IEM (Institut of Electronic Music and Acoustics)
Inffeldgasse 10/3
8010 Graz
Austria

Project for Audio Engineers, Course-No. 17.0036, SS 2005

Psychoacoustical Investigations in Binaural Sound Localization

August 2005

Carried out at:
IRCAM (Institut de Recherche et Coordination Acoustique/Musique)
1, Place Igor Stravinsky
75004 Paris

Presented by Jens Ahrens
Supervisor: Univ.-Prof. Dr. Gerhard Eckel

Abstract

Three major binaural localization tests have been conducted with subjects using individual head-related transfer functions (HRTFs) as well as subjects using non-individual HRTFs. In the case of non-individual HRTFs the interaural time difference (ITD) was approximated using different models.
The tests I and II yielded a diminution of the lateralization of stimuli in the presence of reverberation. Test III investigated the localization of speech in the presence of two different distracting speech signals. Pure reverberation of voices other than the target serving as distracters did not have any influence on localization. Pure direct sound of those voices degraded the localization performance at the lowest target-to-distracter ratio of -12dB.
Except for a slightly lower localization accuracy of subjects using non-individual HRTFs no systematic difference has been found between the two groups of subjects. The discrepancies in the results of test I and test II are assumed to be caused by the different testing conditions (subjects were blindfolded in test II but not in test I).

Zusammenfassung

Es wurden drei binaurale Lokalisationstests durchgeführt, an denen sowohl Probanden mit individuellen Außenohrübertragungsfunktionen (head-related transfer functions, HRTFs) als auch Probanden mit nicht-individuellen HRTFs teilnahmen. Im Falle der nicht-individuellen HRTFs wurde die interaurale Zeitdifferenz mittels verschiedener Modelle approximiert.
Die Tests I und II ergaben eine Verringerung der Lateralisation der Stimuli in der Anwesenheit von Hall. Test III untersuchte die Lokalisation von Sprache in Anwesenheit zweier verschiedener störender Sprachsignale. Purer Hall von zusätzlichen Stimmen als Störsignal hatte keinerlei Einfluss auf die Lokalisation. Purer Direktschall dieser Stimmen führte aber zu einer Verringerung der Genauigkeit der Lokalisation für das niedrigste getestete Target-zu-Distracter-Verhältnis von -12dB.
Außer einer verminderten Genauigkeit der Lokalisation bei Probanden mit nicht-individuellen HRTFs wurde kein systematischer Unterschied zwischen den beiden Gruppen festgestellt. Es wird vermutet, dass die Diskrepanzen zwischen den Ergebnissen von Test I und Test II durch die Testumstände verursacht wurden (Die Probanden in Test II hatten verbundene Augen, jene in Test I aber nicht).

Contents

Chapter 1

Introduction

The work presented in this report was carried out in the Room Acoustics group of Olivier Warusfel at the research institute IRCAM (Institut de Recherche et Coordination Acoustique/Musique) Centre Pompidou in Paris, France, in the period of November 2004 to July 2005 and is part of the project OPERA (Optimisation PErceptive du Rendu Audio).

This report is organized as follows:

Chapter 2 - *Mechanisms of auditory localization* - gives an overview over the principle mechanisms involved in human sound localization. For reasons of completeness perception of the distance of a sound incidence as well as motional theories (those descriptions taking into account the effects of head movements on perception) are explained as well although they are not subject of the present investigation.

Chapters 3, 4, and 5 describe the three main perceptive tests which have been conducted.

In test I (chapter 3) subjects using individual HRTFs localized filtered bursts of pink noise in an anechoic enviroment as well as in the presence of reverberation. Results revealed that the lateralization decreases with decreasing direct-to-reverberant ratio (i.e., the more reverberation is present, the more cumulate the perceptions around the directions straight ahead and straight behind.

Test II (chapter 4) is basically a repetition of test I with people using non-individual HRTFs as subjects. It was given a separate chapter since there are obvious discrepancies in the localization performance of subjects in test I and subjects in test II. Furthermore, it was conducted after test I whose results were already known.
The tendencies (diminution of the lateralization) are in principle similar to those in test I whereas subjects in test II over-lateralized much stronger and showed less accuracy. No explicit explanation has been found. The circumstance that subjects were blind-folded in test II but not in test I is suggested to contribute to the observations. (Confer to chapter 6)

Test III (chapter 5) investigates the localization of speech in the presence of several distracting speech signals. Two different conditions were investigated: Pure direct sound of a number of distracting voices positioned around the head, and pure reverberation of these distracters.
The pure reverberation of the distracters did not have any influence on localization whereas localization degraded in the presence of the direct sound of the distracters for the lowest target-to-distracter ratio (-12dB).
There were no obvious systematic differences in the perfomance of subjects using individual HRTFs and those using non-individual ones. All were blind-folded and all over-lateralized in an order of magnitude similar to test II.

Chapter 6 provides suggestions on how the influence of the method of the answer report on the results could be investigated so that a more thorough interpretation of the obtained results of the three tests would be possible.

All implementations of the accomplished tests were carried out on the Max/MSP platform (IRCAM/Cycling '74) and base on patches prepared by Pierre Guiu in summer 2004. The virtual acoustic environment in which the tests took place was generated by means of IRCAM's *Spatialisateur* (a real-time application for sound spatialization).

The statistical analyses were performed using SuperANOVA (scientek).

Samples of the test stimuli as well as an electronic version of this report can be found on the accompanying CD-ROM.

Chapter 2

Mechanisms of auditory localization

2.1 Localization of the direction of a sound incidence

Unlike in visual perception, where light from different directions falls upon different parts of the retina, sounds from all directions stimulate a common membrane, the eardrum. While visual localization may work by rather straight forward use of the geometric arrangement of the respective scene, the mechanisms of auditive localization are much more complex and yet not entirely understood (Blauert 1999, Mills 1972).

The subject began to occupy investigators toward the turn of the past century, and the importance of having an ear on either side of the head was soon established (Mills 1972). One of the pioneers was John William Strutt (better known as Lord Rayleigh) who established the so-called *Duplex Theory* (Lord Rayleigh 1877, 1904, 1907). He modelled the influence of the head on the propagation of the sound wave, which leads to interaural differences which are described in sections 2.1.1 and 2.1.2.

These binaural time (respective phase) and level differences are powerful cues for the localization of a source, but they have important limitations. They only occur for source positions outside the median plane (the vertical plane cutting the head in straight forward/backward direction), and they can be ambiguous. This brought the attention of investigators toward the influence of the external ear (and the rest of the body as well) on the spectral shape of a sound reaching the eardrums (Hebrank and Wright 1974), leading to the *Head-Related Transfer Functions* (HRTFs). See section 2.1.3.

2.1.1 Interaural time difference (ITD)

The first most obvious interaural difference for sound sources outside the median plane is the *interaural time difference* (ITD). It results from a difference in distances a sound has to travel to either ear. Two subclasses are distinguished:

time shifts in the carrier and in the envelope (Blauert 1999).

Considering a sound source far away enough to produce a plane wave front and modelling the head as a rigid sphere, the difference Δd of the travel paths of the sound wave is given by

$$\Delta d = d_1 + d_2 = r \cdot (\theta + \sin\theta) \tag{2.1}$$

where θ denotes the azimuth of the source in radians, r the radius of the head. Figure 2.1 illustrates the deviation of (2.1). Note that equation (2.1) is restricted to positions on the horizontal plane. Confer to section 4.1.3 for an extension to arbitrary directions.

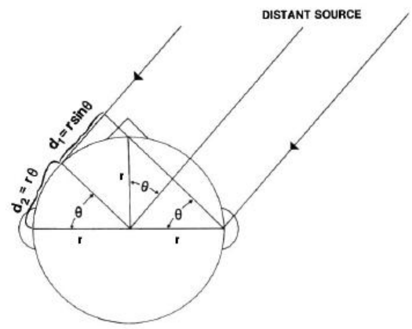

Fig. 2.1: Interaural time difference (ITD) for a distant source; (Begault 2002, p. 13)

As a consequence, the ITD can be calculated according to

$$\Delta t = \frac{\Delta d}{c} \tag{2.2}$$

where c denotes the speed of sound. Considering a head radius of $r = 8.75cm$ and $c = 343\frac{m}{s}$, this simple formula is surprisingly accurate[1] as can be seen in figure 2.2.

This time difference is an important factor for localizing short, impulse sounds and is contributing to the precedence effect ("Haas effect"/"law of the first wave front"): This effect states that it is the first acoustic information that arrives at a listener that determines the location of a sound (Blauert 1999, Clifton 1987). The exact mechanisms of this effect are not entirely understood yet (see e.g. (Damaschke 2004)).

For sound events with slow attacks (those without strong transient characteristics) the hearing mechanism shows the ability to compare the phase angles

[1]In fact, the measurement of the ITD is not staight forward. The resulting accuracy of the estimation based on (2.1) compared to a measurement varies substantially with the measurement method (see Minnar et al. (2000) or Katz et al. (2005) for a review). Woodworth (1938) does not mention which method is used in figure 2.2.

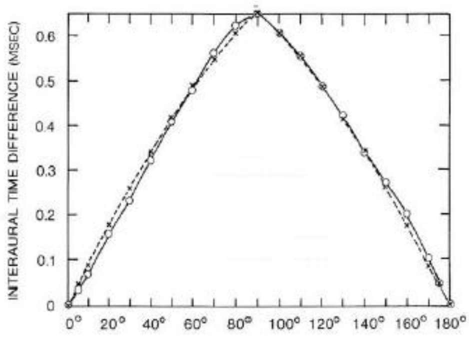

Fig. 2.2: ITD as a function of the azimuth of the sound source; solid lines: measured values for five subjects (Woodworth 1938); dashed lines: values computed from sphere (Feddersen et al. 1957); (Mills 1972, p. 303)

between the ears.
An ITD Δt corresponds to an *interaural phase difference* (IPD)

$$\Delta\phi = \Delta t \cdot 2\pi f \tag{2.3}$$

for a frequency f.

If sounds were localized on the basis of ITD/IPD alone, it would be very difficult to localize sounds with rather high frequency content: For frequencies above about 1.5 kHz, the wavelengths can be shorter than the path differences, leading to ambiguous IPD. And furthermore, only low frequencies show the ability to bend around the human head fairly easily. Higher frequencies are shadowed significantly by the head (1.5 kHz can be considered as boundary as well). In these higher frequency ranges human sound source localization makes use of this circumstance in a way described in the following section.

2.1.2 Interaural level difference (ILD)

In studying the influence of the head on the sound pressure at the ears, the approximation of the head as a rigid sphere has led to useful results as well (Blauert 1999). Since a detailed deviation would go far beyond the scope of this work, only a brief description of the results follows.

Some representative results are depicted in figures 2.3 and 2.4. The head (sphere) radius is 8.75cm; the ears are represented as two points on the surface at azimuths of 100° and -100° (0° is straight ahead) in the horizontal plane. The elevation φ of sound incidences is 0° degrees throughout.
Figure 2.3 shows the difference of the sound pressure level at the left ear point (100°) compared to the sound pressure with the sphere removed (undisturbed plane wave), at the point where the middle of the sphere was, for different angles of incidence.
Note that sound pressure can be higher in the contralateral ear (opposite the

direction of sound incidence) than in the undisturbed sound field condition. In this case the sphere has an amplifying rather an attenuating effect, which is a well-known phenomenon in diffraction (Blauert 1999).

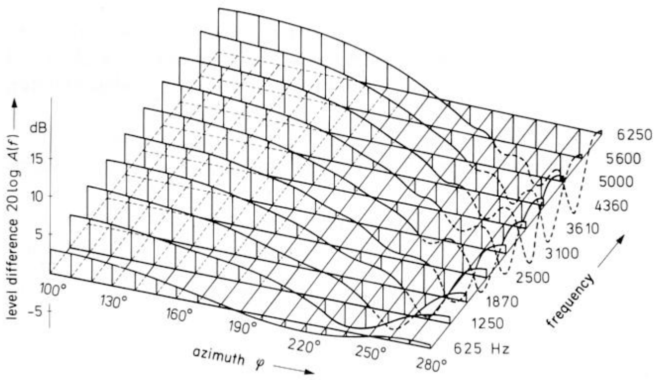

Fig. 2.3: Difference between sound pressure level at the left ear point ($\theta = 100°$) and sound pressure level at the middle of the sphere with the sphere removed as a function of sound incidence (Blauert 1999, p. 72)

The differences between the sound signals at the two ears are of special interest in the study of localization mechanisms.

The interaural difference in sound pressure level (*interaural level difference* ILD) is given in figure 2.4. The left ear point ($\theta = 100°$) is ipsilateral (shows towards the direction of sound incidence). At sound incidence azimuths around 100° the ILD shows a distict dip. It is to be expected that this dip is less marked for the natural head since the influence of the neck and torso are disregarded in the spherical head model.

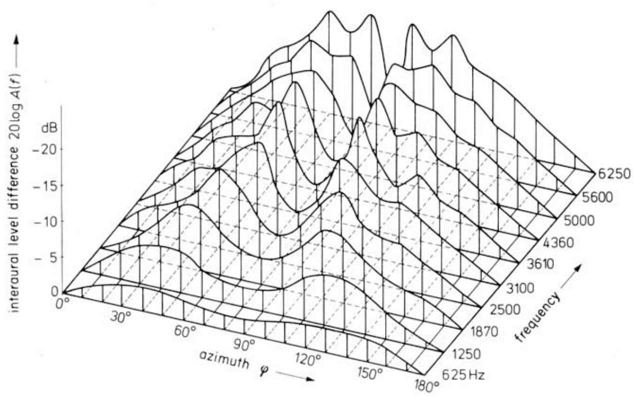

Fig. 2.4: ILD as a function of sound incidence azimuth (Blauert 1999, p. 73)

As can be seen from figure 2.4, ILDs become smaller with decreasing fre-

quency. Especially for very low frequencies they can be below the detectable threshold which can be considered around 1dB (Blauert 1999). In return the ITDs/IPDs become more effective in the low frequency range, thus, gaining more perceptual weigth. This interaction of ILDs and ITDs/IPDs is called the *Duplex Theory*, which was first described by Lord Rayleigh (1907).
The exact mechanisms for the weighting of the interaural differences are still not entirely understood (Blauert 1999). 1.5 kHz (Mills 1972), respectively 1.6 kHz (Blauert 1999), is considered as a boundrary of the two major ranges of function.

Of course, these mechanism alone cannot solve all localization problems. For example, sound sources at any position in the median plane, lead to not a single binaural difference. As a result, the listener with the hypothetical spherical head cannot distinguish between sources in back, in front, or overhead in this case.
Outside the median plane, relying only on ITD and ILD cues for localizing sounds results in a *cone of confusion* (Blauert 1999, Mills 1972), a cone whose surface contains all possible sound source positions which lead to identical ITDs respective ILDs, which is illustrated in figure 2.5. For reasons of simplicity, the model of von Hornbostel and Wertheimer (1920) is used which assumes a spherical head and where the head does not affect the propagation of the sound waves. Figure 2.5 (a) shows the resulting hyperbola regarding only the horizontal plane. For all three dimensions of space the calculations result in a conical shell, the *cone of confusion*.

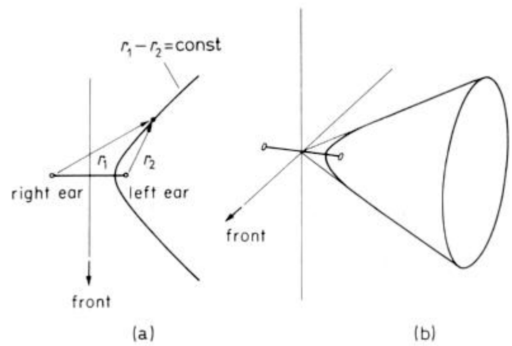

Fig. 2.5: The locus of all points at the same distance from both ears according to the model of von Hornbostel and Wertheimer (1920); (a) hyperbola in the horizontal plane; (b) conical shell in three dimensions (the cone of confusion); (Blauert 1999, p. 179)

As will be seen later in chapters 3-5, confusions along the cone of confusion can indeed be observed, especially when headphones are used (Begault 2000). When forward and backward (or backward and forward) directions are confused, one speaks of front-back confusions or reversals.

2.1.3 Head-Related Transfer Functions (HRTFs)

The shape of the outer ear as well as the reflective properties of the shoulders and upper torso play an important role in spectrally shaping a sound wave that reaches the eardrums (Blauert 1999). Unlike ILDs/ITDs and the resulting cone of confusion, the filtering effects of the outer ear are unique for each position in the three-dimensional space relative to the listener. While this (linear) filtering is present in the perception of sound from any direction, it is a critical aspect, especially of vertical localization.

The alterations a variable undergoes in a linear system (linear distortions) can be described by the system's transfer function $H(f)$, defined as the complex ratio of the Fourier transform $Y(f)$ of the output variable $y(t)$ to the Fourier transform $X(f)$ of the input variable $x(t)$.

$$H(f) = \frac{Y(f)}{X(f)} \tag{2.4}$$

f denotes the frequency.

Consequently, as a first step toward understanding spectral cues in directional hearing, researchers have tried to physically model (e.g. Shaw 1974), empirically measure (e.g. Wightman and Kistler 1989), or more recently computationally simulate (e.g. Kahana et al. 1999) the direction-dependent transfer function of the ear.

These transfer functions of the ear are called *Head-Related Transfer Functions* (HRTFs), and summarize the direction-dependent acoustic filtering a free field sound undergoes due to the head, torso, and outer ear (pinna).
Since each individual has unique physical shapes and characteristics, each individual also has a unique set of HRTFs. These HRTFs can be measured, e.g. by placing small microphones in the ear canal and playing an appropriate sound event at different spatial locations around the fixed head.
Note that an HRTF includes both ITD and ILD information: time delays are encoded into the filters phase spectrum, and ILD information is related to its amplitude spectrum (Cheng and Wakefield 1999).

HRTF data are commonly represented in three different ways:

1. In the frequency domain
2. in the time domain (Head-Related Impulse Response, HRIR)
3. or in the spatial domain, mostly as surfaces (see (Cheng and Wakefield 1999) for further description)

Figures 2.6 and 2.7 show an example of a set of HRTFs measured from a human subject. Figure 2.6 compares the magnitude response of both ears for different azimuths of sound incidence in the horizontal plane (elevation $\varphi = 0°$). Figure 2.7 compares the magnitude response of both ears for different elevations of sound incidence in the median plane ($\theta = 0°$).

The properties of the HRTFs (or similar representations) that contribute to localization are not entirely decoded. Some attributes are to be mentioned representatively:

1. Diffraction of low frequencies can be clearly seen in figure 2.6: The low frequency response on the contralateral side is less disturbed than the high frequency response.
2. In figure 2.7 there is a notch at 7 kHz that migrates upward in frequency as elevation increases (especially in the right ear). There is also a shallow peak at 12 kHz which flattens out at higher elevations.

One major application involving HRTFs is the creation of virtual three-dimensional sound scenes via headphones, as it is done in the listening tests described in chapters 3-5.
The signal of the virtual sound source is filtered with the particular HRTFs corresponding to the desired positions of the sound source for each ear individually to create a respective perception.

2.2 Localization of the distance of a sound incidence

Although a large amount of investigations is available, the knowledge of distance hearing is deficient compared to the knowledge of directional hearing. This deficiency is due to the extraordinary complexity of the subject (Blauert 1999). The particular interest of investigators is concerning the relationship between the distance and the corresponding attributes of a sound event. Reflecting briefly on which attributes of the ear input signals depend on the distance from the sound source, the following classification scheme can be derived (Blauert 1999):

1. The simplest description can be used at intermediate distances from the sound source, approximately 3-15m for point sources. Here the sound pressure level of the ear input signals can be considered as dependent only on the distance from the source. Since the sound pressure p in a free sound field is inversely proportional to the distance, the sound pressure falls 6dB for every doubling of the distance r (the $1/r$ law).
2. At greater distances from the sound source - more than approximately 15m - the frequency dependent attenuation caused by the air path between the sound source and the subject can no longer be neglected. It occurs additionally to the $1/r$ dependency of the overall sound pressure and is stronger for higher frequencies than for lower. Thus, not only does the sound pressure level depend on the distance from the sound source, but the shape of the spectrum also depends on it - or, more precisely, the relative phase curves as a function of frequency depend on it.

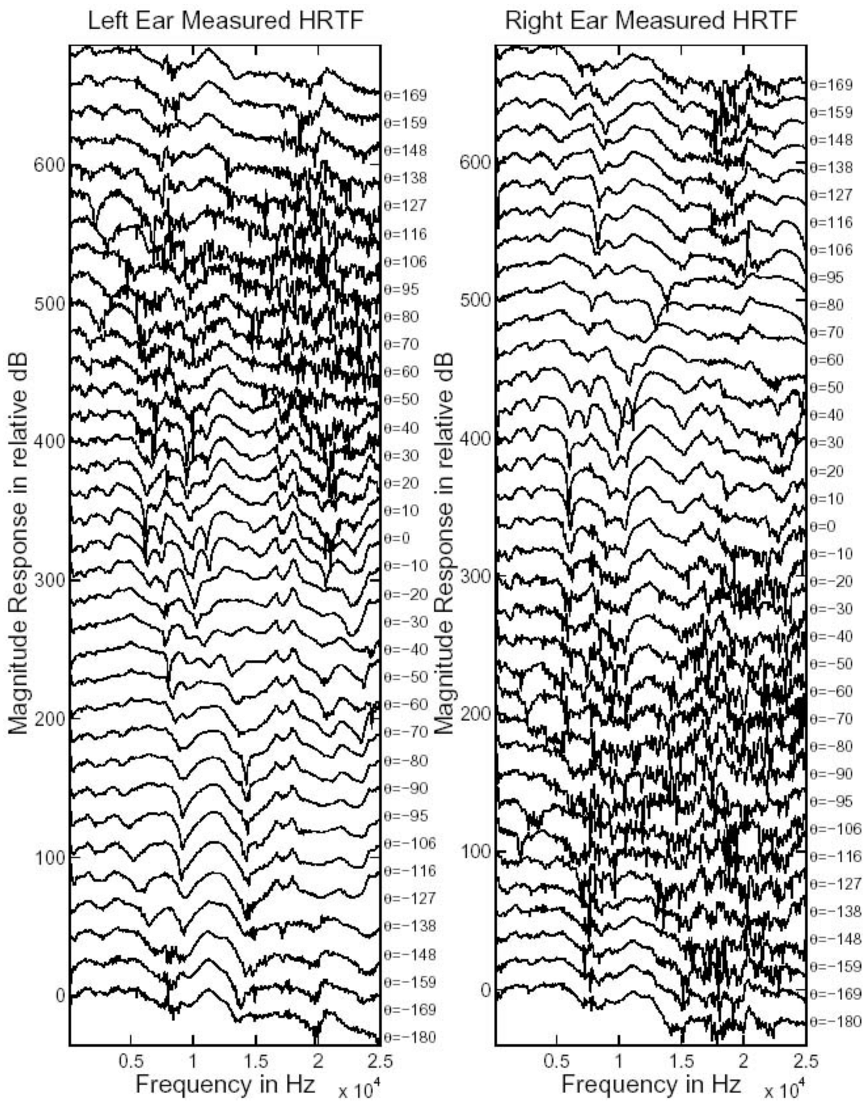

Fig. 2.6: Comparison of the magnitude response of measured HRTFs in the frequency domain as a function of azimuth in the horizontal plane ($\varphi = 0°$). Diffraction effects for low frequencies can be seen on the contralateral side at azimuths $+90°$ and $90°$ for the left and right ears, respectively; (Cheng and Wakefield 1999, p. 19)

3. Close to the sound sources - at distances less than approximately 3m for point radiators - the distance dependent linear distortions of the signals due to the head and the external ears become prominent (HRTFs). The shape of the spectrum of the ear input signal changes distinctively with distance, though in a different way than for large distances. Additionally the 1/r dependency has to be taken into account.

Besides the three already mentioned conditions some further observations can be made for signals presented over headphones. In this case the normal

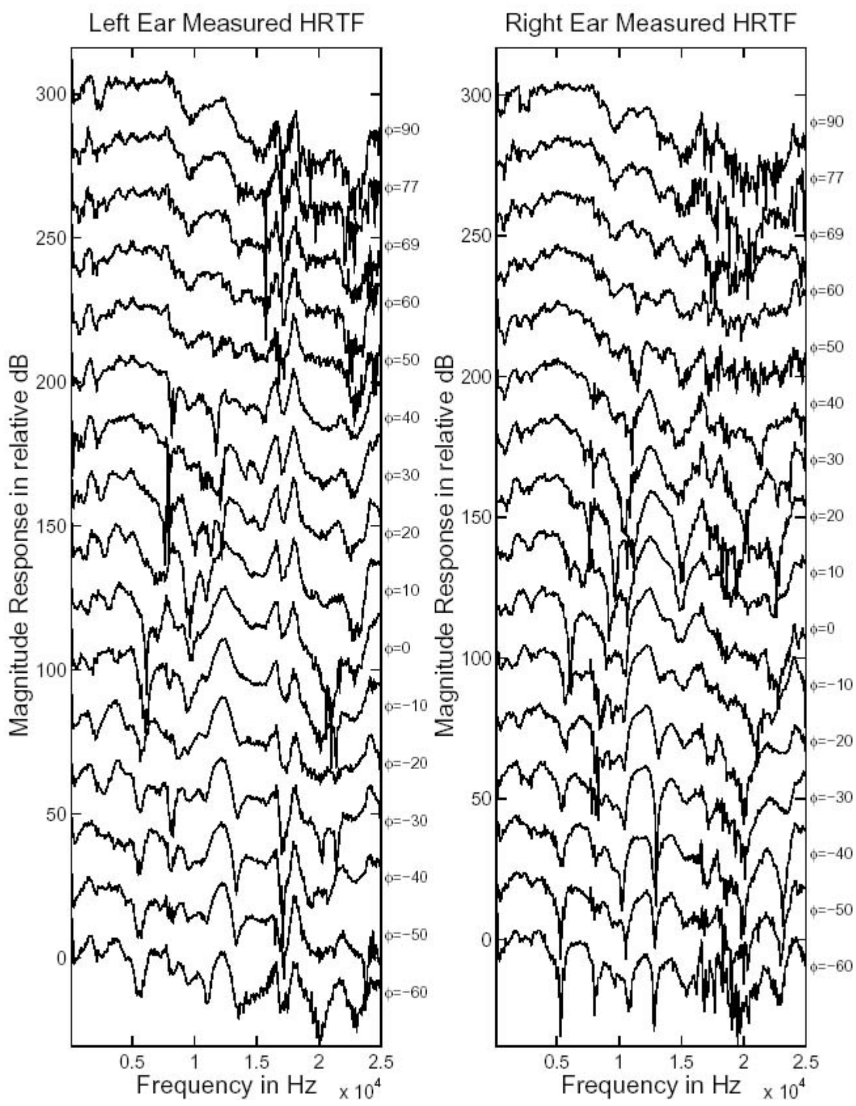

Fig. 2.7: Frequency domain comparison of measured HRTFs as a function of elevation in the median plane ($\theta = 0°$). There is a notch at 7 kHz that migrates upward in frequency as elevation increases. There is also a shallow peak at 12 kHz which flattens out at higher elevations; (Cheng and Wakefield 1999, p. 20)

filtering effect of the pinnae is eliminated, since the headphone loudspeakers lie on the axis of the ears very close to the entrances of the auditory canals. These conditions frequently lead to a perception of the distance of the auditory event smaller than the radius of the head. In other words, the auditory event occurs inside the head. This observation is often referred to as *inside-the-head locatedness* or *intracranial locatedness*.

Unfortunately, the above mentioned properties do not fully cover the mechanisms of distance hearing. Several more parameters are taken into account which lead to the already mentioned complexity of the subject.
A representative example is the following: The loudness and the tone colour of a sound source are not independent quantities. As the loudness increases the tone colour becomes darker. The darkening may be examined in connection with the well-known curves of equal loudness level (see figure 2.8). As the level is increased, the low-frequency components of a broadband signal gain more and more perceptual weight relative to the high-frequency components.
A decrease of the distance of a sound source now leads to an increase of its sound pressure level. The increase of the sound pressure level increases the loudness and thus in turn, changes the tone colour. This change of the tone colour is superimposed on a possible change of the tone colour due to the variation of the distance (confer to point 2 above).
A further description of distance hearing can be found in Blauert (1999).

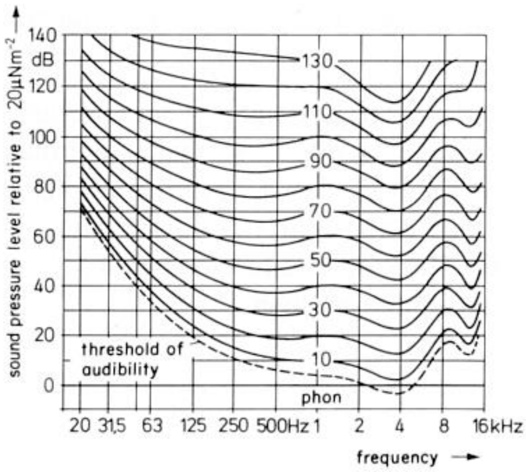

Fig. 2.8: Curves of equal loudness for sinusoids; sound incidence from the front; (Blauert 1999, p. 120)

2.3 Motional theories

Especially in the case of unfamiliarity with a certain sound it is obviously difficult to interpret its spectral shape in terms of localization. Among others this is one reason why we move our head - consciously or unconsciously - when being exposed to a sound.
If the head moves relative to the sound source while a sound signal is being presented, the monaural and interaural attributes of the signals at the eardrums will change in a significant way according to the position of the sound source, which makes our localization much more accurate. Theories of spatial hearing that describe relationships between the position of a sound source and the

changes to the ear input signals during head movements are called *motional* or *motoric theories* (Blauert 1999).

If the capability of moving the head relative to the sound source is prevented (e.g. as it is in binaural hearing, if no tracking is used), the reliability of the localization is degraded, and confusions (especially front/back confusions are most obvious) occur more likely (Begault et al. 2001b).

Chapter 3

Test I

The localization performance of seven adults (three female/four male) in a virtual reverberant environment in the horizontal plane was tested. Filtered bursts of pink noise embedded in a diffuse reverberation were used as stimuli. The playback was accomplished via headphones using the individual Head-Related Transfer Functions (HRTFs). Two more subjects participated with non-individual HRTFs.
Results revealed that an influence on the performance is significant at direct-to-reverberant ratios below -6dB.

3.1 Methods

3.1.1 Subjects

Seven adult (three female/four male) unpaid listeners were tested with their individual HRTFs. All of them had considerable experience in binaural hearing. In addition to those, the author and one more male subject accomplished the test with non-individualised HRTFs since their own had not been measured at that time.
In this instance the HRTFs were chosen from IRCAM's Listen HRTF Database by means of some initial trial to make sure that some externalisation is perceived and the front-back confusion rate is set to the minimum. An alignment of the ITD was not employed.

3.1.2 Stimuli

The stimuli were real-time generated bursts of pink noise, either low pass (f_{-3db} = 500Hz), band pass ($f_{1,-3db}$ = 500Hz,$f_{2,-3db}$ = 4000Hz), or high pass (f_{-3db} = 4000Hz) filtered, with a total duration of 220ms (10ms linear on-set, 20ms linear off-set ramps) played at a sound pressure level around 65dB(A). The initial off-set ramp of 10ms tended to cause an unappreciated perception for some of the stimuli, thus it was prolonged.
The stimuli were presented in the horizontal plane at azimuths of 30°-steps around the head at 6 different direct-to-reverberant ratios (dir/rev) of 40dB, 0dB, -6dB, -9dB, -12dB, and -15dB. 40dB was chosen to approximate an

anechoic condition. Higher dir/rev ratios lead to unnaturally sounding environments because most of the stimuli were internalized.
First reflections and the following cluster (a quasi-stationary part of densely arriving early reflections) were totally removed from the reverberation so that maximum diffusity is provided. Reverberation time was set to 1s. The impulse response of one of the two channels of the reverberation system for the dir/rev ratio of -15dB is shown in figure 3.1.
The stimulus was delayed such that its on-set occurred at the same time as that of the reverberation.

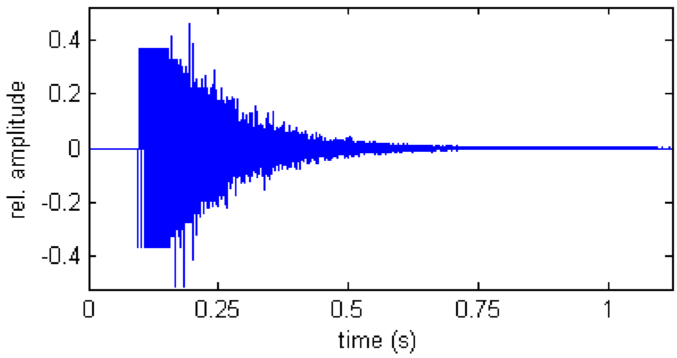

Fig. 3.1: Impulse response of one channel of the reverberation system; the impulse occurs at 0s with amplitude 1; dir/rev = -15dB;

3.1.3 Accomplished modifications on the prepared patches

The found 2nd-order filters did not yield the desired stimuli so they had to be replaced. 6th-order Chebychev Type II filters were chosen and implemented as a cascade of biquads. The cluster was removed from the reverberation. A calibration procedure for the tracking system was employed to compensate for individual ways of wearing the headphone to guarantee a reliable estimate of the head centre position and orientation.
The test as found used the head position at the moment of the validating of the subject's answer to calculate the relative position of the perception. Since prior to the test one could not have been sure which of the head positions (at the time of the stimulus play-back or of the answer) provides the more reliable estimate for the direction, adequate modifications were introduced in order to record both head positions to be able to decide posterior. Despite thorough observation of the subjects this question has not been clarified completely. The decision came towards the head position at the time of the validating of the answer.
Some improvements affecting the handling of the test and the preparation of the files containing the subjects answers for the evaluation of the results were employed. Furthermore several less critical bugs were corrected. However, not all of them could be fixed since this might have involved a major redesign of the patches.

3.2 Procedure

The test was accomplished in a sound treated room with a background noise level of around 24dB(A). The subjects were seated on a chair in front of a computer screen which displayed the number of stimuli already heard. The head was not fixed but subjects were asked to keep their heads as stable as possible.
The experiment was divided into two runs with a brief pause in-between. In an initial reference phase 72 quasi-non-reverberant stimuli (dir/rev = 40dB) were presented. The conclusive session consisted of 360 reverberant stimuli (dir/rev ratios of 0dB, -6dB, -9dB, -12dB, -15dB). In both sessions the stimuli were played in arbitrary order and all possible stimuli configurations occurred exactly twice.
A magnetic tracking system with one sensor fixed on top of the headphones and another hand-held sensor provided the framework to record the subjects' answers. Those indicated their perception by pointing the hand-held sensor towards the perceived direction of a certain stimulus and validating the answer using a foot controller. No feedback was provided. Subjects did not have any information about possible stimulus locations.
The trigger for the subsequent stimulus to be played was initialised by the subject him-/herself by bringing the hand sensor into a predefined area around his/her stomach. Thus, the interval between to successive stimuli was controlled by the subject his-/herself to avoid unnecessary rush or waiting.

3.3 Results

All subjects accomplished the 72 non-reverberant reference stimuli in a good seven minutes. Six of them spent about 30 minutes on the 360 stimuli of the reverberant session, one spent around 37 minutes. That gives an average interval of a good 5 seconds between two successive stimuli.
The data of the subjects participating with non-individual HRTFs are not included. See the summary in section 3.5 for further explanation.

3.3.1 Non-reverberant condition (resp. dir/rev = 40dB)

Figure 3.2 shows the perceived versus presented stimulus azimuths for the dir/rev = 40dB situation. The values are displayed as the means over all the answers with the standard deviations shown by the error bars. In figure 3.2 occurred front-back confusions have been corrected to allow a more interpretable description. The correction rates are shown in figure 3.7 (a).
Positive angles denote the left hand side, negative angles denote the right hand side. $0°$ is straight in front of the listener.
For all frequency bands (low pass (LP), band pass (BP) and high pass (HP)) the plots show a slight but clear excessive lateralisation of the answers rather than a perfect straight line. Figure 3.3 shows this more in detail for the LP stimulus. It uses front-back corrected values as well. All other figures of this section show the raw data.

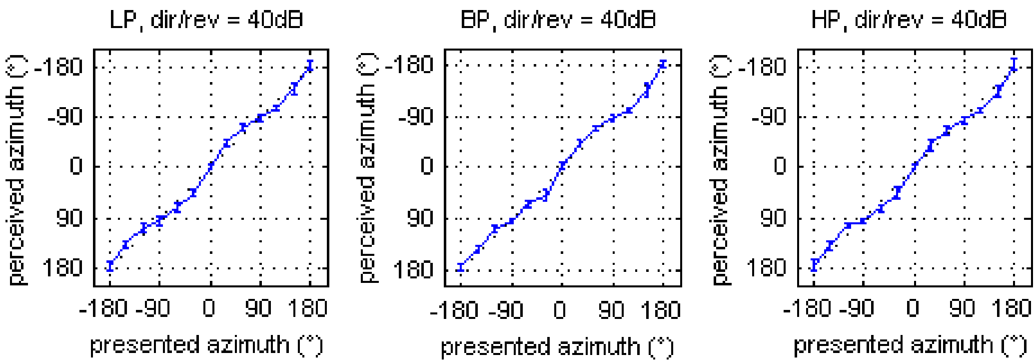

Fig. 3.2: Perceived stimulus azimuths versus presented, mean values with standard deviation, front-back confusions corrected; low pass (LP), band pass (BP) and high pass (HP) stimuli in separate plots;

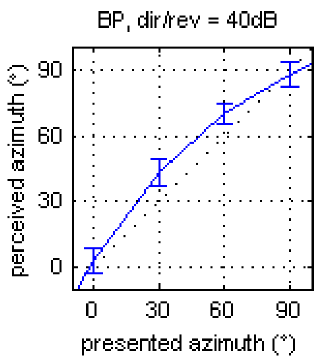

Fig. 3.3: Zoom into left-most subplot of figure 3.2;

Figure 3.4 shows the perceived versus presented angles between the presented/ perceived stimulus directions and the median plane (the vertical plane at 0°/180°) according to

$$\lambda = \arcsin(\sin\theta \cdot \cos\varphi) \tag{3.1}$$

where θ denotes the azimuth and φ denotes the elevation of the stimulus (respective the answer) direction.

Positive angles denote the left hemisphere, negative angles the right one. Thus, a stimulus presented from 90° azimuth (exactly left of the listener) subtends an angle of 90° towards the median plane. This figure displays mean values and standard deviations. Solid lines denote stimuli in the frontal hemisphere; dotted lines denote stimuli from the rear hemisphere.

The presented description was chosen because it is a measure for the lateralization of a certain direction. In binaural hearing elevation perception is not balanced for all directions. Positions near the median plane (especially in front) tend to be perceived higher elevated than lateral positions (Begault 2000) leading to a diminution of the lateralization of a stimulus (confer as well to section

3.3.2). Furthermore, in the given illustration the influence of possible front-back confusions is completely eliminated.
At a first glimpse one might interpret a slight saturation effect when the presented stimulus azimuth goes towards 90° (respectively -90°). The answers seem not to follow the stimuli to the most lateral directions.
One has to keep in mind that 90° and -90° are the limits of the observed quantity, allowing only one-sided distribution functions at these points. No angles larger than 90° (respectively smaller than -90°) can occur. Thus, every inaccuracy of the answers decreases the angle (respectively increases it) so that the most extreme values are less likely to appear. And furthermore, the slopes in figures 3.2 and 3.3 get flatter around 90° and -90°, which is contributing to the observations.
Besides the already observed excessive lateralisation the results show a reliable ability of the listeners referring to left/right localization.

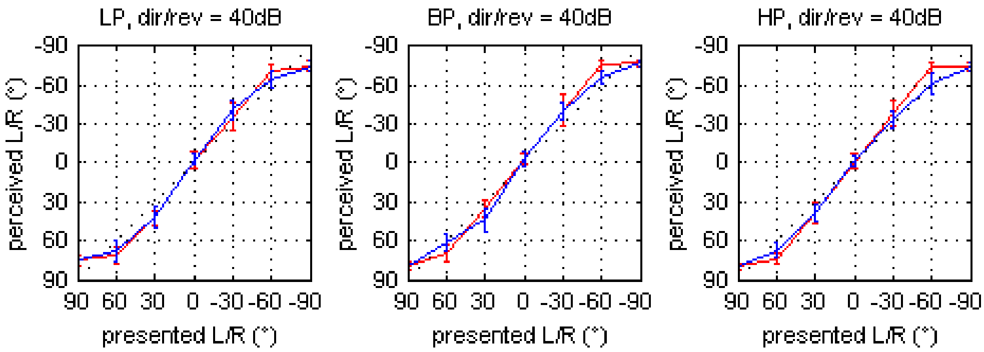

Fig. 3.4: Angle subtended by the presented/perceived direction and the median plane; mean values with standard deviation; LP, BP ,and HP stimuli in separate plots; blue lines: frontal hemisphere; red lines: rear hemisphere;

The subsequent quantity evaluated is the perceived versus presented angles between the presented/perceived stimulus directions and the frontal plane (the vertical plane at 90°/-90°) according to

$$\psi = \arcsin(\cos\theta \cdot \cos\varphi) \tag{3.2}$$

Positive angles denote the frontal hemisphere, negative angles the rear one. Thus, a stimulus presented from in front of the listener (0° azimuth) subtends an angle of 90° towards the frontal plane.
This description reveals detailed information about the tendency if a subject experiences perceptions more likely in the front or in the back hemisphere. Since the individual results are not consistent mean values will not be displayed. 4 of the 7 subjects (S1, S3, S5 ,and S6) did hardly perceive any stimulus in the frontal hemisphere although half of them were presented there. The other 3 subjects (S2, S4 ,and S7) showed a more balanced performance. Figure 3.5 gives an impression about the circumstances. It shows some of the individual

data with the above mentioned characteristics plotted as means of the two identical stimuli with the standard deviation denoted by the error bars. Solid lines denote stimuli from the left; dashed lines denote stimuli from the right.
A more detailed discussion about this issue can be found below in the section about the results in the reverberant condition.

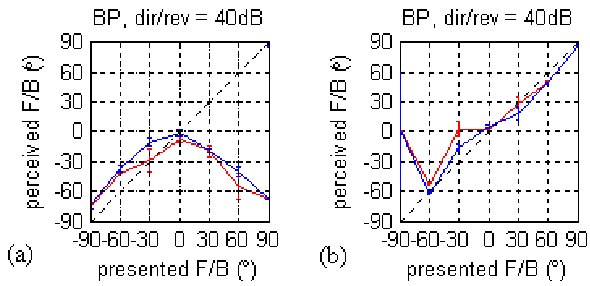

Fig. 3.5: Angle subtended by the presented/perceived direction and the frontal plane, mean values of the two identical stimuli with standard deviation; blue lines: right hemisphere; red lines: left hemisphere; (a) excerpt of the results of S6 with certainly no perception from the front (BP stimuli) (b) excerpt of the results of S2 with a rather balanced distribution (BP stimuli);

3.3.2 Reverberant condition

Figure 3.6 corresponds to figure 3.2 in the non-reverberant condition. It shows the perceived versus presented stimulus azimuths for 5 different dir/rev ratios reaching from 0dB in the left-most plots to -15dB in the right-most plots. The top line shows LP stimuli, the middle line BP stimuli and the bottom line HP stimuli. Front-back corrected values are used. Correction rates are shown in figure 3.7 (b). S2 and S7 show distinctly higher rates in the presence of reverberation, S4 and S5 show only a minor increase, and S1, S3 and S6 show a slightly better performance. In general, reversals toward the rear hemisphere are more frequent than in the other direction.
Figure 3.6 is meant to give a first idea about the results since it provides the most familiar description. The observed changes towards the non-reverberant condition will be discussed by the means of the following figures. The reader may independently return to this figure to equally find the observations there.

Figure 3.8 shows the perceived versus presented angles between the presented/perceived stimulus directions and the median plane according to (3.1). The corresponding figure showing the non-reverberant data is figure 3.4.
In the presence of reverberation lateralisation is reduced. The lower the dir/rev ratio the more cluster the answers around the median plane (respective around 0° and 180° azimuth). For dir/rev ratios below -9dB the observation becomes most obvious.
Thus, for low dir/rev ratios listeners experience perceptions more likely from straight ahead or straight behind rather than from the sides.

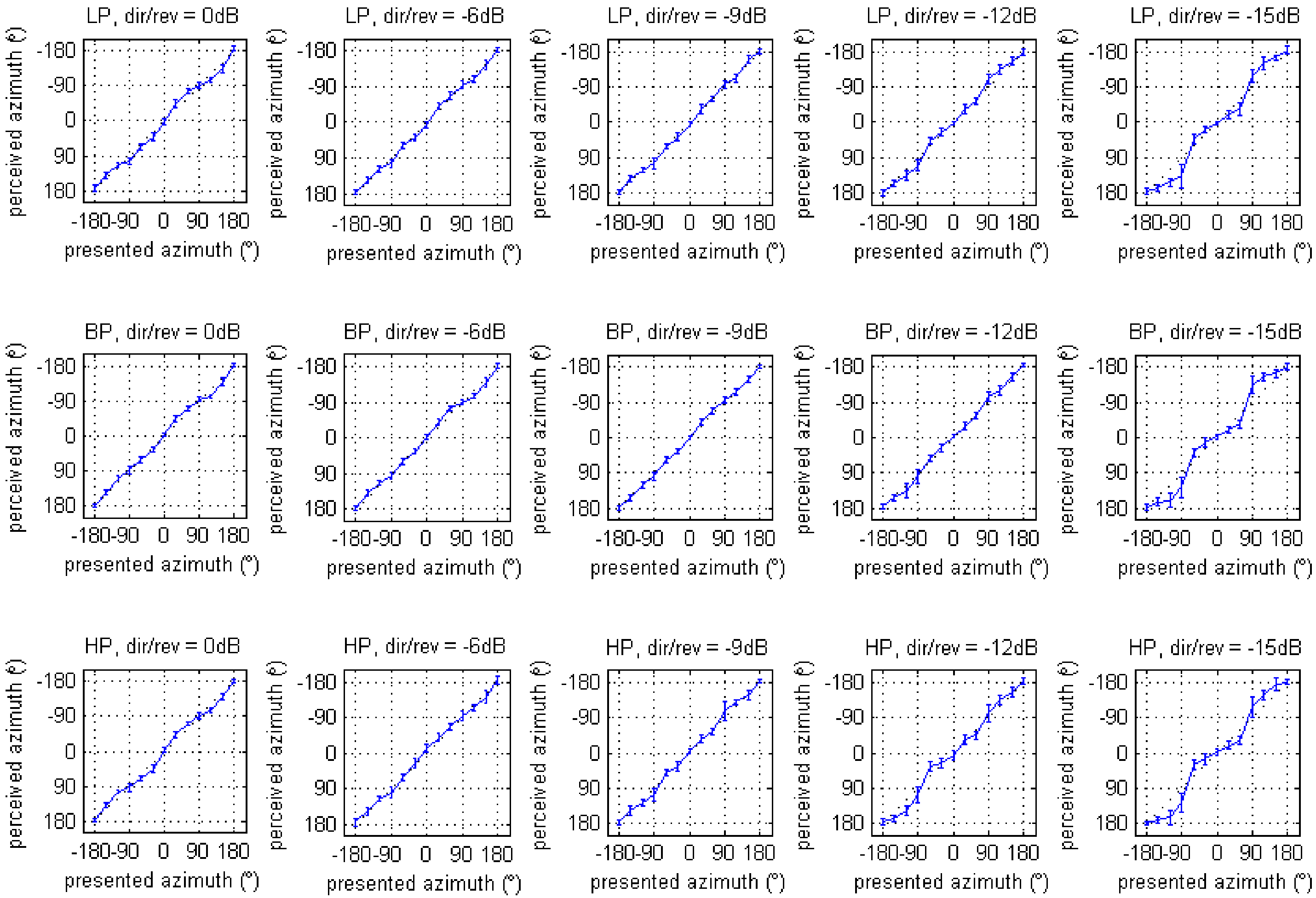

Fig. 3.6: Peceived stimulus azimuths vs. pesented; mean values with standard deviation; front/confusions corrected; different dir/rev ratios in separate columns; LP, BP and HP stimuli in separate lines;

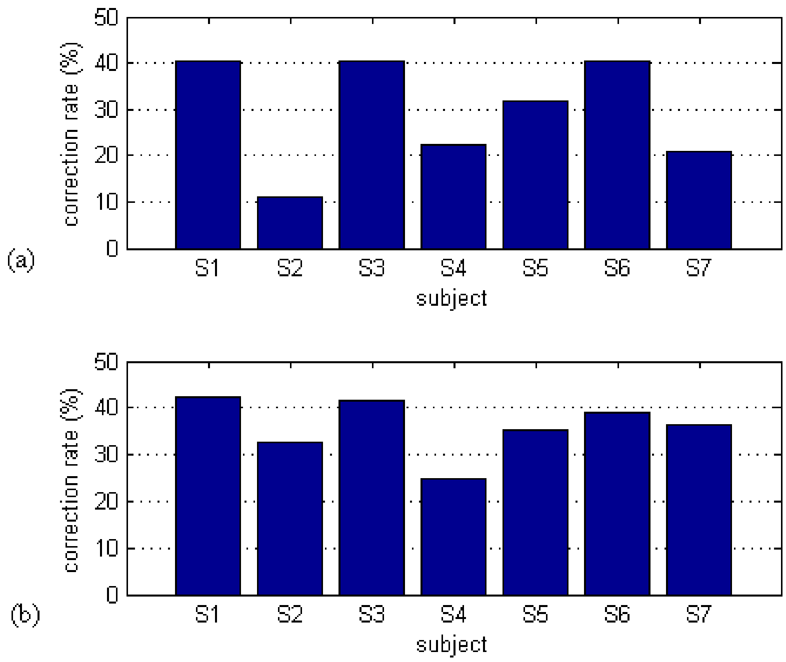

Fig. 3.7: Rate of corrections of front-back confusions for the individual subjects; (a) anechoic condition; (b) reverberant conditions;

Again, as for the non-reverberant condition, the evaluation of the angles between the presented/perceived stimulus directions and the frontal plane (3.2) does not show consistency between the different subjects. Those subjects who tended to experience perceptions more likely in the back in the reference condition did so as well in the reverberant. Those subjects having tended to balanced results did so again whereas a lot more perturbation occurred in the last mentioned case. Surprisingly, S2 showed hardly any perceptions from behind.
In consequence of the observations of figure 3.8 the tendencies of perceptions from a certain hemisphere become more extreme for lower dir/rev ratios, which can be seen in figure 3.9. It shows the answers of S3 for the HP stimulus. Values are the means of the two identical stimuli with the standard deviation denoted by the error bars.

Unfortunately the conditions in terms of background noise were not optimal in the location where the test was conducted. The magnetic tracking system produced some slight steady hum (the tracking system was running when the background noise measurement mentioned in section 3.2 was made). Due to an exhaustive equipment of the location with materials interfering with the tracking system its distance to the sensors had to be kept short to ensure high precision. It was placed right behind the subjects.

Some of those hardly felt affected by the noise, some (S2, S5 ,and S6) found it rather disturbing. Since no particular dependence has been detected these observations can not be explained here.

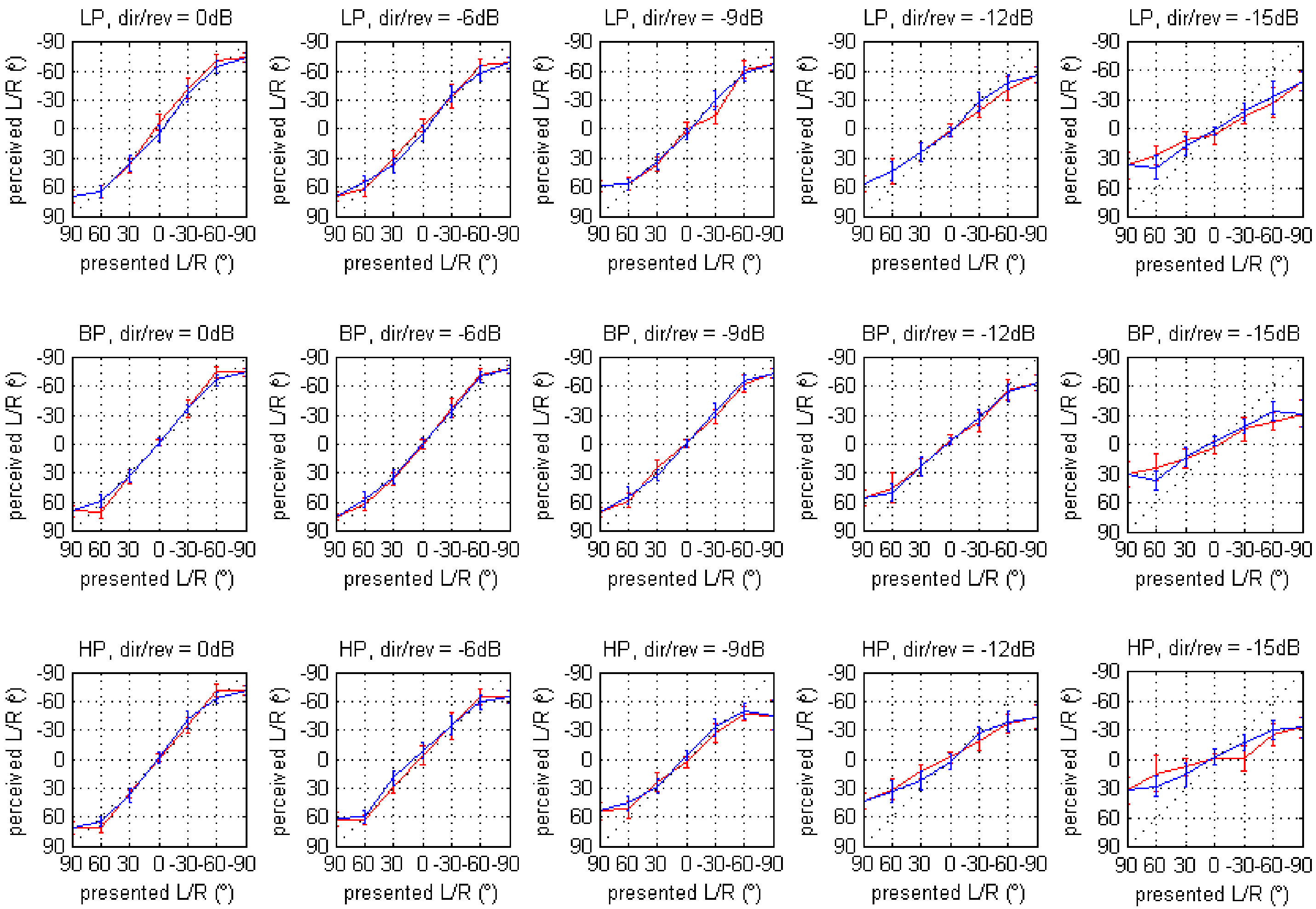

Fig. 3.8: Angle subtended by the presented/perceived direction and the median plane; mean values with standard deviation; different dir/rev ratios in separate columns; LP, BP and HP stimuli in separate lines; blue lines: frontal hemisphere, red lines: rear hemisphere;

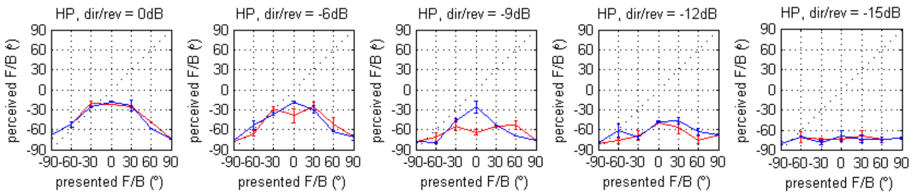

Fig. 3.9: Angle subtended by the presented/perceived direction and the frontal plane; data of S3 for the HP stimuli for the different dir/rev ratios; blue lines: left hemisphere; red lines: right hemisphere;

Another observation to mention is the increase of the elevation of the answers with higher frequency content of the stimuli. This is the only evaluated quantity in this test which shows a dependency on the spectral properties of the stimulus rather than on the reverberation. This can be obtained from figure 3.10. It displays the elevation of the answers versus the stimulus azimuths as means of all answers for the particular direction and for a particular stimulus type. The standard deviations are not shown since they are huge and disturb the illustration. No explicit dependency on the dir/rev ratio was observed, thus, data are divided only into non-reverberant (dir/rev = 40dB) and reverberant (dir/rev $\leq$0dB) stimuli.
The peaks around the median plane are less pronounced in the reverberant condition.

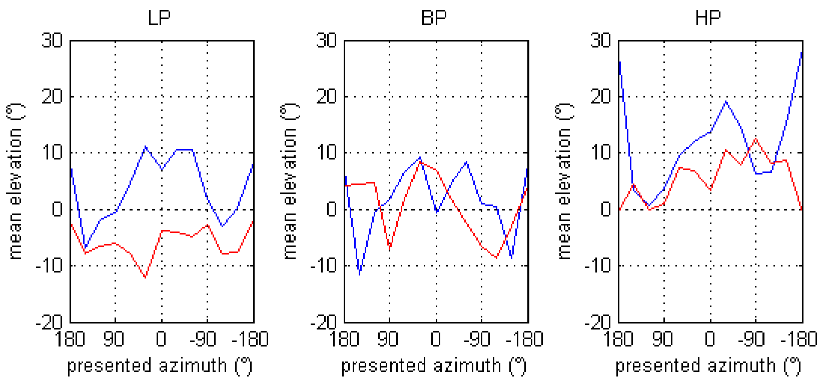

Fig. 3.10: Elevation of the answers for different stimulus azimuths and frequency bands; values are means over all respective data; standard deviations are not shown; blue lines: non-reverberant condition; red lines: reverberant conditions;

3.3.3 Statistical analysis

An ANalysis Of VAriance (ANOVA) was performed on the data to determine below which dir/rev ratio the influence of the reverberation on localization is significant. The spherical distances between the answer directions and the stimulus directions (i.e., the localization error) were taken as dependent variables. Note that the answers were front-back corrected before the calculation of the

spherical distances.
Since ANOVA assumes the dependent variables to be normally distributed the data have to be separated into the different directions of the presented stimuli[1]. For reasons of simplicity the data are divided into only four different groups according to the unsigned lateralization of the direction of the presented stimuli (i.e. 0°, 30°, 60°, and 90°).
For every group the localization error for every reverberant condition was tested separately against the anechoic condition. The resulting p-values are depicted in fig 3.11.

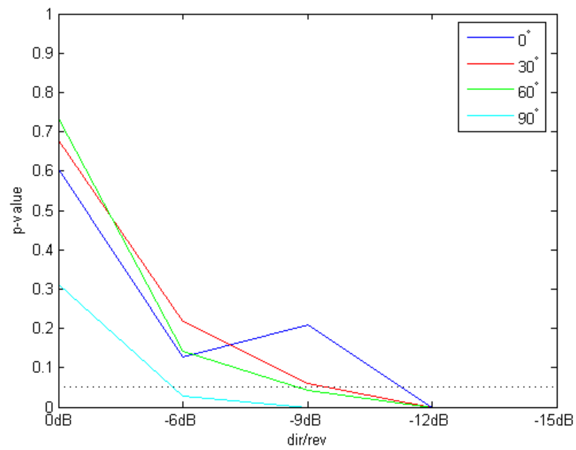

Fig. 3.11: p-values for all dir/rev ratios tested separately against the anechoic condition; each colour represents one group of stimuli according to the legend

Assuming a significance level of 0.05, the influence of the reverberation becomes significant at dir/rev ratios somewhere between -6dB and -12dB, depending on which stimulus direction is considered. In terms of virtual reality applications this means that localization is affected for sound sources at distances farther than about twice the critical distance. The critical distance is defined as the distance from the sound source where the energy levels of the direct and the reverberant sound are equal, i.e., where the dir/rev ratio is 0dB. The critical distance in mid-size rooms is around 2 m and can go up to and beyond 10 m for large concert halls.

The slope of the decrease of the lateralization is depicted in figure 3.12 which shows the mean magnitude of the lateralization against the dir/rev ratio for the different directions. The 0° stimuli are not included since the use of unsigned values shifts their mean.

[1]E.g., the localization error of the answers to the 90°-LP-stimuli at dir/rev = -15dB is normally distributed around its mean of approximately 45°; but the localization error of the answers to the 60° LP stimuli at the same dir/rev ratio is normally distributed around 30°. Adding these two distributions will not yield a normal distribution.

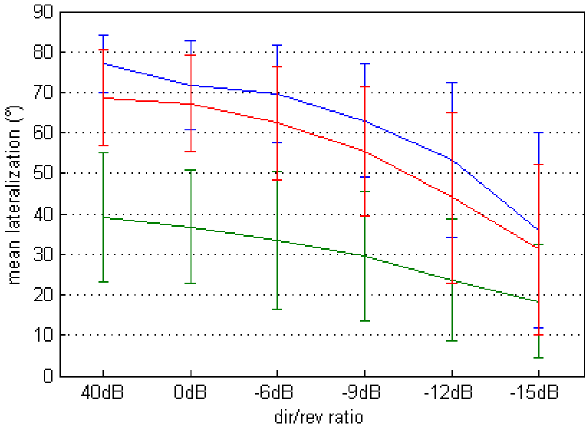

Fig. 3.12: Mean lateralization with standard deviation error bars plotted dir/rev; blue line: stimuli from 90°, red line: stimuli from 60°, green line: stimuli from 30°; note that abscissa is not scaled;

3.4 Comparison with previous investigations

Not many investigations whose procedures are comparable to the accomplished one have been found in the literature. Most of the experiments mentioned below show overlap only for specific conditions. Thus, if a certain component of the results from above is not compared or not compared to all investigations, the corresponding information could not have been extracted from the particular publications. Some previous investigations tested the localisation in the presence of distracting noise which led to quite similar results like the presence of reverberation. They are included below.

All below mentioned studies, as well as (Hartmann 1983) and (Rakerd and Hartmann 2003), yielded a degradation of the performance for lower direct-to-reverberation (respective target-to-distracter) ratios.
Only Begault et al. (2001b) found a higher accuracy in the azimuth judgements for the reverberant condition than for the non-reverberant one. In the latter case the localization performance for speech stimuli in a virtual environment was tested. Since many of their subjects reported a lack of externalisation in the non-reverberant condition, the authors propose to take the higher precision of judgements due to a better externalisation in the reverberant condition as an explanation.

The excessive lateralisation of the answers in the non-reverberant (respective quiet) condition is observed in virtual as well as in real environments. It can be found in the results of Braasch and Hartung (2002) (binaural), Gigure and Abel (1993) (loudspeakers) and Good and Gilkey (1996) (loudspeakers); but not in the results of Rakerd and Hartmann (2003) (loudspeakers) or in those of Lorenzi et al. (1999) (loudspeakers). The results of Lorenzi et al. (1999) even show the opposite, a reduction of lateralization.
Wightman and Kistler (1989) conducted an investigation to compare the local-

ization performance of the same subjects listening to a free-field and a virtual environment (all anechoic). In both situations basically all subjects show a slight over-lateralization which tends to be more pronounced in the virtual sound field. The order of magnitude of the over-lateralization reported in the literature (where present) is comparable to the present one.

The accumulation of the answers around the median plane for low dir/rev ratios appears in the results of Gigure and Abel (1993).
Braasch and Hartung (2002), Lorenzi et al. (1999), and Good and Gilkey (1996) investigated the localization performance in the presence of a noise distracter. In the first two studies an accumulation of the perceptions around the median plane with decreasing signal to noise ratio, as well as the opposite - even more excessive lateralisation with hardly any perception near the median plane - are reported. Whereby the authors have not found a general rule that describes the observations.
Good and Gilkey (1996) display explicit data only for one subject. In that case, an accumulation of the perceptions around the median plane for the low signal/noise ratios is not obvious. It is rather such that the judgements are only weakly related to the presented direction.
Good and Gilkey (1996) used loudspeakers to test the localisation of a click-train signal in the presence of noise. Their results show only few front-back reversals for the condition without distracter, but reversal rates increase with decreasing signal-to-noise ratio (up to 50% at the lowest signal/noise ratio of -10dB).

Gigure and Abel (1993) used 1/3-octave bands of noise as stimuli and found a clear frequency dependency of the localization performance for some particular directions.

Begault et al. (2001b) observe higher elevations for the reverberant condition.

Note that some of the above mentioned studies are revisited in section 5.4.

3.5 Conclusion

The localization performance of adult subjects in a reverberant environment has been examined and compared to the non-reverberant condition. Bursts of pink noise were used as stimuli.
The results revealed that an influence on the performance becomes obvious for dir/rev ratios below -6dB. This influence arises in a diminution of the lateralization, thus a cluster of the perceptions around directions from straight in front respectively straight behind the listener. A dependence on the frequency band (LP, BP or HP) has not been clearly verified in this case.
Concerning the front-back distinguishability no thesis can be stated since the results are not consistent between different listeners and furthermore, since the influence of the noise of the tracking system could not have been judged. A slight increase of the elevation of the answers for BP and HP stimuli towards the LP stimuli has been observed.

Two subjects participated additionally with non-individual HRTFs. Basically the same tendencies can be reported in this case as well, however, these measurements were not considered as contributing meaningfully to the results since on ITD adaptation was accomplished and severe perturbation limits their usefulness. Furthermore, subjects tended to excessively over-lateralize.

In general the results are quite consistent with those of previous investigations, although some discrepancies are present. Unfortunately, the comparison with those previous investigations is only limitedly possible since different methods and procedures were used.

Chapter 4

Test II

Test II is basically a repetition of test I, whereas subjects using non-individual HRTFs were tested. The motivation was to prove if there is a difference in terms of localization between people using individual HRTFs and people using non-individual ones.
The ITD of the subjects was approximated using a spherical head model and subjects accomplished an HRTF adaptation training prior to the test. Results show basically the same tendencies as in test I (diminution of the lateralization at low dir/rev ratios) but a very strong over-lateralization is present at higher dir/rev ratios for all of the subjects.

4.1 Methods

4.1.1 Subjects

7 unpaid adult subjects (2 female/5 male) participated in the test. All subjects reported not to have any known hearing degradation. None of them was experienced in binaural hearing and none of them had participated in test I.

4.1.2 Stimuli

See section 3.1.2.

4.1.3 ITD approximation

The *Spatialisateur* which was used to process the stimuli for playback decomposes the HRTFs into a minimum phase part and a pure delay, as it is proposed by Minnaar et al. (2000). This allows ITD - which is a critical parameter for localization (confer to section 2.1) - to be individually adjusted for every subject. The formula of Woodworth (Woodworth and Schloesberg 1962) gives an approximation of the ITD in the horizontal plane (confer to (2.1) and (2.2) in section 2.1.1):

$$ITD(\theta) = \frac{r}{c} \cdot (\theta + \sin\theta) \tag{4.1}$$

where r is the sphere radius, c the speed of sound, and the azimuth angle in radians. The ears are assumed to be diametrically disposed.

Larcher (2001) proposes an extension of Woodworth's formula for positions outside the median plane via the expression

$$ITD(\theta, \varphi) = \frac{r}{c} \cdot (\arcsin(\sin\theta\cos\varphi + \cos\theta\cos\varphi)) \tag{4.2}$$

which was derived by simple geometrical considerations. r is the sphere radius, c the speed of sound, θ the azimuth angle and φ the elevation angle (both in radians). The spatial shape of equation (4.2) is depicted in figure 4.1.

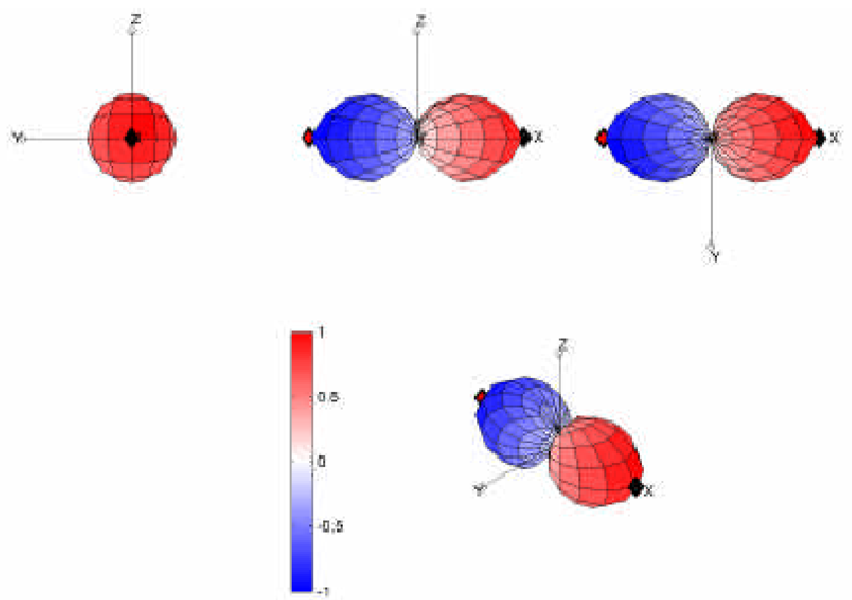

Fig. 4.1: Spatial function for equation (4.2) (Katz et al. 2003, p. 22)

Larcher (2001) performed a least squares fit between the measured ITD and the ITD produced by equation (4.2) to compute the optimal radius of the sphere for each subject. These optimal sphere radii were then related to anthropometric parameters by empirical regression formulas. She used a linear model to estimate the equivalent head radius as a weighted sum of the head dimensions, as it was proposed by Algazi et al. (2001):

$$r = \omega_1 X_1 + \omega_2 X_2 + \omega_3 X_3 + b \tag{4.3}$$

where X_1 is head half width, X_2 is the head half height, and X_3 is the head half length. A regression on the anthropometric data of her subjects results in the following values for the weights ω_i and the bias b (for all quantities measured in cm]:

$\omega_1 = 0.66$
$\omega_2 = -0.04$
$\omega_3 = 0.11$
$b = 3.33$

ω_2 can be set to zero without introducing significant errors (Larcher 2001, Algazi et al. 2001).
Although the accuracy of the 3D model is less than that of the 2D model, this approach was chosen in this investigation to be able to train the subjects' elevation perception to investigate its behaviour for stimuli in the horizontal plane.

4.2 Procedure

The general procedure follows that of test I. This time the test was carried out in a different room with less metallic equipment so that a larger distance between the subject and the humming tracking emitter could be kept. On the other hand the background noise entering the room from outside was somewhat stronger than in test I leading to a background noise level around 28dB(A). Subjects were blindfolded.
The minimum phase components of the used HRTFs were chosen individually for every subject by listening to amplitude modulated pink noise rotating on the horizontal plane around the subject's head and choosing the one that gives the most satisfying perception.
Prior to the localization test every subject completed a 12-minute HRTF adaptation training according to a procedure proposed by Blum(2003, Blum et al. 2004).

4.3 Results

4.3.1 Raw data

The answers of the subjects participating in this test show basically the same tendencies as the answers of the subjects in test I (diminution of the lateralization with decreasing dir/rev ratio). Front/back confusions occurred more frequently as depicted in figure 4.2.
The obvious difference is that the excessive lateralization observed in test I is even stronger here. Figure 4.3 compares the mean results for the anechoic condition for test I (figure 4.3 (a)) and test II (figure 4.3 (b)). As depicted in figure 4.3 (b), e.g. the HP stimuli presented from 30° are perceived at azimuths up to 60° in test II . Furthermore, the standard deviation of the mean results is noticeably larger which suggests that localization accuracy is inferior when using non-individual HRTFs.
Some subjects show biases which vary in magnitude and direction.

The implementation of the ITD-estimation as well as the other parts of the test were thoroughly examined to check if a conceptional fault or a malfunction had led to an over-estimation of the ITD. No error was found.

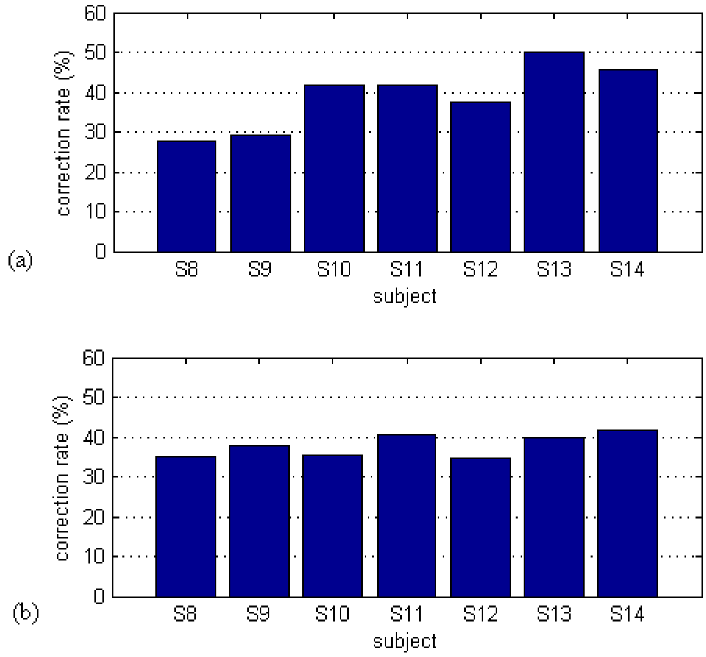

Fig. 4.2: Rate of corrections of front-back confusions for the individual subjects; (a) anechoic condition; (b) reverberant conditions;

In fact, ITD-estimation based on a spherical head model is known to underestimate the ITD (Larcher 2001, Busson et al. 2005) for most positions which should have lead to a diminished lateralization.
Interestingly, the over-lateralization becomes stronger the higher the frequency content of the stimuli is. This suggests that spectral cues may contribute to this since ITD is critical for localization mostly in low frequency regions (cf. to section 2.1.2 for a general description as well as to section 4.3.2 for a more detailed analysis of the present observations).

The raw data of Blum's localization tests (Blum 2003, Blum et al. 2004) were analysed to allow a comparison of the performances of the subjects. The analysis yielded that all of his subjects, i.e., subjects using individual HRTFs as well as subjects using non-individual HRTFs, before as well as after the adaptation training, over-lateralized in the same order of magnitude as subjects in test II (i.e., non-individual listeners). Confer as well to table A.1 in appendix A; Blum's subjects showed a mean localization error around 30° even after the adaptation training!
Examination of the raw data of another unpublished localization test using non-individual HRTFs carried out at Pitié - La Salpétrière yielded the same very heavy over-lateralization for all participating subjects.

To test if the use of non-individual HRTFs is the cause for the extra over-lateralization a series of informal tests was carried out to compare the performance of the same subject listening through individual as well as non-individual

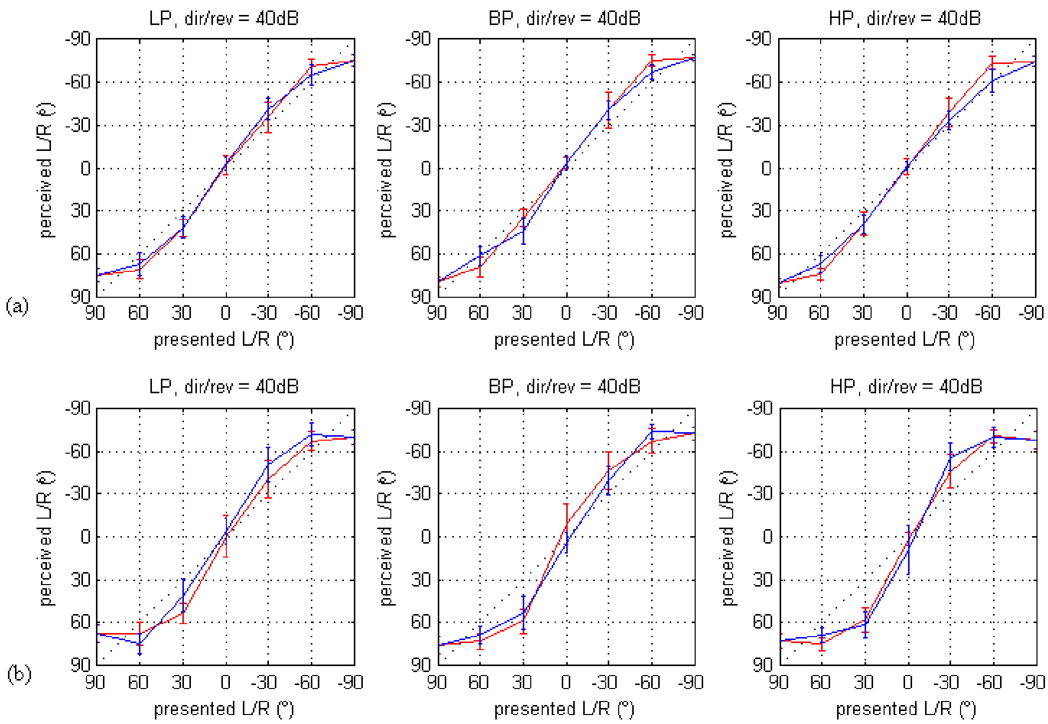

Fig. 4.3: Comparison of the lateralization (the angle of the answer direction to the median plane) in test I (a) and test II (b); anechoic condition; LP, BP, and HP stimuli in separate columns;

HRTFs. However, the obtained data do not allow a reliable interpretation. It was even such that the results of test I could not have been reproduced. In all conditions and HRTF configurations the over-lateralization was much stronger than in test I. Note that subjects passed these runs blindfolded in contrast to test I.

The elevation perception is similar to that in test I as depicted in figure 4.4. The peaks around the median plane are strongly pronounced in the anechoic condition and elevation perception is more balanced in the presence of reverberation.

4.3.2 Statistical analysis

The ANOVA was performed in a manner described in section 3.3.3. Testing the localization error at the different dir/rev ratios against the anechoic condition for all four groups of stimuli (i.e. 0°, 30°, 60°, and 90° , confer to section 3.3.3) gives the results shown in figure 4.5. Values are comparable to those obtained for test I with some more perturbation[1].

[1]Note that the results depicted in figure 4.5 have to be treated with caution. Due to the perturbation of the answers the variances of the different distributions can not be assumed to

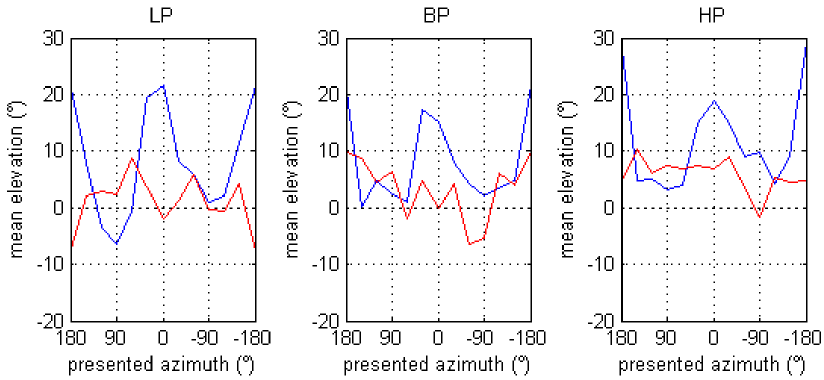

Fig. 4.4: Elevation of the answers for different stimulus azimuths and frequency bands; values are means over all respective data; standard deviations are not shown; blue lines: non-reverberant condition; red lines: reverberant conditions;

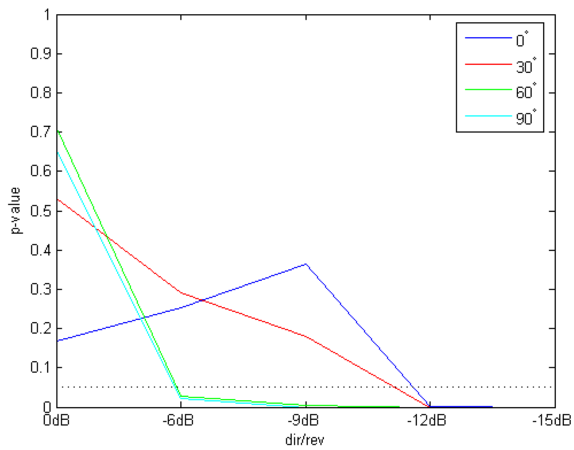

Fig. 4.5: p-values for all dir/rev ratios tested separately against the anechoic condition; each colour represents one group of stimuli according to the legend

The slope of the decrease of the lateralization is depicted in figure 4.6 which shows the mean of the magnitude of the lateralization against the dir/rev ratios. The 0°-stimuli are excluded.

4.4 Comparison with previous investigations

Literature about localization with non-individual HRTFs is sparse. Since the studies investigating the presence of reverberation were already described in sec-

be constant as would be necessary for ANOVA

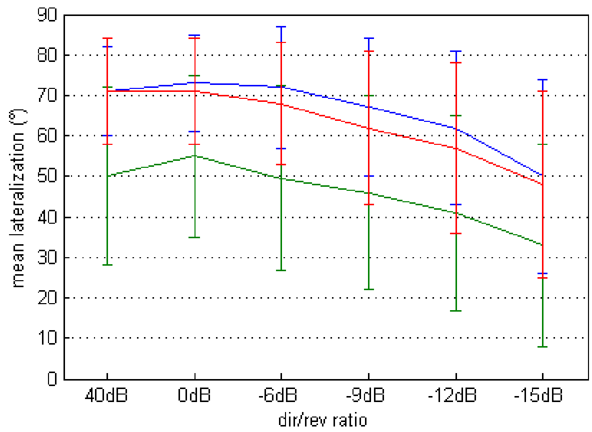

Fig. 4.6: Mean lateralization with standard deviation error bars plotted dir/rev; blue line: stimuli from 90°, red line: stimuli from 60°, green line: stimuli from 30°; note that abscissa is not scaled;

tion 3.4. they are not included here. The only study amongst them employing non-individual HRTFs is Begault et al. (2001). However, the published data are restricted to quite generic localization errors.

Wenzel et al. (1993) compared the localization performance of the same (inexperienced) subjects listening in a real free-field and a virtual one (both anechoic). All subjects used the same set of non-individual HRTFs, no respect of the ITD was taken.
A comparable performance for both conditions for azimuth judgements is reported. In the virtual condition front/back reversal rate was higher and accuracy of elevation judgement was poorer for some of the subjects.

Mention of the very heavy over-lateralization observed in the present test has not been found in the literature.

4.5 Conclusion

According to the above described observations it cannot be assumed that the perception of people listening through non-individual HRTFs is a priori comparable to that of people listening through individual ones although an investigation of Wenzel et al. (1993) suggests this. For all subjects a very heavy over-lateralization is present and accuracy of the answers is inferior. Spectral cues might play a major role in this over-lateralization since it is stronger for stimuli with higher frequency content.

The better performance of the subjects in test I may be due to their more considerable experience in binaural hearing. However, the subjects in test II were blind-folded but those in test I were not. The results of test III (chapter 5) corroborate the hypothesis that this might have influence on the answer reports of the subjects (Confer as well to chapter 6).

Chapter 5

Test III

Test III is another binaural localization test using similar procedures as in tests I and II, whereas speech was used as stimuli (targets as well as distracters), thus providing a "cocktail party" listening task. Subjects using non-individual HRTFs as well as subjects using individual HRTFs were tested. The ITD of the subjects was approximated by scaling a pre-calculated spatial function with respect to the head circumference of the subjects.
Pure reverberation of voices as distracter did not have any influence on loalization; the pure direct sound of those voices as distracter degraded the performance. This observation was only made at the lowest target-to-distracter ratio at the limit of detectability.

5.1 Methods

5.1.1 Subjects

Since the results of test II do not allow to state a thesis about the difference in localization performance of listeners using non-individual HRTFs compared to listeners using individual HRTFs both groups of listeners were tested.
8 adult subjects (4 male/4 female) using individual HRTFs and 8 adult subjects (6 male/2 female) using non-individual HRTFs were tested. Amongst the subjects were S1 and S3 from test I and S8 from test II.
All reported not to have any known hearing degradation.

5.1.2 Stimuli

Stimuli were speech recordings in French language taken from a database of LIMSI-CNRS as well as from an IRCAM database. All recordings were done by native French speakers.

Targets

One sentence of a male speaker and one sentence of a female speaker, each with a balanced pronunciation were chosen as targets. Due to the limited time to accomplish the test only the male voice could be tested.
The male target sentence was: "C'était dans l'enclos du fond où ça devait se passer." ("It was in the enclose in the back where this must have happened.").
The target positions were identical to those used in test I and II whereas $\pm 15°$ and $\pm 165°$ were added to better resolve locations near the median plane.

Distracters

The distracter was composed of twelve different native French speakers (six male and six female) positioned in the median plane at intervals of 30°. Absolute positions of $[\ \ldots, -45°, -15°, 15°, 45°, \ \ldots]$ (with alternating sex) were chosen in order to create a symmetric arrangement minimizing the overlap with target locations.
The material was quite different in terms of semantic content. This was not assumed to be critical since it was found to be very difficult to understand a major part of a sentence during the short time interval during which the signal was presented.
Two different conditions were tested: Pure direct sound of distracters as well as pure reverberation of distracters. Reverberation time was set to 1s and a slight room filter (mid and high frequency attenuation) was applied to yield a natural sound.

5.1.3 Preparation of the stimuli

Since a simple root-mean-square alignment of the individual speakers lead to a very unbalanced distribution in terms of prominence of the voices different methods were tried.

Loudness

Instantaneous loudness of the stimuli was calculated according to ISO 532 B/DIN 45 631 (Zwicker and Fastl 1999) using MATLAB scripts provided by Churchill (2004).

The overall loudness of the speakers was approximated on the basis of the observation of Zwicker/Fastl (1999) that the long-term loudness of a single speaker can be approximated by the percentile loudness N_7, i.e. the loudness value reached or exceeded in 7% of the measurement time.
All signals (targets as well as all the individual signals the distracter was composed of) were aligned to have equal N_7 (for a given playback level).

Loudness alignment gave satisfying results for the distracter voices, whereas the targets were found to be less balanced. Apparently, target signals were too short (3-4s towards 10-15s for distracters) to allow a reliable estimation of the long-term loudness.

Active speech level

All speakers were then aligned to have equal active speech level ASL (i.e. a measure proportional to the power that takes into account only those parts of the signal that are perceived as active; see ITU (1993) for further explanation). A MATLAB implementation is provided by Brookes (2003). ASL alignment gave better - but still not perfect - results than loudness alignment. It was found that despite equal ASL prominence of different speakers varies noticeably. Especially emotion, articulation quality, and content of phonemes (e.g. /t/, /s/, or /sh/ are most prominent) within a sentence are critical.

5.1.4 ITD approximation

The method of approximating the individual ITD the approach of Katz et al. (2003) was chosen which has shown to be more favorable than the method described in section 4.1.3 (Katz et al. 2003).

(Katz et al. 2003) carried out a principle component analysis (PCA) of the measured ITDs of all subjects of the LISTEN HRTF Database to obtain an estimation of the general spatial variation of the ITD that is more closely linked to the actual ITD of humans than a trigonometric function. The individual ITDs of the subjects in the database were calculated using the MaxIACC method (Katz et al. 2005 (comparison of interaural time delay estimation methods), see Minnar et al. (2000) or Katz et al. (2005) for a review of ITD models). The obtained spatial function of the ITD is illustrated in figure 5.1.
The differences to the trigonometric approximation of figure 4.1 are apparent. Whereas the trigonometric approximation is exactly symmetric to all axes, the present spatial function shows obvious abnormalities which are linked to morphological properties of the human body (especially the head, neck, and location of the ears).

Linked to the PCA spatial function is a series of scaling factors for each individual. These factors were correlated to measured morphological data. The subject's head circumference $circ_{head}$ was chosen since it is simple to measure and provides the best correlation. A least-mean-squared approximation lead to the following expression for the weighting factor ω:

$$\omega = 6.4586 + 0.0162 \cdot circ_{head} \qquad (5.1)$$

5.2 Procedure

In general, procedures were similar to those in tests I and II (see section 3.2.). Every subject completed two sessions.

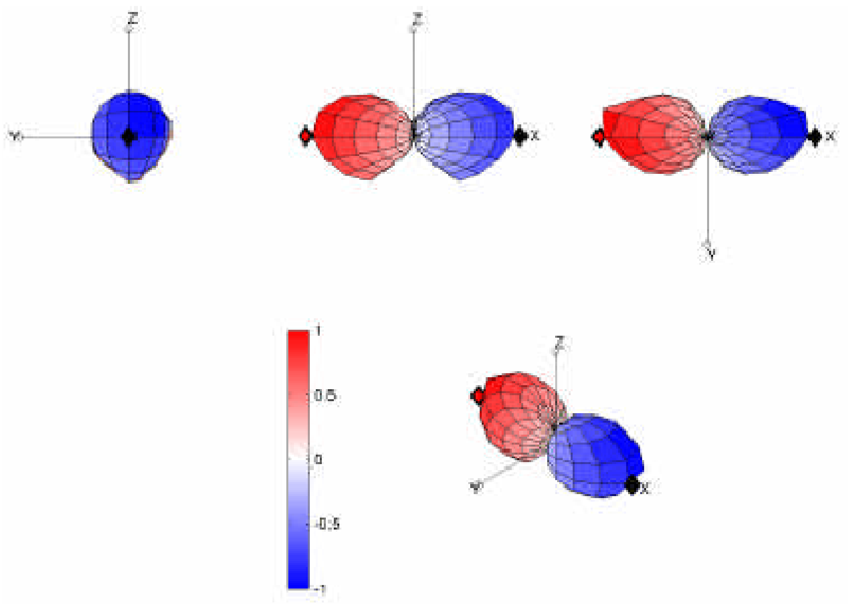

Fig. 5.1: Spatial function for the PCA of the MaxIACC ITD calculation (Katz et al. 2003, p. 23)

The first session was composed of two runs with a brief pause in between. The reference (non-distracter) condition and one of the conditions with distracters present were tested. In the second session implied only the distracter condition which was not been tested. In the initial reference phase 48 stimuli (16 directions x 3 presentations) were presented without distracters. The conclusive session consisted of 192 stimuli with distracters present (target/distracter ratios of 0dB, -6dB, -9dB, -12dB; 3 presentations of each direction). In all stages of the test the stimuli were played in arbitrary order and all possible stimuli configurations occurred exactly three times. The distracter signals were perfectly identical throughout the test; the level of the target was adjusted according to the desired ratio.

Prior to the run with discrete distracters (pure direct sound) every subject was given five minutes of training to get used to the distracter. Without training subjects tended to have problems of detecting the target at low target/distracter ratios. After a couple of minutes detecting was not a problem anymore. The training ensured that the detection performance stayed constant over the entire duration of a run.
For the reverberation distracter the detection was not a problem.

Since it might occur nevertheless that a target is not detected (especially at low dir/rev ratios) the subjects had the possibility to signalize whenever no detection occurred by pressing the space bar of the computer keyboard after the answer. Subjects where forced to indicate a direction nevertheless to make sure that the no-detection procedure employs more effort than a usual answer to prevent subjects from using this option more often than necessary.

Subjects were instructed to indicate a valid direction whenever they had perceived any indication of the presence of the target.

The minimum phase component of the used HRTFs was selected according to a procedure prepared by Brian Katz (2004) who carefully selected the HRTFs of 7 subjects out of the LISTEN Database. The perception of noise bursts rotating around the head on the horizontal plane and on the vertical plane has to be judged on a continuous scale between “mauvais” and “excellent” (i.e. “bad” and “excellent”). The set of HRTFs providing the most satisfying perception was then chosen for the localization test.

Due to the long duration of the test an HRTF adaptation training for non-individual listeners was waived. Furthermore, the gain in azimuth accuracy is rather small anyway (less than 2°). The more thorough HRTF selection procedure was expected to compensate for this.

A new box for the emitter of the tracking system was constructed which almost perfectly eliminated the hum. The background noise level in the testing room was 22dB(A).

5.3 Results

5.3.1 General observations

Most of the subjects tended to be quite obviously biased by up to 15°. Interestingly, subjects with individual HRTFs tend to show a stronger bias. Biases are more commonly toward the left hand side but not systematic for a given subject or a given distracter condition.
Thorough measurement of all involved devices did not yield any indication for a cause. The difference in sound pressure between the two loudspeakers was less than 0.3dB(A) for a sound from 0° at the test playback level.
After the calibration procedure of the tracking system (blindfolded) subjects were asked to indicate straight ahead so that the calibration could be verified. Surprisingly, the accuracy of those indications was quite low. Very few of the subjects indicated straight ahead with a deviation of less than 10° , in contrast to test I where the indication of straight ahead with open eyes was accurate suggesting that the blindfold makes indication somewhat less accurate. This hypothesis is supported by the fact the S1 and S3 were less biased in test I than in the present one.

Data of subject S26 were excluded. S26’s answers were strongly biased by up to 25° and he consequently under-lateralized by up to 60° , i.e., he perceived the 90° stimuli at 30° (reference condition, bias corrected). S26 reported never to be convinced of the perception when using his HRTFs which suggests that they have not been measured or processed correctly.

Reversal rates are quite similar throughout the groups and conditions and comparable to those of tests I and II, whereas two individual HRTF listeners

(S23 and S25) had extraordinarily low reversal rates below 10% in the reference condition. The rates of non-individual HRTF listeners tended to be slightly higher.

5.3.2 Non-distracter condition

The lateralization of the answers of the subjects using individual HRTFs is depicted in figure 5.2. It shows an obviously larger over-lateralization than in the anechoic condition in test I. The performance of subjects using non-individual HRTFs is quite similar.

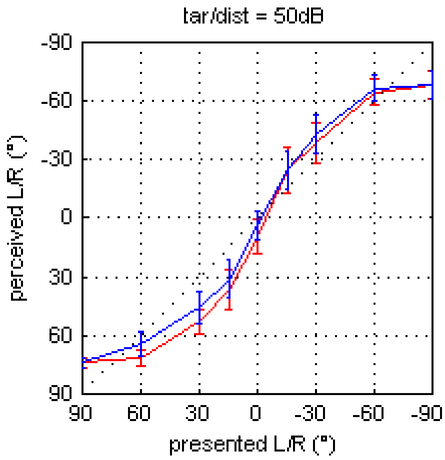

Fig. 5.2: Lateralization of the answers of the subjects using individual HRTFs in the non-distracter condition; blue lines: stimuli from the frontal hemisphere; red lines: stimuli from the back; the bias toward the left hand side is clearly visible;

5.3.3 Reverberant distracter

For both groups of listeners, there is no substantial influence of the reverberant distracter on localization as can be seen in figure 5.3.

5.3.4 Direct sound distracter

The influence of the direct sound as distracter is negligible at tar/dist ratios of -9dB or higher. At tar/dist = -12dB the accuracy of the answers is substantially inferior. Lateralization is slightly diminished.
Braasch and Hartung (2002) found in their study of localization in the presence of a noise distracter that subjects group their answers around the directions outward left, directly frontal, and outward right at low tar/dist ratios.
These tendencies did not occur in the present study as can be seen in figure 5.4. This figure shows the lateralization of the answers versus the lateralization of

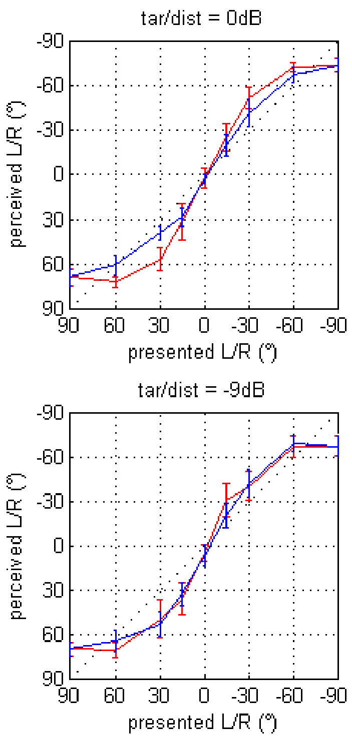

Fig. 5.3: Lateralization of the answers of the subjects using individual HRTFs in presence of the reverberant distracter for the different tar/dist ratios; blue lines: stimuli from the frontal hemisphere; red lines: stimuli from the back;

the directions of presentation. The more answers fell into a sector the darker is its shading. The resolution for the answers is 5°; the resolution of stimulus directions is according to the spherical distance between two neighbouring stimuli (30° respectively 15° close to the median plane).
This illustration was chosen since the observations of Braasch and Hartung (2002) might be smeared due to the averaging which takes place in figures similar to e.g. figure 5.3.
Figure 5.4 shows that at the lowest tar/dist ratio of -12dB in the present study, the answers are just less accurate rather than grouped. Note that the tar/dist ratio of -12dB was on the limit of detectability of the target.

The elevation perception was slight more balanced than in tests I and II throughout the conditions, as depicted in figure 5.5.

5.4 Comparison with previous investigations

Although the "cocktail party" is the subject matter of a vast number of investigations, almost all of them focus on selective attention. Localization plays a less important role.

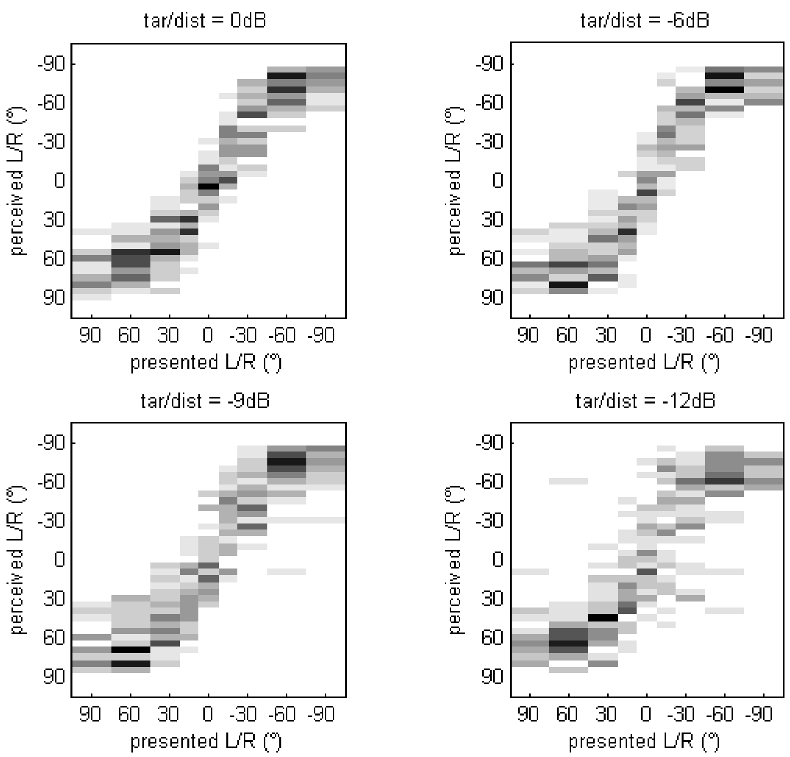

Fig. 5.4: Lateralization of the answers of the subjects using individual HRTFs in presence of the discrete distracter for the different tar/dist ratios; the darker the shading of a sector the more answers fell into it;

The only localization investigation found that used speech as stimuli and that gives detailed information about the performance is that of Begault and Wenzel (1993) where subjects show good performance with a slight over-lateralization in an anechoic evironment. It includes as well a comparison of the performance with noise and speech stimuli Wenzel et al. (1993) (see section 4.4). The performance of the subjects those two investigations are similar.

In other investigations such as Hawley et al. (1999) or Yost et al. (1996) the localization task was rather an identification task (of a loudspeaker or other discrete distributed location) with an angular resolution too rough to allow explicit interpretation of the results.

Other studies such as Good and Gilkey (1996), Lorenzi et al. (1999), Langendijk et al. (2001), or Braasch and Hartung (2002) investigated the influence of a small number (one respectively two) of discrete distracters on localization of a train of clicks/noise bursts. In all studies localization performance was inferior in the presence of distracters. All revealed pulling effects toward the distracters as well as dispersion.
In the studies of Good and Gilkey (1996) and Lorenzi et al. (1999) the front-back confusion rate increased substantially in the presence of a distracter which was not so obvious in the study of Langendijk et al. (2001).

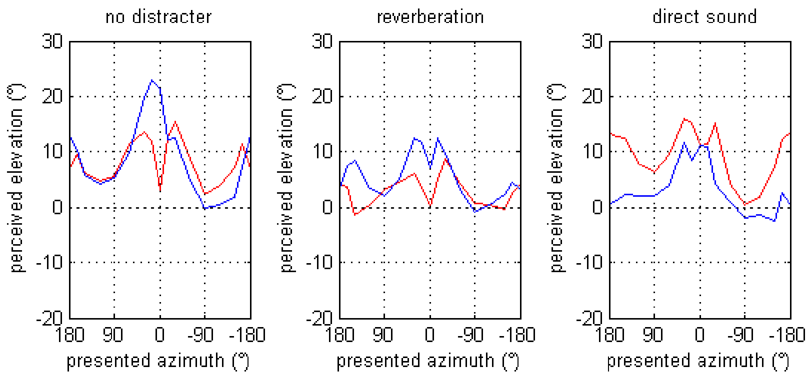

Fig. 5.5: Elevation of the answers for different stimulus azimuths and conditions; values are means over all respective data; standard deviations are not shown; blue lines: subjects with non-individual HRTFs; red lines: subjects with individual HRTFs;

5.5 Conclusion

In principle, localization of speech stimuli in an anechoic environment does not seem to be substantially different from localization of noise stimuli. In the presence of other speech signals as distracers the performance degrades for very low target-to-distracter ratios. The front-back distinguishability is more reliable for some of the subjects.
Since the only condition where the distracter influenced the performance was at the limit of detectability, it is suggested that some interaction between detectation and localization of a traget takes place.

This time - when all procedures were identical for all subjects - no systematic differences in the loclization performance of subjects using individual HRTFs and subjects using non-individual HRTFs have been observed.

Chapter 6

Perspectives

Since the answer reporting technique is one of the linchpins of the investigations, its influence on the results has to be known.
A number of means for the answer report in auditive localization tests are proposed by researchers. Amongst many more, there is the verbal answer report (e.g. by calling out numerical estimates of the azimuth and elevation of the target as in (Wenzel et al. (1993)), the GELP method (subjects have to indicate the direction on a sphere in front of them, e.g. in (Braasch and Hartung 2002)), or the answer report via a pointing device with which the subjects point into the direction of the perception (e.g. in (Langendijk et al. (2001))), all with certain advantages and disadvantages.

The discrepancies between the results of test I and II in the present investigation might be due to the use of a blindfold in test II but not in test I. To be able to entirely interpret the results of the accomplished tests a number of issues concerning the answer report will have to be assessed:

1. The accuracy of the answers given with blindfold compared to the accuracy of answers given with open eyes.

2. The accuracy of answers given in the frontal hemisphere compared to the accuracy of answers given in the rear hemisphere.

3. The accuracy of answers given with the right hand in the left and right hemispheres compared to the accuracy of answers given with the left hand in the left and right hemispheres.

Furthermore, to allow a more reliable interpretation of the results of test III the interaction between auditory detection and localization has to investigated more thoroughly.

Appendix A

HRTF adaptation training

A.1 Introduction

Blum (2003, Blum et al. 2004) conducted a test to elicit adaptation of subjects to non-individual HRTFs. The idea is based on an observation by Hofman et al. (1998) who showed that subjects with modified pinnae (using inserts), following a dramatic degradation of sound elevation perception, steadily reacquired localization accuracy over time.

A.2 Procedure

Blum (2003, Blum et al. 2004) used a proprioceptive feedback of auditory spatial information by giving blindfolded subjects control over a virtual sound source spatialized at the position of a handheld tracking sensor. This allowed the subjects to associate the source position with the acoustic cues used for binaural rendering through the constant awareness of his/her hand position.
His investigation was composed of three stages: A localization test to evaluate the subject's initial performance, an adaptation procedure, and the localization test again. The procedure of the localization test was similar to that of test I whereas 40 Hz amplitude modulated pink noise bursts were used as stimuli and stimulus positions were not limited to the horizontal plane.
The subjects' task in the adaptation procedure was to search for animal sounds hidden around them. In this phase, alternating pink/white noise was continuously played at the hand position with the frequency increasing as the spherical distance between subject's hand position and the target point decreased. When the target was found, the noise was replaced by an animal sound which had to be placed at a reference point directly in front of the subject. The adaptation training was limited to 12 min.

The HRTFs used were decomposed into a minimum phase component (for spectral cues) and a pure delay (for ITD cues). This separation enabled the use of hybrid HRTFs, where the individual ITD was kept and combined with a non-individual minimum phase component.

10 subjects participated in the test. They were divided into two groups: One

using hybrid HRTFs, and a control group using individual HRTFs to be able to separate possible learning effects due to the training.

A.3 Results

The evaluation of the results of the investigation is shown in table A.1. It suggests that the localization performance of subjects using hybrid HRTFs indeed improves due the adaptation training and that it is afterwards even comparable to the performance of a control group using individual HRTFs.

results test1 | test2	spherical angle (°)		elevation (°)		azimuth (°)		f/b confusion rate (%)	
test group	35	29	27	21	17	14	35	25
variance	4.2	2.0	4.1	2.2	3.5	2.0	9.2	7.5
control group	31	30	22	22	16	15	31	29
variance	5.7	7.1	5.9	6.5	4	4.6	10.8	11.7

Tab. A.1: Mean localization errors before (test1) and after (test2) adaptation training; data from (Blum 2003)

Bibliography

V.R. Algazi, C. Avendano, R. Duda (2001): *Estimation of a Spherical Head Model - from Anthropometry.* J. Aud. Eng. Soc., Vol. 49, No.6, p.472-478 (June, 2001)

Arbogast, T.L., Mason, C.R., Kidd, G. (2002): *The effect of spatial separation on informal and energetic masking of speech.* Journal of the Acoustical Society of America (JASA) 112(5), pp. 2086-2098

Begault, D.R. (1997): *Virtual acoustics: evaluation of psychoacoustic parameters for Auralization.* Proc. of IWHIT '97, 3rd Int. Workshop on Human Interface Technology, NASA Ames Research Center

Begault, D.R. (2000): *3-D Sound for Virtual Reality and Multimedia.* Academic Press Professional, Cambridge, MA

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 Virtual Speech Source.* Journal of the AES 49 (10), 904-916

Begault, D.R. (2002): *Introduction to Psychoacoustics. Part II. Binaural Hearing.* Workshop presentation, AES 112th meeting (Munich) and 113th meeting (Los Angeles)

Begault, D.R. and Wenzel, E.M. (1993): *Headphone localization of speech.* Human Factors 35(2), pp. 361-376

Blauert, J. (1999): *Spatial Hearing: The Psychophysics of Human Sound Localization.* MIT Press, Cambridge, London

A. Blum (2003) : *Etude de la plasticité du système auditif en localisation sonore - Application au problème de l' individualisation en synthèse binaurale.* Rapport de DEA, Université d'Aix Marseille II, DEA d 'Acoustique 2003

A. Blum, B.F.G. Katz, O. Warusfel (2004): *Eliciting adaptation to non-individual HRTF spectral cues with multi-modal training.* CFA/DAGA 2004. Congrès commun : 7ème congrès de la Société Franąise d'Acoustique et 30ème congrès de la Société Allemande d'Acoustique. Strasbourg, France, 22-25 mars 2004

Braasch, J. and Hartung, K. (2002): *Localization in the Presence of a Distracter and Reverberation in the Frontal Horizontal Plane.* I. Psychoacoustical Data. Acoustica 88, 942-955

Bregman, A.S. (1990): *Auditory Scene Analysis: The perceptual organization of sound.* MIT Press, Cambridge, London

Brookes, M. (2003) : *VOICEBOX: Speech Processing Toolbox for MATLAB.* Department of Electrical and Enlectronical Engineering, Imperial College.
URL=http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html

Brungart, D.S. and Simpson, B.D. (2002): *The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal.* JASA 112(2), 664-676

S. Busson, R. Nicol, B.F.G. Katz (2005): *Subjective investigations of the interaural time difference in the horizontal plane.* Presented at the 118th AES Convention 2005 May 28(31 Barcelona, Spain)

Cheng, C.I. and Wakefield, G.H. (1999): *Introduction to Head-Related Transfer Functions (HRTF's): Representations of HRTF's in Time, Frequency, and Space* (invited paper). Proc. of the 107th AES 107th Conv., New York, preprint

Clifton, R. K. (1987): *Breakdown of echo suppression in the precedence effect.* JASA 82(5), 1834-1835

Culling, J.F., Hodder, K.I., Toh, C.Y. (2003): *Effects of reverberation on perceptual segregation of competing voices.* JASA 114(5), 2871-2876

Damaschke, J. (2004): *Towards a neurophysiological correlate of the precedence effect: from psychoacoustics to electroencephalography.* Oldenburg, Univ., Dissertation

Devore, S., Shinn-Cunningham, B. (2003): *Perceptual consequences of including reverberation in spatial auditory displays.* Proc. of the Int. Conf. on Auditory Display, Boston, MA, USA, July 2003

Divenyi, P.L. and Oliver, S.K. (1989): *Resolution of steady-state sounds in simulated auditory space.* JASA 85(5), 2042-2052

Feddersen, W.E., Sandel, T.T., Teas, D.C., Jeffres, L.A. (1957): *Localisation of high frequency tones.* JASA 29, 988-991

Fletcher, H. (1940): *Auditory patterns.* Rev. Mod. Phys. 12, 4765

Gardner, M.B. (1969): *Distance estimation of 0° or apparent 0°-oriented* *speech signals in anechoic space.* JASA 45, 47-53

Giguère, C. and Abel, S.M. (1993): *Sound localisation: Effects of reverberation time, speaker array, stimulus frequency, and stimulus rise/decay.* JASA 94(2), 769-776

Good, M.D. and Gilkey, R.H. (1996): *Sound localization in noise: The effect of signal-to-noise ratio.* JASA 99(2), 1108-1117

Grantham, D.W., Hornsby, B.W.Y., Erpenbeck, E.A. (2003): *Auditory spatial resolution in horizontal, vertical, and diagonal planes.* JASA 114(2), 1009-1022

Guiu, P. (2004): *OPERA: Optimisation Perceptive du Rendu Audio [Perceptive Optimization of Audio Rendering].* Intern paper of IRCAM, Paris

Hartmann, W.M. (1983): *Localization of sound in rooms.* JASA 74(5), 1380-1391

Hawley, L.M., Litovsky, R.Y., Colburn, H.S. (1999): *Speech intelligibility and localization in a multi-source environment.* JASA 105(6), pp. 3436-3448

Hebrank, J. and Wright, D. (1974): *Spectral cues used in the localisation of sound sources on the median plane.* JASA 56, 1829-1834

Hofman, J.A.G. Riswick, A.J. van Opstal (1998): *Relearning sound location with new ears.* Nature Neuroscience 1(5)

ITU (1993) : *Objective measurement of active speech level.* ITU-T recommendation P.56 (03/93)

Kahana, Y. et al. (1999): *Numerical Modelling of the Transfer Functions of a Dummy-Head and of the External Ear.* Proc. of the 16th AES International Conf. on Spatial Sound Reproduction, Rovaniemi, Finland

Katz, B., Vandernoot, G., Warusfel, O. (2003): *Listen Project Deliverable D4.3* - Individual Adaptation. Version 1.0

Katz, B. (2004): *HRTFselectiontest.* Perception située, LIMSI-CNRS
URL=http://www.limsi.fr/Individu/katz/www/HRTFselectiontest/

B.F.G. Katz, S. Busson, O. Warusfel (2005): *Comparison of Inter-aural Time Delay estimation methods in the signal and perceptive domain.* Presented at the 118th AES Convention 2005, May 28-3, Barcelona, Spain

Kidd, G., Mason C.R., Burghera, A., Hartmann, W.M. (2004): *The role of reverberation in release from masking due to spatial separation of sources for speech identification.*
url = http://labrosa.ee.columbia.edu/Montreal2004/papers/...
.../kiddetal_acustica1.pdf

Kopco, N. and Shinn-Cunningham, B.G. (2003): *Spatial unmasking of nearby pure-tone targets in a simulated anechoic environment.* JASA 114(5), 2856-2870

Langendijk, E.H.A., Kistler, D.J., Wightman; F.L. (2001): *Sound localization in the presence of one or two distracters.* JASA 109(5), 2123-2134

Langendijk, E.H.A., Bronkhorst, A.W. (2002): *Contribution of spectral cues to human sound localization.* JASA 112(4), 1583-1596

Larcher, V. (2001): *Techniques de spatialisation des sons pour la réalité virtuelle.* These de doctorat de l' Université de Paris VI

Lord Rayleigh (1877): *Acoustical observations.* Philosophical Magazine 3, 6th series, 456-464

Lord Rayleigh (1904): *On the acoustic shadow of a sphere.* Phil. Transact. Roy. Soc. London 203A, 87-99

Lord Rayleigh (1907): *On our perception of sound direction.* Phil. Mag. 13, 6th series, 214-232

Lorenzi, C., Gatehouse S. and Lever, C. (1999): *Sound localization in noise in normal-hearing listeners.* JASA 105(3), 1810-1820

Mehrgardt S, Mellert V.: *Transformation characteristics of the external human ear.* J Acoust Soc Am. 1977 Jun;61(6):1567-76

Mills, A.W. (1972): *Auditory localization.* In Tobias, J. (editor), *Foundations of Modern Auditory Theory*, Volume II, Academic Press, New York, pp. 301-348

P. Minnaar, J. Plogsties, S.K. Olesen, F. Christensen, H. Møller (2000): *The Interaural Time Difference in Binaural Synthesis.* Presented at the 108th AES Convention, Paris, France, Feb 2000

Moore, B.C. (1997): *An introduction to the psychology of hearing.* Academic Press, 4th edition, San Diego

Moore, B.C.J., Glasberg, B., Baer, T. (1997): *A model for the prediction of thresholds, loudness, and partial loudness.* JAES 45(4), pp. 224-240

Morse, P.M. (1948): *Vibration and sound.* McGraw-Hill, New York.

Rakerd, B. and Hartmann, W.M. (2003): *Localization of noise in a reverberant environment.* Proc. of the 13th Int. Symposium on Hearing, Dourdon, France

Roman, N., Wang, D., Brown, G.J. (2002): *Location-based sound segregation.* Proc. of the Int. Joint Conf. on Neural Networks, Honolulu, Hawaii, USA, May 2002

Scharf , B. (1970): *Critical Bands.* In Tobias, J.V. (editor), *Foundations of Modern Auditory Theory, Volume I*, Academic Press, New York, pp. 157-202

Shaw, E.A.G. (1974): *The External Ear. In Handbook of Sensory Physiology V/1: Auditory System, Anatomy Physiology (Ear)*, Springer-Verlag, New York

Shinn-Cunningham, B., Ihlefeld, A. (2004): *Selective and divided attention: Extracting information from simultaneous sound sources.* Proc. of ICAD 04-Tenth Meeting of the Int. Conf. on Auditory Display, Sydney, Australia, July 2004

Schroeder, M.R. (1954): *Eigenfrequenzstatistik und Anregungsstatistik in Räumen.* Acustica 4, pp. 456-468, English translation in: *Statistical parameters of the frequency response curves of large room.* JAES 35, pp. 299-306

von Hornbostel, E.M., Wertheimer, M. (1920): *Über die Wahrnehmung der Schallrichtung [On the perception of the direction of sound].* Sitzungsbericht Akad. Wiss. Berlin, 388-396

Wenzel, E.M., Arruda, M., Kistler, D.J., Wightman, F.L. (1993) : *Localization using nonindividualized head-related transfer functions.* JASA 94(1), pp . 111-123

Wightman, F.L., Kistler, D.J. (1989a): *Headphone simulation of free-field listening. I: Stimulus synthesis.* JASA, 85(2)

Frederic L. Wightman, Doris J. Kistler (1989b) : *Headphone simulation of free-field listening. II : Psychophysical validation.* JASA 85(2), pp. 868-878

Woodworth, R.S. (1938): *Experimental psychology.* Holt, New York

R.S. Woodworth and H. Schloesberg (1962): *Experimental Psychology.* Holt, Rinehard, and Winston, New York, pp. 348-361

Yost, W.A., Dye, R.H., Sheft, S. (1996): *A simulated "cocktail party" with up to three sound sources.* Perception and Psychophysics 58(7), pp. 1026-1036

Zwicker, E. and Fastl, H. (1999): *Psychoacoustics: Facts and Models.* Springer-Verlag, Berlin Heidelberg, New York