Duraiswami, R., Grassi, E., & Gumerov, N. A. 2006. Fast head-related transfer function measurement via reciprocity. *Journal of the Acoustical Society of America*, **120**(4), 2202–2215.

Appendix A

HRTFs measured on a Neumann KU 100 dummy-head

The following figures show all the measured and post-processed HRTFs from the dummy-head Neumann KU 100 reported in Chapter 2. The reader is referred to that Chapter for further details on the measurement and post-processing procedures. The coordinate system has been explained in Chapter 1, 1.5. In the following figures, black lines correspond to the left ear signals while grey lines correspond to the right ear signals.

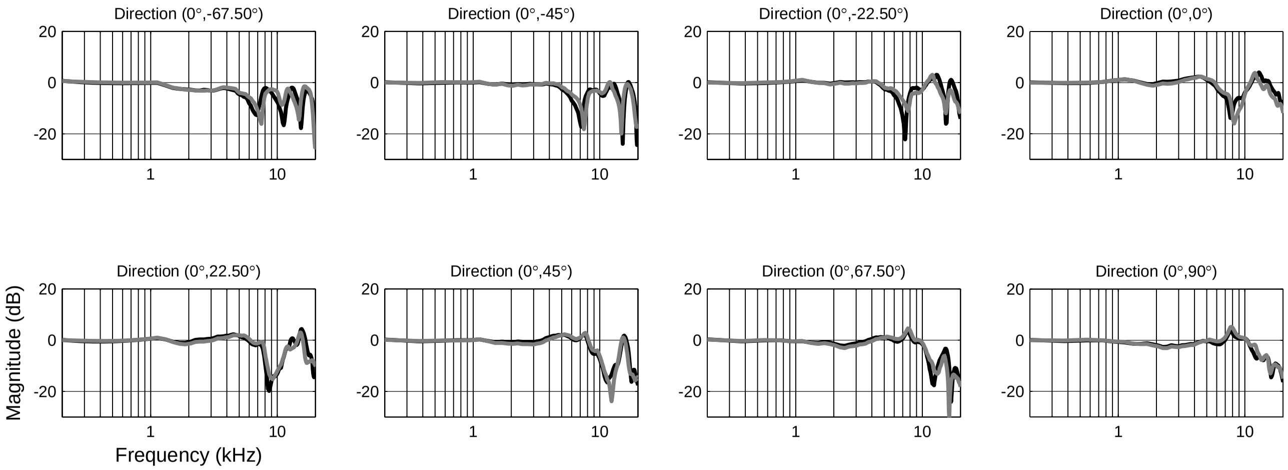

Figure A.1: Neumann KU 100 dummy-head HRTFs measurements from the MSP.

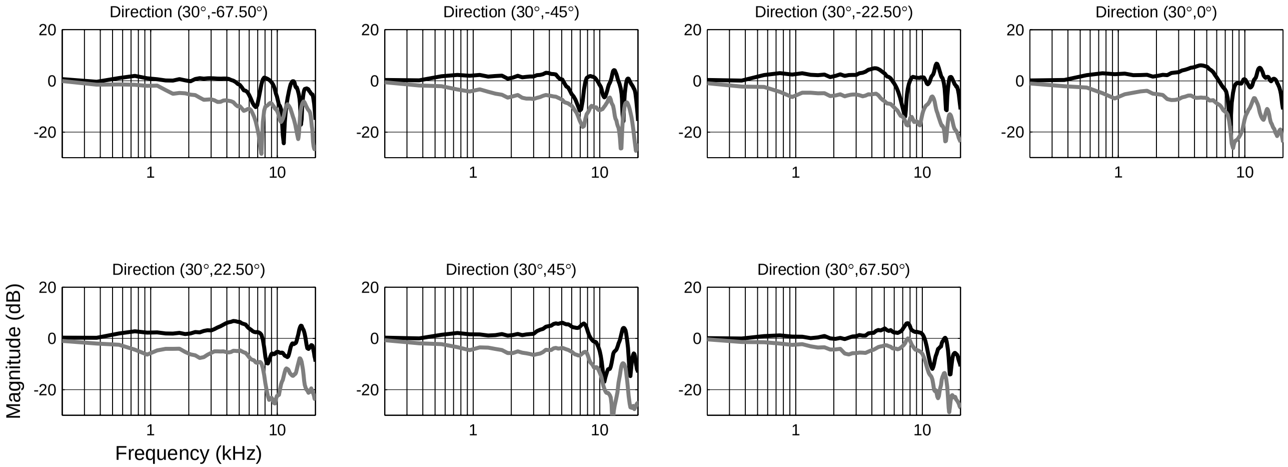

Figure A.2: Same as Fig. A.1, but for the 30° azimuth plane.

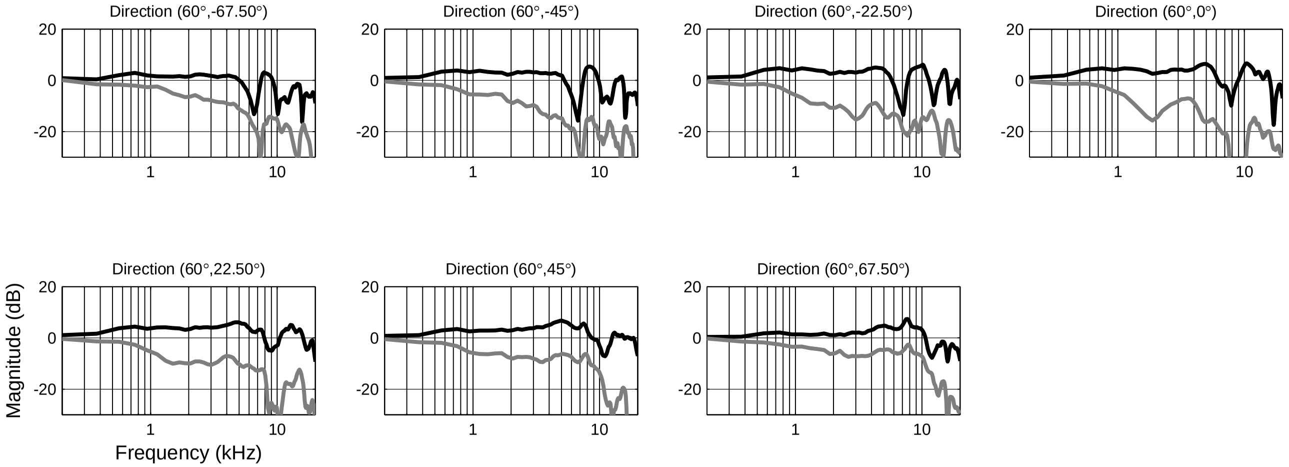

Figure A.3: Same as Fig. A.1, but for the 60° azimuth plane.

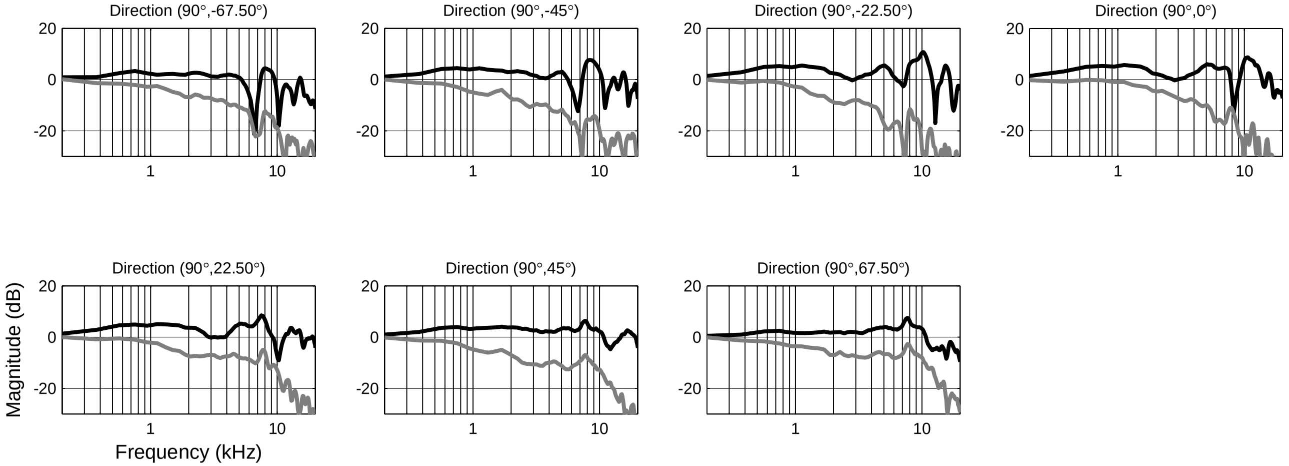

Figure A.4: Same as Fig. A.1, but for the 90° azimuth plane.

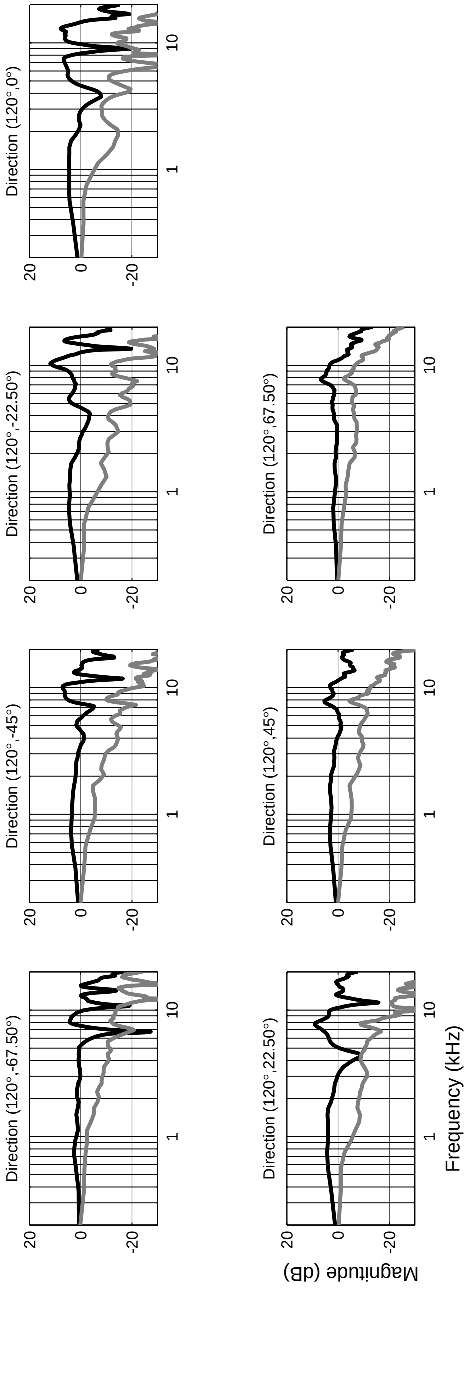

Figure A.5: Same as Fig. A.1, but for the 120° azimuth plane.

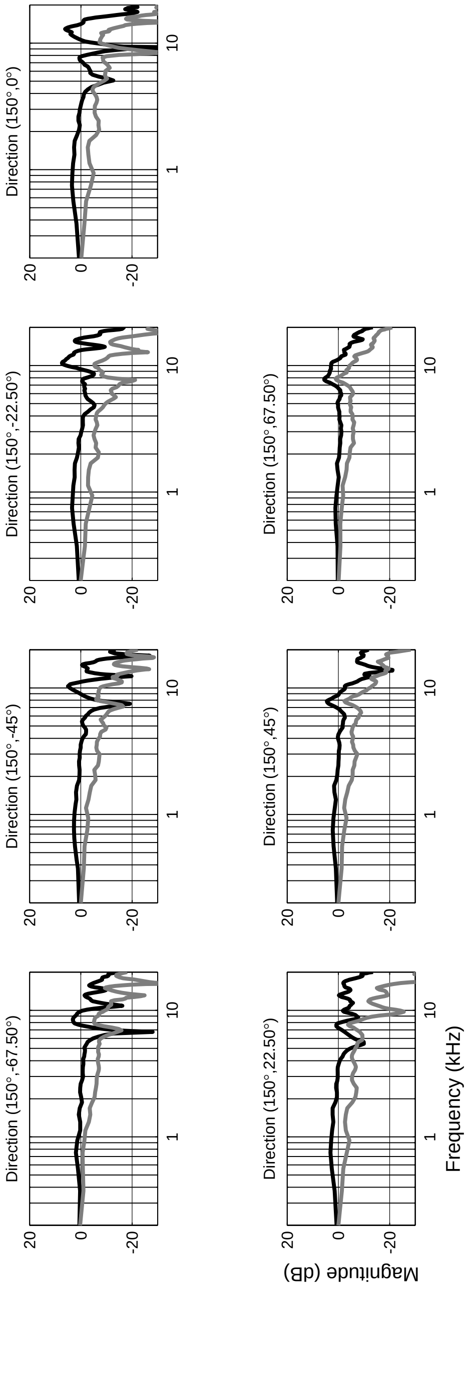

Figure A.6: Same as Fig. A.1, but for the 150° azimuth plane.

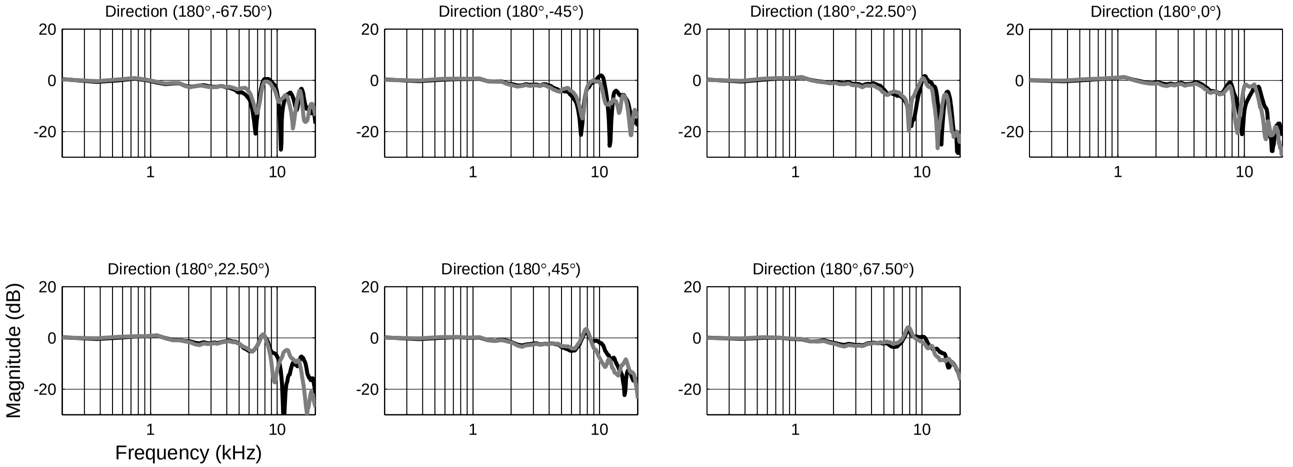

Figure A.7: Same as Fig. A.1, but for the 180° azimuth plane.

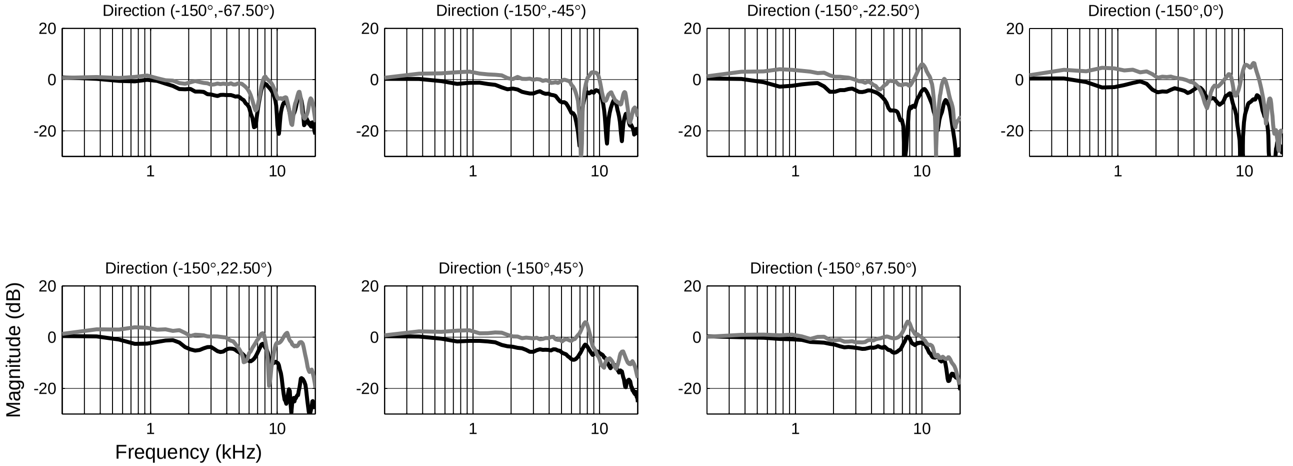

Figure A.8: Same as Fig. A.1, but for the $-150°$ azimuth plane.

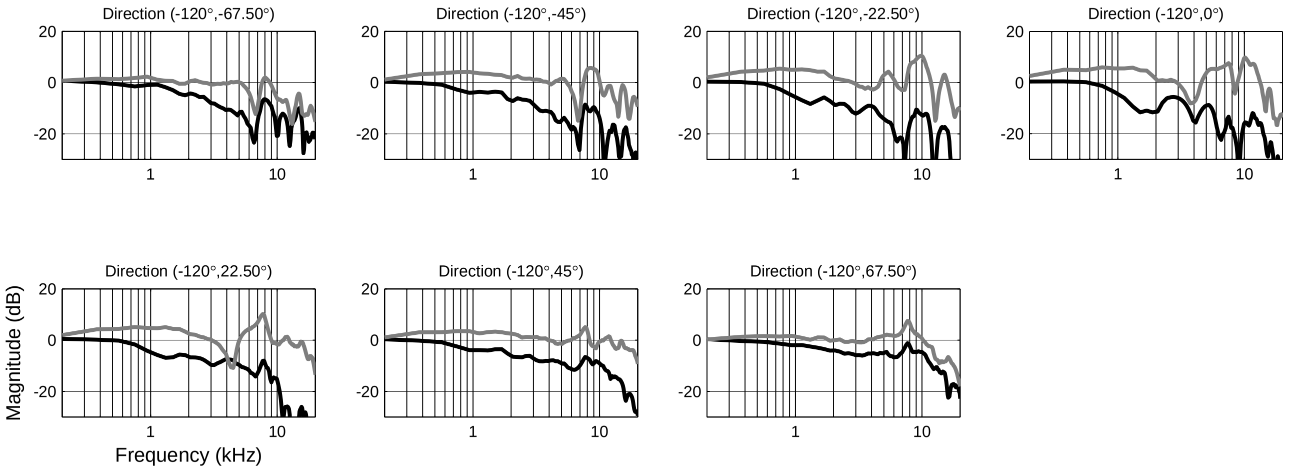

Figure A.9: Same as Fig. A.1, but for the -120° azimuth plane.

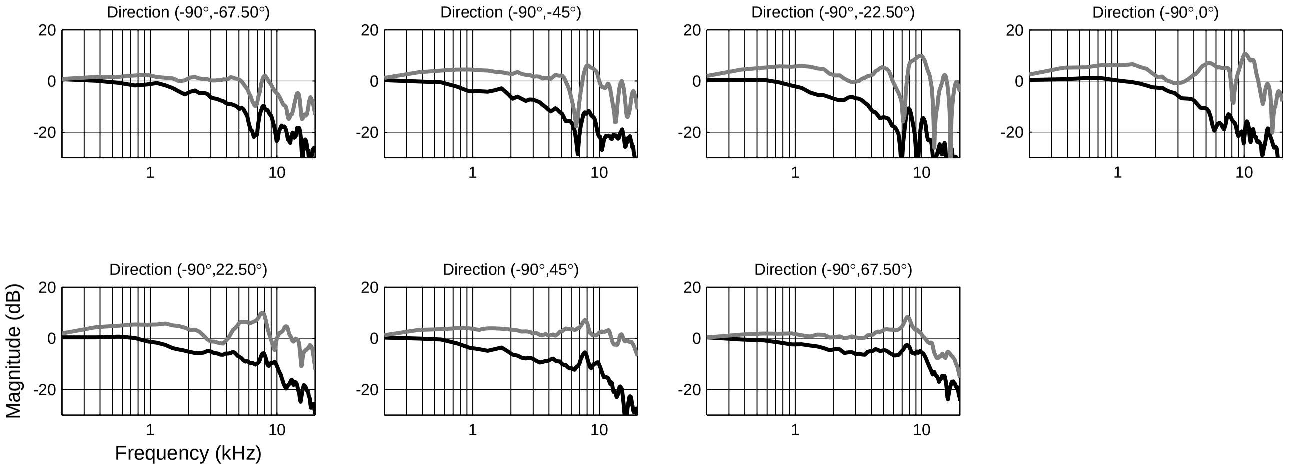

Figure A.10: Same as Fig. A.1, but for the -90° azimuth plane.

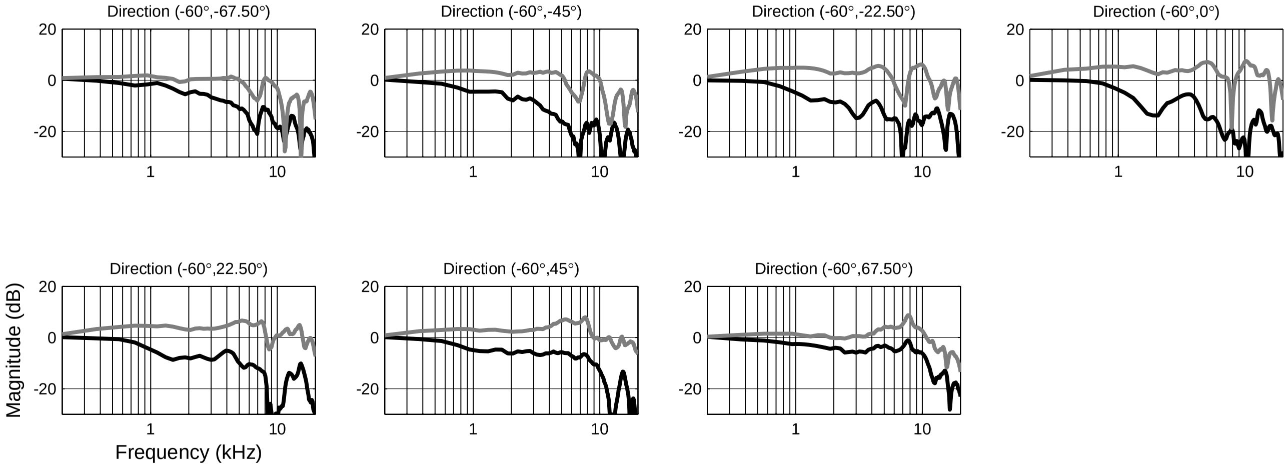

Figure A.11: Same as Fig. A.1, but for the -60° azimuth plane.

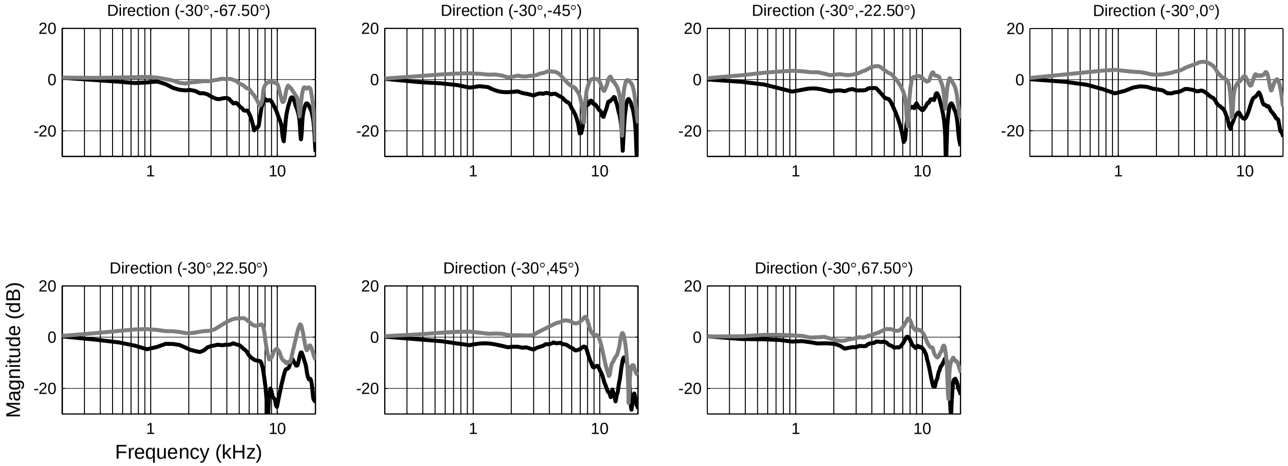

Figure A.12: Same as Fig. A.1, but for the -30° azimuth plane.

Appendix B

Results not included in Chapter 5

Extensive results were obtained from the localization experiments, HRTFs matching procedures and parameterization of spectral features reported in Chapter 5. Therefore, only examples from four subjects (two from Group A and another two for Group B) were given in the main text of Chapter 5 for localization experiments and HRTFs matching procedures. The results for the remaining six subjects (three from Group A and another three for Group B) are given here.

Figures B.1 to B.3 show localization judgements for the remaining three subjects categorized in Group A, for the three conditions tested. Subplots in each panel are labeled according to the corresponding condition: *Real Life*, *Individual* and from *Non-individual 1* to *Non-individual 10*. Subject codes are given in the legend of each Figure panel: JC (Fig. B.1), BG (Fig. B.2) and AM (Fig. B.3). In these figures, each main subplot corresponds to the perceived elevation under one condition, while the small subplots contained in the main ones correspond to the perceived azimuth. For further details about the figures, the reader is referred to 5.4.1.

Figures B.4 to B.6 show the results of matching HRTFs for Group A, which were obtained as explained in 5.3.1. Subject codes are given in the legends. Figs. B.4 to B.6 are organized so that there are 15 subplots in each Figure panel, each of them corresponding to one individual pair of HRTFs. In each subplot, the individual pair of HRTFs is plotted along with all the non-individual HRTFs which matched it. The title of each subplot reads the coordinates of the centroid of the evoked distribution for the individual pair of HRTFs which constitutes the basis of the subplot. This is marked as *'Evoked centroid'* and it is given in (azimuth θ, elevation ϕ) coordinates, unless the distribution was found to be uniform or isotropic. In that case, the title reads *'Evoked centroid: Isotropic'* and the plot is left empty. The characteristics of the distributions can easily be seen in Figs. B.1 to B.3. The title of each subplot also reads the original direction for which the individual HRTFs pair in question was measured, and it is given under *'Measurement direction'*. This is also shown in (azimuth θ, elevation ϕ) coordinates.

Similar results for the three remaining subjects in Group B are given in Figs. B.7 to B.9 (localization experiments) and Figs. B.10 to B.12 (matching HRTFs). For more details, see 5.4.2.

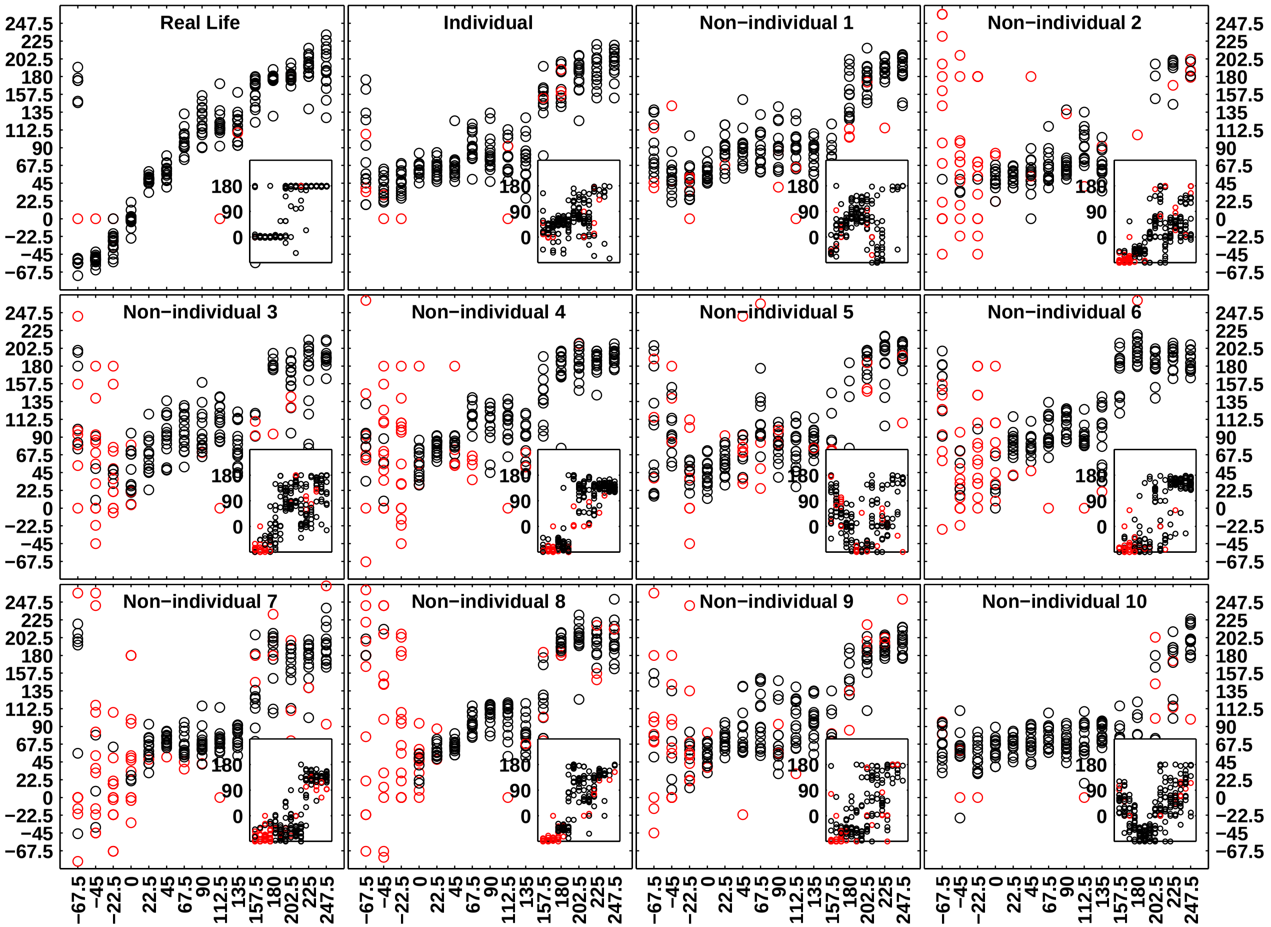

Figure B.1: Localization performance results for the different experimental conditions tested, for subject JC. The computed correlation value between real sound sources and individual HRTFs binaural synthesis centroids is $\hat{\rho}v = 0.2$ and the null hypothesis $\hat{\rho}v = 0$ is rejected in favor of $\hat{\rho}v > \hat{\rho}_{\alpha}$.

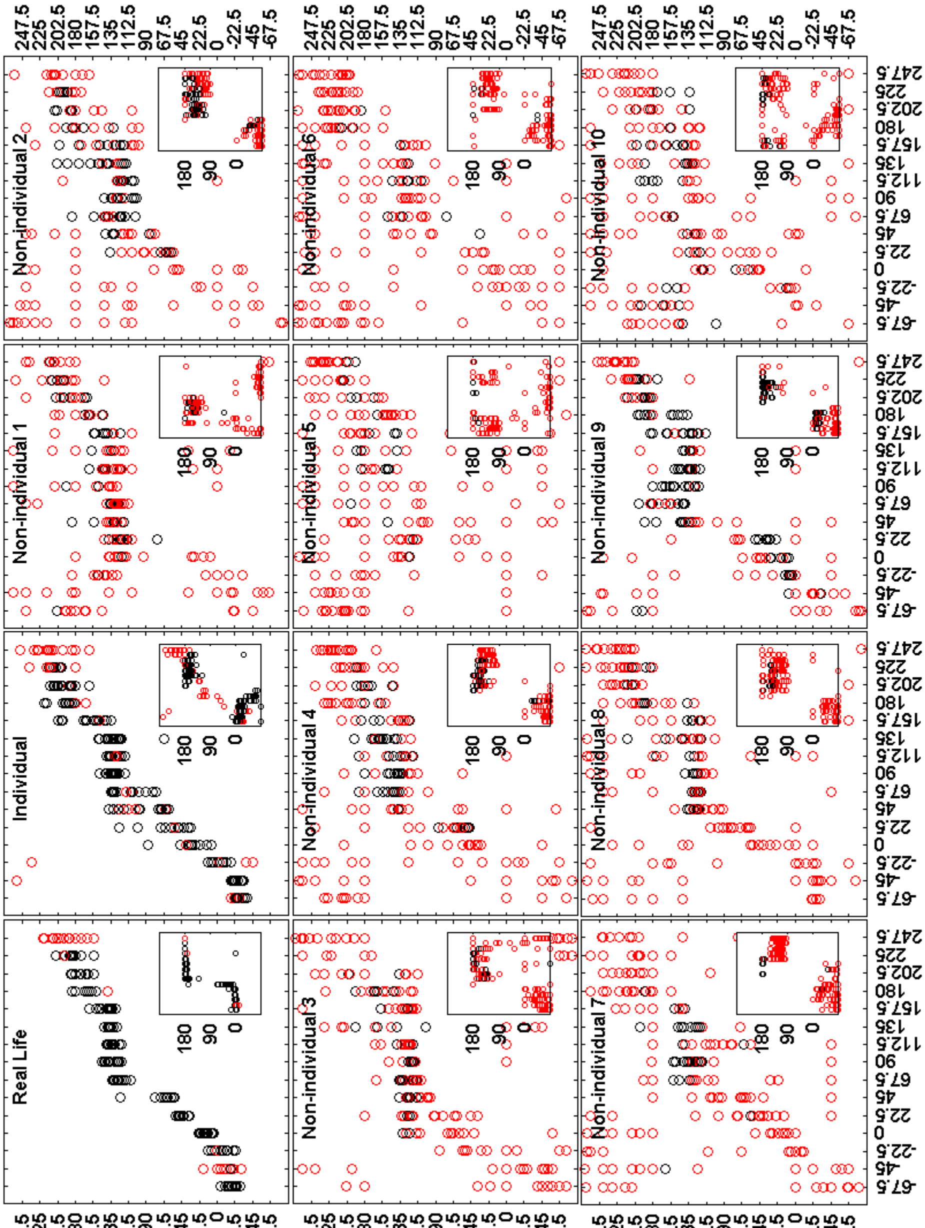

Figure B.2: Same as Fig. B.1, but for subject BG. The computed correlation value between centroids is $\hat{\rho}\nu = -0.4$ and the null hypothesis $\hat{\rho}\nu = 0$ is rejected in favor of $\hat{\rho}\nu < \hat{\rho}_{1-\alpha}$.

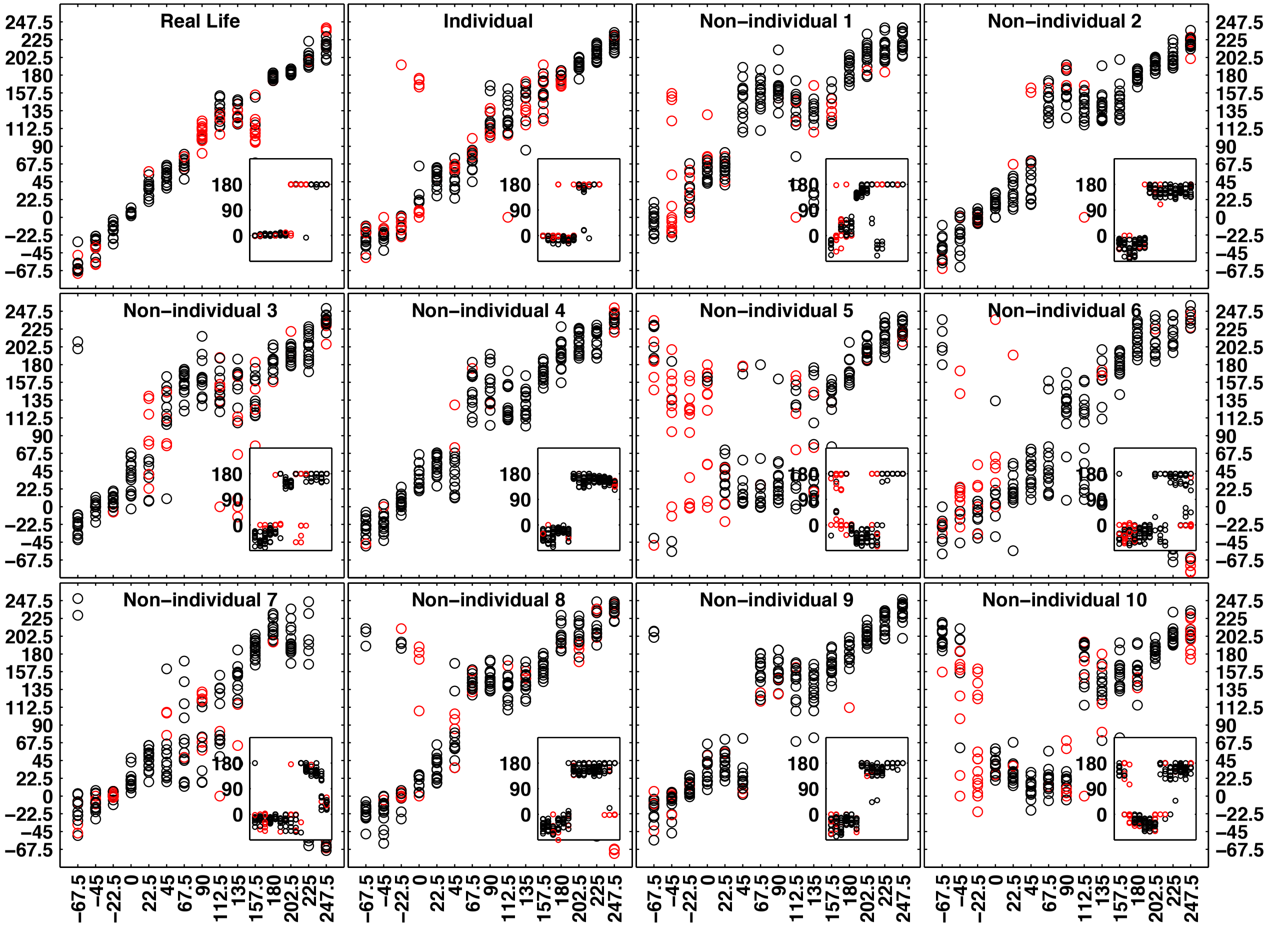

Figure B.3: Same as Fig. B.1, but for subject AM. The computed correlation value between centroids is $\hat{\rho}\nu = -0.4$ and the null hypothesis $\hat{\rho}\nu = 0$ is rejected in favor of $\hat{\rho}\nu < \hat{\rho}_{1-\alpha}$.

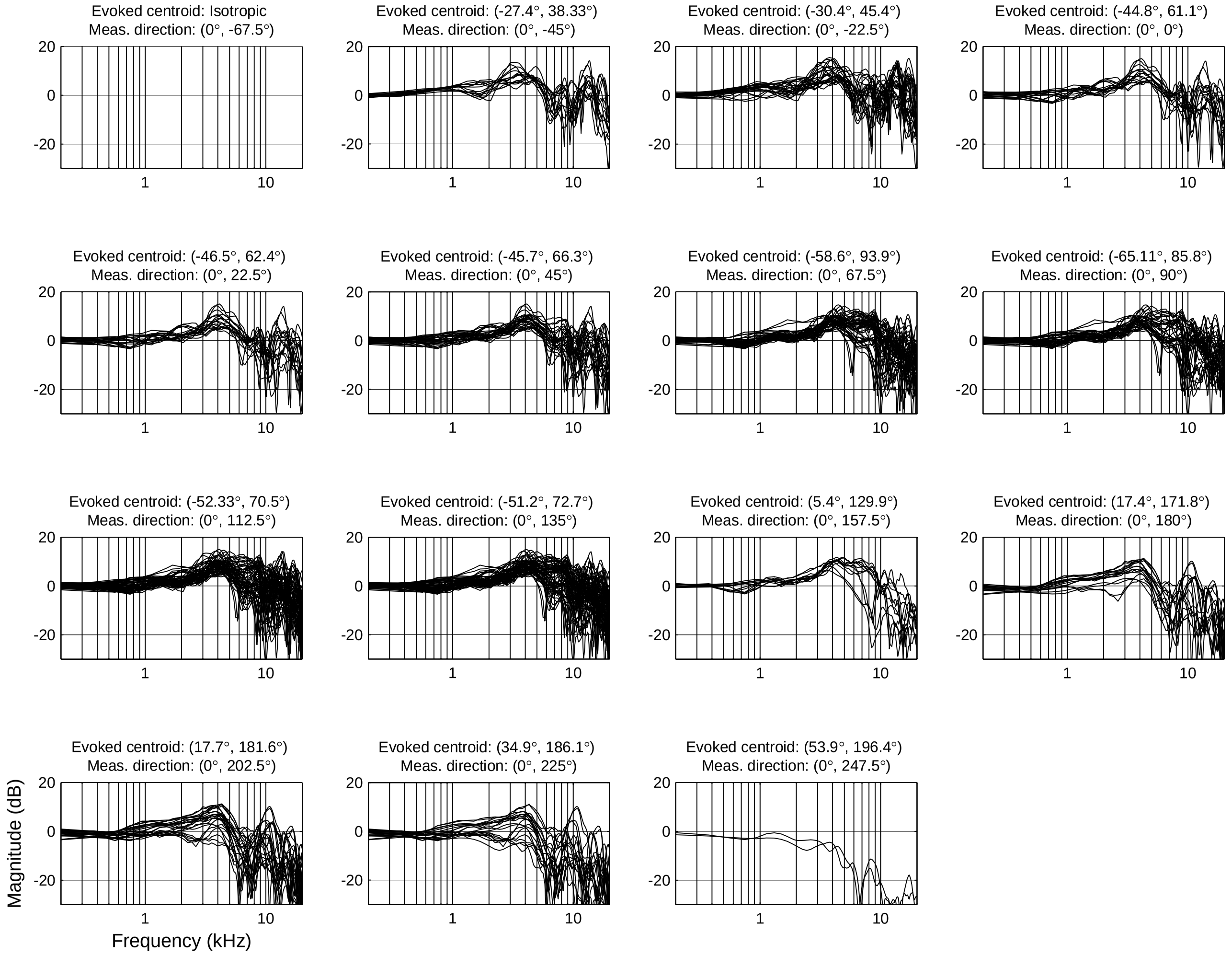

Figure B.4: Results from the HRTFs matching procedure, for subject JC. The figure shows the non-individual HRTFs that evoked a localization that matched those of the individual HRTFs, as obtained from the behavioral data presented in Fig.B.1.

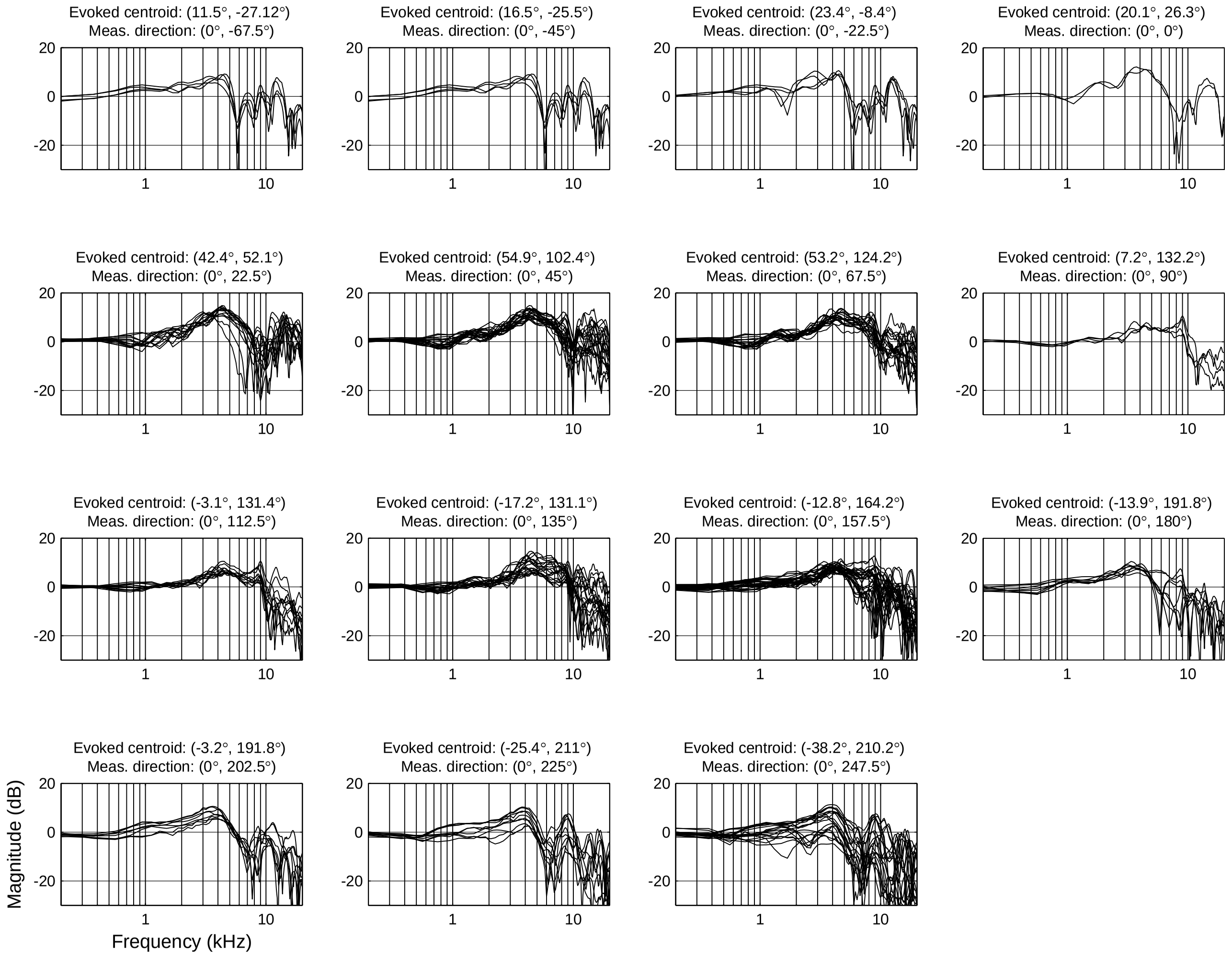

Figure B.5: Same as Fig. B.4, but for subject BG. The HRTFs matching procedure was based on the behavioral data presented in Fig. B.2.

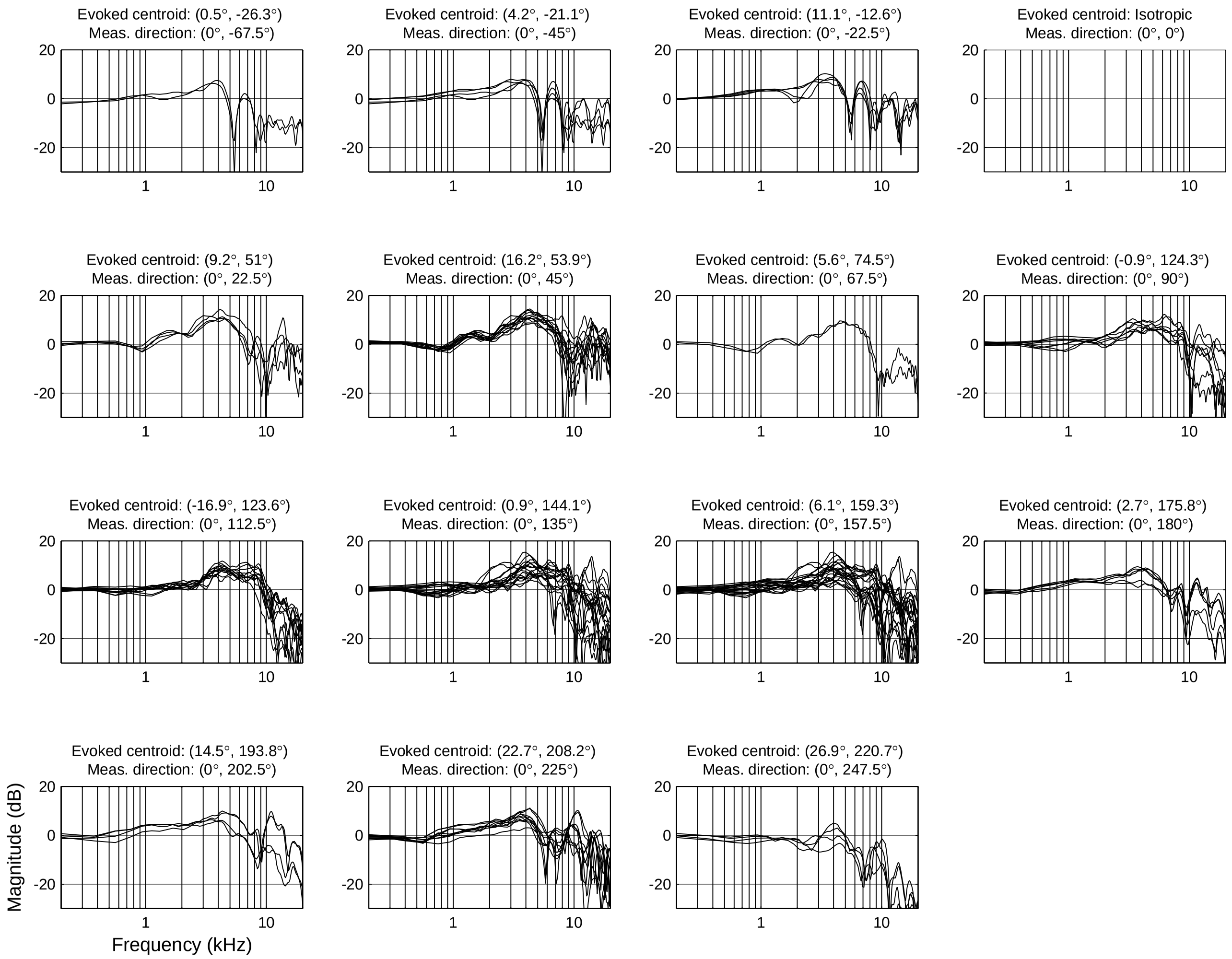

Figure B.6: Same as Fig. B.4, but for subject AM. The HRTFs matching procedure was based on the behavioral data presented in Fig. B.3.

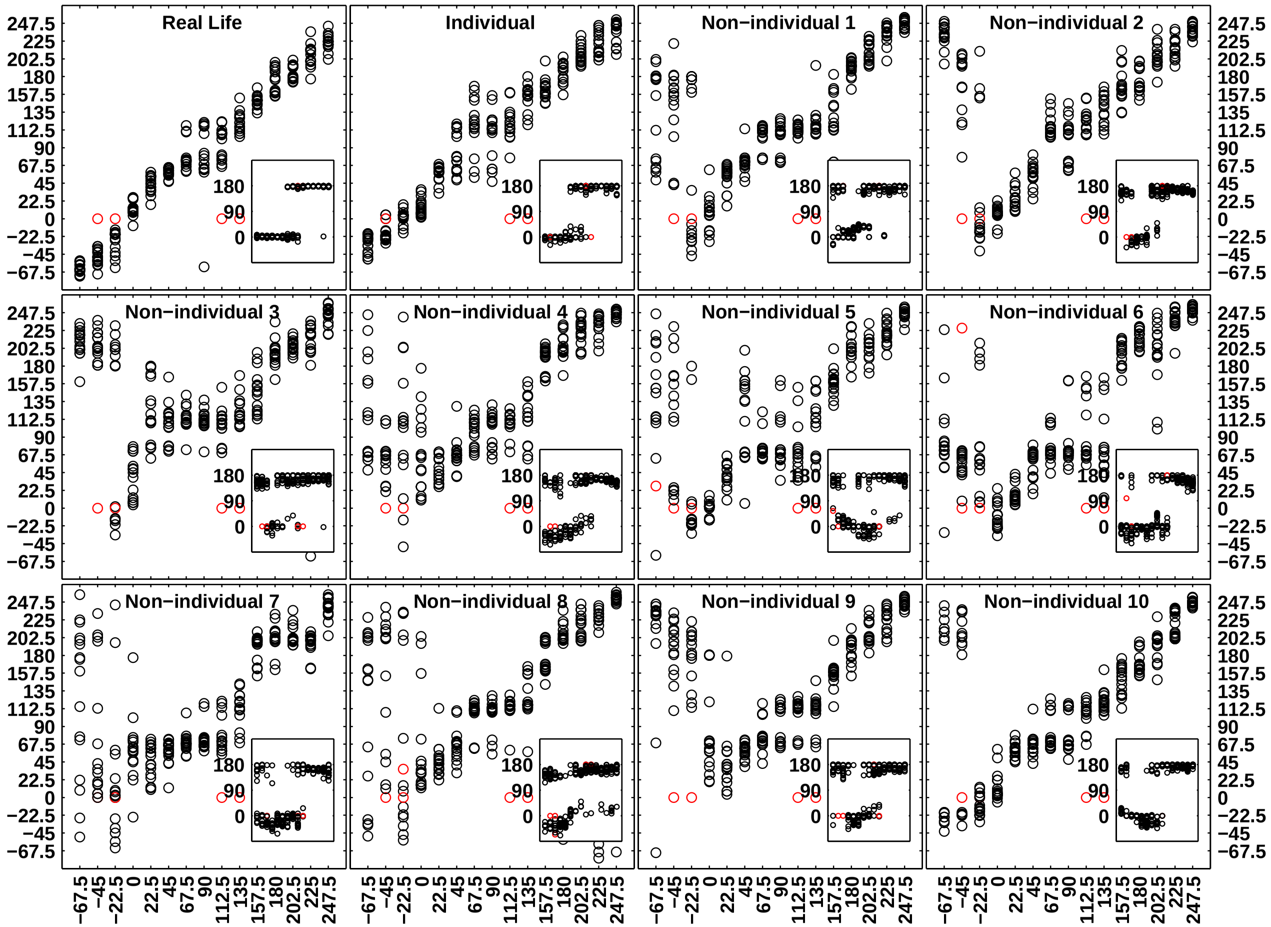

Figure B.7: Localization performance results for the different experimental conditions tested, for subject AW. The computed correlation value between real sound sources and individual HRTFs binaural synthesis centroids is $\hat{\rho}v = -0.2$ and the null hypothesis $\hat{\rho}v = 0$ is rejected in favor of $\hat{\rho}v < \hat{\rho}_{1-\alpha}$.

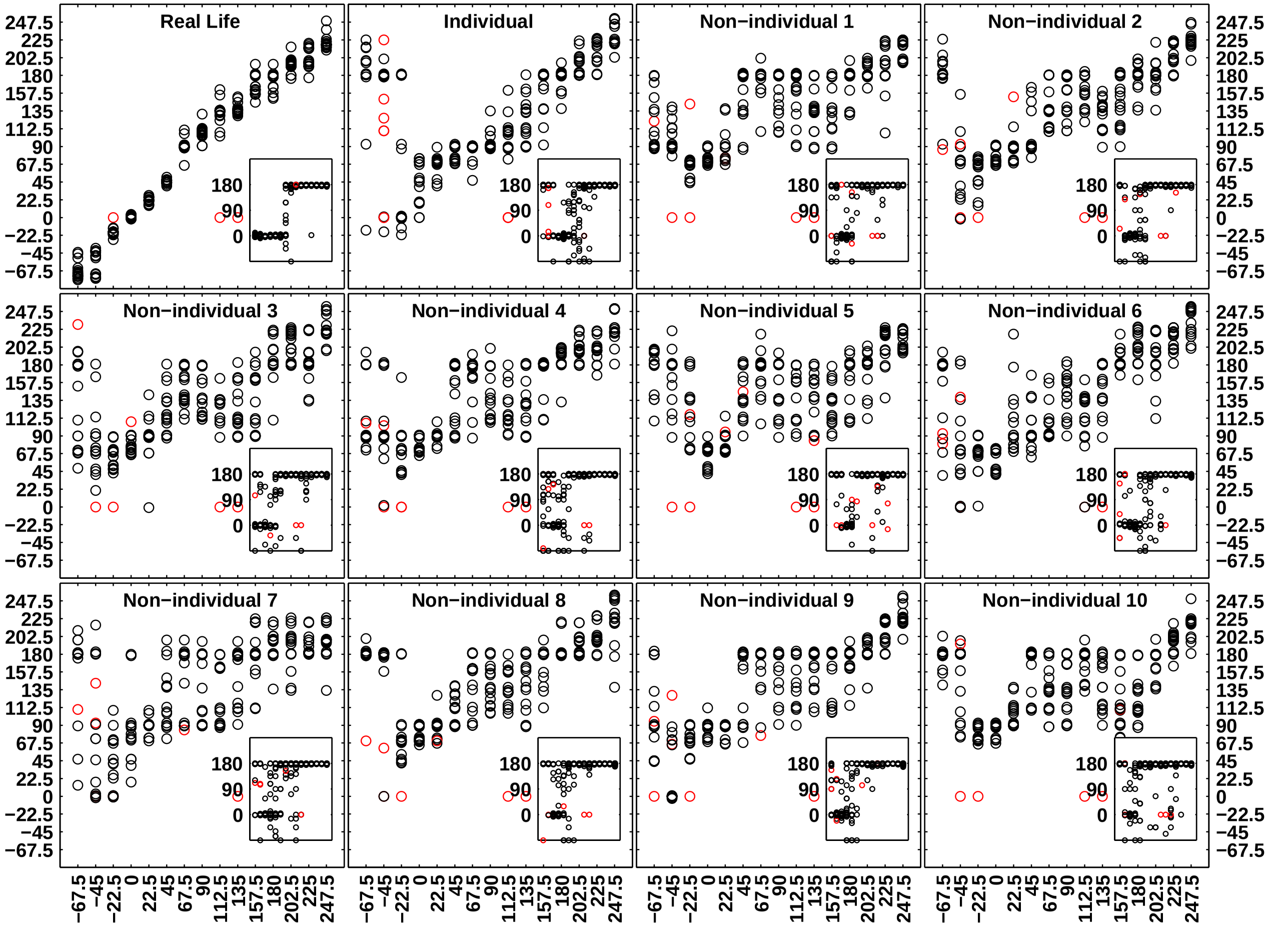

Figure B.8: Same as Fig. B.7, but for subject SR. The computed correlation value between centroids is $\hat{\rho}\nu = -0.3$ and the null hypothesis $\hat{\rho}\nu = 0$ is rejected in favor of $\hat{\rho}\nu < \hat{\rho}_{1-\alpha}$.

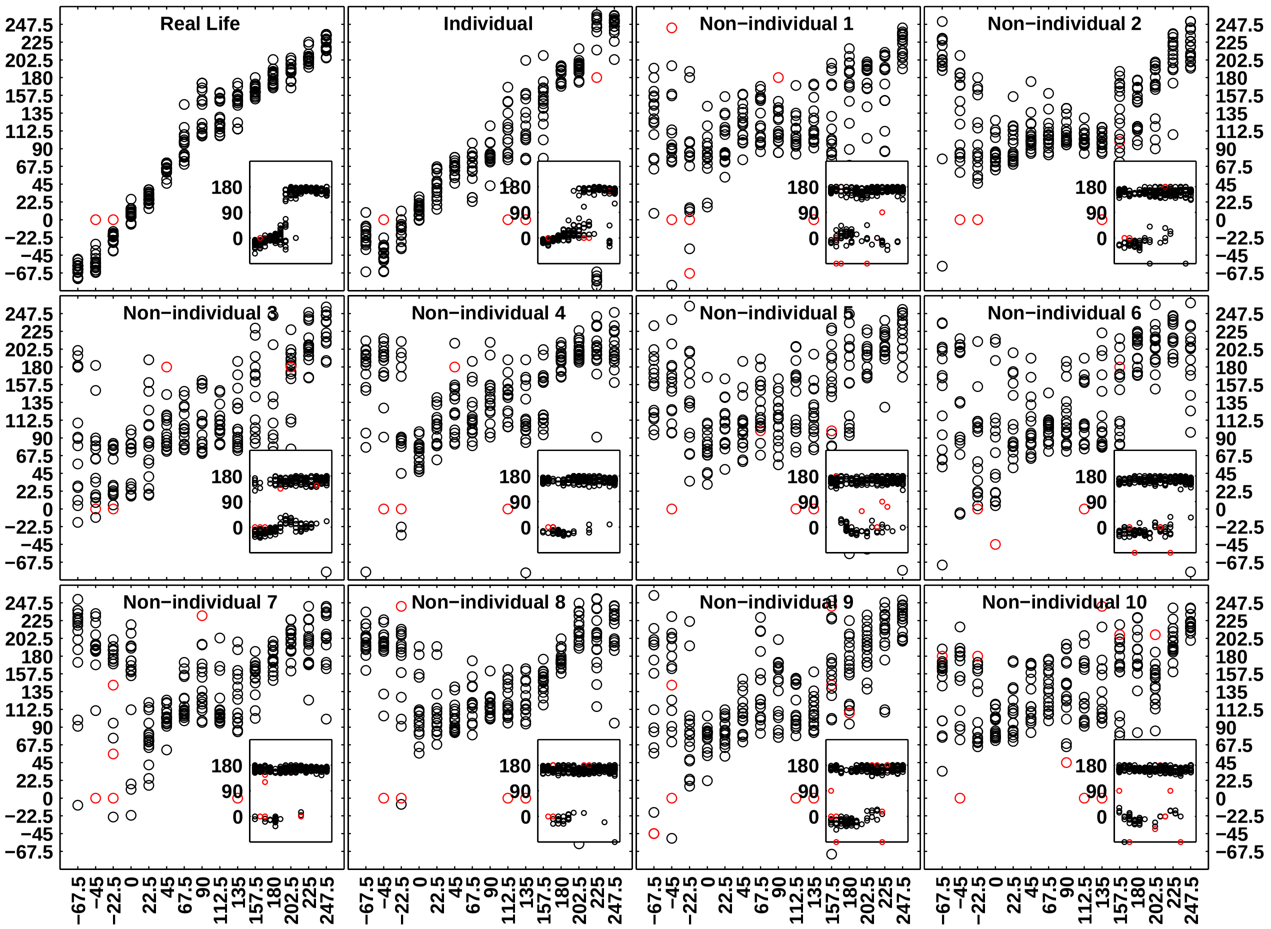

Figure B.9: Same as Fig. B.7, but for subject LA. The computed correlation value between centroids is $\hat{\rho}v = 0.3$ and the null hypothesis $\hat{\rho}v = 0$ is rejected in favor of $\hat{\rho}v > \hat{\rho}_\alpha$.

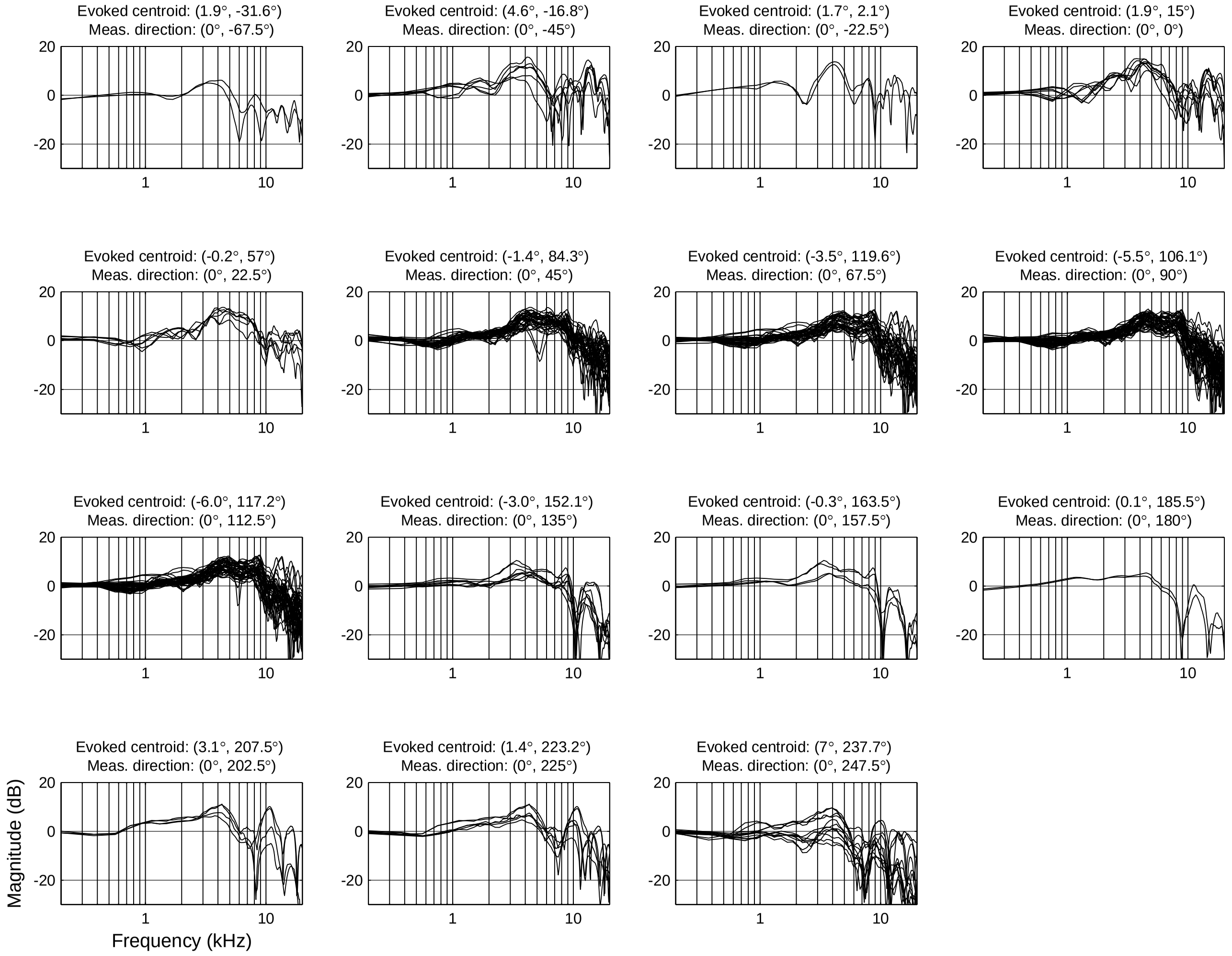

Figure B.10: Results from the HRTFs matching procedure, for subject AW. The figure shows the non-individual HRTFs that evoked a localization that matched those of the individual HRTFs, as obtained from the behavioral data presented in Fig.B.7.

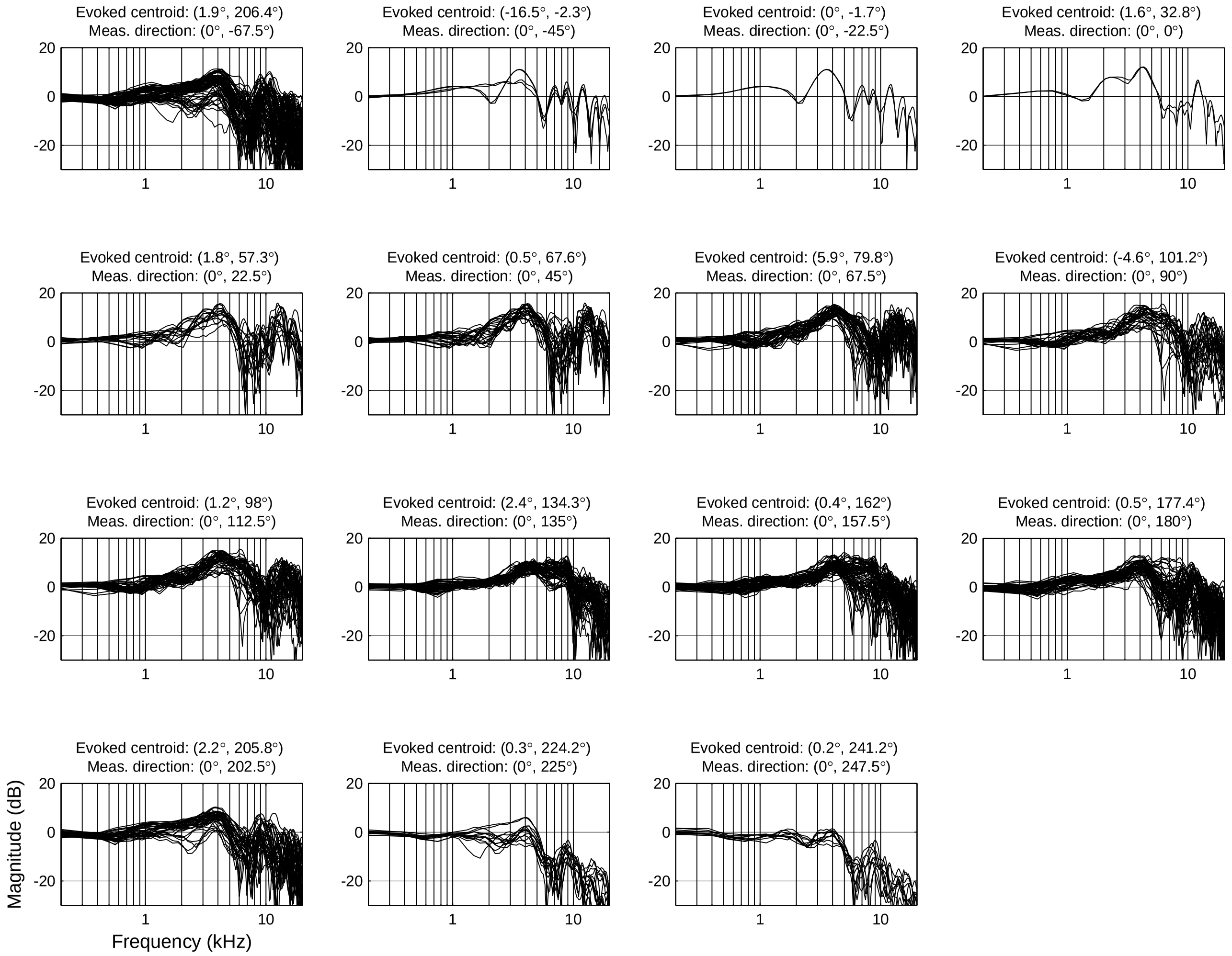

Figure B.11: Same as Fig. B.10, but for subject SR. The HRTFs matching procedure was based on the behavioral data presented in Fig. B.8.

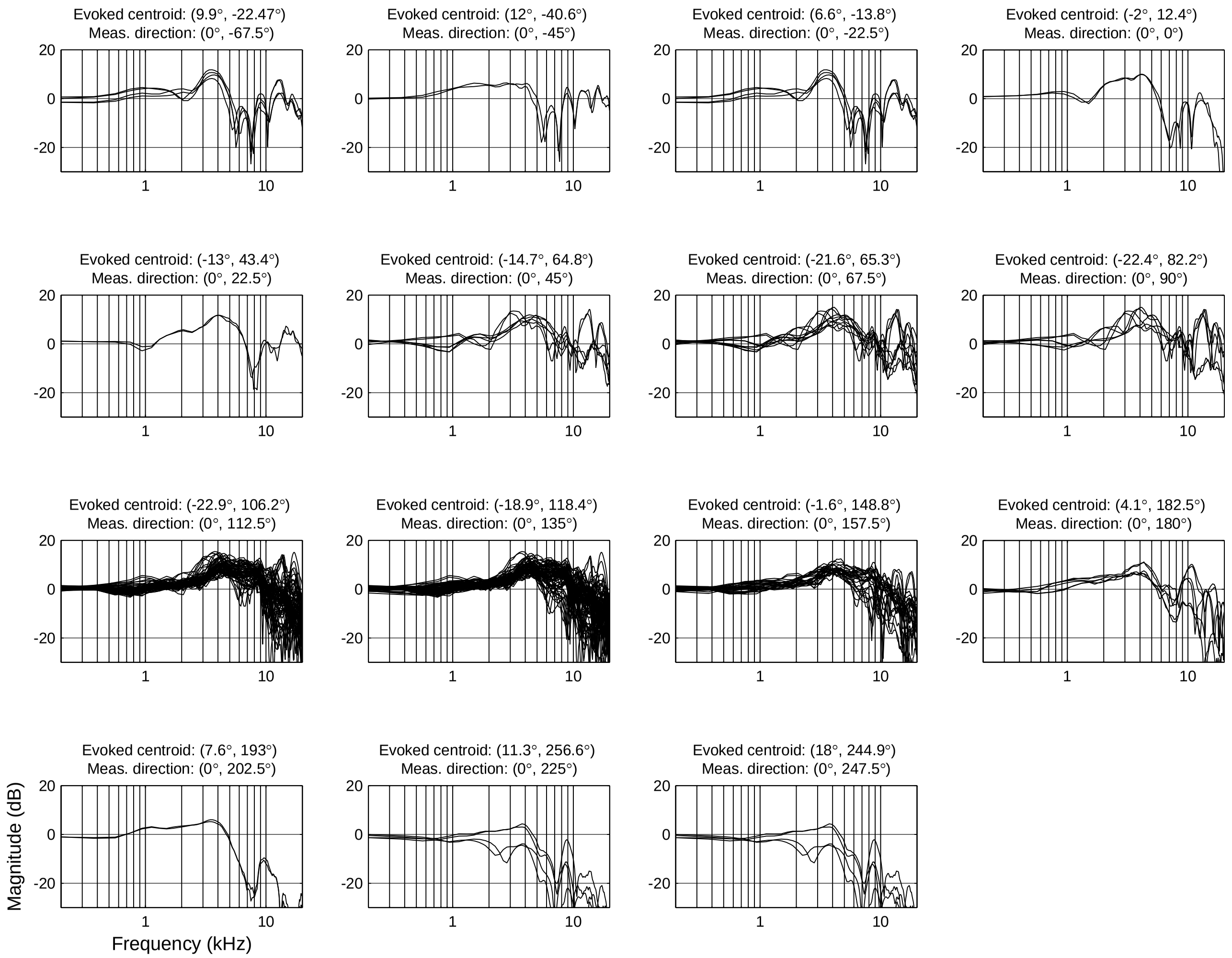

Figure B.12: Same as Fig. B.10, but for subject LA. The HRTFs matching procedure was based on the behavioral data presented in Fig. B.9.

Appendix C

Results not included in Chapter 7

This Appendix includes further results from Chapter 7. The ratio between non-individual to individual HRTFs was performed for all the subjects participating in the experiments of Chapter 5, for those non-individual HRTFs that perceptually matched individual HRTFs. In the main text of Chapter 7, only results for four subjects were shown. Results for the rest of the subjects are included here. It has to be noted that some of the subplots in each figure panel showing the results are empty. Those are either cases in which there were not non-individual HRTFs pairs perceptually matching the individual HRTFs pair, or in which the perception with the individual HRTFs pair led to an isotropic distribution. The reader is referred to Chapter 7 for an analysis of the Figures presented here.

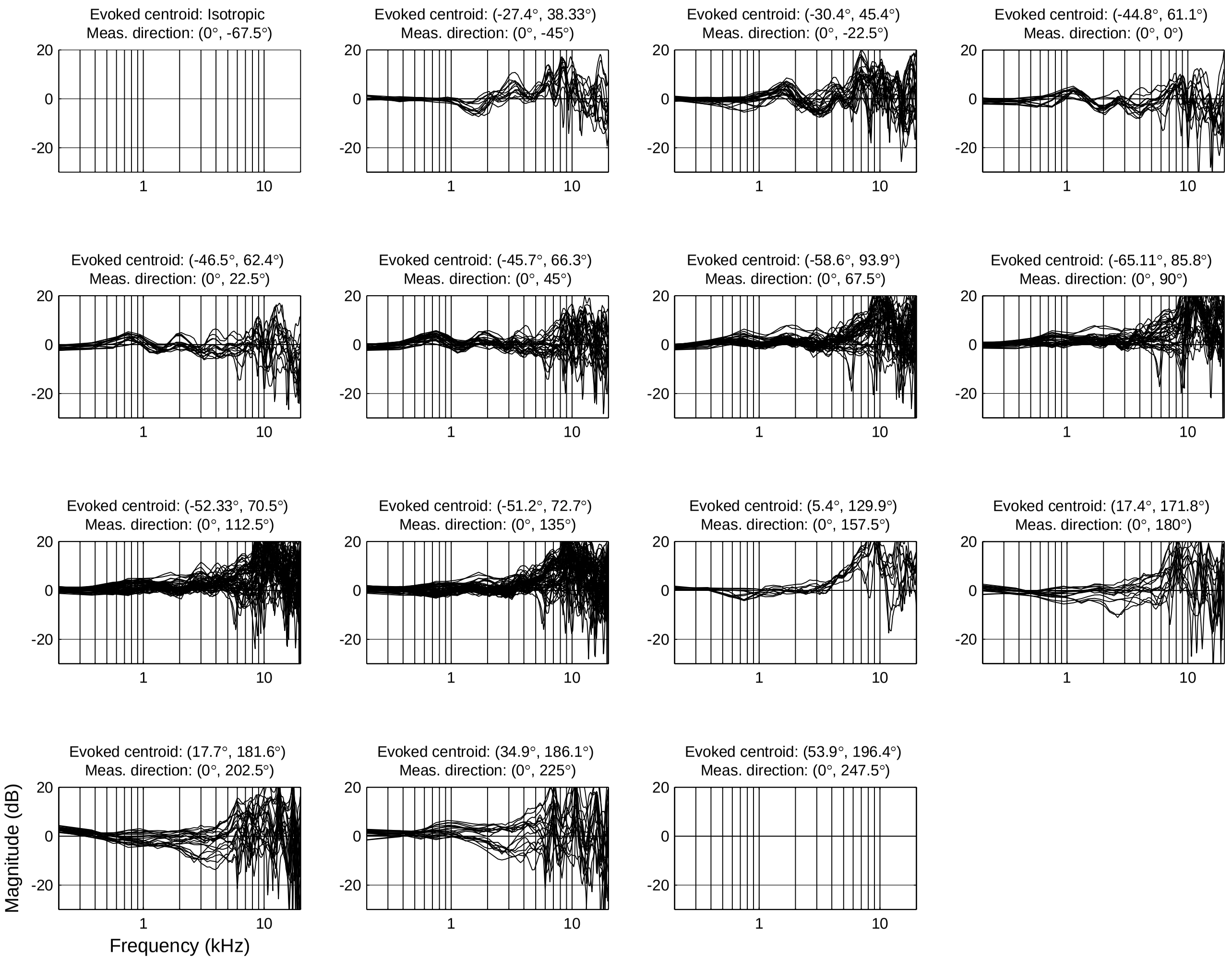

Figure C.1: Subject JC. Computed ratio $HRTF_{non-individual}/HRTF_{individual}$, based on those matching HRTFs already shown in Figure B.4.

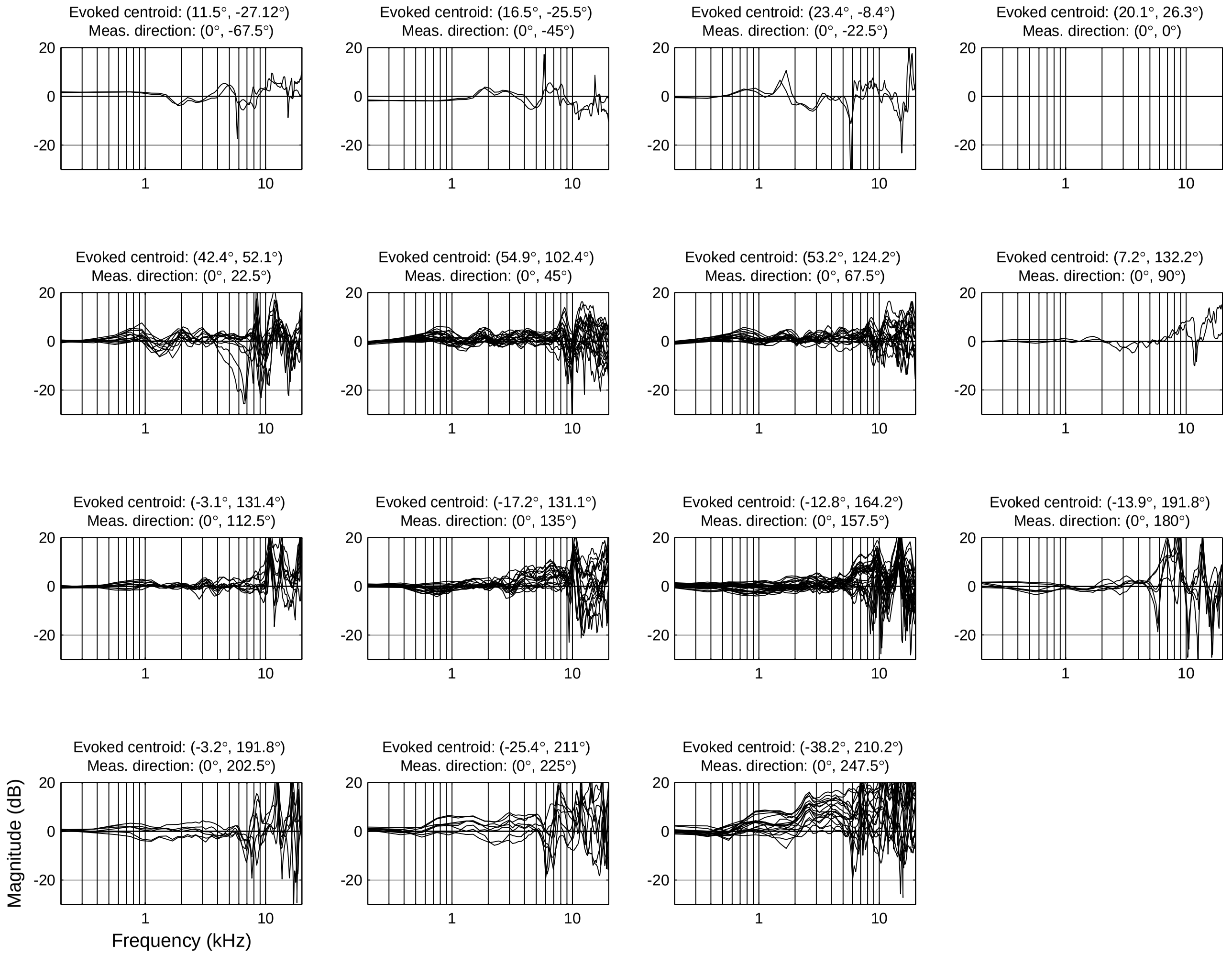

Figure C.2: Subject BG. Computed ratio $HRTF_{non-individual}/HRTF_{individual}$, based on those matching HRTFs already shown in Figure B.5.

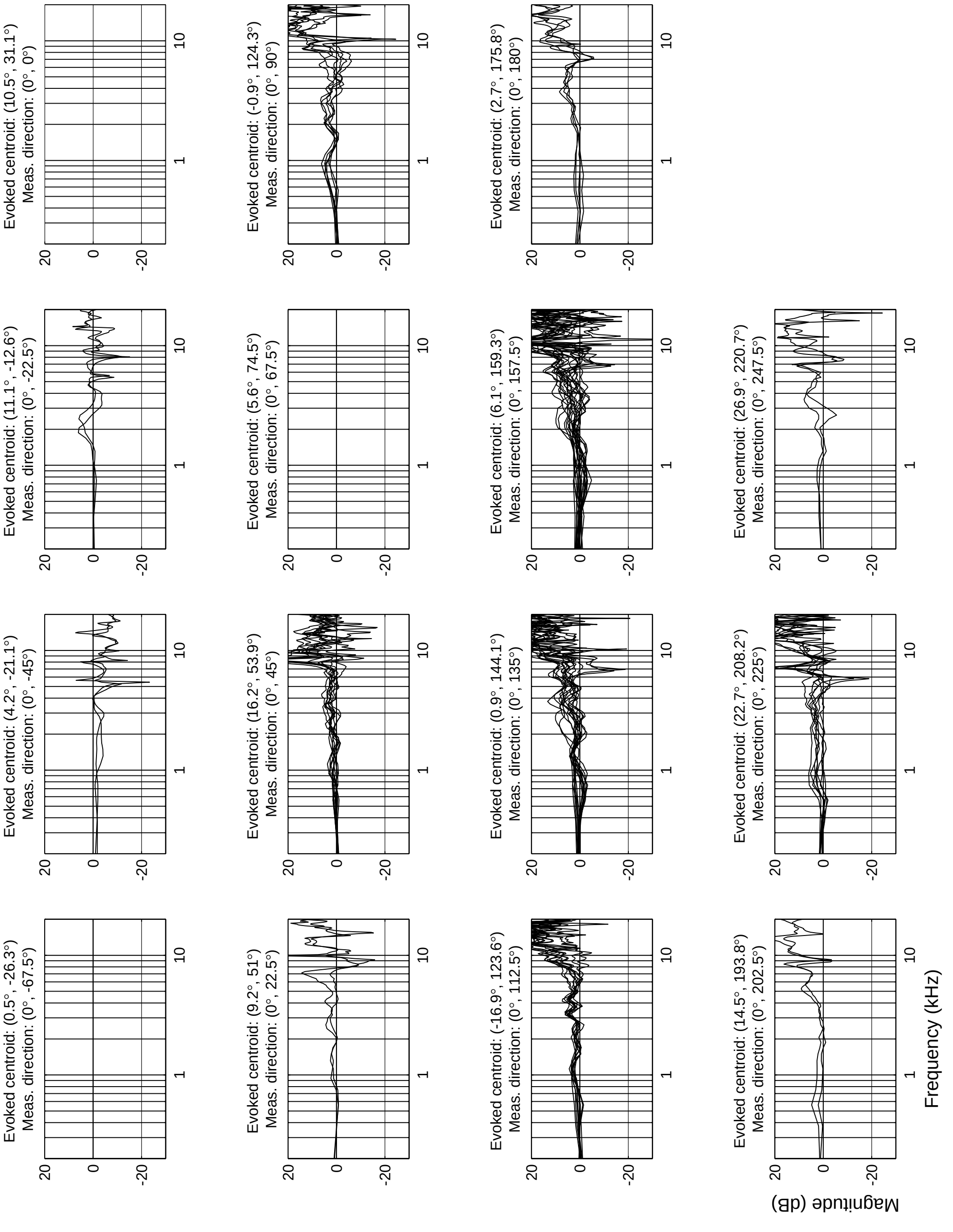

Figure C.3: Subject AM. Computed ratio $HRTF_{non-individual}/HRTF_{individual}$, based on those matching HRTFs already shown in Figure B.6.

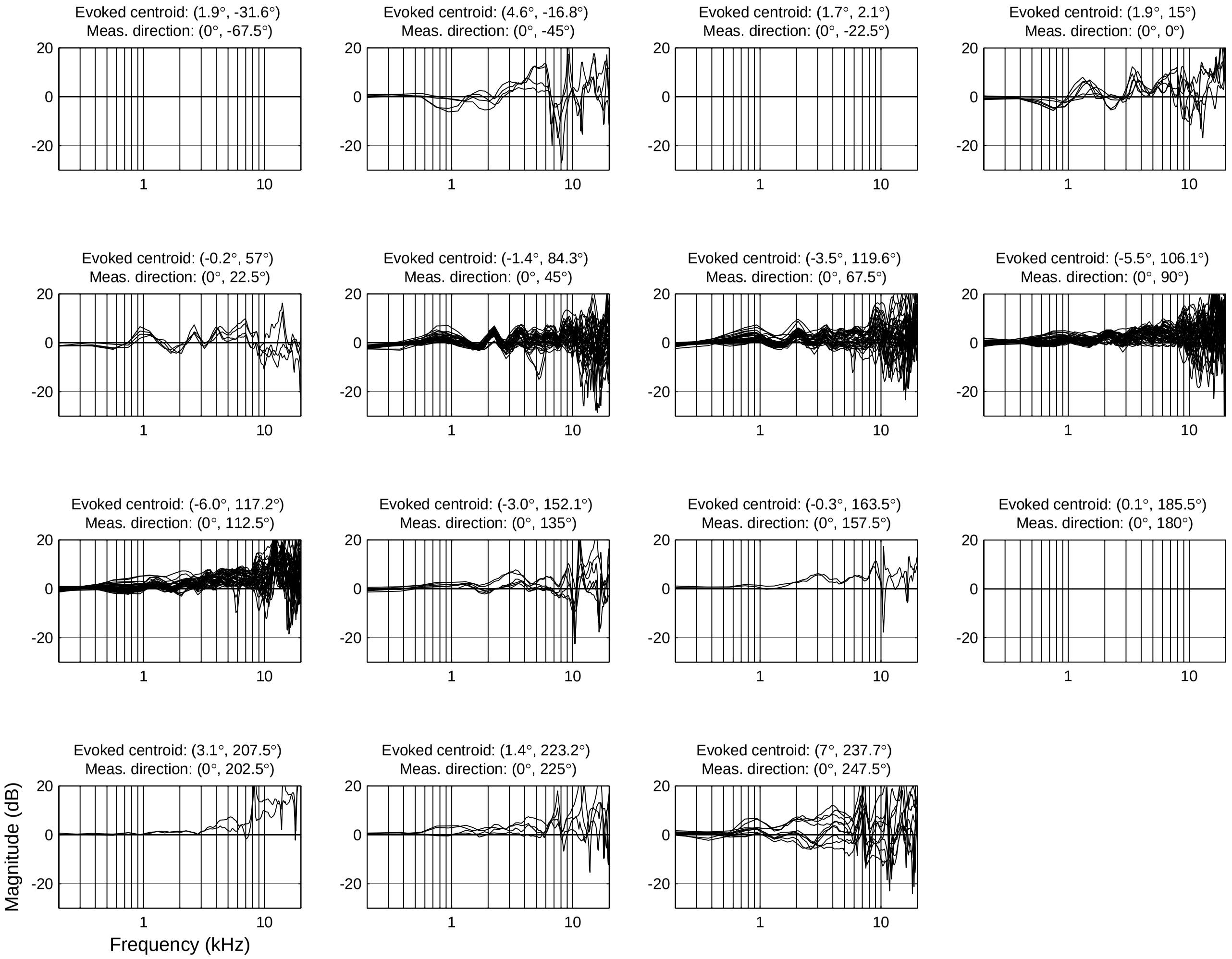

Figure C.4: Subject AW. Computed ratio $HRTF_{non-individual}/HRTF_{individual}$, based on those matching HRTFs already shown in Figure B.10.

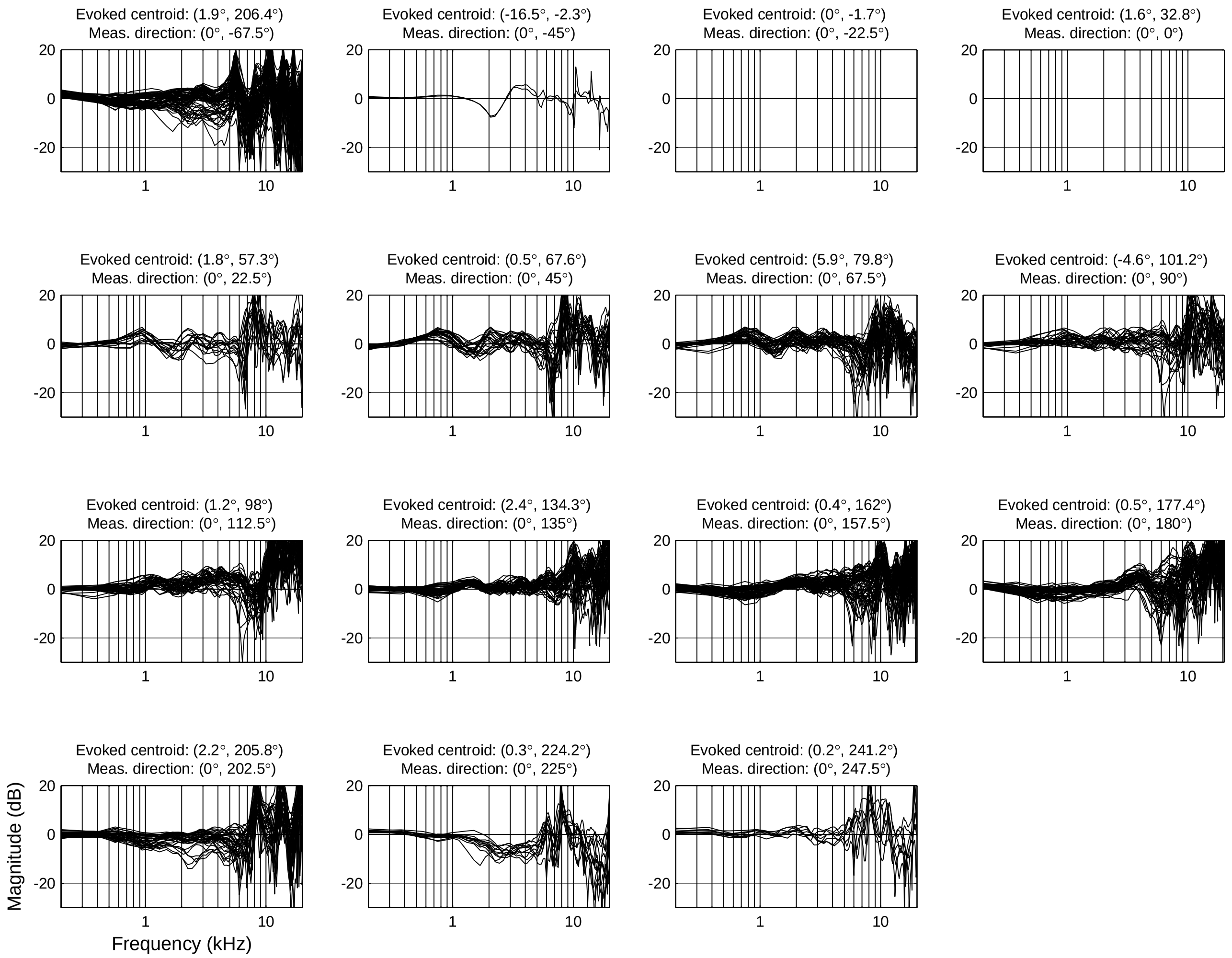

Figure C.5: Subject SR. Computed ratio $HRTF_{non-individual}/HRTF_{individual}$, based on those matching HRTFs already shown in Figure B.11.

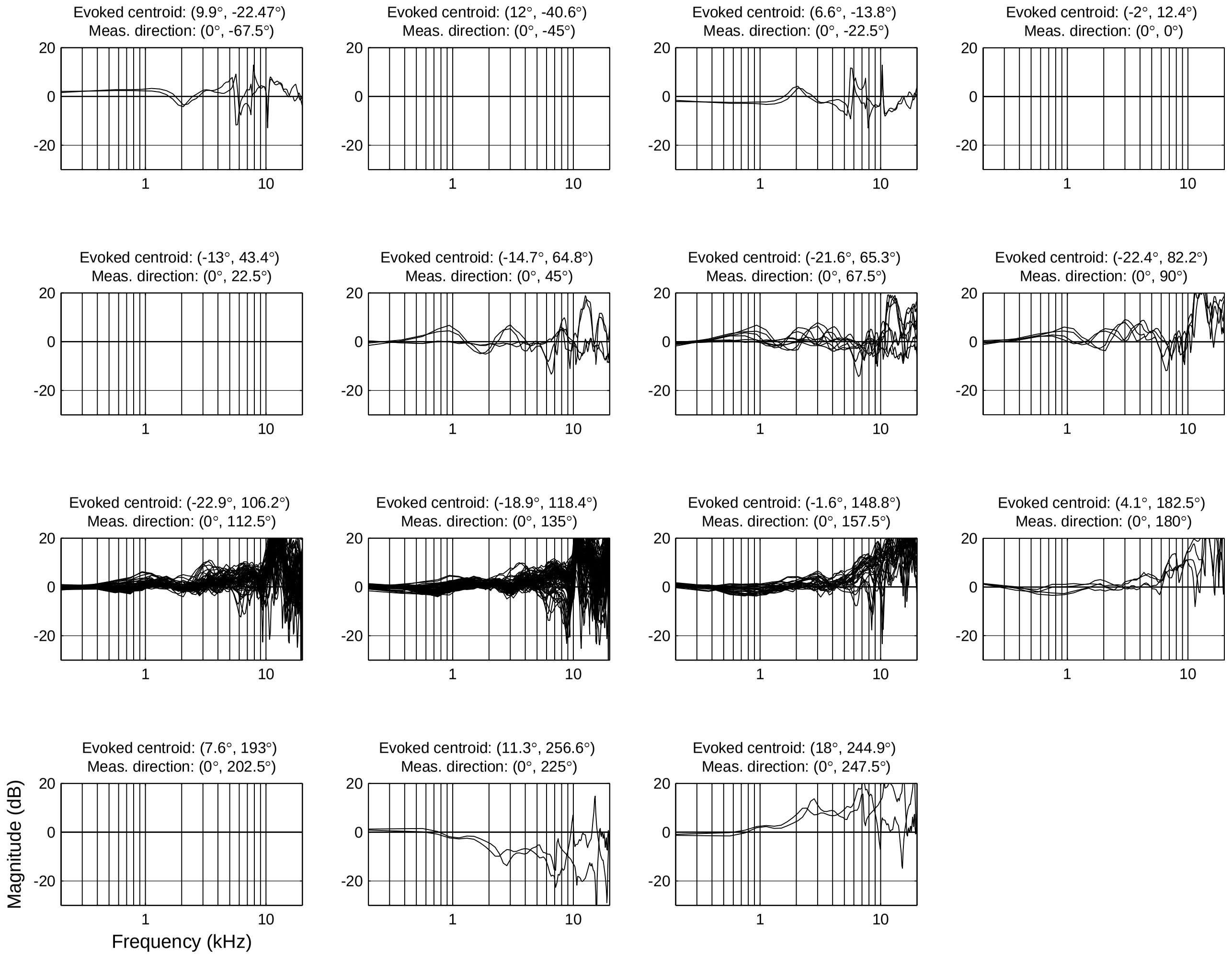

Figure C.6: Subject LA. Computed ratio $HRTF_{non-individual}/HRTF_{individual}$, based on those matching HRTFs already shown in Figure B.12.