Factors influencing acoustical localization

Vom Fachbereich Physik der Universität Oldenburg
zur Erlangung des Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
angenommene Dissertation

Tag der Disputation: 21.09.2001

Erstreferent: Prof. Dr. Dr. B. Kollmeier

Korreferent: Prof. Dr. V. Mellert

Jörn Otten
geb. am 22. Juni 1970
in Leer / Ostfriesland

Factors influencing acoustical localization

Vom Fachbereich Physik der Universität Oldenburg
zur Erlangung des Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
angenommene Dissertation

Tag der Disputation: 21.09.2001

Erstreferent: Prof. Dr. Dr. B. Kollmeier

Korreferent: Prof. Dr. V. Mellert

Jörn Otten
geb. am 22. Juni 1970
in Leer / Ostfriesland

Contents

Chapter 1

General introduction

The ability of the auditory system to determine the spatial position of a sound source is essential for the orientation in our daily life environment. Due to a comprehensive analysis of the sound field generated by the source, human listeners are able to assess the direction, the distance and the spaciousness of a sound source. In contrast to the visual system, this capability is not restricted to a limited spatial range and, thus, the auditory sense does not only extend the perception of the environment to the acoustical modality but also extends the range of spatial cognition to the whole range of spatial directions. This extension allows us to be completely enveloped in the environment and it is, therefore, not surprising that we often close our eyes (for instance, in a music concert or even on a silent meadow) if we do not want to focus our attention to the spatially restricted range provided by the eyes.

The spatial information that is used by the auditory system to localize a sound source in a non-reverberant environment is captured by head related transfer function (HRTFs). They describe the transformation of the sound from its source location in the free-field to the microphone in the left or right ear canal. HRTFs can be measured by recording a sound emanating from a speaker at a certain location in space by small probe microphones in the ear canal of a subject. The auditory system uses two different kinds of cues that can be extracted from the HRTFs to estimate the source location. The *binaural* cues are calculated from a comparison of the HRTFs of the left and right ear. The interaural level difference (ILD) is caused by head shadowing and interference effects and describes the differences in level at the left and right ear as a function of frequency. The interaural time difference (ITD) reflects the differences in the path length (for lateral source positions) from the sound source to the left and right ear, respectively. The ITD and ILD are proposed by Lord Rayleigh (1907) to be the localization cues that characterize the spatial position of a sound source in the horizontal plane. However, there is no unique relation between the binaural cues and the position of a sound source in space because a whole cone of source positions can

be specified for which the ILD and ITD are almost constant (see (Woodworth, 1954) for a description of the 'cone of confusions'). In the 70^{th} the role of the pinnae (the outer ear) in sound localization began to emerge (see Blauert (1974) for a review). Shaw and Teranishi (1968) were able to show that the pinna cavities have a variety of resonance modes at characteristic frequencies. The amplitudes and the frequencies for which the resonances occur depend on the direction of sound incidence. Hence, the spectrum of the sound source is transformed by the resonances of the pinnae in a way that is characteristic for the source position of the sound. The spectral cue is denoted as 'monaural' since it is introduced independently at each ear. In addition to the binaural cues it represents the second group of spatial information. This 'spectral fingerprint' generated by the spectral transformation is different for sound incidence from each point on a 'cone of confusion' and, hence, monaural spectral cues are used to estimate the sound elevation as well as to decide if the sound is coming out of the frontal or rear hemisphere ((Hebrank and Wight, 1974; Butler and Belendiuk, 1977; Morimoto and Aokata, 1984; Asano, 1990)).

Since all spatial information that can be used by the auditory system to estimate the position of a sound in a non-reverberant environment is given by HRTFs, they provide the capability to simulate a free-field presentation of a sound. By presenting the signal convolved with head related impulses responses (HRIRs, that are the time domain representations of the HRTFs) of the left and right ear over headphones, a perception similar to a free-field condition can be achieved. This technique is called 'virtual acoustics' and allows to present externalized sound sources over headphones with an localization accuracy that is comparable to the acuity for real free-field presentations (e.g. (Wightman and Kistler, 1989a; Wightman and Kistler, 1989b; Hammershoi, 1995; Otten, 1997)). Virtual acoustics can be used to build computer controlled virtual auditory displays (VADs), that are capable of projecting sounds to any desired location in space, for instance, as a component of a virtual environment generator.

Two major problems emerge for VADs. First, the source positions could be distributed on a whole sphere of possible source locations and, hence, HRTFs are needed from a high number of source locations. Therefore, a measurement setup is needed that allows for flexible positioning of a physical sound source on a spherical surface. To reduce the measurement effort, fast and accurate positioning is required and the procedure to measure the HRTFs should introduces only small time delays.

Furthermore, it is not sufficient to measure a comprehensive set of head related transfer functions for only one selected listener. To achieve the same perceptual impression for each subject, individual HRTFs have to be used in VADs. If non-individual HRTFs are used to generate virtual sounds, the main problems that occur are an increased localization blur for the elevation perception and an increased occurrence of front-back confusions (i.e. the source position is perceived on a point on the appropriate cone of confusion that is opposite to the source location where the HRTFs were measured from.)

(Wenzel *et al.*, 1993). Both kinds of localization errors are introduced by deviations between the HRTF spectra of the listener and the HRTF spectra provided by the VAD. Because of the need for individual HRTFs, VADs are very costly to implement and, therefore, they are far away from being applicable for the common run of mankind. However, there are lots of potentialities for VADs to improve communication in our daily life, for instance for man-machine communication (especially for blind people) or for each application for which the distribution of information in a 3D space could be useful (for instance, telephone conferences or to improve communication in aircrafts). Thus, further research is needed to understand which aspects of individual HRTFs (providing the most basic localization cues) are needed for an accurate spatial perception.

This thesis deals with both the experimental needs for measuring HRTFs and the need for individual information in the localization cues to achieve an accurate perception of spatially localized objects.
Thus, the thesis is structured as follows: In Chapter 1 a mechanical setup is introduced (TASP, Two Arc Source Positioning) that allows for a rapid and accurate positioning of physical sound sources to almost any point on a spherical surface. The usability of the TASP system is investigated by free-field localization experiments and the results are validated by a comparison to data from the literature. The GELP technique (God's eye view Localization Pointing) introduced by Gilkey et al. (1995) is used to collect the subjective responses and it is investigated, furthermore, in which way the use of the GELP technique affects the recorded localization data.
In order to analyze inter-individual differences between HRTFs the TASP system is used to measure HRTFs from 11 subjects and one dummy head. This investigation is presented in Chapter 3 of this thesis. The HRTFs are described in terms of individual binaural and monaural localization cues and differences between HRTFs. For virtual acoustics, HRTFs are often realized as digital minimum phase finite impulse response (FIR) filters with smoothed spectra. The effects of spectral detail reduction on minimum phase HRTFs is investigated in the second section of Chapter 3.
While in Chapter 3 the investigation is focused on the effect of smoothing on the physical localization cues, in Chapter 4 the scope of the study is extended to perceptual consequences of HRTFs manipulations. By conducting discrimination experiments perceptual thresholds for spectral and temporal (variations of the ITD) manipulations of the individual physical localization cues are obtained to assess deviations of the individual localization cues that are not noticeable for human subjects.
In reverberant environments the direct sound emanated by the sound source is followed by reflections from objects surrounding the listener. The auditory system suppresses the spatial information in the reflections and estimates the source position mainly by means of the spatial information in direct sound. This effect is called 'precedence effect' because the auditory system gives precedence to the spatial information in the direct

sound. However, it can be assumed that the evaluation of the localization cues in the direct sound is influenced by reflections. To test this hypothesis it is investigated by discrimination experiments in Chapter 5 if the perception of changes in the localization cues of the direct sound differs under reverberant and non-reverberant conditions.
A common method to restrict the perceptual dimension in localization experiments to spatial cues is to rove the source spectrum ('spectral scrambling') before filtering the stimulus with HRTFs. Without this technique, subjects would be able to use the stimulus timbre as a cue. The scrambling procedure is also applied to stimuli in the parts of the experiments in Chapters 4 and 5. However, it could be that spectral scrambling introduces spatial information to the virtual stimuli that affects the localization of the stimuli. Thus, in Chapter 6 it is investigated by using an absolute localization paradigm, if spectral scrambling can vary the perceived stimulus positions.

Chapter 2

Influence of procedural factors on localization in the free-field using a two-arc loudspeaker system

Abstract

A computer controlled mechanical loudspeaker positioning system (TASP, two arc source positioning) is presented. It allows for continuous sampling of source positions in azimuth and elevation. To validate the system, free-field localization measurements in the horizontal plane ($\phi = 0° - 180°$, $\Delta\phi = 15°$) and in the median plane ($\theta = -40,° -60°, \Delta\theta = 20°$) were conducted. The stimulus was a 300 ms click train. A comparison to localization measurements from the literature revealed that consistent results are achieved even though the setup presented here deviates in several aspects from those described in the literature. However, to capture the improved localization performance for frontal sound incidence a head monitoring technique to center the head seems to be necessary. The GELP technique (Gilkey et al., 1995) was used to collect the localization data. To validate the use of the GELP technique in a darkened room the free-field localization performance was compared to data obtained from three control experiments with non-acoustical localization tasks. In the first control experiment, numerical values of the target azimuth and elevation were presented. In the second and third experiment, visual stimuli were presented in a lighted or darkened room. A comparison of the control experiments with the acoustical free-field localization experiment showed that the localization accuracy in the free-field setup employed here is not restricted by using the GELP technique in a darkened room.

2.1 Introduction

Study of localization ability has gained considerable interest in recent years (e. g. (Oldfield and Parker, 1984a; Makous and Middlebrooks, 1990; Good and Gilkey, 1996; Lorenzi *et al.*, 1999)), even though a variety of studies in this area have been conducted since the beginning of the 20^{th} century (see Blauert 1974 for a review). For measuring the localization ability in an anechoic chamber (free-field condition), either a fixed array of speakers has been employed or one or two speakers that can mechanically be positioned at certain locations. However, the mechanical setups for positioning the sound sources were not able to cover the whole range of spatially relevant source positions with high resolution. Since this has not yet been achieved in a satisfactory way, this contribution presents and evaluates a new setup that overcomes some of the restrictions of the systems known from the literature.

Different approaches for positioning a sound source on a virtual spherical surface of source locations with high resolution were used in the recent literature. Gilkey et al. (1995) used a static sphere of 272 loudspeakers evenly distributed on a surface of a sphere with a diameter of about 4.3 m. This construction allows for a rapid collection of localization data because there is no need to move a sound source between stimulus intervals. The main disadvantages are the fixed resolution of possible source locations and the considerable amount of reflecting surfaces of the metal construction and the speakers itself. Another possibility is to use only one arc of speakers that is rotating around a fixed axis (e.g (Wightman and Kistler, 1989a; Makous and Middlebrooks, 1990)). This concept reduces the reflecting surface by a considerable amount compared to the localization dome and increases the maximal resolution in at least one dimension (azimuth or elevation). If the rotation axis lies within the interaural axis, the construction is optimal for a double pole system of coordinates (Morimoto and Aokata, 1984), whereas a vertical rotation axis through the center of the head prefers a single pole system of coordinates. However, in both cases the resolution is restricted by the fixed location of the speakers either in elevation or in azimuth. A disadvantage of this approach compared to a fixed array of speakers lies in the time delay needed for a rotation of the arc. A setup similar to the one presented here was realized by Bronkhorst (1995). The subject is seated in the center of a rotatable arc (diameter 1.4 m). The rotation axis coincides with the interaural axis. However, the leverage of the arc at 0° elevations makes it difficult to control and effectively limits the diameter of the arc.

The measurement setup introduced in the current paper is termed TASP: Two Arc Source Positioning system. It is capable of positioning one of two speakers at nearly every point on a spherical surface with a diameter of approx. 4 m. The TASP system consists of two rotating hemi-arcs, with a vertical rotation axis going through the center of the head of the subject. The angle of azimuth of the sound source is adjusted by a

rotation of the arcs. Two sledges moving along the arcs position the elevation of the sledges (see Figure 2.1). The usability of the TASP system is verified by conducting a free-field localization experiment in the horizontal and median plane. The results of this experiment are compared to data from the literature (Section 2.3).

A prerequisite for the collection of localization data is that the subject can transform the subjective acoustical perception into an objective recordable variable. In the literature, a variety of different methods was used to collect localization data. For instance, Wightman and Kistler (1989b) asked their subjects to make a verbal report of the source location in terms of azimuth and elevation angle. The subjects had to train the report intensively before data collection began. In a study of Makous and Middlebrooks (1990) the subjects had to point to the stimulus location with their nose. The data was collected by monitoring the head orientation. Oldfield and Parker (1984a) used a pistol-like input device and asked the subject to 'shoot' the stimulus position. Langendijk used a virtual acoustic pointer controlled by a joystick-like input device (Langendijk and Bronkhorst, 1997). It was shown that the virtual pointer technique is more accurate than the verbal report technique. The GELP (God's Eye View Localization Pointing technique [1](Gilkey *et al.*, 1995)) uses a little globe in front of the subject that represents the sphere of stimulus locations surrounding the subject. The subject has to point to the location on the sphere that corresponds to the stimulus location on the sphere of possible source locations (see Section 2.4.1 for a detailed description). The study of Gilkey et al. showed that the input accuracy was as accurate as for the verbal report technique but not as accurate as the technique used by Makous and Middelbrooks. However, data collection was much faster by using the GELP technique (16-20 trials per minute) compared to the verbal report (2-3 trials per minute) and the 'nose pointing' (3-4 trials per minute) technique.

In the current study the GELP technique is used to record the subjective localization data because it is easy to implement and allows for a rapid collection of data. Furthermore, it turned out that the subjects did not need any training.

In the experiments conducted by Gilkey et al. to validate the GELP technique, the subjects were able to see the surface of the GELP globe. In contrast, the localization experiments presented in this study had to be performed in a darkened room to prevent the subject from seeing the moving parts of the TASP system. Therefore, the subject has to use his/her tactile instead of visual sense to point to the correct input location. It can be assumed that the capability of the subject to handle the GELP technique is reduced if only the tactile sense can be used.

Consequently, two different experiments are described in this paper to validate the usability of the TASP system in combination with the GELP technique. The suitability of the TASP system for serving as a positioning system in localization experiments is

[1]The GELP technique is similar to a technique developed by Blauert et al., called 'Bochumer Kugel', (Blauert, 1998)

examined in the first section of this study (Section 2.3). Since an investigation of the localization ability for each possible location on a sphere is beyond the scope of this study, the source locations were distributed only in the horizontal and median plane to reduce the overall measurement time and the size of the data set.
In the second section (Section 2.4), it is investigated if the GELP technique can also be used in a darkened room where the subjects were not able to see the spherical surface of the GELP technique. Three control experiments were conducted in which non-acoustical stimuli were presented to the subjects. In the first experiment, stimulus locations were presented numerically on a screen in terms of azimuth and elevation ('numeric' condition). In a second experiment the subject had to estimate the position of one of the TASP sledges in a lighted room (visual I). The third measurement was conducted in the darkened anechoic chamber. A little diode fixed in the center of the loudspeaker served as a target (visual II). The general assumption is that if the input performance (i.e. the capability of the subjects to point to the desired locations on the spherical surface in the non-acoustical stimulus conditions) is higher in the control conditions than in the free-field localization experiment, the localization accuracy is not restricted by using the GELP technique in a darkened room.

2.2 Technical description of TASP

Figure 2.1: The TASP (Two Arc Source Positioning) system inside the anechoic room.

The apparatus presented here was constructed under the constrains of maximum resolution in azimuth and elevation and as little positioning time delay and amount of surface reflections as possible. Furthermore, the setup can only be installed temporarily in the

anechoic room so that the mechanical installation procedure is required to be as short as possible. Consequently, the construction was chosen to consist of a fixed part at the ceiling and a removable part being attached to it. The mechanical installation procedure of the removable part takes about two and a half hours.
Figure 2.1 presents a photograph of the TASP (Two Arc Source Positioning) system within the anechoic room of the University of Oldenburg. Figure 2.2 depicts a scheme of the main functional parts. The removable part consists of two opposed hemi-arcs with a moveable loudspeaker sled attached to it. A sound source is positioned to the desired location by dragging the sledge into the correct elevation and turning the arc to the desired azimuth. The two opposed arcs divide the sphere of possible source locations into two hemispheres. Hence, the frontal and the rear hemisphere are covered by the two hemi-arcs.
The dimensions of the Oldenburg anechoic room hosting the TASP system (Figure 2.1) is 8,5m x 5m x 4m (width, depth, height) with a 1.3 m absorber depth and a lower cutoff frequency of 50 Hz. The TASP system itself is mounted at the ceiling by a double cross consisting of four iron double T profiles (1). A metal plate in the center of the double cross carries the main rotational axis (2) and the stepping motor (3, Positec VRDM31122) is responsible for the azimuthal rotation. The rotating system itself (4) is constructed as an open circle with a dihedral angle of 90°.

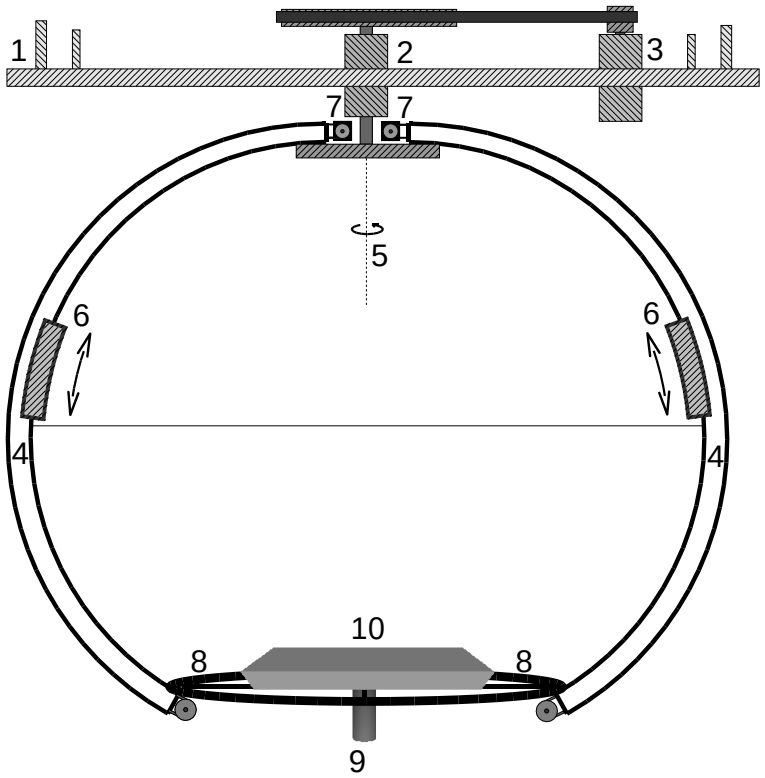


Figure 2.2: Scheme of the TASP system. See text for a detailed description of the numbered parts.

The rotation axis (5) corresponds to the axis of symmetry of the open circle. Two little sledges (6) using the inner part of the double T profile of the arc as tracks, serve as transports for the sound sources. Two stepping motors (7, Positec VRDM 3913 LWC), one for each sledge, are mounted directly at the rotation axis of the arc. They allow for an independent movement of the sledges on both hemi-arcs. A toothed belt is affixed to the sledge. Driven by the gear wheel of the respective stepping motor, the sledge is dragged into the desired direction. In this way the elevation of the sound source is adjusted. To prevent the hemi-arcs from oscillating around the rotation axis, their lower ends are connected via a metal ring (8) which is pivoted at its center point by a solid cylinder (9). This cylinder also serves as a stand for the platform (10) which carries the subjects chair or the dummy head to be positioned in the center of the sphere.
Figure 2.3 depicts the connection of the controlling software to the stepping motors. An IBM compatible 486 PC controlled by the WinShell[2] command line is connected via the serial port to a programmable stepping motor control device (Positec WPM 311). Two power devices (Positec WD3-004 and WD3-008) drive the stepping motors for positioning of the source.

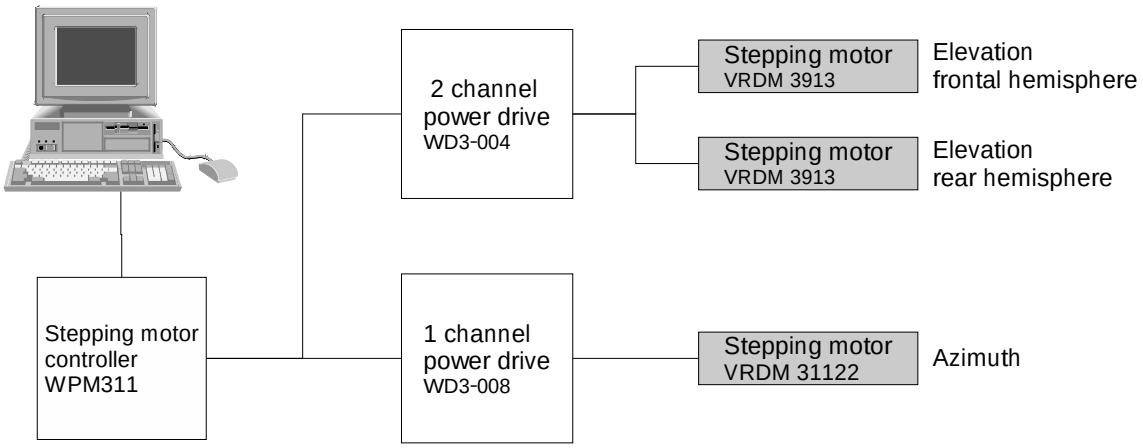


Figure 2.3: Controlling of the stepping motors.

2.2.0.1 Performance

The performance of the TASP system can be described by a) the time delay between the presentation of stimuli at different source locations b) the overall range of possible source locations c) the maximum resolution in azimuth and elevation d) the amount of reflecting surface disturbing measurements in anechoic conditions e) cues of the source position generated by TASP system itself.
With respect to the positioning time delay it turned out that the positioning in azimuth

[2]The WinShell is a command line experiment control system, developed by members of the work group 'Medizinische Physik' at the Universität Oldenburg which is capable of linking libraries providing control commands for hardware devices.

is much more critical than the movement in elevation. A non-continuous alteration of the rotation velocity causes the arc to oscillate around its axis of rotation. Therefore, onset and offset ramps have to be used to allow for a smooth movement and to limit the angular momentum to be applied. These ramps slow down the process of positioning the arc to the correct azimuth. To prevent the subject from using the delay as a cue for the relative distance between subsequent stimuli, a variable angular velocity of the rotation was introduced in a way that the delay is nearly independent of the relative distance of two successive stimuli. Hence, a fixed time delay can be specified with 6 s for the azimuth positioning and 2 s for the elevation positioning.

The mechanical constrains do not restrict the range of azimuth but limit the elevation to the range between $-40°$ and $+80°$. This should be sufficient for all kinds of investigations that are related to directional hearing. The minimal distance between two source positions is nearly arbitrarily small. It was limited by software to one degree in azimuth and elevation.

Although the reflecting surface of the TASP technique is quite small compared to other setups (like a localization dome, for instance) the environment of the subject is not without reflections. The sound wave generated by the speaker of one hemi-arc is reflected by the opposite hemi-arc and its speaker as well as the whole construction under the ceiling that carries the rotating part.

The rotation of the hemi-arc around the z-axis provides no hint to the speaker location because the driving motor is mounted overhead and the movement of the hemi-arc in the air is very silent. However, the sliding of the sledges along the arc is not noiseless. If one sledge is moved to a certain elevation, the toothed belt driving the sledge grates along the surface of the arc. The originating noise is not correlated to the speaker elevation but allows the subject to identify the azimuth of the arc. Hence, in measurement conditions were stimuli positions are distributed in azimuth and elevation, the noise from the sledge movement has to be masked by an external sound source. Another possibility would be to first position the elevation and afterwards the azimuth.

The localization measurements described in this study used only movements with a fixed azimuth or elevation providing no mechanical localization cue. Therefore, no masking noise was needed.

2.3 Free-field localization

2.3.1 Method

2.3.1.1 Subjects

Eight normal hearing subjects, six male and two female aged from 27 to 34 participated voluntarily in the free-field localization task. All subjects were members of the faculty and had extensive experience in psychoacoustic tasks but none of them was involved in localization experiments before. However, subject 'JO' is one of the authors.

2.3.1.2 Stimuli

The stimuli used for the presentation were click trains with a duration of 300 ms presented at a level of approx 60 dB(A). Clicks were repeated at a rate of 100 Hz. The onsets and offsets were gated by 25ms squared cosine ramps. The stimuli were equalized by the transfer function of the speaker within the frequency range of 100Hz to 14 kHz. After positioning the speaker to the desired position the stimulus was presented only once. The subject had no limitation in time to convey the perceived source location to the computer using the GELP technique (s. Section 2.4 for a comprehensive description of the GELP technique implementation).

2.3.1.3 Procedure

The localization performance in the horizontal plane and in the median sattigal plane was measured in two separate sessions. The measurements were conducted in the darkened anechoic room using the TASP technique for positioning the sound source. The subject was seated on a chair and adjusted in height so that the interaural axis lies within the horizontal plane. The head was not fixed by a chin rest or an equivalent method. Instead, the subject was told to focus the straight forward position (where the speaker was located during the instruction to the subject when the light were still on) and to re-establish this position after the input of the localization perception to the GELP technique. Before the beginning of each measurement the room was darkened and the speakers were moved at random three times without emanating a stimulus. Because the movement of the speaker in the horizontal plane does not give any cue for the detection of the speaker location in the darkened room, the subjects reliably lost the speaker location after the threefold positioning. Three seconds after the last movement the stimulus was presented. After recording the localization data by the computer, a 200 ms gated sine wave was presented from a speaker mounted under the subject's chair platform to acknowledge the recording. This signal was normally localized inside the

head and should not influence the localization task.
Each subject conducted two sessions. In the first session the source location was randomly chosen from 24 positions in the horizontal plane at 0° elevation (15° spacing). The subjects were not informed about the discrete distribution of the possible stimulus locations. Each position was measured three times resulting in 72 trials per session.
In a second session, source elevations in the median plane were distributed from −30° to +60° in the frontal and rear half-plane with a constant distance between locations of 10 degrees. The measurement routine was the same as the previous one except for the different source locations.
Data collection began with the first presentation of the stimulus. Thus, the subject were untrained and had only experience in using the GELP technique by participating in the validating experiments described in Section 2.4, which were conducted with each subject before the free-field localization measurements.

2.3.1.4 Localization data analysis

To compare the outcome of the free-field localization experiment to data from the literature, the analytical methods used by Wightman and Kistler (1989b) and Gilkey et al. (1995) were adapted.
The judgement centroid describes the mean judgment of subjects response to a stimulus from one certain location. It is calculated by summing up the normalized vectors from the center of the GELP sphere to the position on the surface indicated by the subject.
The mean absolute error is calculated by computing the absolute difference between the target and the judgement angle either in azimuth or elevation. The angle of error is computed for each response individually and then averaged across source locations and subjects.
The spread of responses for one target location is described by the parameter κ^{-1}. The concept of κ^{-1} was adapted from the statistics of spherical distributions and is similar to the standard deviation (Fisher *et al.*, 1987). A detailed description how to calculate κ^{-1} for localization data is given by (Wightman and Kistler, 1989b) and (Gilkey *et al.*, 1995).
A special problem in the investigation of localization data is the appearance of front to back reversals (e.g. (Makous and Middlebrooks, 1990; Wightman and Kistler, 1989b; Gilkey *et al.*, 1995)). The binaural localization cues are only capable to determine the source position on a 'cone of confusion' (Woodworth, 1954) for which the binaural parameters are constant. The position within each cone of confusions is resolved by utilizing monaural spectral cues. Hence, applying the former methods to the raw localization data would result in large azimuth errors which do not reflect the binaural localization accuracy. One way to resolve the confusion is to mirror the judgement to the (front-back) hemisphere where the distance between target location and judged location is smallest.

This concept can introduce errors for target locations near 90° of azimuth because it is possible that judgments are 'resolved' which were genuine localization errors. It is assumed that the number of errors introduced is small compared to the benefit of the mirroring procedure which avoids an overestimation of the errors due to front-back confusions.
In addition, a linear regression function was calculated for the localization data and the correlation coefficient between the presented and the judged locations was computed.

2.3.2 Results

Azimuth

The results for eight subjects participating in the localization experiments are shown in Figure 2.4. The judgment centroids are plotted as a function of the target angles in azimuth. The dotted line marks the ideal performance of correct responses. A linear regression function is plotted as a solid line in each diagram. To provide more information on the spread of data, the centroids are stretched along the judgement dimension, if κ^{-1} for the actual angle is greater than the mean value of κ^{-1} averaged across all azimuthal angles for this subject. In this case, the stretching is proportional to κ^{-1}. If κ^{-1} is less than the mean, the diameter of the centroid is set to a lower limit. Therefore, the ellipses mark an increased variability of the judgements relative to the mean spread of data. The centers of the ellipses still coincide with the original centroid.
In each sub-plot of Figure 2.4 additional information on the inter-individual differences in localization performance is provided. The bars in the lower right quadrant of each sub-plot give the individual performance normalized by the mean performance averaged across subjects. The dark gray bar in the left half of the surrounding box shows the mean error angle and the light gray bar on the right side reflects κ^{-1} for each individual subject. The bar heights were calculated by the same general procedure for the mean angle of error and κ^{-1}. The top of the surrounding box is the maximum value across all subjects and the dotted vertical line represents the mean value across subjects. In this way the diagram shows the individual performance expressed by the individual mean angle of error and κ^{-1} relative to the mean performance across subjects.
Although the localization performance is quite high, the subjects show the same pattern in those localization errors that still occur. The localization acuity is near optimum for frontal (0°) and rear (±180°) sound source incidence. If the source is positioned at more lateral angles ($\varphi < 90°$), the subjects tend to project the source to the side. However, the effect is small for subjects MK and JO. The localization uncertainty marked by κ^{-1} indicates that stimuli coming from angles between ±130° and ±180° are more difficult to localize than sounds in the frontal hemisphere. Again, this effect is quite small and not shown by each subject.

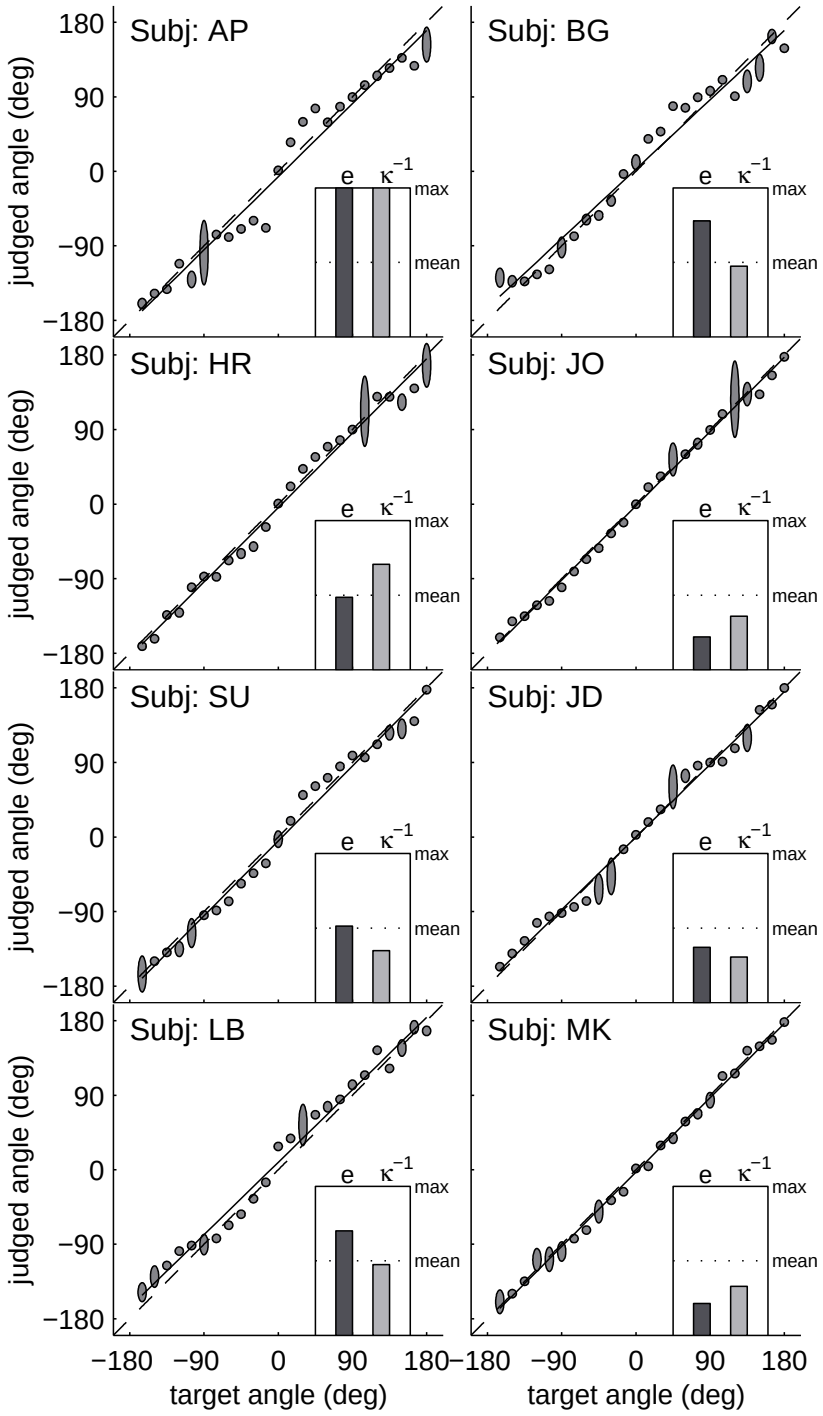


Figure 2.4: Extended centroid diagrams of the localization performance in azimuth. The stretched centroids identify a spread of the data that is higher than the mean across all locations for that subject. The bar diagrams in each plot represent the inter-individual differences in localization performance. The left, dark gray bar shows the mean absolute error $\bar{e}$ for the presented subject relative to the mean across all subjects (dotted horizontal line). The top of the box represents the maximum values across subjects. On the right side the light gray bar represents the same for κ^{-1}.

Subject	m	b	r	$\bar{e}$	κ^{-1}	F/B [%]
AP	0.981	-6.15	0.982	19.69(13.37)	0.065(0.036)	23
BG	0.931	2.84	0.985	16.11(13.05)	0.023(0.017)	17
HR	0.993	-4.05	0.994	11.38(10.78)	0.041(0.013)	7
JO	0.996	-2.11	0.998	07.08(06.42)	0.015(0.004)	0
SU	1.000	-4.71	0.994	11.83(11.34)	0.014(0.009)	7
JD	0.974	-0.05	0.995	09.52(09.42)	0.011(0.004)	6
LB	0.973	9.01	0.992	14.85(12.67)	0.024(0.011)	7
MK	0.998	-2.83	0.998	06.99(07.92)	0.012(0.007)	4
Ø	0.980	-1.00	0.992	12.18(10.62)	0.026(0.013)	9

Table 2.1: Results from the localization measurement in azimuth. Listed are the slope m and intercept b of the linear regression function, the correlation coefficient r, the mean absolute angle of error $\bar{e}$ (median values are shown in parenthesis), κ^{-1} and the percentage of front-back confusions.

The bar diagrams show that the angle of error and the spread of the input (κ^{-1}) are positively correlated ($r = 0.74$). This indicates, that under the present conditions the absolute localization uncertainty is well described by one of the measures. Subject 'AP' shows the poorest localization performance with the greatest angle of error and κ^{-1}.
It should be noted that subject 'JO', being one of the authors, shows a better than normal performance. It is likely that the lower errors are due to the a priori knowledge of the author that the source positions are discretely distributed in azimuth. This would allow for a substantial decrease in localization error because the absolute localization task changes to a identification task across different locations. However, the accuracy is still restricted by the accuracy of the GELP technique (see Section 2.4).
Table 2.1 summarizes the quantitative parameters of the localization results. Presented are the slope and intercept of the linear regression function (m, b), the correlation coefficient r between the target and judged angle, the mean absolute error $\bar{e}$, κ^{-1} and the number of front-back confusions in percent. Mean values across all subjects are presented in the last row of the table. These values will be used to compare the result of the current study to data from the literature.
Figure 2.5 shows the mean absolute error (solid line) and the signed error angle (dash-dotted line) averaged across subjects as a function of the source azimuth. The absolute error is a measure of the general localization uncertainty and the signed error reflects the bias to a certain direction. The absolute error varies slightly around the average of 12.3° with minima at 0° and ±90°. A prominent maximum can be seen at 45°. The positive values of the signed error for azimuthal source positions less than 90° indicate that the subjects tend to overestimate the angle in the frontal hemisphere. The opposite is true for the rear hemisphere. The negative values for target angles greater then 90° show an

underestimation of the azimuth position. It can be concluded, that subjects have a bias towards the more extreme lateral positions.

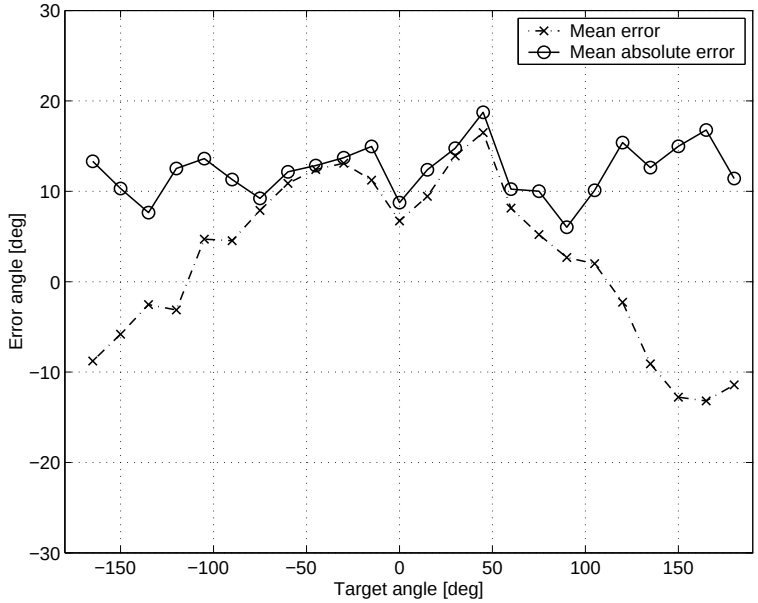


Figure 2.5: Mean absolute error (solid line) and mean error (dashed-dotted line), averaged across subjects, plotted as a function of azimuth.

Elevation

Figure 2.6 shows the localization data for source positions in the median plane. The spread of the distribution for each elevation is marked by stretching the centroid proportional to κ^{-1}. Data for source positions in the frontal hemisphere (0° azimuth) are plotted as light gray centroids and the corresponding centroids in the rear hemisphere (180° azimuth) are dark grey. Linear regression functions have been computed independently for the two hemispheres and are plotted within each subplot (frontal hemisphere: solid line, rear hemisphere dash-dotted line). The bar diagram is calculated in the same way as for the azimuth condition. However, the values were separately calculated for frontal and rear sound incidence and then averaged across hemispheres.
In contrast to the azimuth condition, localization performance varies considerably across subjects. In general, elevations greater than 20° are overestimated. This effect is more prominent for rear sound incidence. Targets at elevations lower than 0° are well localized by nearly all subjects. Only subject SU shows greater deviations from the target angles in this situation. Subject AP shows a large localization uncertainty with high errors and a wide spread of data. Higher elevations are strongly overestimated and only frontal locations near the horizontal plane are correctly localized. The subject stated that she had a high uncertainty on the stimulus position and felt like she was only guessing most locations. The data from subject BG for rear elevations is also remarkable.

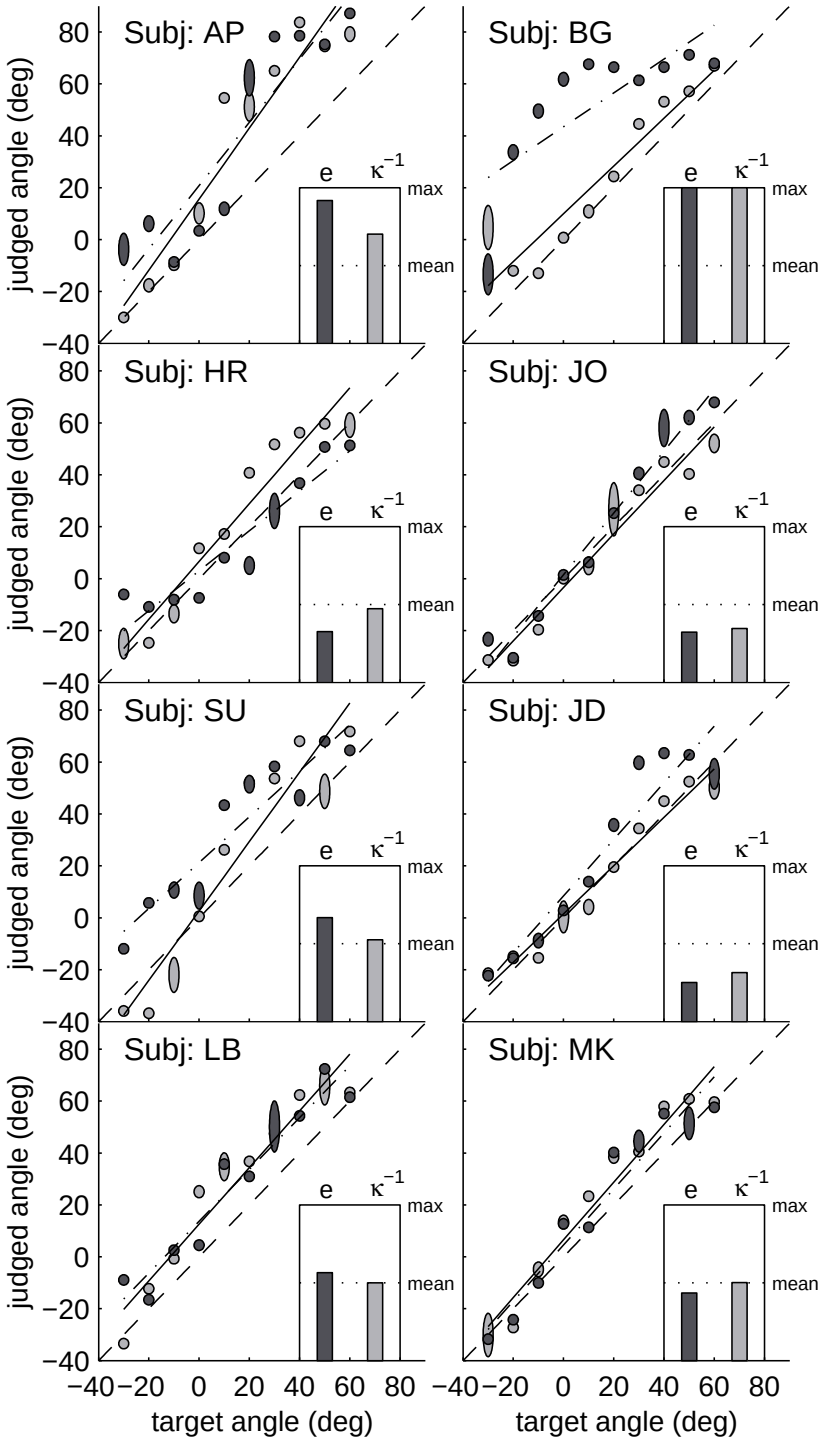


Figure 2.6: Localization performance for source locations in the median plane. The light gray centroids represent source positions for frontal sound incidence and the dark gray source positions for rear sound incidence. Regression functions are calculated for both hemispheres separately (solid line: frontal hemisphere, dash-dotted line: rear hemisphere). The bar diagrams show the individual localization ability relative to the mean across all subjects.

Although the perception of the stimulus location is very accurate for frontal sound incidence, the rear elevations are highly overestimated. However, the judged elevation is limited to 60° elevation. The subject reported that she knew the limitation of target locations to 60° and, therefore, did not judge higher elevations. If she had not known this, she would have judged higher elevations resulting in a more linear behavior of the localization data at higher elevations.
The results from a quantitative investigation of the localization data are summarized in Table 2.2. Each parameter was computed independently for the frontal and rear hemisphere. Mean values across subjects are presented in the last row of that table. It can be seen from the data, that the localization accuracy in the rear hemisphere is reduced compared to the frontal hemisphere. The inter-individual differences in localization performance represented by the bar diagrams in Figure 2.6 in each sub-plot are similar to the diagrams for the horizontal plane. Subjects AP and BG show a poorer localization performance than the mean and subject JO (one of the authors) a better than normal. However, the localization accuracy of the subjects JD, MK, HR is comparable to JO 's data, indicating that the a priori knowledge of the possible source positions and their discrete distribution is not as important as in the azimuth condition.
The mean absolute error (dark gray bar) and κ^{-1} are highly correlated (r=0.92).

2.3.3 Comparison with data from the literature

The results of the localization experiments described in the former sections are compared to data from the literature in Table 2.3 [3,4]. The derived parameters 'm' (slope of regression line), 'b' (intercept of y-axis), 'r' (correlation between target and judged locations) that are listed in Table 2.3, are separately computed for azimuth and elevation, whereas '$\bar{e}$' (mean absolute error), 'κ^{-1}' (spread of judgement) and 'fb' (number of front-back confusions in percent) are averaged across both dimensions.

[3]A brief description of the free-field localization experiments that were used for a comparison of the localization accuracy is given in Appendix A.

[4]The data from the literature was obtained as follows: If data was given for each subject, the mean across subjects was computed. The row 'Gilkey' represents data of experiment I in (Gilkey *et al.*, 1995). The next row 'Gilkey (W & K) ' shows data from Wightman & Kistler (1989a), subject SDO and SDE re-analyzed by Gilkey et al.. The row 'W & K' represents native data from (Wightman and Kistler, 1989a) taken from their Table II (correlation and reversals) and Table III (mean angle of error and κ^{-1} averaged across 0° and 18° for source positions in the azimuth and across all elevations in the frontal and rear quadrant for the comparison in the elevation). The last row shows the mean values from the current study taken from Tables 2.1 and 2.2. The average angle of error and the mean of κ^{-1} were computed differently by Gilkey and Wightman & Kistler. In the former study median values were calculated, whereas the latter presented mean values. To account for these deviations, both median and mean values were computed for these parameters in the current study. Median values are listed in parenthesis.

Subject	m_f	m_r	b_f	b_r	r_f	r_r	$\bar{e}_f[°]$	$\bar{e}_r[°]$	κ_f^{-1}	κ_r^{-1}	f/b [%]
AP	1.37	1.22	15.6	20.8	0.951	0.921	22.5(21.86)	27.6(27.01)	0.034(0.027)	0.153(0.036)	23.3
BG	0.92	0.65	9.93	43.49	0.938	0.761	16.24(14.29)	38.35(38.93)	0.094(0.070)	0.046(0.029)	08.3
HR	1.11	0.77	6.58	3.02	0.967	0.943	11.51(08.76)	10.96(9.30)	0.012(0.006)	0.036(0.029)	11.7
JO	1.03	1.19	-3.41	-1.46	0.978	0.985	07.24(09.35)	08.65(08.58)	0.012(0.003)	0.018(0.017)	00.0
SU	1.33	0.88	2.62	21.23	0.946	0.939	16.75(16.47)	20.89(19.43)	0.021(0.010)	0.041(0.032)	18.3
JD	0.93	1.09	1.47	8.35	0.982	0.955	05.72(05.30)	11.92(08.83)	0.008(0.005)	0.013(0.012)	13.3
LB	1.09	1.00	12.52	13.69	0.964	0.962	16.67(17.08)	15.70(13.44)	0.045(0.018)	0.033(0.021)	10.0
MK	1.11	1.08	6.51	4.48	0.973	0.967	11.73(12.09)	11.46(11.61)	0.020(0.014)	0.031(0.025)	00.0
Ø	1.11	0.95	6.55	14.25	0.962	0.929	13.55(13.15)	18.19(17.14)	0.030(0.019)	0.046(0.025)	10.0

Table 2.2: Localization performance in the median plane. The analyzing parameters are the same as in Table 2.1 but separately calculated for the frontal and rear hemisphere. Values in parenthesis are median values. All other values are mean values across source locations. The indices indicate if the parameters were calculated for frontal (f) or rear (r) sound incidence.

Paper	m_a	m_e	b_a	b_e	r_a	r_e	$\bar{e}$	κ^{-1}	fb
Gilkey	0.97	0.703	-2.47	6.87	0.996	0.889	(18.2)	(0.035)	
Gilkey (W&K)	1.01	0.77	0.85	8.45	0.995	0.829	(20.95)	(0.047)	
W & K					0.982	0.903	21.04	0.052	6
Otten	0.98	1.03	-1.00	10.4	0.992	0.945	14.64(13.64)	0.034(0.019)	9.5

Table 2.3: Comparison of parameters from this study with data from the literature (slope m, y-axis intersection b, correlation coefficient r, mean absolute error $\bar{e}$, mean spread κ^{-1}) and front-back confusions in percent fb. Indices denote those values that are computed separately for azimuth and elevation. Values in parenthesis are median values[4].

A detailed comparison of the values listed in Table 2.3 is not appropriate because of differences in the methods between studies and the inter-individual differences between subjects. However, it is obvious that similar results are obtained in the current study compared to the data from the literature. Although the subjects of this study were completely untrained, there is a tendency for higher localization accuracy in the current study, marked by a comparatively low mean angle of error $\bar{e}$ and spread of the data as indicated by κ^{-1}. This can be related to the restricted range of source positions in the current study because the highest localization uncertainty occurs at higher elevations for rear sound incidence (e.g. (Oldfield and Parker, 1984a)). In the current study only few source positions are located in this region. As expected, the localization uncertainty is increased in this region but the mean localization performance is dominated by the higher accuracy at the remaining source positions. Furthermore, in the study of Wightman and Kistler scrambled white noise stimuli were used to prevent the subject from using monaural cues, whereas unscrambled stimuli were used in the current study. Hence, the subjects were also able to use monaural spectral cues for estimation of the source location. It is likely that the acuity is increased by additional spatial information provided by monaural cues.

A comparison across studies of the mean absolute localization error in azimuth is given in Figure 2.7 [5]. The overall shape of the error as a function of azimuth is very similar across studies. There is a trend towards smaller errors for frontal source positions in the data of the cited literature that can not be observed in the results of the current study. This could be caused by the lack of some kind of head fixation (e.g. a bite bar (Gilkey et al.) or an acoustical reference location (Makous and Middlebrooks)). Subjects were allowed to move their head between trials and had no reference point to re-establish the head orientation before the next stimulus was presented.

[5]The data for Gilkey et al. and Makous & Middlebrooks was obtained from Table 2 in (Gilkey *et al.*, 1995) by averaging across $\pm 5°$ elevation. The data from this study is a re-plot of the data from Figure 2.5 collapsed over the left/right hemispheres.

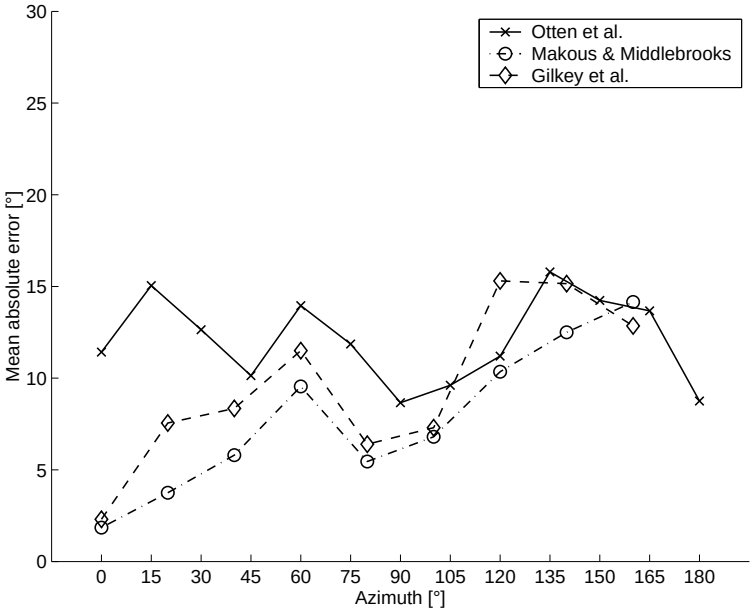


Figure 2.7: Comparison of the mean absolute error measured by Makous & Middlebrooks and Gilkey et al. with the current study.

Hence, the increased spatial resolution for frontal sound incidence could be concealed by changes of the orientation of the listeners head between stimulus trials.

2.4 Validation of the GELP technique

The experiments presented in this section were conducted to validate the different implementation of the GELP technique and its use in a darkened room.

2.4.1 Method

2.4.1.1 Implementation of the GELP technique

The general idea behind the GELP technique is that the spherical surface of possible source locations surrounding the subject is mapped to a globe with a much smaller diameter in front of the subject. This mapping is done by projecting the center of the subjects' head to the center of the sphere in front of the subject. The subject has to point to the corresponding point on the globe as if the subject was sitting inside.
The globe employed here has a diameter of 30 cm and consists of polystyrene. To facilitate the orientation on the sphere in a darkened room, the horizontal plane, the median plane and the planes with a constant elevation of $-30°$, $30°$ and $60°$ were carved into the surface. The sphere is placed on a wooden stand at height of 80 cm, which

makes it comfortable for the subject to reach any point on the sphere. The subject was seated on a chair that could be adjusted in height. To measure the position indicated by the subject, a Polhemus inside track pointer was used. The emitter (Model 3A06906) is mounted at the stand of the sphere and a normal receiver (Model 4A0332) was used to point to the source locations[6]. If the receiver had a distance greater than 1 cm from the surface of the sphere recording of data was not possible. The position of the pointer was recorded by the computer if the receiver was placed on the surface for one second. A short tone, transmitted by a loudspeaker mounted under the subjects' chair, acknowledged the recording of the data. The inside track was controlled by the computer that was also responsible for the movement of the TASP system.

2.4.1.2 Subjects

A total of 15 subjects participated voluntarily in the experiments. At least seven subjects participated in each experiment. The subjects were aged from 27 to 34 years and had normal hearing. All subjects were members of the physics and psychology department of the University of Oldenburg. Except for subject 'JO', none of the subjects received any training or had pre-knowledge about the measurements.

2.4.1.3 Procedure

Three control experiments where conducted in separated sessions. Two numerical values, representing azimuth and elevation coordinates, were displayed on a monitor screen in front of the subject ('numeric' condition). The task of the subject was to point to the corresponding location on the spherical surface of the GELP technique. In this experiment subjects were sitting in a normal reverberant room because no acoustical stimulus was presented. After recoding the response of the subject, feedback in terms of the judged azimuth and elevation angles was given. The stimuli locations were equally distributed in azimuth (15° spacing). However, for each azimuth a different angle of elevation was randomly chosen from $-30°$ to $60°$ in steps of $10°$. The positions were presented three times in random order. Note that only one randomly distributed elevation at each azimuth was chosen.
The general measurement procedure in the conditions 'visual I+II' was equivalent to the free-field localization measurement (see above). The task of the subject in the 'visual I' condition was to judge the location of one sledge of the TASP system in the lighted anechoic room. A different set of source positions but distributed in the same way as in the 'numeric' condition was chosen. To identify which of the two speakers of the TASP system was the target, a short click train was emitted.

[6]Although it has no nib as the Stylus (compare Gilkey et al. (1995)) we felt that direct contact with the surface of the sphere facilitates the use of the pointer.

To examine if subjects are able to handle the GELP technique in a darkened room, they had to judge the location of a little diode, mounted in the center of the speaker ('visual II'). It was not possible for the subjects to see the globe of the GELP technique. Hence, the subjects had to use their tactile sense to find the desired location on the spherical surface.

2.4.2 Results

In Figure 2.8 results from the control experiments and the free-field localization experiment are shown for two representative subjects. In the left column data for subject 'JO' is shown and in the right column the results for subject 'MK' are given for source positions in the horizontal plane. The centroids were stretched proportionally to κ^{-1}, if κ^{-1} for the current azimuth is greater than the mean across all azimuth positions for that subject. A linear regression function is plotted in each panel.
In the first row data obtained in the 'numeric' condition is plotted. The judgements are near to the optimum performance for every angle of azimuth. Both subjects are able to position the pointer of the GELP technique very accurately.
In the 'visual I' condition more spread can be seen in the response pattern. Furthermore, the centroids are slightly more distant from the optimum performance. This tendency remains for the two other experiments and the highest error can be seen in the free-field localization task. This qualitative description is quantified in Figure 2.9. Here, the mean absolute error (averaged across the left and right hemisphere) as a function of the stimulus azimuth is plotted for the four different experimental conditions. In addition, data from 'experiment II' from Gilkey et al. (1995), averaged across $\pm 5°$ elevation is shown (dashed lines). This experiment is very similar to the 'numeric' condition in the current study. It deviates only in the presentation of the azimuth and elevation angles, which were reported verbally to the subjects.
The absolute error is lowest in the 'numeric' experiment and highest for the acoustical free-field presentation. The input performance in the 'visual II' condition (darkened room) is substantially reduced in comparison to the 'visual I' condition (lighted room). Hence, the handling of the GELP technique in the darkened room seems to be complicated. However, the main constraint given in the introduction was that the localization error in the free-field localization experiment is higher in comparison to the error in the control experiments. A non-parametric ANOVA (Kruskal-Wallis) performed on the mean localization errors for the 'visual II' condition and the free-field localization experiment shows that mean localization error in the acoustical localization experiment is still higher ($p < 0.01$).

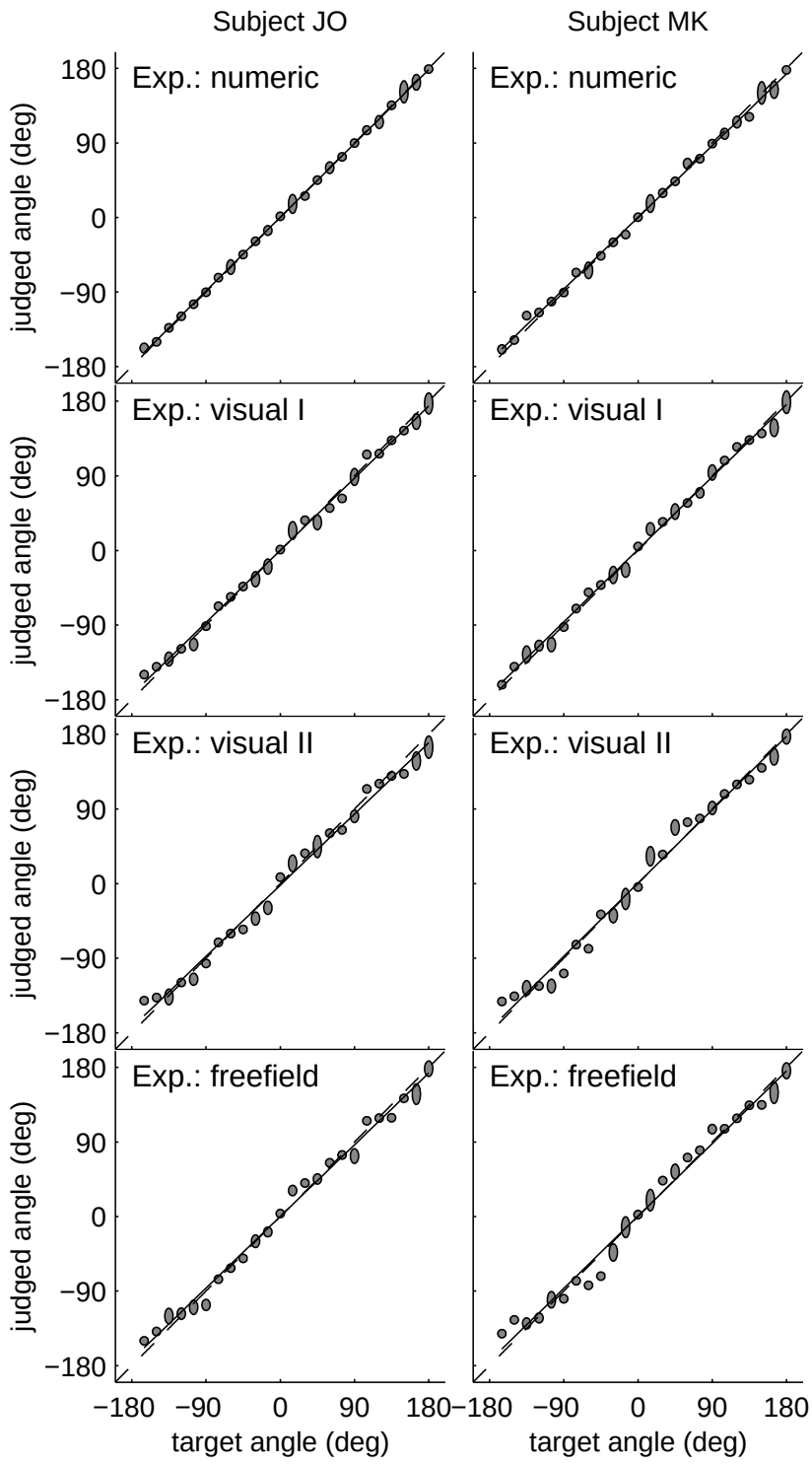


Figure 2.8: Validation of the GELP technique: Judged azimuthal angles for various conditions ('numeric', 'visual I': lighted room, 'visual II': darkened room, acoustical) as a function of the azimuthal target angle. Results for two representative subjects (left and right side) are shown for three control conditions (row 1-3) and for the free-field localization experiment.

Minima can be seen in the regions around 0°, 180° and 90° for each condition. These regions were marked by carves on the surface of the GELP sphere and this seems to ease the handling of the technique. A comparison of the results from the 'numeric' condition to the data from Gilkey et al. reveals that the performance is comparable in the frontal hemisphere. An increase of the mean absolute angle of error for increased azimuths can be observed for the data from Gilkey et al. This might be due to fixation of the subject's head by a bite bar that makes it more difficult to point to rear positions on the surface of the GELP sphere. The smaller error in the current study could also be caused by the greater size of the sphere (30 cm compared to 20 cm in the study of Gilkey et al.) because a small displacement of the pointer on the surface of the globe generates smaller errors if the diameter of the sphere is increased.

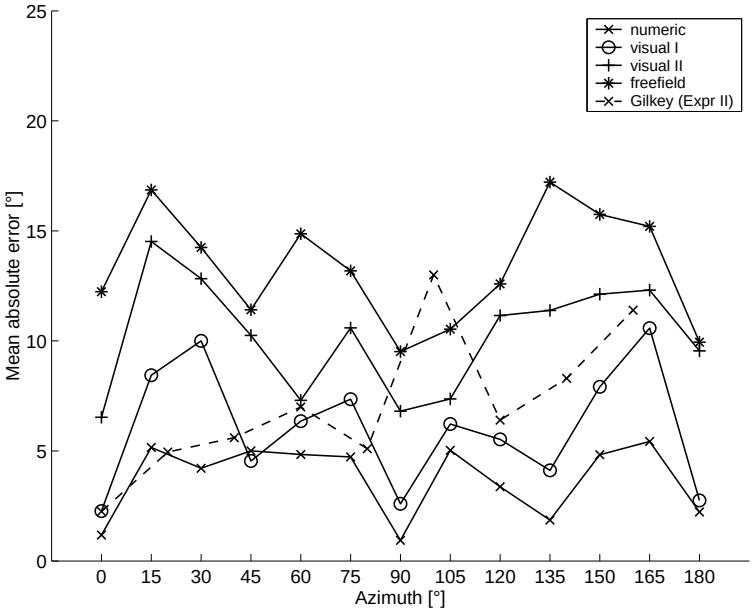


Figure 2.9: Mean absolute error averaged across subjects and left/right hemispheres under four conditions are shown as solid lines (see legend). The dashed line shows data of the verbal presentation experiment II from Gilkey et al. (1995).

2.5 Discussion

2.5.1 TASP and free-field localization

In the current study a method for positioning a sound source on a spherical surface was presented. The TASP system allows almost continuous sampling of the virtual sphere of source positions. The upper and lower limits of the elevation angle are $-40°$ and $+80°$ and the whole azimuth range is covered. The average time interval between two stimulus presentations is about approx. 6 s.

In a free-field localization experiment eight subjects were requested to localize a click train stimulus presented from positions out of the horizontal and median plane. The localization performance for positions in the horizontal plane were very accurate for most subjects. However, two subjects showed a considerably lower localization performance in azimuth as well as in elevation. Inter-individual differences in the localization performance have also been found in the literature. For instance, subject 'SDE' in the study from Wightman and Kistler (1989b) showed an accuracy that was considerably lower than the average. The lower localization performance was mainly found for judgements of the elevation and is also expressed by the number of front-back confusions. Wightman and Kistler related the lower localization performance to the physical spectral cues provided by the head related transfer functions (HRTFs). They showed that for subject 'SDE' less spectral information was contained in the spectral cues compared to other subjects. Hence, the decreased localization accuracy for subject 'SDE' might have been caused by less spatial information provided by the HRTFs. In the current study, the lower localization performance of the two subjects also occurred in the horizontal domain. It is unlikely that the HRTFs for the subjects with a decreased acuity provide less *binaural* information. Hence, it can be assumed that the low localization performance is not only caused by a lack of spatial information contained in the HRTFs but by a decreased utilization of the physical cues available to the subjects.

In the current study no method was used to center the head of the subject to the center of the hemi-arcs of the TASP system. Hence, it was possible that the position of the head was changed slightly between stimulus presentations. An analysis of the mean absolute error in azimuth showed that the error was higher for frontal sound incidence in comparison to studies in which a head fixation (Gilkey *et al.*, 1995) or a reference position given by an acoustical stimulus from 0° azimuth and elevation (Makous and Middlebrooks, 1990) was used. Therefore, to be able to measure the higher localization accuracy for frontal sound incidence, the head of the subject has to be centered to the middle of the sphere of possible source locations before each stimulus presentation. However, a fixation of the head reduces the flexibility of the subjects and a stimulus from a reference location could change the absolute localization task to a discrimination task for the reference location. Hence, it seems to be suitable to center the head by a head monitoring technique that gives verbal or visual instructions to the subject to center the head. Such a technique has been used, for instance, by Kulkarni and Colburn (1998) to center the head for measuring head related transfer functions.

A comparison of the mean localization performance (averaged across subjects and source positions) revealed that despite of the differences in the methods (reduced set of source positions, click train stimulus, recording technique and untrained subjects) the acuity is comparable across studies. It can be concluded that the use of the TASP system for positioning the sound source did not influence the localization data.

GELP in a darkened room

The GELP technique was used to collect the localization data in the free-field localization experiment. Although the technique was already validated by Gilkey et al., a re-examination was necessary because the ability of subjects to handle the sphere in a darkened room was unclear. Therefore, three different control experiments were conducted with non-acoustic spatial stimuli. In the first experiment the coordinates of the stimulus position were given in terms of azimuth and elevation angles to the subject. The subjects were able to point to the corresponding positions of the surface of the GELP sphere in the lighted room very accurately (mean angle of error: approx. 4°). In two further control experiments subjects' capability to handle the GELP technique in a lighted and a darkened room was investigated. In the 'visual I' condition subjects had to judge the position of a sledge of the TASP system (mean angle of error: approx. 6°) and in the 'visual II' condition a little diode in the center of the sledge served as a target in the darkened room (mean angle of error: approx. 9.5°). The differences between the mean angle of error in these two conditions can be related to two properties: First, the geometry of the anechoic room and the visual cue of any reference direction could be used by the subjects in the 'visual I' condition. The absence of this aid in the 'visual II' condition could complicate the allocation of source positions to positions on the GELP globe. Second, the subjects were not able to see the surface of the GELP sphere in the darkened room. This also seems to increase the input uncertainty. However, a comparison of the 'visual II' condition to the free-field experiment shows that the mean absolute error for the presentation of an acoustical stimulus in the free-field is still above the error obtained in the 'visual II' condition.

In order to validate the GELP technique, Gilkey et al. conducted an experiment which was similar to the 'numeric' condition in the current study. A comparison of the data from both experiments showed that the subjects in the current study were able to handle the technique with a higher accuracy. This can be related to the bigger size of the GELP sphere and the lack of a head fixation. Although the head fixation increases the localization accuracy for frontal sound incidence, it seems to reduce the input accuracy for positions on the rear surface of the GELP sphere. Therefore, it can be concluded that an adjustment of the head position by emanating a stimulus from a reference position (as used by Makous and Middlebrooks) or by monitoring the head position should increase the localization accuracy for frontal sound incidence. These techniques should be preferred because they do not reduce the flexibility of the subjects to handle the GELP technique.

The influence of using the GELP technique in the dark could be further investigated by conducting the 'numeric' condition in a darkened room and presenting the azimuth and elevation coordinates by a verbal report. A comparison of the input accuracy in the lighted and darkened room could directly show the error that is introduced by using

only the tactile sense for handling the GELP technique.
A main advantage of the GELP system is that it enables to collect localization data at a high rate. The handling of the GELP technique in the dark substantially lowers the collection rate. Gilkey et al. stated that they were able to measure 16-20 source positions per minute (by using a static loudspeaker array). This rate can not be achieved if the subject can only use the tactile sense for handling the GELP technique. Although the collection rate was not measured explicitly, it can be specified with 3-5 stimulus positions per minute for the measurement setup presented here.
However, the GELP technique seems to be a suitable method for collecting localization data. Its implementation is less expansive than the head monitoring technique (Makous and Middlebrooks, 1990), at the same time it is as accurate as the verbal report (Wightman and Kistler, 1989b) and even without any training a high accuracy can be accomplished by subjects.
In general, it can be concluded that the combination of the GELP technique with the TASP system is a suitable setup for measuring the localization accuracy. This method can be enhanced by using a head monitoring technique to re-establish the position of the subjects head before the localization stimulus is presented.

Chapter 3

Head related transfer functions and the effect of spectral smoothing on individual localization cues

Abstract

Head related transfer functions (HRTFs) were measured from 11 subjects and one dummy head with high resolution in azimuth and elevation. The head related impulses responses (HRIRs) were obtained at the blocked ear canal entrance by using maximum length sequences (MLS). Binaural and monaural localization cues are calculated from the HRTFs and presented for selected source positions. The inter-individual differences of the localization cues are investigated by their standard deviations across subjects as a function of azimuth and elevation. Furthermore, the individual HRTFs are compared to the HRTFs from the dummy head. The results show, that both the binaural and monaural localization cues of the HRTFs strongly vary across subjects at low elevations and are less individual at high elevations. A comparison between the individual HRTFs and the dummy head HRTFs revealed, that the dummy head can not serve as an average listener, if spatially correct perception is needed. In order to reduce the amount of data required for an individual spatial auralization, the effect of cepstral and $1/N$ octave spectral smoothing is investigated on I) the inter-individual standard deviation of the spectra across subjects, II) the interaural level difference (ILD), III) the interaural time difference (ITD) and IV) the length of the HRIRs. $1/N$ octave smoothing introduces high ILD deviations to the smoothed HRTFs and is, therefore, not recommended for spectral HRTF smoothing. Cepstral smoothing with 16 coefficients, on the other hand, introduces only perceptually irrelevant changes to the binaural and monaural localization cues. Note, that this is only true if the ITD of the minimum phase HRTFs are computed from low-pass filtered impulses responses. An further advantage of cepstral smoothing is that is reduces the length

of the impulse responses more effectively than $1/N$ *octave smoothing.*

3.1 Introduction

The physical properties that are exploited by the auditory system to estimate the position of a sound source are captured by head related transfer functions (HRTFs). They described the directional dependent transformation of a sound from its source location to a point within the ear canal. The cues that are provided by HRTFs and which are characteristic for each source position can be divided in two groups. The binaural localization cues (interaural level difference, ILD and interaural time difference, ITD) are obtained from a comparison of the left and right ear HRTFs, whereas the spectral filtering of the source spectrum due to interferences effects and pinna filtering is introduced at each ear individually and is, therefore, called monaural cue (see (Blauert, 1974; Middlebrooks and Green, 1991) for comprehensive reviews of localization cues). If the HRTFs of the left and right ear for a certain source direction are known, they can be used to introduce the localization cues for that spatial direction to an arbitrary sound source by convolving it with the head related impulse responses (HRIRs), which are the corresponding time domain representations of HRTFs. Thus, a set of HRTFs, sampled from the whole spatial range of directional perception provides the possibility to project a sound source by headphones to any of the sampled locations. This technique is called virtual acoustics. It has been shown, that a virtual source presentation, based on individual measured HRTFs, is capable of producing an acoustical perception with an accuracy that is near to the free-field condition (Wightman and Kistler, 1989a; Wightman and Kistler, 1989b; Hammershoi, 1995; Otten, 1997; Kulkarni and Colburn, 1998).

Although HRTFs from different subjects have similar shapes in ITD, ILD and spectral filtering, the details of each cue are highly individual (e.g. (Møller *et al.*, 1995)). Therefore, it is not sufficient to use non-individualized HRTFs to yield a localization performance that is comparable to the free-field acuity (Wenzel *et al.*, 1993). However, it is a major effort to measure individual HRTFs for a number of source positions covering the whole range of spatial directions. The use of dummy heads that provide the localization cues of an average subject would facilitate the generation of virtual displays. Therefore, a comparison of dummy head HRTFs and individual HRTFs provides valuable information about the needs for creating dummy heads and appropriate auralization methods for virtual acoustic environments.

The aim of the first section in this investigation is to describe HRTFs measured from 11 subjects and one dummy head. These HRTFs are used in the subsequent chapters of this thesis to create individual virtual stimuli. The HRTFs are described by extracting the monaural and binaural localization cues and by presenting standard deviations across

subjects. Furthermore, the capability of the dummy head to serve as an average subject is investigated by comparing dummy head HRTFs to individual HRTFs.
Virtual acoustic displays are created by realizing HRTFs as digital filters. To reduce the computational effort of the digital filters, the filter order is often reduced by roughly approximating the HRTF spectra. Furthermore, if finite impulse response (FIR) filter are used, the filter length is reduced by applying minimum phases to the HRTFs because they have a minimal energy delay (Oppenheim and Schafer, 1975). However, HRTFs that are approximated by digital filters have to provide the same directional properties as the original HRTFs. That implies, that each manipulation (e.g. all pass filtering, smoothing) may not alter perceptually relevant localization cues.
To gain further insight into the effects that smoothing has on minimum phase HRTFs, the effect of smoothing is analyzed on I) the inter-individual standard deviation of the HRTF spectra, II) the interaural level differences and III) the interaural time differences. Furthermore, to asses the computation time that is saved by smoothing, the length of the impulse responses is analyzed as a function of spectral detail reduction.

3.2 HRTF measurements

HRTFs were described by a variety of studies (see (Møller *et al.*, 1995) for a summary). The main difference in the method between studies is the type of microphone and its location within the ear canal. The position of the microphone within the ear canal or the cavum concha and its influence on the HRTFs was investigated by several researchers (e.g. (Wiener and Ross, 1946; Mehrgardt and Mellert, 1977; Hammershoi and Møller, 1996)). The investigations show that, within the frequency range of interest, all spatial information is present at any point within the ear canal because the transition of the sound from the entrance of the ear canal to the eardrum does not add spatial information. Hence, it is not necessary to place a probe tube near to the ear drum. A recording location at the entrance is sufficient for capturing all spatially relevant information. In the present study, the position of the microphones was several millimeters inside the ear canal, which was blocked by the microphones.
Impulse responses can be obtained from a variety of measurement techniques (e.g. single clicks, sweeps, noises, etc.). Because a high number of impulse responses have to be measured, the method has to allow for accurate measurements in a short time. By using maximum length sequences (MLS) a high signal to noise ratio is obtained by only few measurement repetitions (Alrutz, 1983; Rife and Vanderkooy, 1993).

3.2.1 Theory

In the following it is described in which way the HRTF $A(\omega, \phi, \theta)$ can be obtained by using a MLS stimulus. The angles ϕ and θ denote the position of the sound source in spherical coordinates.
The MLS stimulus is acoustically radiated by a loudspeaker and recorded by microphones in the ear canal of the subject. The sound pressure $y(t, \phi, \theta)$ recorded in the ear canal consists of the MLS stimulus $m(t)$ convolved with the impulse response of the complete electroacoustical transducer system $h(t, \phi, \theta)$.

$$y(t, \phi, \theta) = h(t, \phi, \theta) * m(t) \tag{3.1}$$

One important feature of the MLS sequence is, that its auto-correlation function is a delta impulse with a small DC offset, which is inverse proportional to the length of the sequence. The DC offset is neglected in the present case because it is assumed that the sequence is sufficiently long. Thus, by a cross correlation of $y(t, \phi, \theta)$ with the MLS sequence the impulse response of the complete system including the HRIRs can be extracted[1].

$$m(t) \otimes y(t, \phi, \theta) = m(t) \otimes m(t) * h(t, \phi, \theta) = \delta(t) * h(t, \phi, \theta) \tag{3.2}$$

The impulse response $h(t, \phi, \theta)$ can be split into in the impulse response of the electroacoustical system $e(t)$ (that is directionally independent) and the HRIR $a(t, \phi, \theta)$.

$$h(t, \phi, \theta) = e(t) * a(t, \phi, \theta) \tag{3.3}$$

The impulse response of the measurement system $e(t)$ can be obtained by recoding the MLS stimulus at the position corresponding to the center of the head with the head absent. If $e(t)$ is known it can be used to extract the HRIR by deconvolving $h(t, \phi, \theta)$ with $e(t)$. The easiest way to accomplish this is to transform Equation 3.3 into the time domain and to solve for $A(\omega, \phi, \theta)$:

$$A(\omega, \phi, \theta) = \frac{H(\omega, \phi, \theta)}{E(\omega)}. \tag{3.4}$$

The inverse transfer function $E^{-1}(\omega)$ is only stable if it is minimum phase (see (Oppenheim and Schafer, 1975)). If it is not minimum phase the calculation given by Equation 3.4 can be performed on the absolute spectra and a minimum phase phase or a linear phase can be applied to the HRTFs. In this case the absolute spectrum of $E(\omega)$ may not contain zeros. The construction of minimum phase HRTFs is discussed in Section 3.3.4.

[1]The computational effort to calculate a cross correlation is proportional to the square of n (n denotes the length of both correlation sequences). To reduce the computation time a Fast-Hadamard transformation was used. In principal, it applies a butterfly algorithm to the correlation process that reduces the computation time to $n \times \log(n)$.

3.2.2 Methods

3.2.2.1 Subjects

Eleven subjects aged from 27 to 34 served as subjects. In addition, the HRTFs of a dummy head (Trampe, 1988) were measured. The dummy head has a rubber like surface with a shape of a normal head without hair. The outer ears have been modelled from the actual ear impression of an 'average' person and were constructed by a computer controlled cutter. At the entrance of each ear canal a B&K microphone (1/2 inch, 4165 capsule) records the sound pressure. The head has no shoulders and torso.

3.2.2.2 Experimental setup

A DSP board (AT&T DSP32C) hosted in a 486-IBM compatible PC was used for the output of the MLS stimulus. The signal was transmitted through an amplifier (Alesis RA 100) to the TASP system (Two Arc Source Positioning, see Chapter 2) for a description of the TASP system). The TASP system was used to position the electro - acoustical transducer (Manger MSW in a self constructed closed enclosure) to the desired location. Two microphones (Sennheiser KE4-211) were used to record the stimulus several millimeters inside the blocked ear canal. The recorded signal was amplified by a microphone amplifier (Unides Design, Model MPA10D) and directly fed into the AD converter entrance of the DSP Board. Stimuli were averaged by summing them up in the DSP memory.

3.2.2.3 Stimuli

Maximum length sequences with a length of 4095 samples were used. The sampling frequency was 50 kHz for subjects and 100 kHz for the dummy head. The stimuli were calculated off-line and stored in the DSP memory. Each position was measured five times (ten times for the dummy head) and averaged in the time domain. The whole range of azimuth positions with a resolution of 5° degree for subjects and 1° for the dummy head was measured. For each azimuth the HRTFs at elevations from $-40°$ to $+70°$ (dummy head: $-30°$ to $+60°$) were recorded with a resolution of 5°.

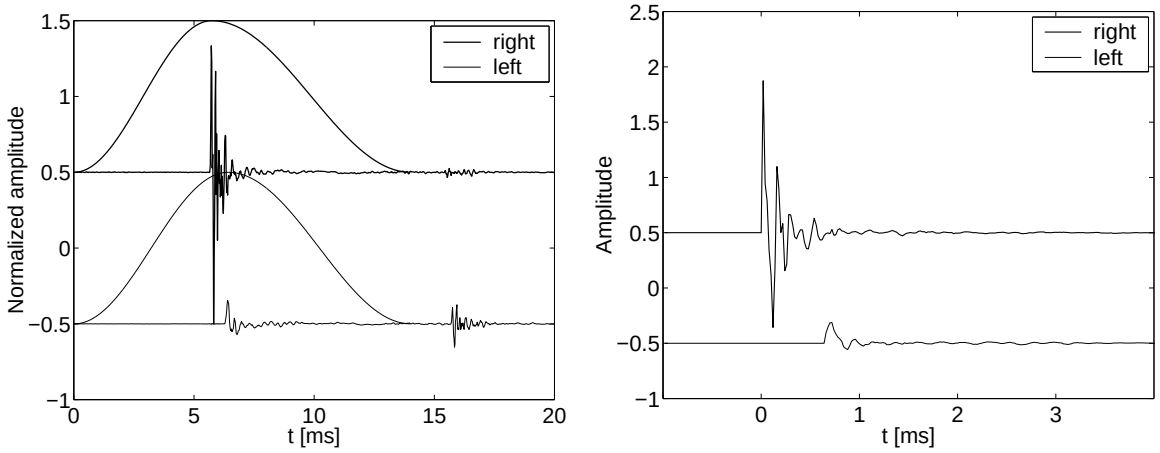
3.2.2.4 Procedure

The subject was seated in a modified bureau chair. A chin rest was fixed to the chair providing a comfortable deposit for the subjects head. The rod of the chin rest was led near to the chest of the subject to prevent it from disturbing the stimulus sound field. A simple method was used to adjust the subject's head to the center of the two

rotating arcs of the TASP system. The two speakers of the TASP setup were initially positioned in the horizontal plane at 0° and 180° azimuth, respectively. A long cord connecting the enclosures of the speaker at 0° elevation was stretched from the rear to the frontal speaker. A knot in the middle of the cord marked the center of the TASP system. The head of the subject was adjusted in a way, that the entrance of the right ear canal was exactly next to the knot of the cord. Before this procedure was performed, it was assured that the median sagittal plane of the subject coincided with the median plane of the TASP system. This was already determined by the geometrical position of the chair and the head rest in the center of the TASP system. The subjects head position was finally checked by eye. After the head had been centered, the cord was removed and the subject was told to remain as still as possible.
The dummy head, which was mounted on a stand, was adjusted in the same way. To remove specular reflections from the platform, the ground was covered with foam.
A measurement at 60° azimuth was taken to avoid an overload of the AD converter on the DSP board. To reduce the delay time between two measurements, the elevation of both the frontal and the rear sledge were adjusted at the same time. For each orientation of the arc in azimuth, first the measurement with the source in the frontal hemisphere and then the measurement in the rear hemisphere was performed. All measurements in elevation were conducted before the arc was moved to a new position in azimuth. The HRTFs were recorded in four separate sessions with overlapping azimuths at −40°, 0°, and +40°. After each session a reference measurement was performed with the microphones located at the center of the TASP system (resp. the center of the subjects head).

3.2.2.5 Data manipulation

The raw microphone signal, consisting of the MLS sequence convolved with the impulse response of the complete electro-acoustical setup was recorded. To obtain the impulse response, a Fast-Hadamard-Transformation was computed (Alrutz, 1983; Borish and Angell, 1983). The HRTFs were extracted by deconvolving the impulse responses with the reference transfer function measured at the center of the head with the head absent. The absolute level differences between sessions were resolved by comparing the levels of the double recordings of the overlapping regions in azimuth. For instance, all elevations for −40° azimuth were measured in the first and second recording session. The overall level of the impulses in the second session was adjusted by the level of the measurements in the first session for the same source locations. Figure 3.1 shows the windowing of the impulse responses $h(t)$ that was used to eliminate reflections from the TASP system before the HRTFs are extracted. The window was constructed by two squared cosine ramps. The sound source was located at 90° azimuth and 0° elevation and the impulse responses are shifted in amplitude for visibility. Reflections can be identified in the left panel of Figure 3.1 at 16 ms. This corresponds to a distance from one ear to the



(a) Measured impulse responses.

(b) Head related impulse responses.

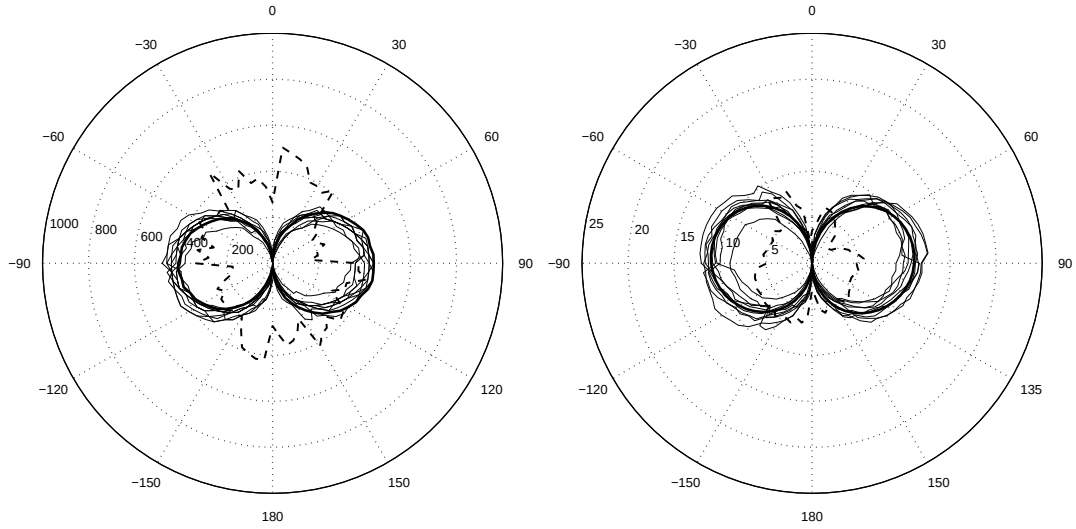
Figure 3.1: The window used for an elimination of reflections is depicted in the left panel. It eliminates reflections at approx. 16 ms. In the right panel the corresponding head related impulse responses are shown. The HRIRs were recorded for a sound source located at 90° azimuth and 0° elevation. In both figures the impulses responses are shifted in amplitude for visibility.

reflecting surface of about 3.44 m, being approximately the diameter of the arc of the TASP system. In the right panel of Figure 3.1 the corresponding HRIRs $a_{r,l}(t)$ for the left and right are shown that were extracted from $h_{r,l}(t)$.

3.2.3 Results and Discussion

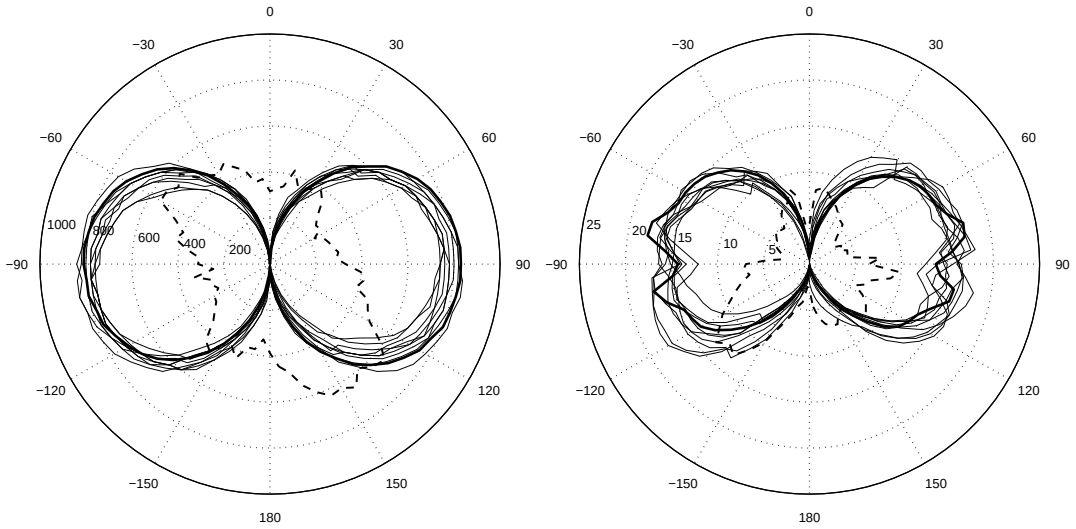
3.2.3.1 ITD and ILD

Interaural time and interaural level differences calculated from HRIRs of nine subjects and one dummy head are shown in Figures 3.2 and 3.3. The ITD is computed as the time shift of the maximum of the cross correlation function of the left and right ear impulse responses. The impulses were low-pass filtered at 500 Hz edge frequency before cross correlation. The ILD is computed as the absolute level difference between the unfiltered left and right ear HRIRs. This is equivalent to a calculation of the *signed* level differences of the HRTF spectra averaged across frequency.
Each polar plot shows absolute values of the interaural parameters as a function of azimuth. The thin dark lines show data for the subjects and the thick lines data for the dummy head. In addition, the standard deviation σ across all subjects, but not the dummy head, is plotted as a dashed line.



(a) ITD at 60° elevation. $\bar{\sigma} = 32.4\mu s$

(b) ILD at 60° elevation. $\bar{\sigma} = 1.06dB$

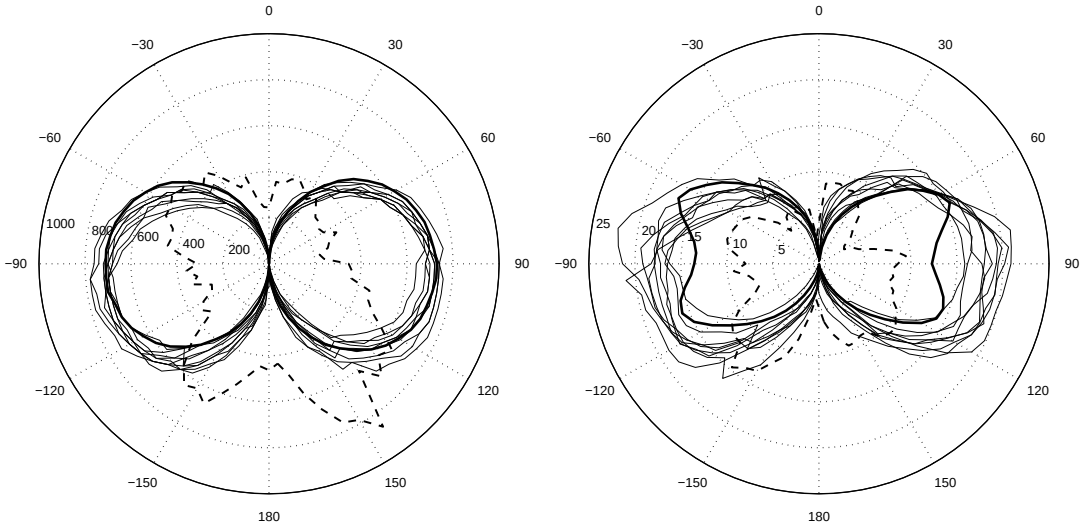


(c) ITD at 0° elevation. $\bar{\sigma} = 40.6\mu s$

(d) ILD at 0° elevation. $\bar{\sigma} = 1.29dB$

Figure 3.2: Polar plots of ILD (left column) and ITD (right column) calculated from HRIRs of 9 subjects (thin lines) and one dummy head (thick lines). Standard deviations of ILDs and ITDs are plotted as dashed lines and zoomed by a factor of 10 for ITDs and 5 for ILDs. Mean standard deviations across azimuth are given in the caption of each figure.

To make it visible within the axis range, the standard deviation was scaled by a factor of 10 for the ITD and a factor of 5 for the ILD. In the left half of Figures 3.2 and 3.3 the ITDs for the elevations of 60°, 0° and −30° are shown, whereas in the right column the ILDs for the same elevations are depicted.



(a) ITD at$-30°$ elevation. $\bar{\sigma} = 43.6\mu s$

(b) ILD at $-30°$ elevation. $\bar{\sigma} = 1.62dB$

Figure 3.3: Same as Figure 3.2 but at $-30°$ elevation.

The maximum ITD value across locations can be observed at extreme lateral source positions at 0° elevation (3.2(c)). Below and above the horizontal plane the maximum ITD decreases as the absolute distance between the source position and the median plane decreases (3.2(a) and 3.3(a)). The circle like structure of the ITD reflects the sinusoidal behavior of the ITD as a function of azimuth. A prediction of the ITD, obtained from theoretical considerations for low frequency components, is given by $\tau = \frac{3a}{c} sin(\varphi)$ (Kuhn, 1977) where a is the radius of the head and c is the velocity of sound in air. This theoretical ITD would result in two circles in the polar plot, one for each hemisphere. The standard deviation across subjects ($\bar{\sigma}$) averaged across all azimuths shows, that for higher elevations the inter-individual differences are significantly smaller than for low elevations ($p < 0.01$). The highest values of σ can be observed, especially at low elevations (Figure 3.3(a)), for azimuthal angles around $\pm 30°$ and $\pm 150°$.
The ILD pattern is similar to the ITD. For higher elevations the ILD values are smaller (3.2(b)), increasing for elevations near to the horizontal plane (3.2(d)). In contrast to the ITD, the ILD for $-30°$ elevation (3.3(b)) are partly larger than in the horizontal plane. The more complex interference pattern of the left and right ear HRTFs, induced by the torso and the shoulders, accounts for this effect. Because the interference is very sensitive to different shapes of the head and the torso, the inter-individual standard deviation is higher for low elevations ($\bar{\sigma}$ = 1.06dB at 60° elevation and $\bar{\sigma}$ = 1.62dB at 0° elevation, p<0.01). The maximum standard deviation across subjects can be seen around $\pm 30°$ and $\pm 150°$ azimuth.

The dummy head ITDs and ILDs are represented by the thick line in each plot of Figures 3.2 and 3.3. Although the range of the dummy head ILDs and ITDs is within the one for individual listeners, substantial differences between subjects and the dummy head can be observed. Furthermore, the deviation pattern varies across elevations. For instance, the dummy head ILD at 60° and 0° elevation is within the individual range, but is substantially smaller at −30° elevation. These differences may occur due to the lacking shoulders and torso of the dummy head. It can be seen, furthermore, that the dummy head cues are much more symmetrical with respect to the median plane and the interaural axis, resulting from the symmetrical geometrical design of the dummy head.

3.2.3.2 Spectral cues: Azimuth

In the following section the logarithmic power spectra of the HRTFs are presented for selected azimuths (0°, 45°, 90°, 135°) at constant elevations (−30°, 0°, 60°). Each subfigure of the Figures 3.4 - 3.6 presents the spectra for five subjects (thin lines) and the dummy head (thick lines). The spectra of the left and right ear HRTFs are plotted in the left and right columns, respectively. Additionally, the standard deviation across subjects is plotted as a thin solid lines at the bottom of each panel with the corresponding axis at the right side.

The HRTF spectra of the subjects and the dummy head are characterized by means of the following properties: a) spectral shape as a function of the source position b) differences between subjects (described by the standard deviation) and c) differences between the dummy head and the individual spectra.

Normally, two prominent resonances can be observed in the HRTF spectra (Shaw, 1997). The first one at approx. 2-3 kHz corresponds to the ear canal resonance. However, this resonance can only be seen if the HRTFs were measured in the open ear canal. The blocked meatus method used in our study, only shows the second prominent resonance at around 4-5 kHz, which belongs to the first mode of a concha resonance (Teranishi and Shaw, 1968). It is exited at both ears from nearly all directions.

The level in the frequency range above 10 kHz is influenced by the head shadow effect. Hence, for lateral sound incidence the level is decreased in the high frequencies (e.g. Figure 3.5, $\varphi = 90°$). The complexity of the HRTF spectra is highest at the contralateral ear for low source elevations and lowest for high elevations at the ipsilateral ear (comp. Figure 3.4 and 3.6, $\varphi = 90°$).

The level in the frequency range below 500 Hz varies only slightly as a function of the source position and shows little variability across subjects. Because the wavelength of this frequency range is about 70 cm, the level variation can not be assigned to the head shadow effect. However, it is likely that reflections from the chair and the legs of the subject cause the low level variations.

The standard deviation of the HRTF spectra is plotted at the bottom of each panel.

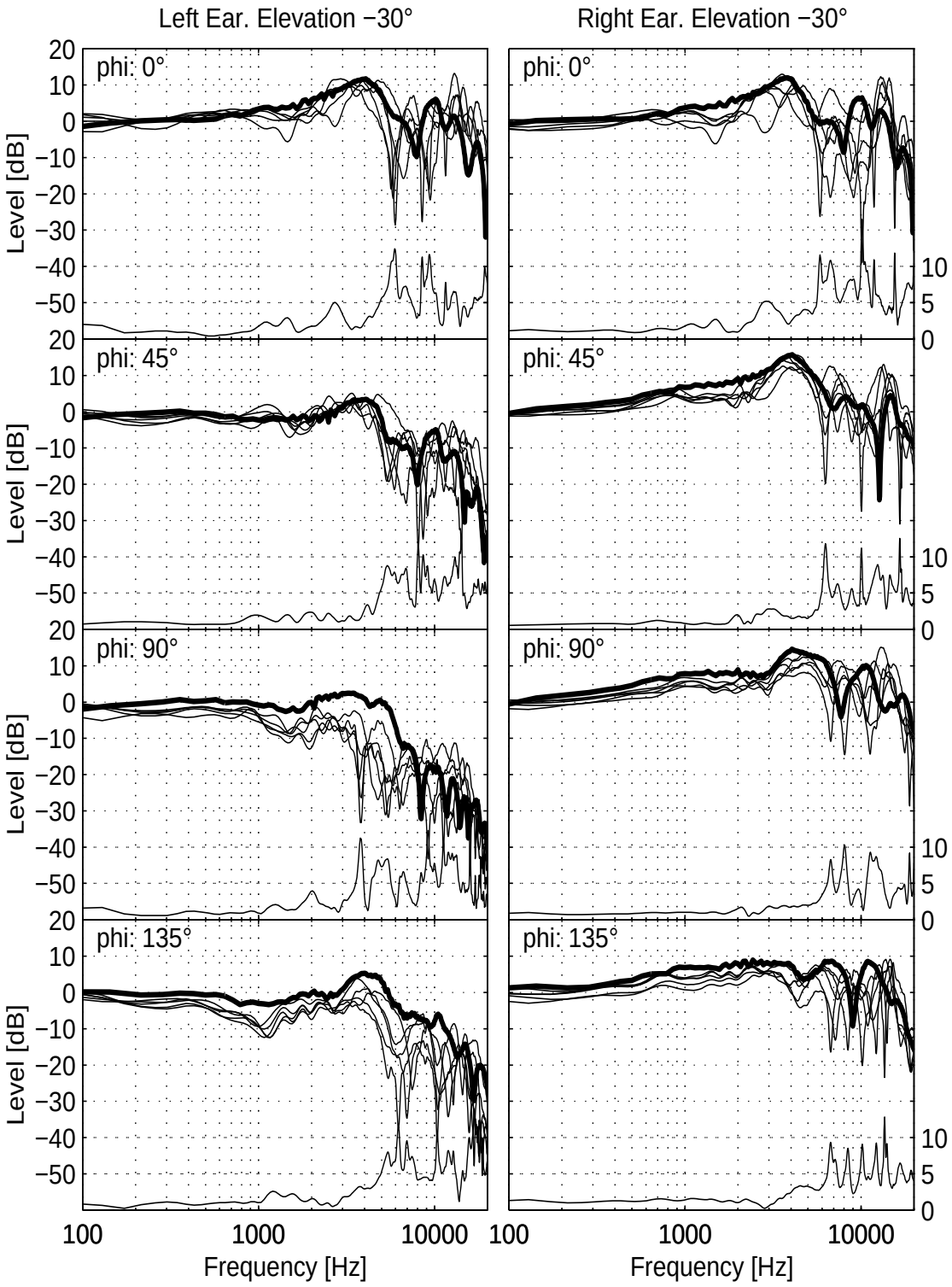


Figure 3.4: HRTFs of the left and right ear recorded from five subjects (thin lines) and one dummy head (thick line) at different azimuths and −30° elevation. The standard deviation of the individual HRTF spectra is plotted as a thin solid line at the bottom of each sub-plot.

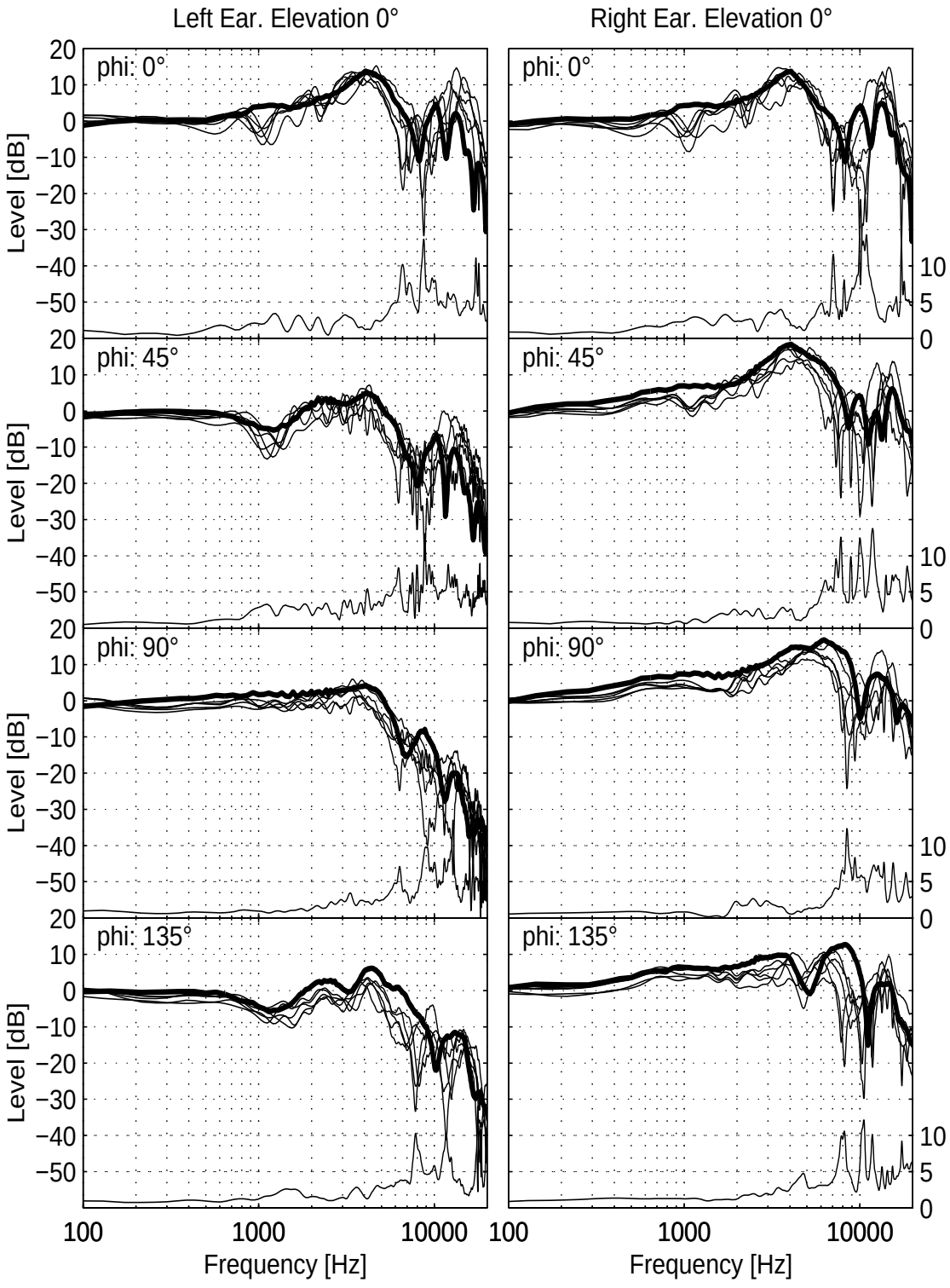


Figure 3.5: Same as Figure 3.4 at 0° elevation.

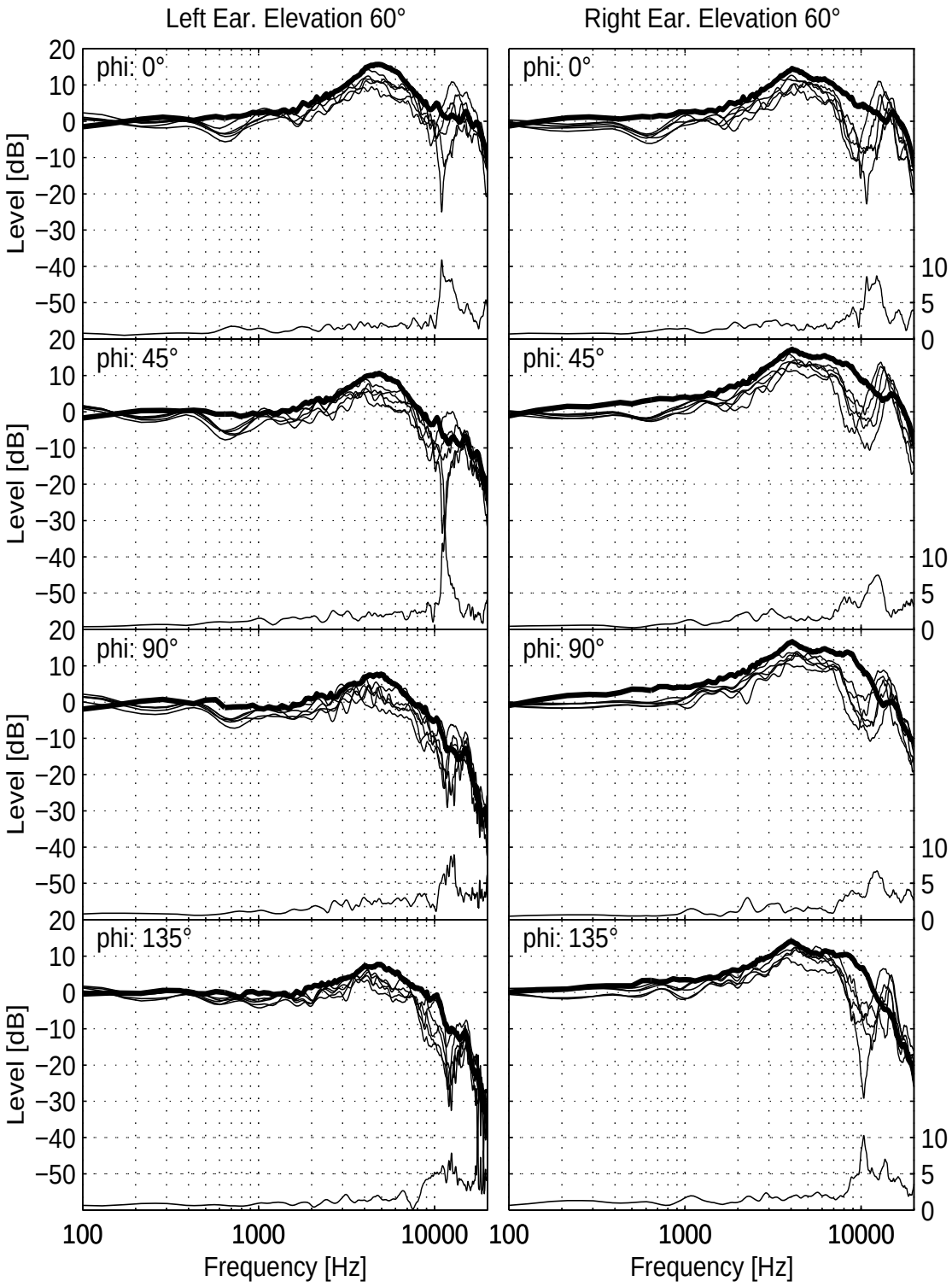


Figure 3.6: Same as Figure 3.4 at 60° elevation.

It is less than 3 dB for frequencies below 5 kHz. If the wavelength is within the dimension of the outer ear ($f > 6$ kHz), the standard deviation has peaky maxima due to spectral notches in this region that differ with respect to their center frequency. The inter-individual differences of the HRTF spectra are at maximum at approx. 10 kHz. The low standard deviation across all subjects at higher elevations is caused by the relatively smooth transfer functions (Figure 3.6). At lower source elevations, the peak around 10 kHz is broadened with a maximum standard deviation of about 10 dB.
The thick line in each panel of Figures 3.4 - 3.6 represents the dummy head HRTF spectra. The differences between dummy head spectra and individual spectra vary across source positions and frequencies. Although the dummy head was positioned on a stand and has no torso or shoulder, it follows the individual dependence for frequencies below 3 kHz within ±3dB but shows large deviations at higher frequencies.

3.2.3.3 Spectral cues: Elevation

In Figures 3.7 and 3.8 the spectral variation as a function of the source elevation (−30°, 0°, 30°, 60°) at constant azimuth (0° and 90°) is presented in the same way as given above.
The spectra of the left and right ear are symmetric up to about 5 kHz at low source elevations (Figure 3.7). The notch frequencies deviate between both ears at higher frequencies. The most prominent feature of the spectra are the concha resonance at 4 kHz and a notch in the area of 8-10 kHz. It can be seen, that the concha resonance is stable in its center frequency for all elevations. The notch shifts to higher frequencies as the source is elevated and is lowered in depth. At 90° azimuth and −30° elevation (Figure 3.8) the spectra of the contralateral HRTF are dominated by sharp notch resonances in a broad frequency range. With increasing source elevation the spectra are increasingly symmetrical across both ears.
The inter-individual differences between the HRTF spectra are represented by the standard deviation at the bottom of each panel of Figures 3.7 and 3.8. At low elevations, the standard deviation for frequencies up to 4 kHz increases for higher frequencies up to 10 dB. If the source is elevated the inter-individual differences are decreasing.
The dummy head spectra (thick line) show less variation as a function of frequency than the individual HRTFs. Especially in the frequency range above 8 kHz less interference effects can be observed. The concha resonance and the notch at 8 kHz are clearly identifiable and clarifies the behavior of the individual spectra. However, only the variation of the dummy head HRTF spectra below 4 kHz are comparable in level to the subjects HRTF. For higher frequencies, the dummy head spectra deviate strongly from the individual ones. Furthermore, the level of the high frequency range above 8 kHz is overestimated by the dummy head HRTFs.

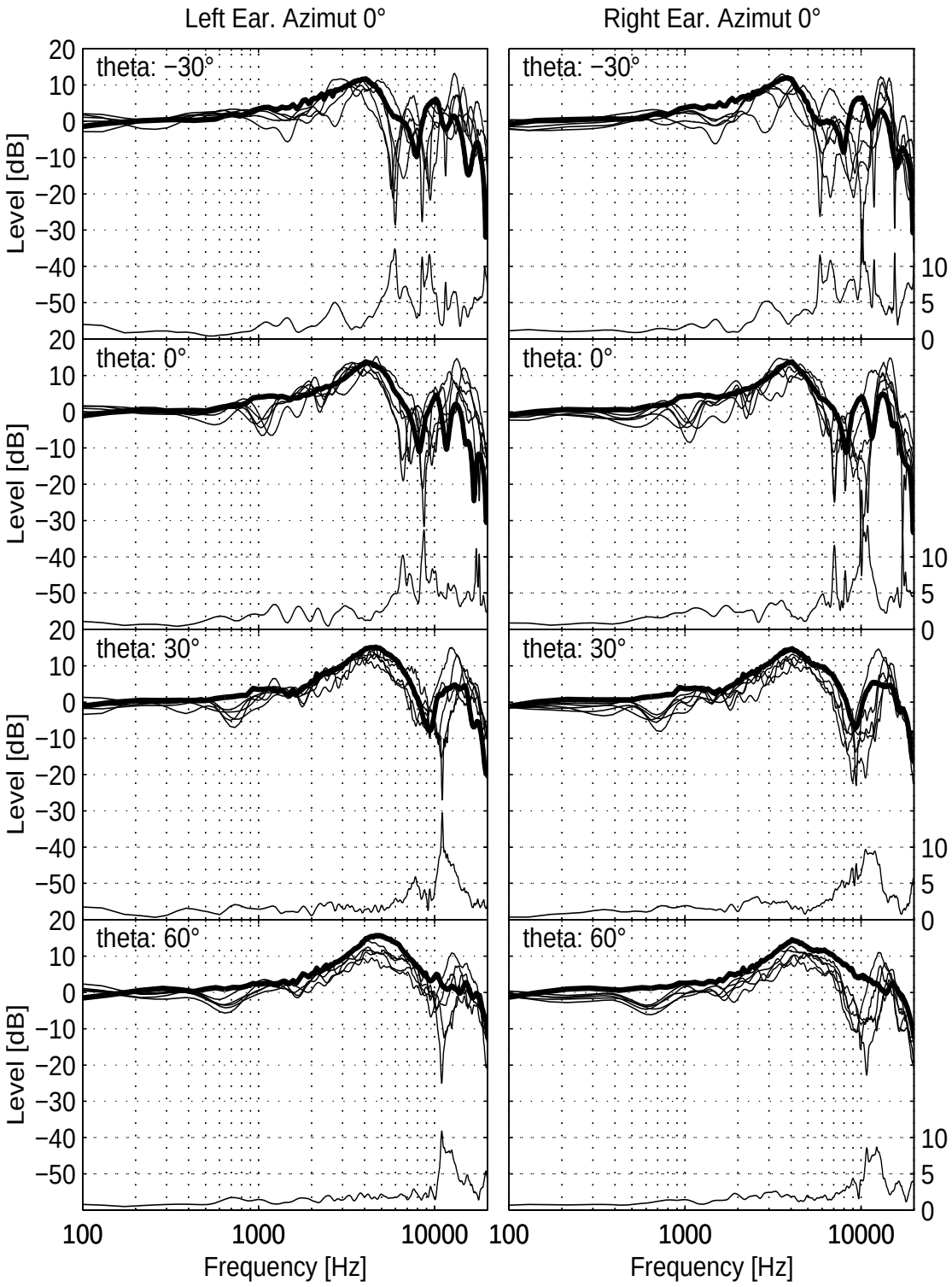


Figure 3.7: HRTF spectra of the left and right ear recorded from five subjects (thin lines) and one dummy head (thick lines) at 0° azimuth for different elevations. Additionally, the standard deviation of the individual HRTFs is plotted as a thin solid line at the bottom of each sub-plot.

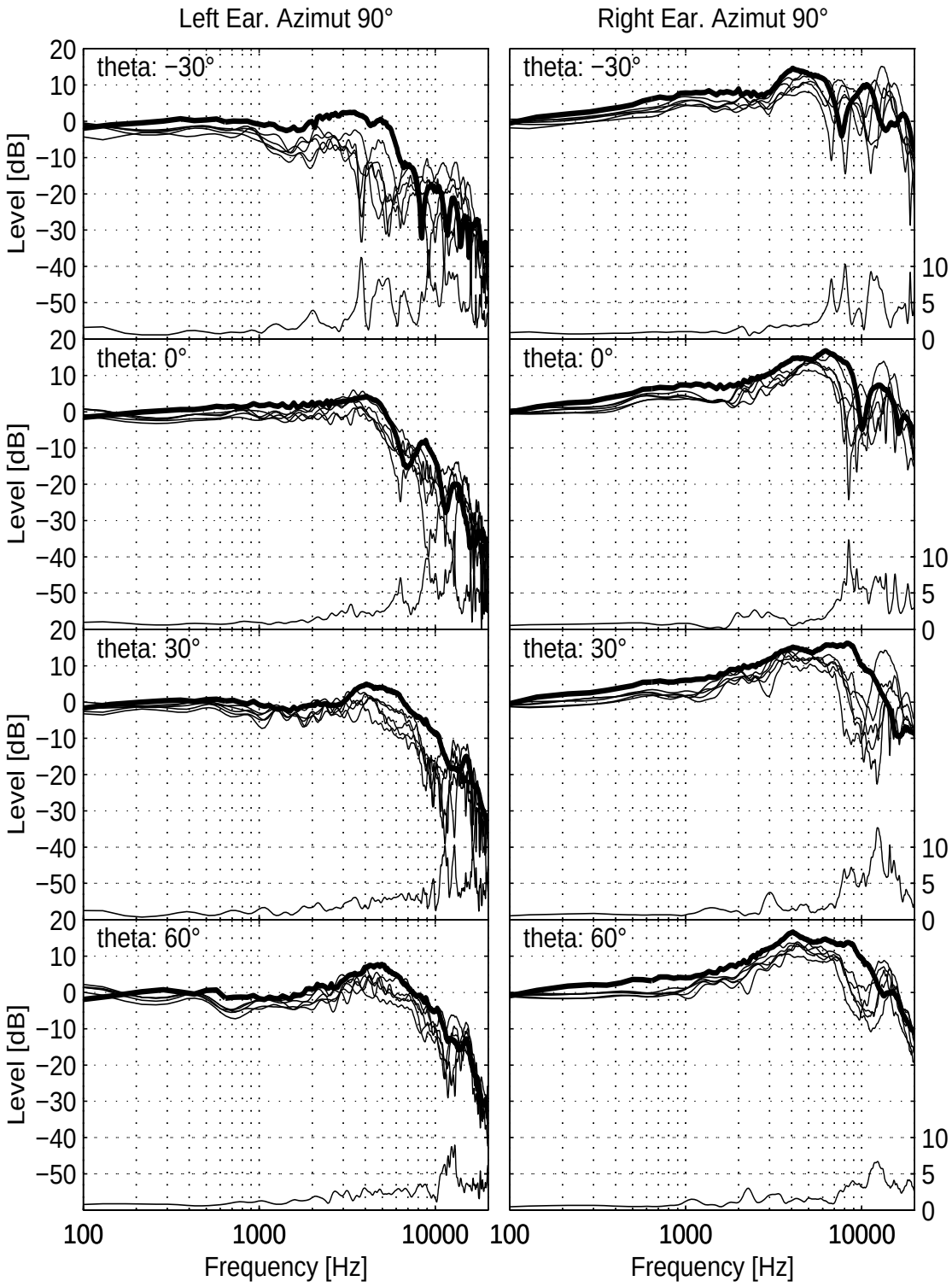
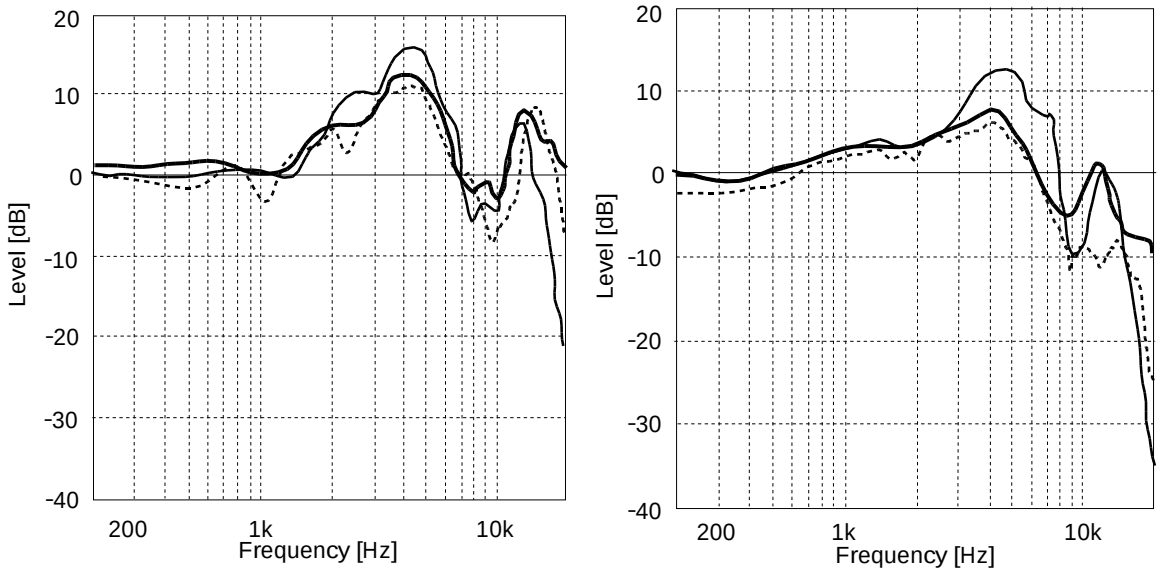


Figure 3.8: Same as Figure 3.4 at 90° azimuth.

3.2.4 Comparison of mean HRTFs

In the high frequency area the spectral variation as a function of frequency is highly variable across subjects, especially at low elevations. The spectral variation in the frequency range below 8 kHz provides less spatial information because the two resonances (i.e. the ear canal resonance and the resonance of the cavum conchae) are stimulated from nearly all spatial positions. Therefore, to compare HRTFs measured in different studies, mean HRTFs averaged across subjects are computed even though the suitability of averaged transfer functions for the presentation of spatial cues in the HRTFs spectra is doubtful: The inter-individual differences in the high frequency area are high and, therefore, by averaging HRTFs most of the spatially relevant details of the HRTFs are eliminated. However, comprehensive comparisons of mean HRTFs have been presented by Shaw (1974) and Møller et al. (1995). In the study of Møller et al. HRTFs were measured at the blocked ear canal and in the open ear canal. The results from both measurements techniques mainly differ in the frequency range below 10 kHz. HRTFs measured at the blocked meatus do not show the ear canal resonance, which is prominent in the open ear canal HRTFs. However, it was concluded by Møller et al. that the blocked meatus measurements still capture all spatially relevant information. Hence, only this kind of data are considered here.



(a) Mean HRTFs at 0° azimuth.

(b) Mean HRTFs at 180° azimuth.

Figure 3.9: Comparison of mean HRTFs across literature. The thick solid lines show mean HRTFs from Møller et al. (1995) averaged across 40 subjects and the thin solid lines represent data from Pösselt et al. (1986) averaged across 11 subjects. Mean HRTFs of the present study are given by thin dashed lines (10 subjects).

In Figure 3.9 mean HRTFs from the studies of Pösselt et al. (1986) (11 subject, thin solid lines) and Møller et al. (1995) (40 subjects, thick solid lines) are shown and compared to mean HRTFs computed from the data of the present study (10 subjects, thin dotted lines). In the left panel mean HRTFs for 0° azimuth are shown and in the right panel the source was positioned at 180° azimuth.
The general shape of the HRTF spectra is consistent across studies, although there are differences in the details. For frontal sound incidence there is a good agreement between the HRTFs of the present study and the HRTFs measured by Møller et al. The data from Pösselt et al. deviate in the amplitude of the peak in the frequency region at 2-5 kHz. Even higher deviations can be see for 180° azimuth in a slightly higher frequency area, whereas the data from Møller and the present study are very similar. However, the data obtained in the present study deviates for rear sound incidence from the cited data at frequencies above 10 kHz. A peak that is present in the data from Møller et al. and Pösselt et al. can not be seen in the mean HRTFs of the present study.
The differences of the mean HRTFs could be due to different groups of subjects and different positions of the microphones in the ear canal. In both, the study of Pösselt et al. and the present study subjects were sitting, whereas in the study of Møller et al. subjects were standing. Influences of the position of the torso on the HRTF spectra should only occur in the low frequency region. However, there is no better agreement between the studies where the subjects were sitting. Hence, the orientation of the torso does not consistently influence the shape of the low frequency HRTF spectra.

3.3 Influences of spectral smoothing on HRTFs

In order to assess the effect of reducing spectral information included in the HRTF by smoothing, the following have to be observed.

1. The standard deviation across spectra of individual HRTFs is a measure of the individual information contained in the spectra. Therefore, by investigating the standard deviation as a function of smoothing the amount of individual information in the HRTFs can be assessed. This is presented in Section 3.3.2.

2. Spectral smoothing is performed independently for the left and right ear HRTF. Therefore, also the ILD is affected by monaural smoothing. The relation between monaural smoothing and the variation of the ILD is considered in Section 3.3.3.

3. The effect of smoothing the HRTF spectra on the interaural time difference (ITD) is not easy to assess. It is a common practice to smooth the absolute spectrum and to model the HRTF phase as minimum phase plus a frequency independent group

delay τ_{Emp}, which has to be obtained from the empirical impulse responses. The minimum phase of the HRTF model is calculated from the logarithm of the absolute spectrum (s. (Oppenheim and Schafer, 1975)). Therefore, different degrees of smoothing also result in different phase spectra and different impulse responses. The ITD, however, is calculated from the impulse responses and can, therefore, vary as a function of smoothing. This is investigated in Section 3.3.4.

4. The last point concerns the length of the impulse responses as a function of smoothing which is analyzed in Section 3.3.5.

3.3.1 Smoothing methods

Two different types of smoothing are applied to the individual HRTF spectra: cepstral smoothing (see Section 4.4) and $1/N$ octave smoothing. The parameter M of the cepstral procedure describes the number of cosine terms used for the Fourier reconstruction of the spectra.
While cepstral smoothing is a linear approach with respect to the frequency axis, the humans ears' frequency resolution might be represented better by using a logarithmic approach with respect to the frequency axis. Hence, the amplitude spectra in $1/N$ octave bands are averaged by a moving average

$$H_M S(2\pi\nu) = \frac{1}{\nu_2 - \nu_1} \sum_{k=\nu_1}^{\nu_2} H(2\pi k\nu) \tag{3.5}$$

where $\nu_1 = \nu * 2^{-(\frac{1}{2N})}$ and $\nu_2 = \nu * 2^{+(\frac{1}{2N})}$ are the edges of the averaging band.
In all investigations presented below cepstral and $1/N$ octave smoothing were applied to the HRTF spectra. However, data for both procedures is only shown separately, if the results deviate between both procedures. Otherwise, results for cepstral smoothing are presented.

3.3.2 Smoothing and inter-individual differences

In Figure 3.10 the frequency dependent standard deviation of individual HRTF spectra across 10 subjects recorded at two positions in the horizontal plane (45° (top row) and 90° of azimuth (bottom row)) is plotted as a function of cepstral smoothing.
It can be seen that an increasing amount of smoothing flattens the standard deviation in the high frequency region. For eight cepstral coefficients the standard deviation is nearly constant across frequency. Hence, the inter-individual differences of the HRTF spectra are reduced to different amounts of energy in broad frequency bands.
In order to investigate the influence of the elevation on the inter-individual standard deviation of the left and right ear HRTF spectra, respectively, the mean standard deviation

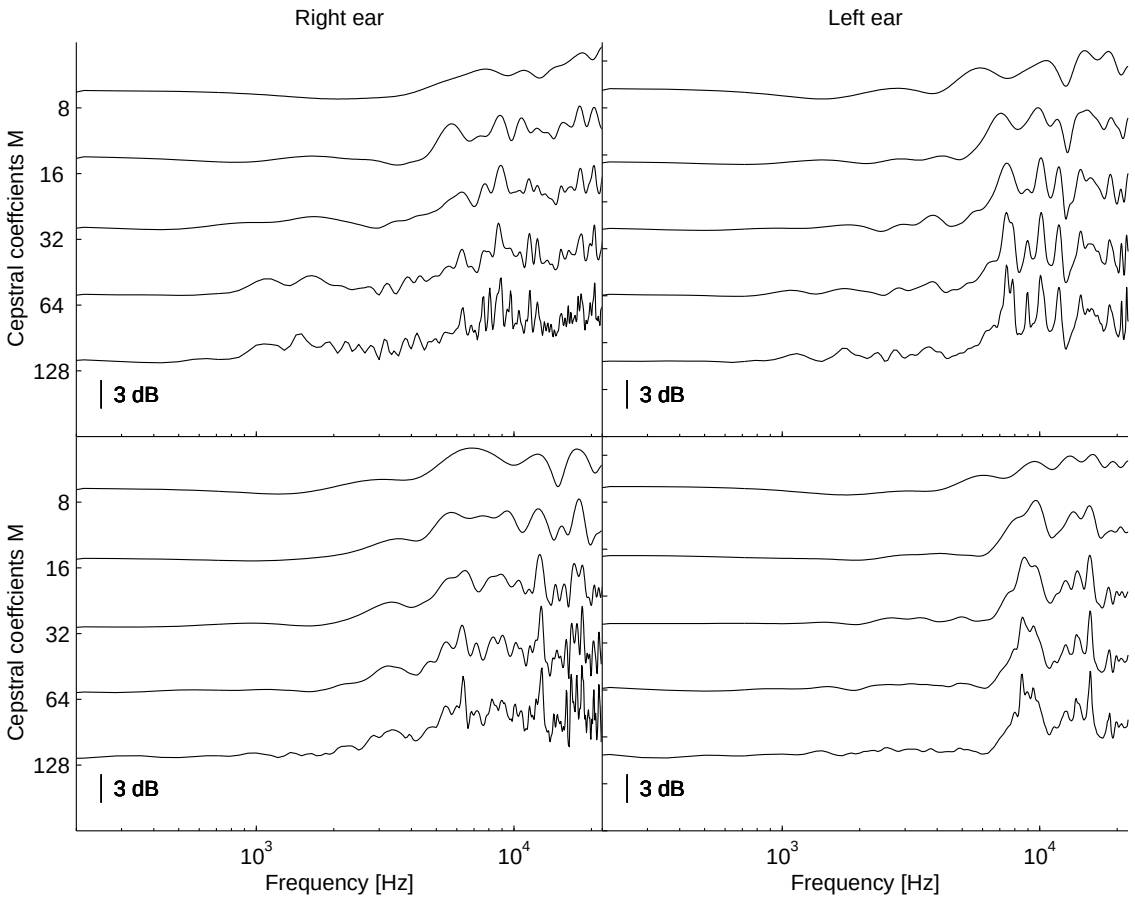
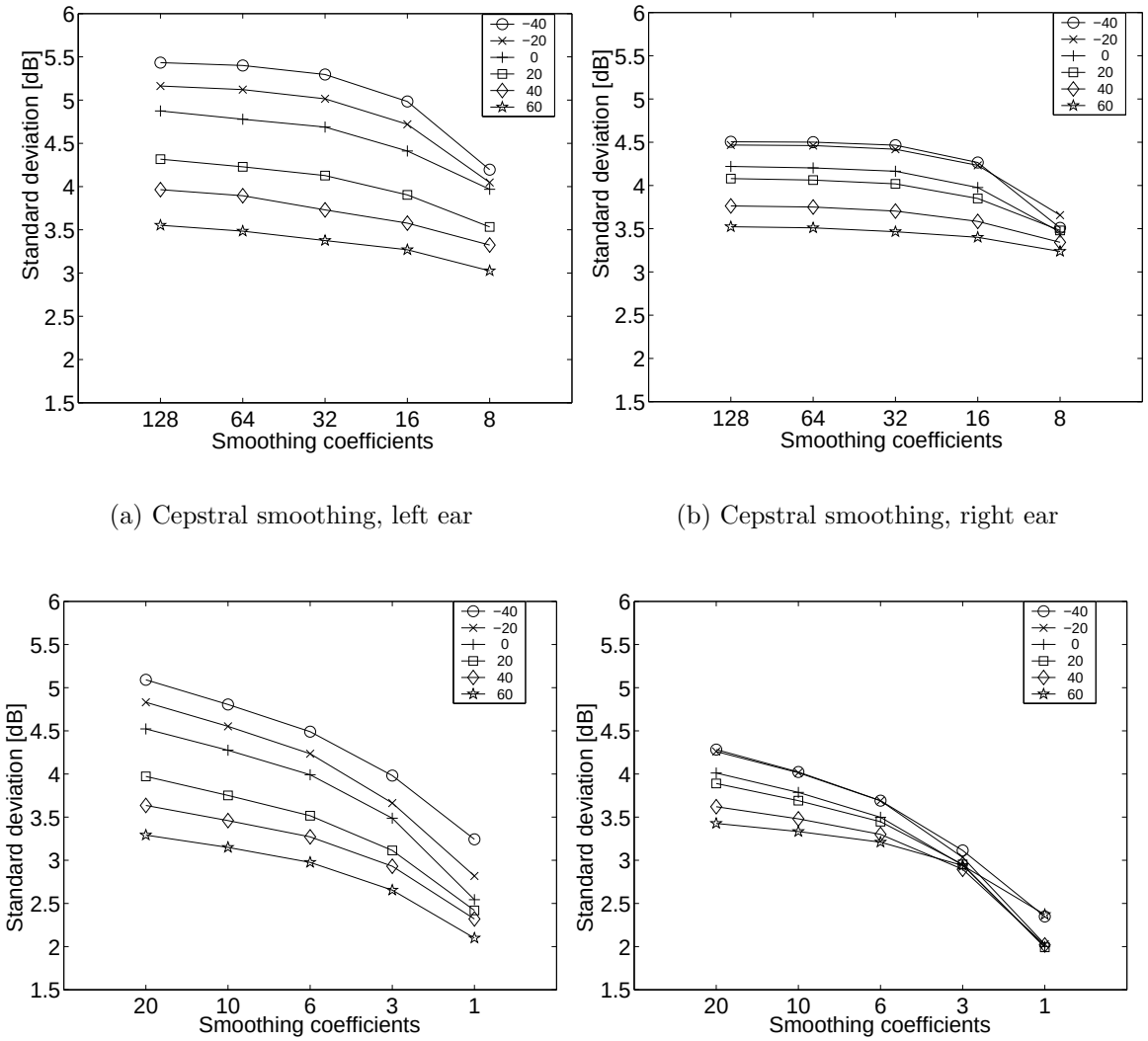


Figure 3.10: Standard deviation of the HRTFs spectra of 10 subjects for a sound source located in the horizontal plane at 45° azimuth (top row) and 90° (bottom row) azimuth for a variable number of cepstral smoothing coefficients (M) are shown for both ears separately.

averaged across frequency is plotted in Figure 3.11. Mean values for 10 subjects were computed across azimuth ($\phi = 0° - 180°, \Delta\phi = 15°$) for different elevations ($\Delta\theta = 20°$) and are given as a function of smoothing coefficients. In the top row cepstral smoothing was used, whereas data for $1/N$ octave smoothing is depicted in the bottom row.
The highest standard deviation can be observed for low elevations and the smallest standard deviation for high elevations. This tendency is more distinct at the contralateral than at the ipsilateral ear. Logarithmic smoothing, depicted in the bottom row of Figure 3.11 reduces the standard deviation more than the linear cepstral approach. This can be explained by the fact, that the linear smoothing conserves the more individual information in the high frequency region, whereas the logarithmic algorithm smoothes it out.



(a) Cepstral smoothing, left ear

(b) Cepstral smoothing, right ear

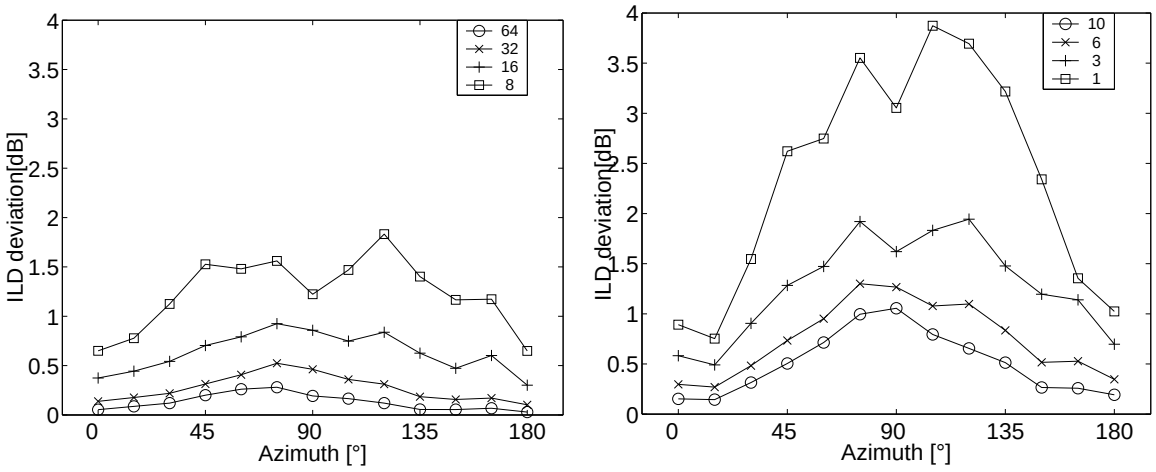
(c) Nth octave smoothing, left ear

(d) Nth octave smoothing, right ear

Figure 3.11: Standard deviation of HRTF spectra of 10 subjects averaged across azimuths plotted as a function of smoothing for different elevations. Cepstral smoothing was used in panels a) and b) and logarithmic smoothing in panels c) and d).

3.3.3 ILD deviations of smoothed transfer functions

In Figure 3.12 the ILD deviation that is introduced by smoothing the HRTF spectra is shown. The ILD deviation is calculated as the absolute level difference between the frequency dependent ILDs computed from the empirical HRTFs and the smoothed HRTFs, averaged across frequency and 10 subjects. In Figure 3.12 the ILD deviation is plotted as a function of azimuth at 0° elevation for different degrees of smoothing. In the left panel (Figure 3.12(a)) the HRTF spectra were smoothed by cepstral smoothing and in



(a) Cepstral smoothing

(b) $1/N$ octave smoothing

Figure 3.12: Deviations between original and smoothed ILDs are plotted as a function of azimuth and smoothing parameters.

the right panel (Figure 3.12(b)) by $1/N$ octave smoothing.
For 32 cepstral coefficients the influence of smoothing on the ILD is rather small (< 0.5 dB). If 16 coefficients are used the ILD deviations are up to 1 dB for sound incidence from the side. The ILD deviation increases to approx. 2 dB at lateral source positions if only eight cepstral coefficients are used.
Higher deviations can be observed in Figure 3.12(b) for $1/N$ octave smoothing. For 1/10 octave smoothing the ILD deviation is up to 1 dB at the side. If the averaging bandwidth is increased, the ILD deviation increases to up to 4 dB.
The head shadowing effect that causes the ILD, vanishes for frequencies for which the diameter of the head is small compared to the wavelength of the sound. Hence, only small ILDs can be observed in the low frequency range and higher ILDs occur mainly in the high frequencies. Furthermore, sharp notches can be observed in the HRTF spectra in the high frequencies that are introduced by interference effects and pinna filtering. Since the averaging bandwidth is increased for $1/N$ octave smoothing in the high frequencies the notches are smoothed out and even the macroscopic spectral shape is affected. The smaller averaging bandwidth in the low frequencies provides no advantage because the ILD is nearly zero in this region.
In contrast, cepstral smoothing reduces the same amount of spectral detail in each frequency band and, hence, the spectral detail in the high frequencies is better conserved by cepstral smoothing. Therefore, cepstral smoothing produces a smaller ILD variation than $1/N$ octave smoothing does.
In auditory models the frequency channels are approximatively separated by 1/3 octaves. The results of this investigation show, that this kind of auditory processing is not ap-

propriate for compressing the information available in HRTFs, since it can be assumed that the ILD deviation of approx. 2 dB is detectable for subjects.

3.3.4 ITD deviations of smoothed transfer functions

If the transfer function $E(\omega, \phi, \theta)$ in Equation 3.4 is not minimum phase the inverse transfer function $E^{-1}(\omega, \phi, \theta)$ can be unstable. In this case the calculation of $A(\omega, \phi, \theta)$ should be restricted to the absolute spectrum of the HRTF and an appropriated phase can be applied to the spectrum.
It has been shown by Mehrgardt and Mellert (1977) and Kulkarni et al. (1999) that the empirical HRTF phase is almost minimum phase plus an frequency independent group delay.
A minimum phase can be obtained from the absolute HRTF spectrum by

$$P_{min}(\omega, \phi, \theta) = \Xi(-ln(|A(\omega, \phi, \theta)|)) \tag{3.6}$$

where Ξ is the Hilbert transform. The complex HRTF is then given by

$$A(\omega, \phi, \theta) = |A(\omega, \phi, \theta)| \times e^{-iP_{min}(\omega,\phi,\theta)}. \tag{3.7}$$

An important property of minimum phase transfer functions is that they have a minimal energy delay ((Oppenheim and Schafer, 1975)). As a consequence, the group delay of the minimum phase impulse response is always nearly zero. Therefore, if both the left and the right ear HRTFs are minimum phase, the ITD is nearly zero independent on source location. To apply an appropriate ITD an frequency independent group delay is introduced to one ear that matches the ITD obtained from the empirical impulse responses. However, the ITD of the pure minimum phase HRTFs is not equal to zero for all source positions. Therefore, the frequency independent group delay that is applied to the minimum phase HRTFs has to be corrected for the inherent time delay of the minimum phase HRTFs. This correction term has to be subtracted from the ITD that is introduced to the minimum phase impulse responses.
Only the low frequency range of the ITD is perceptually relevant, because for high frequencies the phase differences at the two ears are ambiguous. Thus, it is important that the low frequency ITD of the minimum phase plus frequency independent delay HRTFs is consistent with the empirical ITD. Hence, in this study the group delay is calculated in a way that the low frequency ITDs of the minimum phase plus delay HRTFs and the empirical HRTFs are equal.
As pointed out in Section 3.3 the ITD of minimum phase plus delay HRIRs is directly related to the spectrum. Therefore, the smoothed HRTF spectra have to be taken into account when the group delay, that is introduced to the minimum phase HRTFs, is calculated.

Taken together, three different ITDs have to be calculated for introducing an low frequency ITD to minimum phase impulse responses that matches the low frequency ITD of the empirical HRTFs. They are obtained as follows:

$$\begin{aligned} \tau_{Emp} &= argmax(\Gamma(h_l, f) \otimes \Gamma(h_r, f) \\ \tau_{Corr} &= argmax(\Gamma(h_{l,min,S}, f) \otimes \Gamma(h_{r,min,S}, f) \\ \tau_{Min} &= \tau_{Emp} - \tau_{Corr} \end{aligned} \tag{3.8}$$

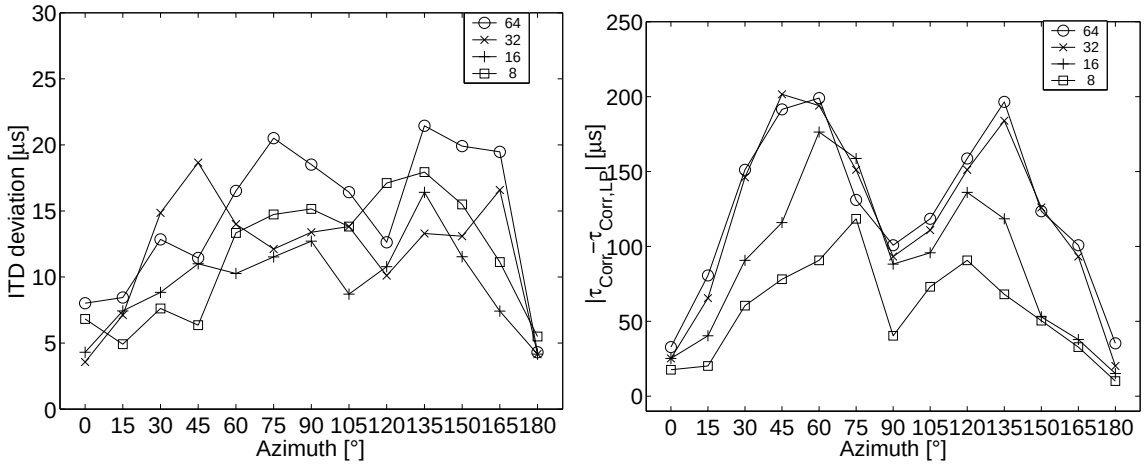
τ_{Emp} is the ITD of the empirical HRTFs calculated as the time shift of the maximum of the cross-correlation function (marked by the symbol $\otimes$) of the left (h_l) and right (h_r) ear HRIRs. '$argmax$' denotes the time shift of the maximum of the cross correlation function. The function $\Gamma(h, f)$ is the low-pass filtered impulse response $h(t)$ with edge frequency f. It is applied to extract the low frequency ITD ($f = 500\ Hz$). The correction term τ_{Corr} is calculated from the minimum phase HRIRs with smoothed spectra ($h_{r/l,min,S}$). The index S denotes the degree of smoothing either for $1/N$ octave or cepstral smoothing. Then the frequency-independent group delay introduced to the minimum phase impulse responses is given by τ_{Min}.
Based on this calculation, the low frequency ITD of the minimum phase plus frequency independent group delay HRIRs should match the empirical low frequency ITD τ_{Emp}, independent of spectral smoothing. To verify this, the ITD τ_{ReCalc} is re-calculated from the minimum phase plus frequency independent group delay HRIRs by

$$\tau_{ReCalc} = max(\Gamma(h_{l,min,S}(t + \tau_{Min}), f) \otimes \Gamma(h_{r,min,S}(t), f)) \tag{3.9}$$

The ITD error between the minimum phase plus frequency independent group delay HRIRs and the empirical HRIRs is then given by

$$\tau_{Err} = \tau_{Emp} - \tau_{ReCalc} \tag{3.10}$$

In Figure 3.13(a) the ITD error τ_{Err} (averaged across 10 subjects) is plotted as a function of azimuth for four different degrees of smoothing. The error is small for sound incidence out of the median plane ($\simeq 5 - 8\mu s$) and increases at lateral angles. Furthermore, the ITD error is not independent from smoothing and is varying in a range of approx. $10\mu s$. The results for $1/N$ octave smoothing are comparable and not shown here. It is important to note, that the perceptually relevant low frequency ITD τ_{ReCalc} only matches the empirical low frequency ITD τ_{Emp} if the minimum phase correction term τ_{Min} is computed from low pass filtered HRIRs. In a study of Kulkarni et al. (where the correction term τ_{Corr} was introduced) the sensitivity of subjects to HRTFs phases was investigated by discrimination experiments. It was shown, that minimum phase plus frequency independent group delay HRTFs were distinguishable from empirical HRTFs at lateral source positions. At these positions the low frequency ITD of the minimum phase plus frequency independent group delay HRTFs deviated from the empirical ITD. It was concluded that these deviations served as a cues for the subjects.



(a) $|\tau_{emp} - \tau_{ReCalc}|$ (b) $|\tau_{Corr} - \tau_{Corr,LP}|$

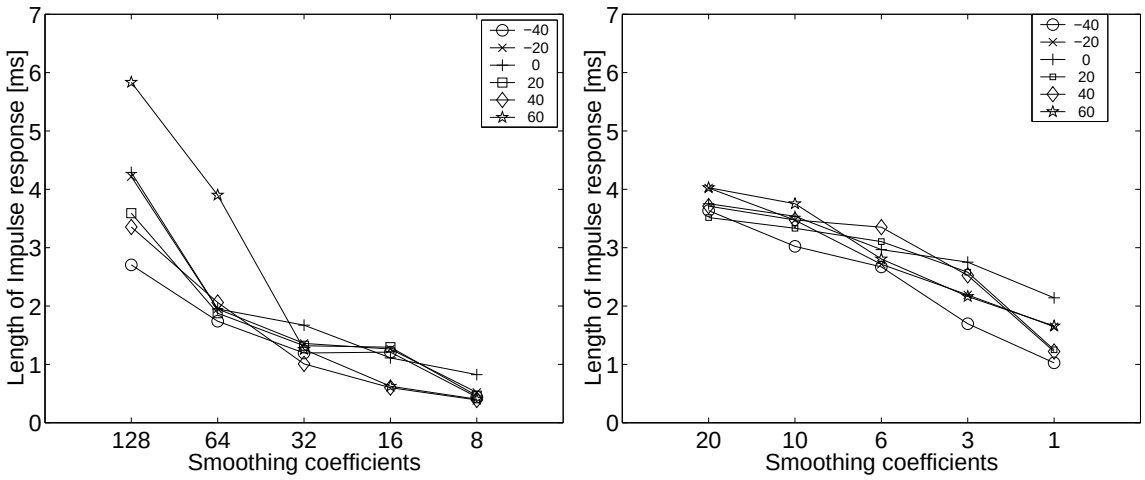
Figure 3.13: Left Panel: The differences between the low frequency ITDs of the empirical HRTFs (τ_{Emp}) and the minimum phase plus frequency independent group delay models (τ_{ReCalc}) are shown for different degrees of cepstral smoothing as a function of azimuth. Right panel: The difference of the correction term τ_{min} calculated from unfiltered and low-pass filtered minimum phase impulse responses for different degrees of smoothing are shown for source positions in azimuth. The data is averaged across 10 subjects for both plots.

However, in their study the ITDs were computed from unfiltered impulses responses (i.e. by removing Γ in equation 3.8). ITDs calculated in this way represent the group delay of the broadband signal. If the correction term τ_{Corr} is also computed from broadband impulses, the low frequency ITD deviates from the empirical low frequency ITD. This is illustrated in Figure 3.13(b). The absolute differences between the correction term τ_{Corr} computed from unfiltered and low-pass filtered minimum phase HRIRs are plotted as a function of azimuth for different degrees of cepstral smoothing. It can be seen from this figure that the low frequency group delay of the minimum phase HRIRs is clearly different from the overall group delay of the unfiltered minimum phase impulse responses. The range of the differences strongly depend on spectral smoothing, whereas the general shape of the curve is conserved.

The differences between the low frequency ITD of the empirical HRTFs and the ITD obtained from minimum phase plus frequency independent group delay HRIRs shown in Figure 3.13(a) are below the detection threshold (Durlach and Colburn, 1979; Kinkel, 1990) and are, therefore, perceptually irrelevant. However, the differences of the minimum phase correction term shown in Figure 3.13(b) are above the detection threshold for lateral sound incidence, both the absolute differences between the empirical and the minimum phase plus delay ITDs and the differences between different degrees of

smoothing.
It can be concluded from this investigation, that only if τ_{Corr} is calculated from low pass filtered minimum phase impulse responses the smoothed minimum phase plus frequency independent group delay HRTFs are perceptually indistinguishable from empirical HRTFs.

3.3.5 Impulse response shortening by spectral smoothing



(a) Cepstral smoothing.

(b) $1/N$ octave smoothing.

Figure 3.14: The length of the HRIRs are plotted as a function of cepstral smoothing coefficients (left panel) and $1/N$ octave smoothing for different elevations averaged across subjects and azimuth.

Smoothing effectively reduces the frequency resolution of the HRTF spectra. The frequency resolution is directly related to the length of the impulse response: An impulse response of length τ can be considered as an impulse response of infinite length that is multiplied with a rectangular function of length τ that is one for $0 < t < \tau$ and zero outside this interval. Therefore, if the impulse response with length τ is Fourier transformed, the spectrum of the finite impulse response can be interpreted as the Fourier transform of the spectrum of the impulse response convolved with the Fourier transform of the rectangular function, which is a 'sinc' function $(sin(x)/x)$. The convolution of the spectrum of the impulse response with the 'sinc' function can be regarded as a weighted moving average of the frequency spectrum. The width of the first maximum of the 'sinc' function is proportional to $1/\tau$ and, therefore, shorter impulse responses have lower frequency resolution. Hence, the frequency resolution is directly related to the length of the impulse response. An analytical derivation of the length of the HRIRs as a function of cepstral and $1/N$ octave smoothing might be very complicated and is beyond the

scope of the study. Therefore, the length of the impulse response is directly measured for different degrees of smoothing.
In Figure 3.14 the effect of spectral smoothing on the length of the impulse responses is investigated. The length of the impulse responses was calculated by considering the squared impulse responses of the right ear HRTFs within time frames of 110 ms. The end of the HRIR was then defined as the point where the energy has decreased to 1.5 times the energy estimated from the noise floor.
The HRIR length of the right ear was averaged across azimuth ($\phi = 0° - 180°$) and 10 subjects and plotted as a function of elevation for different degrees of smoothing. In Figure 3.14(a) cepstral smoothing and in Figure 3.14(b) $1/N$ octave smoothing was applied . It can be seen that for cepstral smoothing the length of the impulse response is reduced by a factor of up to 6 at high elevations, if M is reduced from 128 to 8 cepstral coefficients. On the average, the length is reduced by a factor of approx. 3. Similarly, for $1/N$ octave smoothing a reduction of the impulse response length by a factor of approx. 3 can be observed, for an increase of the averaging bandwidth from 1/20 to 1/1 octave. If the HRTF is realized as a FIR filter, the filter coefficients are identical to the HRIR. The computational effort to process the filter depends linearly on the number of filter coefficients. Hence, by smoothing the HRTFs the computational effort is reduced by a factor of approx. 3.

3.4 Summary and general discussion

HRTF measurements

In the first section of this chapter, HRTF measurements from 11 subjects and one dummy head were presented. The HRTFs were sampled from a high number of source positions (5° resolution for individual and 1° for the dummy head) using the TASP system (see Chapter 2) . The monaural and binaural localization cues of the HRTFs show the typical spatial dependencies that were found in the literature (e.g. (Mehrgardt and Mellert, 1977; Shaw, 1974; Wightman and Kistler, 1989a; Møller *et al.*, 1995)) and, hence, the mean HRTFs obtained here are in good agreement with results from Møller et al. (1995) and PÖsselt et al. (1986).
The standard deviations of the monaural and binaural cues across subjects are highest for low elevations and decrease as the source elevation is raised. The standard deviation of the monaural spectral cues can be separated into two frequency regions. In the low frequency region, the standard deviation across listeners is small (typically below 2 dB). In the high frequency region the standard deviation increases to up to 10 dB. The cross over frequency between both frequency bands is approx. 4 kHz for low elevations and increases to approx. 8 kHz for high elevations. Thus, HRTFs from low elevations

show more individual properties than HRTFs from high elevations. In the literature standard deviations across subjects are only shown for selected source positions mostly in the horizontal plane (Wightman and Kistler, 1989a) or for HRTFs collapsed over a wide range of spatial positions (Møller *et al.*, 1995). Although the general behavior of the standard deviation shown in the literature is consistent with the results of this study, the reduction of the standard deviation for elevated source positions has not been explicitly shown so far.

Dummy heads are intended to represent an average head of an individual subject. The HRTFs of the dummy head employed here ('Oldenburg dummy head') show that the binaural cues are fairly within the range of individual cues for higher elevations. However, due to the lack of a torso and shoulder the binaural cues at low elevations strongly deviate from individual cues. The spread of the binaural cues across individuals clearly limit the use of dummy head HRTFs as a replacement for individual recordings. Since the deviations of the dummy head cues from the individual binaural cues vary considerable across source locations, the dummy head cues would only be suitable for some source positions. Furthermore, the deviations of the monaural cues provided by the dummy head from the ones originating from the subjects' own ears are even larger than in the binaural case. It is known from the literature, that the spectral filtering of the pinna in the high frequencies is responsible for resolving front-back confusions and to estimate the source elevation (e.g. (Oldfield and Parker, 1984a; Oldfield and Parker, 1984b)). However, the monaural dummy head cues strongly deviate from the individual cues especially at high frequencies. These differences depend on frequency and source position, and an extraction of a systematic pattern of these differences is difficult. Therefore, the results of this study suggest that the 'Oldenburg dummy head' can not be used as an average head of a listener if spatially correct perceptual representations of virtual stimuli are required. If no possibility is given to measure HRTFs of an individual listener at least the HRTFs from a different listener should be employed because in this case at least the low frequency spectra are well matched. However, Wenzel et al. (1993) showed that localization performance is reduced by using non-individualized HRTFs. On possibility to overcome this problem is to scale the non-individual HRTF spectra in frequency to match the individual center frequencies of the peaks and notches (Middlebrooks, 1999a; Middlebrooks, 1999b). The scale factor can be obtained by performing a psychoacoustic task that lasts about one hour (Middlebrooks *et al.*, 2000)).

Spectral smoothing

In the second part of this investigation the effect of spectral smoothing on HRTFs was investigated. Spectral smoothing obviously reduces individual information in the high frequencies. If less than 16 cepstral smoothing coefficients are used, individual information in the high frequency region is reduced to level differences in relatively broad

frequency bands. However, the small peaks and notches code individual spatial information and should, therefore, possibly left unchanged in the smoothing process. This consideration is supported by an investigation of Kulkarni et al. and the results of Section 4. Both studies show that differences between original and smoothed HRTFs with regard to localization can be detected, if less than 16 cepstral smoothing coefficients are used.

This smoothing limit is also supported by the analysis of the ILD deviations as a function of smoothing. For 8 cepstral coefficients the broadband ILD deviation exceeds 1 dB. This suggest that the ILD deviation is detectable by subjects (Durlach and Colburn, 1979). Furthermore, the results of this investigation show, that $1/N$ octave smoothing is not appropriate for smoothing HRTF spectra. The increasing amount of smoothing for high frequencies result in ILD deviations that are above the detection threshold even for 1/3 octave smoothing.

Smoothing does not only affect the ILD but also the ITD. However, it is shown in the present study that the ITD deviation is assumed to be perceptually irrelevant if the ITD that is incorporated to the minimum phase impulse responses is calculated from low pass filtered versions of the empirical HRIRs. Furthermore, the correction term that eliminates inherent ITDs of the minimum phase HRIRs has also to be calculated from low pass filter impulse responses. If, however, this correction term is computed from unfiltered impulse responses it can be assumed that the low frequency ITD of the minimum phase plus frequency independent group delay HRTFs deviates from the ITD of the empirical HRTFs in a perceptually relevant range.

This result is consistent with findings of Kulkarni et al. (1999). In their study, the group delay of the minimum phase HRTFs and the minimum phase correction term were computed from unfiltered HRTFs. The results of a discrimination experiment showed, that subjects were able to distinguish minimum phase plus frequency independent group delay HRTFs from empirical HRTFs for sound incidence from the sides. On the basis of the considerations presented in the current study it can be supposed that the subjects would not have been able to detect the minimum phase plus frequency independent group delay stimuli if the incorporated ITD would have been correctly matched in the low frequency range.

In the last investigation presented in this study the length of the HRIRs as a function of smoothing was calculated. For both cepstral and $1/N$ octave smoothing the length of the impulse responses is substantially reduced. The length of the original impulse responses averaged across azimuth positions ranged from 4.5 ms to 6 ms depending on elevation. In a study of Kulkarni and Colburn (1998) it was shown that 16 cepstral coefficients were sufficient for providing all spatially relevant information of the HRTF spectra. The results of this study show that for this amount of smoothing the length of the impulse responses is below 1.5 ms. From the investigation of the effect of smoothing on the ILD it can be concluded that 1/6 octave averaging produces perceptually irrelevant deviations

for logarithmic smoothing. The length of the corresponding impulse responses for this amount of smoothing ranges from 2-3 ms. Hence, by applying cepstral smoothing to the HRTF spectra the resulting impulse responses are shorter in comparison to $1/N$ octave smoothing. Therefore, both the effect of smoothing on the ILD (see above) and on the HRIR length lead to the same conclusion that linear smoothing is more appropriate for smoothing HRTFs than $1/N$ octave smoothing.

3.5 Conclusions

The investigations of this study show that

- localization cues show high inter-individual differences in ITD, ILD and monaural spectral cues. The cues are highly individual at low source elevations and less individual at higher elevations.
- dummy head HRTFs and also non-individualized HRTFs can not replace individual HRTFs, if perceptually correct virtual stimuli are needed.
- smoothing the HRTF spectra by cepstral smoothing (16 coefficients) reduces the inter-individual standard deviation of HRTF spectra by approx. 0.5 dB and approx. 1 dB for logarithmic 1/3 octave smoothing.
- the ITD of minimum phase plus frequency independent group delay HRIRs has to be calculated from low pass filtered empirical HRTFs.
- the length of minimum phase HRIRs can be reduced by a factor of 3 by smoothing the corresponding HRTF spectra.
- linear cepstral smoothing is more appropriate for HRTF spectra than $1/N$ octave smoothing because the ILD is less affected.
- linear cepstral smoothing with 16 coefficients seems to be the best compromise between preservation of localization cues and minimization of computational effort.

Chapter 4

Sensitivity of Human Listeners to Manipulations of the Head Related Transfer Functions.

Abstract

The sensitivity to changes of the individual localization cues, described by head related transfer functions, is investigated in three discrimination experiments by a two interval, two alternative forced choice (2I-2AFC) measurement paradigm. In the first two experiments, the sensitivity to manipulations of the HRTF spectra was investigated. In experiment I the spectral detail of the HRTF spectra was reduced by cepstral smoothing. The smoothed HRTFs were applied to white noise, click train and scrambled white noise stimuli (500 ms length). Hence, in two conditions of experiment I subjects could use timbre cues whereas in one condition only spatial cues were provided. Stimuli were presented from five different directions in azimuth ($0° - 180°$*, $45°$ *resolution). The results of experiment I show that the detection of the smoothed HRTFs strongly depends on the source stimulus. 16-32 cepstral coefficients (depending on source position) are sufficient for providing all spatially relevant information to the subjects but more than 64 cepstral coefficients have to be used to leave the stimulus timbre unaffected. In experiment II the individual HRTF spectra were stepwise transformed to spectra of a dummy head ('spectral morphing') and applied to a scrambled white noise stimulus (500 ms) to affect also the center frequencies of the peaks and notches of the HRTF spectra. The results show that subjects were very sensitive to the 'morphing' procedure for frontal sound incidence and less sensitive for source positions at the side. Both for spectral detail reduction (scrambled whited noise) and for 'spectral morphing' the ILD deviation (averaged across frequency bands) introduced by the spectral manipulations is an appropriate measure for the cues that subjects could have used. ILD deviations of approx. 0.5-0.8 dB for frontal sound*

incidence and of approx. 1.2 dB for lateral sound incidence can be detected by subjects. In experiment III it is shown that subjects are less sensitive to changes of the interaural time difference (ITD), if a plausible frequency distribution of the ILD was applied to the stimuli as opposed to literature experiments with a fixed ILD across frequency. It is assumed that the lower sensitivity is caused by additional spatial information in the natural ILD that stabilizes the perception of the virtual objects and makes it less sensitive to distortions of the ITD.

4.1 Introduction

The physical entities that are used by the auditory system to localize a sound source in a non-reverberant environment are captured by *head related transfer functions* (HRTFs) (Mehrgardt and Mellert, 1977; Wightman and Kistler, 1989a; Møller *et al.*, 1995; Hammershoi and Møller, 1996)). They describe the directional dependent transfer functions of an acoustical object from its source location to a point within the ear canal of the left resp. the right ear. The auditory system analyzes the directional transformation by a comparison of the sound pressures at the two ears (binaural cues) and the spectral filtering of the head and pinna at each ear (monaural cues) to estimate the location of the sound source. The general contribution of binaural and monaural cues to the localization process are well known (see (Blauert, 1974; Middlebrooks and Green, 1991) for reviews). Both, the monaural and the binaural localization cues are highly dependent on the individual subject because the shape of the head and especially the complex structure of the pinna differ from subject to subject (cp. Chapter 3, (Mehrgardt and Mellert, 1977; Møller *et al.*, 1995; Middlebrooks, 1999a)).

A natural and accurate perception of an externalized sound can be achieved by headphone listening, if the time domain representation of the individual HRTFs (the head related impulse responses, HRIRs) are convolved with a monophone sound stream (Wightman and Kistler, 1989a; Wightman and Kistler, 1989b; Bronkhorst, 1995; Langendijk and Bronkhorst, 2000). For an accurate perception of virtual stimuli individual HRTFs are needed because non-individualized HRTFs reduce the localization performance (Wenzel *et al.*, 1993).

Hence, the physical differences between individual HRTFs of different subjects are normally above the detection threshold of deviations from the individual localization cues. The primary goal of this study is to provide thresholds for perceptually irrelevant deviations of the individual localization cues described by HRTFs. Therefore, the sensitivity of subjects to manipulations of the individual HRTFs is investigated.

The measurement paradigm needed to analyze the perceptual effect of HRTF manipulations has to capture all possible perceptual changes of the virtual stimuli. An absolute

localization measurement paradigm is only capable of capturing a change of the spatial stimulus position. However, a manipulation of HRTFs may result in a different stimulus timbre or spaciousness while preserving the spatial centroid. Therefore, all perceptual differences introduced by the HRTF manipulations should be taken into account. Hence, a discrimination task was chosen for the psychoacoustical experiments.

Spectral HRTF manipulations

Only few studies in the literature conducted discrimination experiments of HRTF manipulations. In a study of Langendijk and Bronkhorst (2000) the detection performance of stimuli created from interpolated HRTFs was investigated. The authors concluded that subjects were able to detect a change in stimulus timbre if spectral differences of 1.5 dB to 2.5 dB in one 1/3 octave band of the right ear HRTF spectrum occur. Spatial displacement was detected by differences greater than 2.5 dB. However, data for different source positions is only qualitatively described. Hence, the sensitivity of the auditory system to changes in the HRTFs as a function of source position has not been investigated. Presumably, the thresholds for frontal source positions will be lower and those for more lateral angles will be higher than the mean threshold across source positions.
Kulkarni and Colburn (1998) reduced the spectral details of the HRTF spectra by cepstral smoothing. They showed that the subjects were not able to discriminate a virtual from a real sound source, as long as more than sixteen terms of a fourier series expansion were used for a reconstruction of the HRTF spectra. Although psychometric functions were shown for different angles of azimuth, the stimulus spectrum was scrambled and hence, only spatial displacements of the manipulated stimulus could be detected by the subjects. However, for virtual acoustic displays, for instance in a virtual recording studio, also non-spatial cues like timbre should be unaffected by spectral smoothing.
Hence, in experiment I of the present study the results from Kulkarni and Colburn are extended to non-spatial cues (i.e., detection of spectral cues introduced by smoothing).
A different approach with respect to the manipulation of the HRTF spectra was used in experiment II. The peaks and notches of the HRTF spectra are the primary cues for elevation perception and aid to resolve front-back confusion (e.g. (Roffler and Butler, 1968; Oldfield and Parker, 1984a; Oldfield and Parker, 1984b; Middlebrooks, 1992)). Although the general shape of the HRTF spectra is similar across subjects, the center frequencies of the peaks and notches vary between subjects, representing individual information (see Chapter 3). Smoothing changes the amplitude of the notches and peaks but does not change their center frequencies. Therefore, it is possible that subjects are more sensitive to spectral manipulations that also shift the center frequencies of the peaks and notches. In this case the detection thresholds obtained from the spectral detail reduction experiment do not describe the general sensitivity to spectral variations.
Therefore, in the second experiment of this study a transformation was applied to the

HRTF spectra that also varies the center frequencies of the peaks and notches. This was done by transforming the shape of individual HRTF spectra to the shape of dummy head HRTF spectra. The results of the investigation on HRTF spectra in Chapter 3 show that dummy head HRTFs differ from individual HRTFs in amplitude as well as in the center frequencies of peaks and notches in the spectra. Thus, by a stepwise incorporation of the macroscopic dummy head HRTF spectra the frequency distribution of the individual HRTFs is changed. This process is called 'spectral morphing' further on. A comparison of the results obtained from experiment I (spectral detail reduction) and experiment II ('spectral morphing') can reveal if subjects are more sensitive to a transformation that primarily affects the individual information coded in the HRTF spectra.

ITD variations

Both experiments, reduction of spectral detail and 'spectral morphing', focus on the HRTF spectra and the sensitivity of human listeners to its spectral structure. However, it is known from the literature that the ITD plays an important role in sound localization and even dominates the cues based on HRTF spectra (e.g. (Wightman and Kistler, 1992)). The sensitivity of subjects to changes of the ITD (just noticeable differences, JNDs) have been investigated intensively by headphone experiments (review (Durlach and Colburn, 1979; Kinkel, 1990)). Normally, JNDs are obtained for tone stimuli (e.g. (Hershkowitz and Durlach, 1969; Domnitz, 1968)), broadband noise (e.g. (Tobias and Zerlin, 1959; Mossop and Culling, 1995)) or narrowband noise ((Kinkel, 1990)). The ITD JNDs vary considerably across studies. Depending on method and subjects JNDs are in the range of 6 μs to 60μs for a reference with zero ITD. However, it is a common finding that ITD JNDs are increasing by a factor of approx. 2-3 for higher ITD references.
The stimulus ILD also affects the ITD JND, tending to cause higher values if the ILD is increased (e.g. (Koehnke *et al.*, 1995)).
The ILD-ITD combinations presented to subjects in studies from the literature, are often implausible for the auditory system because they do not occur in the daily listening scenario (except for conditions were ILD and ITD are zero). Furthermore, the ILD of broadband stimuli is normally constant across frequency. This is very implausible for the auditory system because the interaural level differences in the low frequencies are close to 0 dB due to the vanishing head shadow effect. The most plausible ITD-ILD combinations are given by stimuli convolved with individual HRIRs. Two hypotheses can be stated that predict the results of measuring the ITD JND with empirical stimuli in opposed ways: The spaciousness of a virtual stimulus depends on the consistency of the localization cues. A smaller spatial extend of the object is expected if all localization cues point to the same direction. If, however, the spatial information deviates across cues (for instance, the ITD is pointing to the left hemisphere and the ILD is pointing to

the right hemisphere), an increased blur of the virtual object can be observed. Hence, it can be assumed that changes of the localization cues of a focused stimulus are easier to detect than changes of the localization cues of an object with increased spaciousness. Based on this hypothesis the ITD JND would be *smaller* for stimuli with natural ITD-ILD combinations than for unnatural combinations.
On the other hand, it can be assumed that plausible combinations of the localization cues provide a more robust perception of the virtual object because all localization cues point into the same direction. Hence, a change of one localization cue is less important for correct spatial perception if consistent redundant information is given by the remaining cues. Based on this hypothesis the ITD JND would be *larger* for stimuli with natural ITD-ILD combinations than for unnatural ones.
To test both alternative hypotheses two experimental conditions were investigated. In the first condition, detection rates of ITD variations for stimuli convolved with individual HRTFs were measured. In the second condition, the stimuli had the same ITD but the ILD was constant across frequency. However, the ILDs of the individual stimuli and the flat spectrum stimuli were matched with respect to their level difference averaged across frequencies. Hence, the individual stimuli provide more consistent localization cues than the flat spectrum stimuli. A comparison of the results obtained in both conditions tests both alternative hypotheses that predict the variation of the ITD JND differently.

Hence, the following experiments are presented in this study to obtain thresholds for differences in the individual localization cues: In the first experiment of this study in Section 4.4 detection rates were measured for stimuli with smoothed HRTF spectra. In the second experiment in Section 4.5 the HRTF spectra were manipulated by transforming the individual spectral shape to the shape of dummy head HRTF spectra (see (Trampe, 1988) for a description of the dummy head). Finally, the sensitivity to ITD manipulations is investigated in Section 4.6.

4.2 General Method

The subject was sitting in an sound isolated booth (IAC, Model No. 405A) with dimensions of 3x3x2m (length, depth, height). An IBM compatible computer was located outside the cabin and controlled the experiments by running a MatLab script. The subject was seated in front of a window with the computer monitor behind it. The stimuli were presented to the subject by an AKG 501 headphone which was plugged to the audio output of a SoundBlaster 128 sound card. The presentation level was set to approx. 70 dB A[1].

The measurement paradigm was a two interval, two alternative forced choice paradigm (2I-2AFC). The stimulus sequence consisted of two intervals, each containing two virtually presented sounds. The HRTFs for creating the virtual stimuli were individually measured in a separated session (see chapter 3 for a description of individual HRTFs). Within each interval, both stimuli were separated by 300 ms silence. A pause of 500 ms separated both intervals within each trial. One of five different directions was chosen at random for each trial (ϕ =0°, 45°, 90°, 135°, 180°). The task of the subject was to select the interval in which the two stimuli were different in the aspects that were defined in the instructions given before. The instructions differed slightly from condition to condition so that the subject should focus the attention to the predefined differences in the stimulus sequence. The keyboard of the PC was used for recording the interval number. The only differences between the experiments I-III are the source stimulus (white noise, scrambled white noise, click train) and the type of manipulation (spectral detail reduction, 'spectral morphing', ITD variation) that was applied to the HRTFs.

4.3 Subjects

A total number of 10 subjects (eight male and two female) aged from 27 to 34 with clinical normal hearing participated voluntarily in the experiments. The number of subjects participating under each condition is listed in Table 4.1. All subjects were members of the physics and psychology department of the University of Oldenburg and had extensive experience in psychoacoustic tasks. The author participated in all measurements.

	SS I	SS II	SS III	'spectral morphing'	ITD variation
Subjects	8	7	7	6	6
Sessions	4	4	4	3	3
Trials p. Cond.	20	20	40	30	24

Table 4.1: Number of subjects per measurement condition (row I), number of sessions (row II) and number of trials per stimulus condition and measurement situation (row III).

4.4 Experiment I: Cepstral smoothing

In this experiment the sensitivity of the subjects to a reduction of the spectral HRTF detail was investigated [2]. The first two conditions ('SS I & II') were intended to estimate

[1]The presentation level was measured by presenting a virtual stimulus from 0° azimuth and elevation to a dummy head (Trampe, 1988) over headphone. The microphones in the ear canal of the dummy head (B&K 1/2", 4165 capsule) were directly plugged to a sound level meter (B&K 2610, fast averaging)

the sensitivity of the listeners to any changes of the stimulus perception introduced by cepstral smoothing. The instruction given to the subject, therefore, was to identify the interval in which stimulus differed by any spatial or non-spatial cue. In a further condition only spatial cues were provided to the subject by scrambling the source spectrum ('SS III'). However, due to spectral scrambling, every stimulus in the sequence changed its spatial position slightly. Therefore, the subject was instructed to select the interval in which the spatial deviation between stimuli was larger.
Separate measurement sessions were performed for each degree of smoothing. The number of stimulus repetitions for each stimulus condition (degree of smoothing and azimuthal angle) is given in the third row of Table 4.1. The measurement paradigm is described in Section 4.2.

4.4.1 Stimuli

To smooth the HRTF spectra the ceptral smoothing procedure used by Kulkarni et al. (1998) was adapted. The cepstrum of a HRTF spectrum $H(k)$ with N frequency components can be computed by applying the inverse Discrete Fourier Transform to the logarithm of the absolute magnitude spectrum $|H|$

$$C(n) = \sum_{k=0}^{N-1} \log |H(k)| e^{\frac{i2\pi kn}{N}} \tag{4.1}$$

[2]In this study, the spectral detail is defined as the amplitude variation of the frequencies that is removed by cepstral smoothing for the following reasons: The concept of auditory filters is well established to represent the frequency resolution of the auditory system. The bark scale (Zwicker and Fastl, 1990) and the ERB (equivalent rectangular band) (Moore *et al.*, 1990) scale are frequency scales which were developed to represent the properties of the auditory system with respect to frequency resolution. An appropriate method for smoothing the HRTF spectra may be to average the energy within each bark or erb band. This is approximately a logarithmic smoothing. The spectral detail in this case can be described as the amplitude variation of the frequencies within each band. This point of view is supported by the investigation of Asano et al. (1990). In this work it is shown, that only the macroscopic structure of the high frequency HRTF spectra contains spatial information and the details can be smoothed out without influencing the localization performance.
However, logarithmic smoothing reduces spectral information more in the high than in the low frequency regions. The head shadow and interference effects ($f > 2$ kHz) and the relevant monaural spectral transformation of the outer ear ($f > 5$ kHz), are located in high frequency areas. It seems, therefore, to be reasonable that the high frequencies should be accurately reproduced because more spatial information is contained in the high frequencies than in the low frequencies. This consideration would recommend a more linear weighting of the frequency scale. For this reason, the spectral detail was smoothed out here by cepstral smoothing, which is a linear approach with respect to the frequency scale. This consideration is consistent with the results obtained in Chapter 3. One outcome of this Chapter is that $1/N$ octave smoothing is less appropriate for spectral detail reduction because comparably high reduction of individual spectral information and deviation of the ILD is introduced.

where n varies from one to $N/2$. In this way, the variation of the absolute magnitude as a function of frequency is analyzed. The real part of this transformation can be used to reconstruct the logarithmic magnitude spectrum by a Fourier Synthesis.

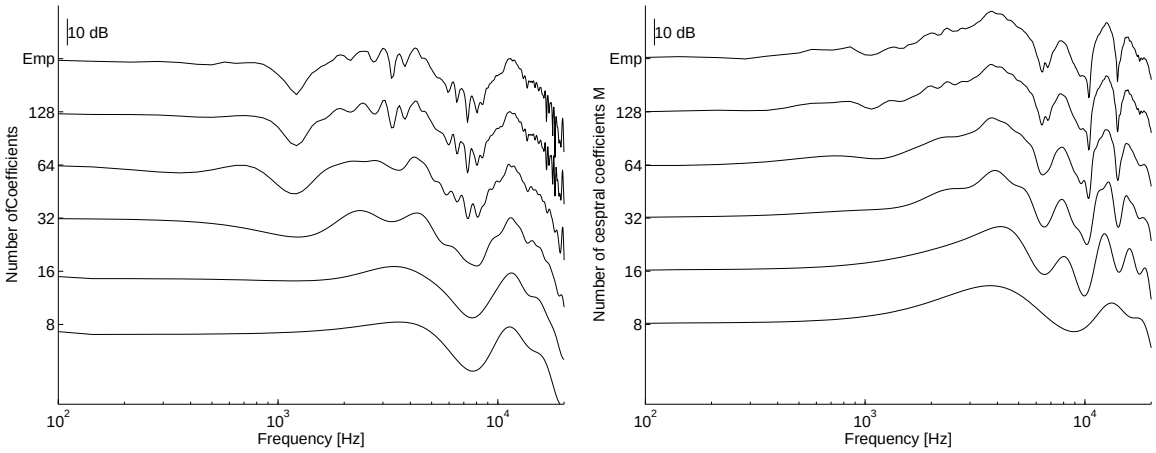
$$log(|\hat{H}(k)|) = \sum_{n=0}^{M} \tilde{C}(n) \cos \frac{2\pi nk}{N} \tag{4.2}$$

where $\tilde{C}(n)$ is defined as

$$\tilde{C}(n) = \begin{cases} \frac{(C(1)+C^*(1))}{2} & : \quad n = 0 \\ (C(n) + C^*(n)) & : \quad 1 <= n <= N/2 \end{cases}$$

For $M = N/2$ the reconstructed spectrum $\tilde{H}(k)$ equals the original one. A smoothed version of $H(k)$ can be obtained for M< $N/2$. In this way the cosine terms representing higher oscillations of the logarithmic magnitude spectrum are not used for a re-synthesis of the spectrum.

In Figure 4.1 smoothed versions of the left and right ear HRTFs (45° azimuth and



(a) Left ear.

(b) Right ear.

Figure 4.1: Smoothed HRTF spectra of one subject at $\phi = 45°$ azimuth and $\vartheta = 0°$ elevation. The empirical HRTF spectrum is plotted at the top of each panel.

0° elevation) of one subject are plotted for $M = 8, 16, 32, 64, 128$. The top line of each diagram represents the empirical HRTF spectrum. The logarithmic scale of the x-axis highlights the effects of linear smoothing on a logarithmic frequency scale. The macroscopic structure in the high frequency area of Figure 4.1(b) is well reconstructed even for $M = 16$. The notch in the mid frequency area around 1.3 kHz of the left ear spectrum (Figure 4.1(a)) is completely smoothed out for M=16. If eight coefficients are used for the synthesis procedure, the macroscopic structure in the high frequency region is only roughly approximated.

The stimulus sequence presented at each measurement trial consisted of three reference stimuli and one target stimulus. The reference stimuli were smoothed HRTFs with M=128. It is known from the work of Kulkarni et al. that this degree of smoothness is not distinguishable from the empirical HRTF (Kulkarni and Colburn, 1998). The targets were smoothed HRTFs with M=16, 32 and 64 in the conditions 'SS I' and 'SS II'. In addition, only eight coefficients were used in the 'SS III' condition.
A minimum phase was estimated from the smoothed HRTF spectra by

$$H_{min}(k) = \Xi(-ln(|\hat{H}(k)|)) \tag{4.3}$$

where Ξ denotes the Hilbert transform. This phase was applied to the smoothed spectrum. After transforming the HRTFs into the time domain, they are convolved with the source stimulus. Three different source sounds were used in separate conditions. In the first condition ('SS I') a 500 ms frozen white noise sample served as a source sound. The on- and offsets were ramped with 5 ms squared cosine ramps. A click train of 500 ms duration was used in the condition 'SS II'. The clicks were repeated at a rate of 100 Hz and onsets and offsets of the train were ramped in the same way as for the white noise. In the third condition the spectrum of the white noise stimulus (from condition 'SS I') was roved in sixth octave bands by up to $\pm$ 5 dB to prevent the subject from using timbre cues. The number of stimulus repetitions for each condition is summarized in the second row of Table 4.1.

4.4.2 Results

Detection rates for smoothed HRTFs in the measurement conditions 'SS I-III' are shown in Figure 4.2. The percentage of correct responses averaged across subjects is plotted as a function of the smoothing parameter M. The dashed lines in each subplot represent the 95% significance level for deviations from chance performance. The subplots are showing data obtained from different source positions in azimuth (denoted by ϕ).
Open squares depict the 'SS I' measurement condition. For azimuth $\phi = 0°$ and $\phi = 45°$ the correct response rate is near 100%. Even with $M = 64$ the smoothed HRTFs stimulus can be discriminated from the reference easily. The detection rates for $M = 32$ are below the threshold only for $\phi = 180°$. If sixteen cepstral coefficients are used for smoothing, the manipulated stimulus is detectable for all sound positions.
The results from the click train condition ('SS II', open diamonds) show much less detectability of the HRTF manipulation than for the white noise situation ('SS I'). Even for frontal sound incidence ($\phi = 0°, 45°$) the detection rates for $M = 32$ and $M = 64$ are near to or below the threshold. If the sound originates from lateral and rear azimuths ($\phi = 90° - 180°$) the smoothing manipulations are not detectable for the subjects, independent from the number of reconstruction coefficients.
In the third measurement condition ('SS III', crosses) spectrally roved white noise was

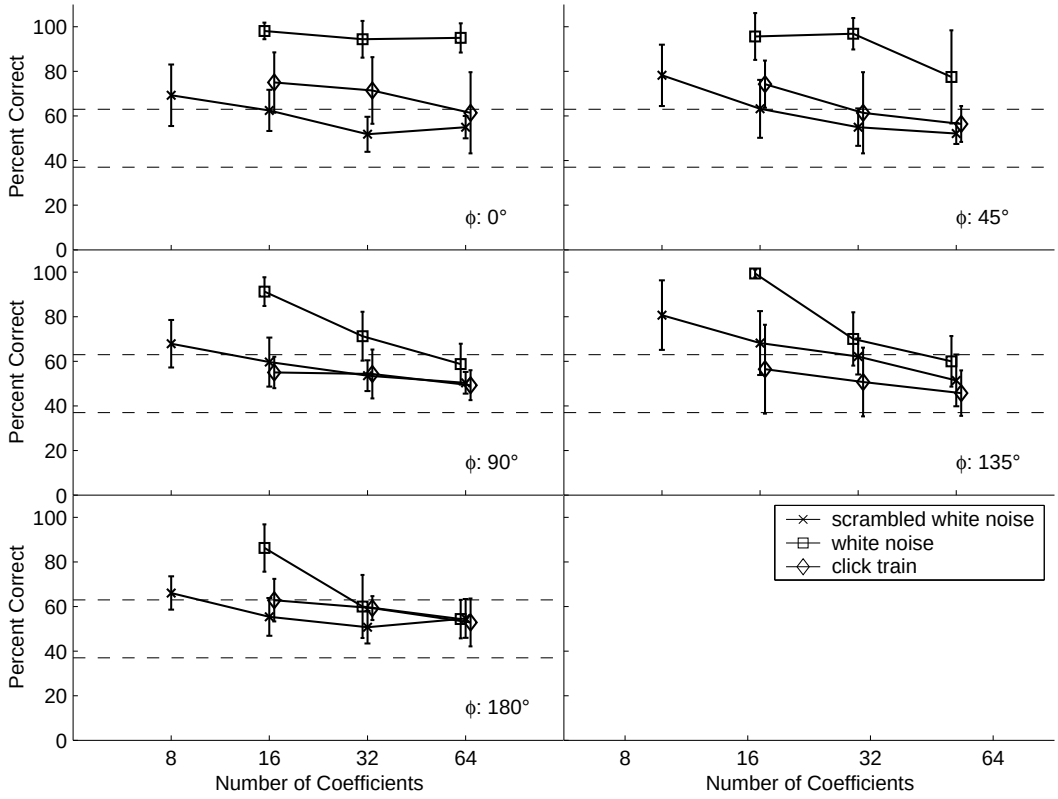


Figure 4.2: Results from the conditions 'SS I - III'. Percent correct responses averaged across subjects are plotted as a function of the number of smoothing coefficients. The error bars represent inter-individual standard deviations. The dashed lines mark the 95% significance threshold for being above chance level. Different angles of sound incidence are depicted in each subplot.

used as a sound source to prevent the subject from using non-spatial cues for the detection task. In general, the detection rates are below the threshold if more than 8 cepstral coefficients are used. Only at 135° of azimuth the detection rate approaches threshold for 32 cepstral coefficients. Except for this azimuth angle the detection rates for the scrambled white noise condition are lowest compared to the other measurement conditions.

Relation to physical stimulus parameters

In order to relate the physical cues that were available to the subjects to their performance, Figures 4.3 and 4.4 give the level differences between the smoothed and the reference HRTFs for the right and left ear, respectively. Each subplot shows the unsigned differences between the HRTF spectra reconstructed with 128 coefficients and the smoothed target spectra with $M = 8, 16, 32, 64$ plotted on a logarithmic frequency scale for one subject and angle of sound incidence ϕ. The subplots differ in the angle of sound incidence. A logarithmic frequency axis is used since it relates better to the perceptual

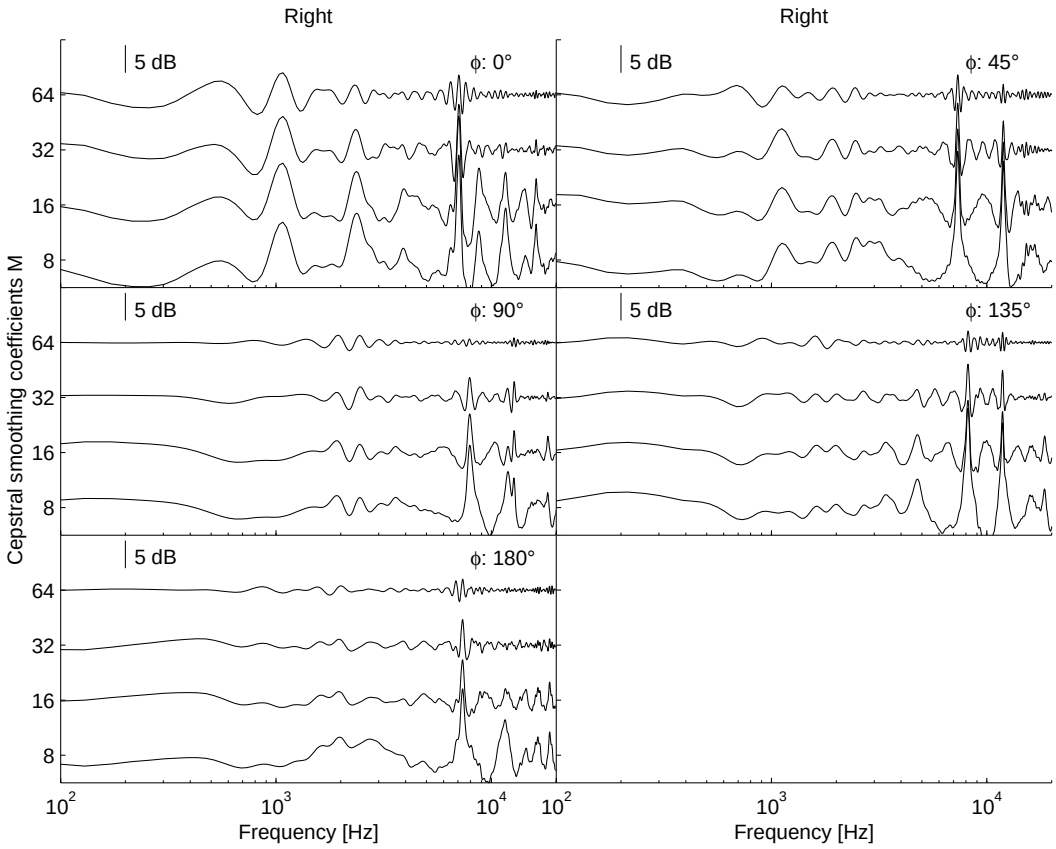


Figure 4.3: Level differences between reference and smoothed HRTF spectra of the right ear for one subject.

cues that can be exploited by the subject.

It can be seen from Figures 4.3 and 4.4 that roughly the same structural differences in spectral shape occur for all degrees of smoothing, while the magnitude of these differences increases with increasing smoothing, predominantly in the high frequency region. The corresponding effect of smoothing on the ILD is given in Figure 4.5. The broad band ILD difference is computed from the absolute level deviation between the smoothed and original interaural transfer function (ITF) averaged across frequencies and subjects. In Figure 4.5(a) level deviations were averaged for frequencies up to 4 kHz and in Figure 4.5(b) for frequencies above 4 kHz. From Figure 4.5(a) it can be seen that the influence of the smoothing process on the low frequency area strongly depends on ϕ. Only small level deviations can be observed for frontal and rear sound incidence ($\Delta ILD < 0.7dB$), but for lateral sources the level deviation reaches values up to 2.6 dB. For frequencies above 4 kHz the ILD deviations depend less on source azimuth. At positions on the cone of confusions ($0°, 180°$ and $45°, 135°$) the ILD deviations are very similar.

To relate the physical cues presented above to the perceptual data, correlation coefficients between percent correct responses and different distance measures of the smoothed and original HRTFs were calculated (see Appendix A.2). Two different distant measures

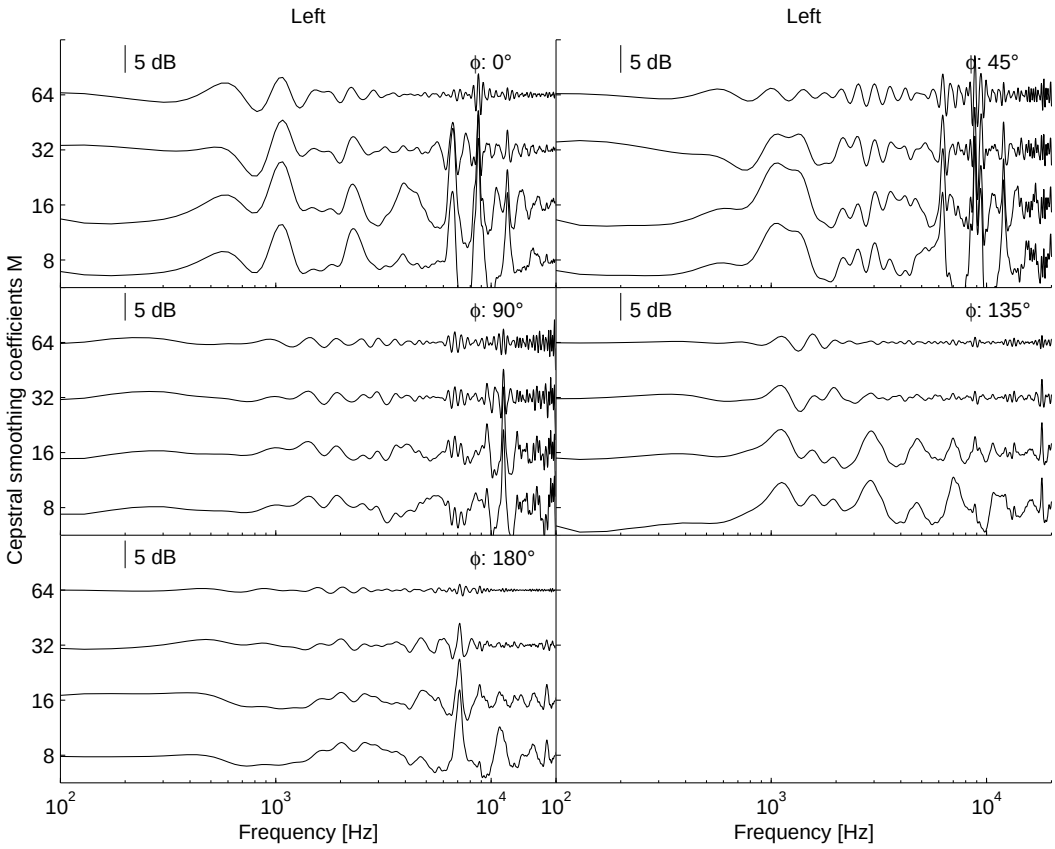
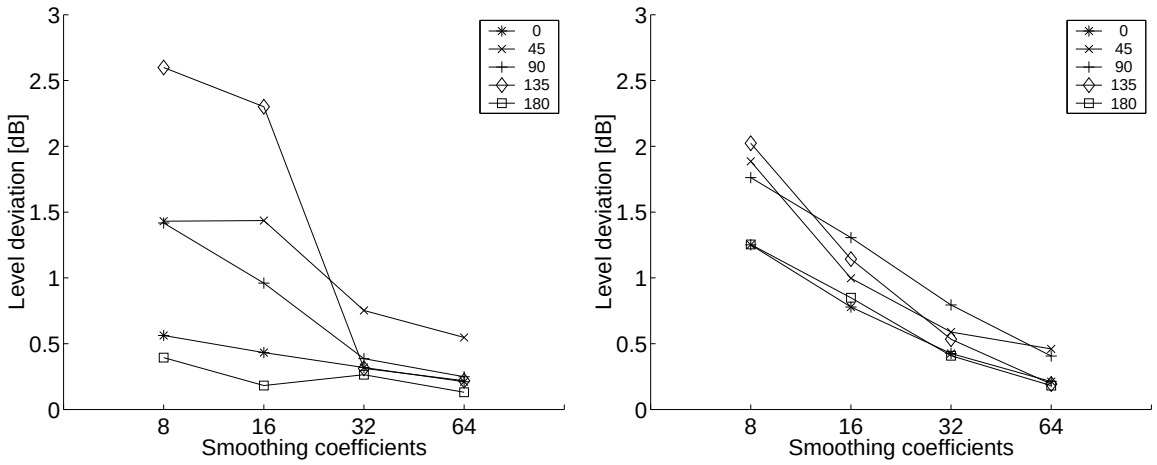


Figure 4.4: Level differences between reference and smoothed HRTF spectra of the left ear for one subject.

that show high correlations for the conditions 'SS I', 'SS II' and 'SS III' are given here. The HRTF spectra were first filtered by a Gammatone filter bank. In the 'SS I' and 'SS II' condition, absolute level differences between the smoothed and original HRTFs of the right ear were calculated for each filter bank channel and averaged across frequency. This distance measure is called D_{mon}. To derive a binaural distance measure D_{bin} for the 'SS III' condition, interaural level differences for each frequency channel were computed both for smoothed and original HRTFs. Then, the level deviations between smoothed and un-smoothed ILDs were calculated in each frequency channel. Finally, the mean across frequencies was computed. Correlation coefficients for the distance measure D_{mon} and the percent correct values in the conditions 'SS I' and 'SS II' are listed in Table 4.2 (see Appendix A.2 for a complete table with correlation coefficients for all distance measures). In the third row the correlation coefficients for D_{bin} and the percent correct responses of the condition 'SS III' are given. The low correlation values for $\phi = 0°$ and $\phi = 45°$ in the 'SS I' condition are due to the ceiling effect of subjects' response. For the other angles of azimuth the correlation coefficients are at least 0.8. Only low correlations can be found for the 'SS II' condition. This can be related to the detection rates that do not deviate significantly from chance performance for $\phi = 90° - 180°$. It



(a) $f < 4kHz$ (b) $f > 4kHz$

Figure 4.5: Level deviation between smoothed and original ILD calculated in two frequency bands.

can be assumed that the correlation rises if the degree of smoothing is increased. The distance measure shows higher correlation to the perceptual data at 45° of azimuth and nearly no correlation to the data for $\phi = 0°$. The correlation analysis shows that some distance measures have higher correlations for 0° azimuth (see Appendix A.2). However, since these correlations are still very low (≤ 0.32) no alternative distance measure for the condition 'SS II' is given here.

Higher correlations can be observed for the 'SS III' condition which takes its maximum value at 45°. In general, lateral source positions show higher correlation values than source positions in the median plane.

Condition	0°	45°	90°	135°	180°
SS I	0.14	0.42	0.81	0.85	0.8
SS II	0.18	0.63	0.19	0.29	0.41
SS III	0.66	0.88	0.71	0.75	0.59

Table 4.2: Correlation values between percent correct responses and the distance measure D_{mon} are listed in the first and second row for the conditions 'SS I' and 'SS II', respectively. In the third row correlation values for the detection rates and the distance measure D_{bin} are shown.

Figure 4.6 displays the number of correct responses in percent as a function of the ILD deviation described by the distance measure D_{bin} for the 'SS III' condition. Each subplot shows data from a different angle of sound incidence. Regression lines are plotted as solid lines for each angle of azimuth. The dashed lines mark the 95% confidence interval for

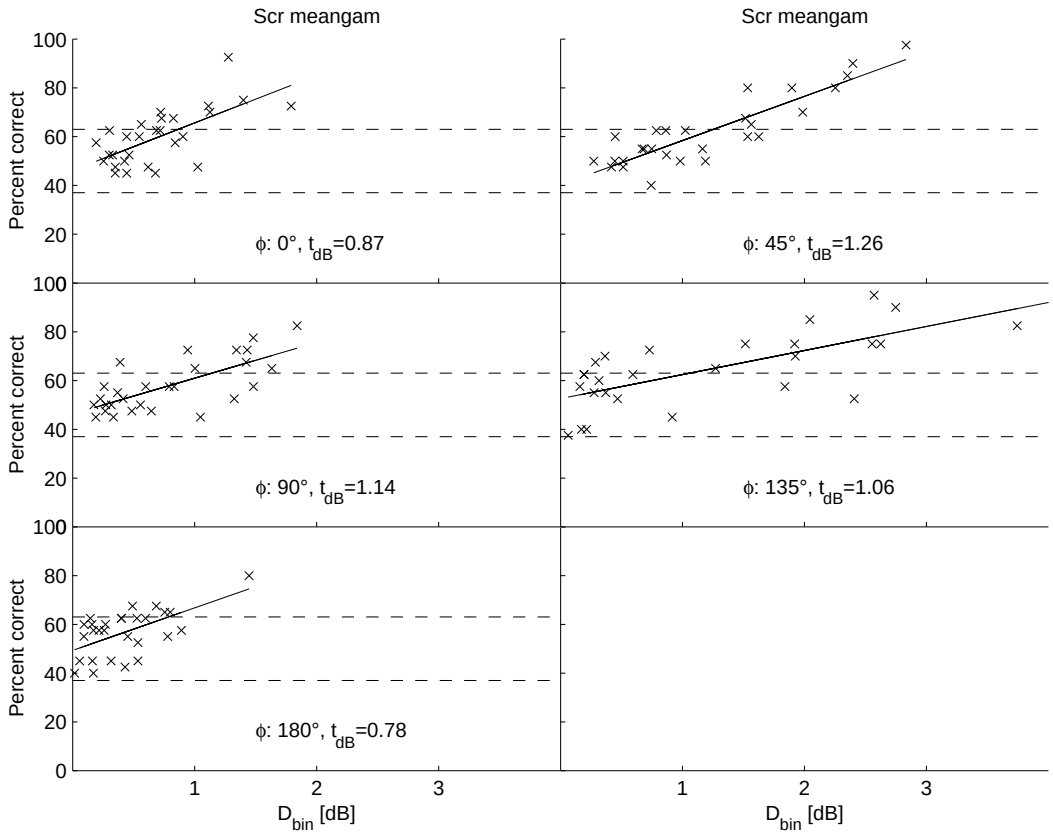


Figure 4.6: Percent correct responses as a function of the acoustical differences between target and reference stimuli. Data for all subjects averaged across sessions is presented. The dashed lines are representing the 95% confidence bounds for chance performance. In each subplot the mean detection thresholds t_{dB} are given. They are computed by calculating the level deviations for which the regression functions intersect the significance threshold.

deviations of the responses from chance performance. From these data, a physical detection threshold can be specified as the level for which the regression function intersects the significance threshold for deviation from chance performance. These thresholds can be interpreted as the average physical value which causes a manipulation in the HRTF to be detectable. The exact thresholds are given in each subplot of Figure 4.6.
The thresholds are approximatively 1 dB with a slight decrease for sound positions in the median plane and an increase for lateral source positions. If the data is plotted for the 'SS I' condition in the same way, it is obvious that subjects were able to detect the target stimulus for level differences greater than 0.5 dB. Because the psychometric functions for $\phi = 0°$ and $\phi = 45°$ are always above threshold, no corresponding physical detection threshold can be presented for these positions. However, the threshold is at least below 0.5 dB.
An estimate of the thresholds for the click train condition can only be given for $\phi = 45°$. A calculation of the threshold from a plot similar to Figure 4.6 shows that it amounts

to about 1 dB.

4.4.3 Discussion

In experiment I the HRTFs of the target stimuli were manipulated by smoothing the spectra. Although the three measurement conditions 'SS I - III' only differed in the source sound (white noise, click train and scrambled white noise), the detection performance of the subject as a function of smoothing differed considerably. Subjects were most sensitive to changes of the HRTFs when a white noise served as a source. The lowest sensitivity was observed for the scrambled white noise stimulus. The detection performance for the click trains is between the other two conditions.
It is remarkable that the detection rate for the white noise stimulus decreases for lateral source positions. The results show that smoothing with 32 and 64 cepstral coefficients was easily detected for sound incidence from 0° and 45° but the detection rate at, for instance, 90° is near to or equal to chance level. Because no spectral scrambling was applied to the stimulus, it is likely that the subjects used timbre variations as a detection cue. To relate this effect to physical cues, monaural level deviations for these positions were computed, averaged across frequency and subjects. The results show that the monaural level deviations are nearly equal for 0° and 90° at the same degree of smoothing. Hence, it seems that subjects were not only using monaural but also some binaural cues for the detection process.
The results show, furthermore, that the detection thresholds for the click train condition ('SS II') are lower than for the white noise condition ('SS I'). Both stimuli contain the same amount of spectral deviation between the reference and target stimulus. However, the click train has, in contrast to the white noise, a tonal component that corresponds to the repetition rate of the clicks (100Hz). This tonal component may dominate the perception of the click train and, hence, reduces the attention of the subject to spectral changes of the stimulus. This consideration could explain the lower rate of percent correct responses for the click train. It can be concluded from this result that HRTFs may be smoothed by a higher degree, if more complex stimuli than white noise are convolved with the smoothed HRTFs. Hence, the detection thresholds for white noise appear to be upper limits.
Two different distant measures of the physical differences of the spectral localization cues introduced by smoothing were correlated to the perceptual data. The correlation analysis revealed that the monaural level differences (described by the distance measure D_{mon}) between the original and the smoothed HRTF spectra correlate well to the perceptual data of the 'SS I' condition. This suggests, that subjects were using mainly monaural cues for the detection process. The thresholds calculated by the distance measure D_{mon} suggest that detectable spectral timbre variations for stimuli at lateral source positions are introduced by cepstral smoothing if D_{mon} exceeds 0.5 dB. For

frontal sound incidence the threshold is even below 0.5 dB.
The binaural distance measure 'D_{bin}' describes the mean ILD deviation introduced by smoothing and correlates well to the perceptual data in the 'SS III' condition. Because timbre cues are excluded in this conditions, subjects had to use spatial displacements of the stimuli for the detection of the manipulated HRTFs. Detection thresholds given by D_{bin} are approx. 0.8 dB for sound incidence from the median plane and approx. 1.2 dB for lateral positions.
An appropriate distance measure that correlates to the perceptual data in the 'SS II' condition have not been found. This is mainly caused by the low detection rates that do not deviate from chance performance for most source positions. The detection threshold is only exceeded for $\phi = 0°$ and 45° azimuth. Consequently, higher detection correlations can be observed at 45°. However, an explanation for the low correlation of the physical cues to the perceptual data for 0° azimuth can not be given here.

Comparison to the literature

The measurement task and the kind of manipulation introduced to the HRTF spectra are based on a study of Kulkarni and Colburn (1998). In their study, the reference stimuli were delivered in the free-field via a loudspeaker and compared to headphone presented virtual stimuli in a discrimination task. For the virtual stimuli, HRTFs with different degrees of spectral smoothing (8, 16, 32, 64, 128, 256, 512 cepstral coefficients) were convolved with a white noise stimulus (80 ms length) which was randomly scrambled in its spectrum (1/3 octave bands, ± 5 dB range). It was found that 16 cepstral smoothing coefficients are sufficient for providing all spatially relevant spectral information for the investigated source directions. This result is consistent with the findings of the corresponding experiment ('SS III') in the current study. However, Kulkarni and Colburn excluded the 135° azimuth position. For this source direction the results of our study show that the detection performance is slightly above chance level even for 32 cepstral coefficients. This is caused by a comparatively high ILD deviation in the frequency region below 4 kHz. Hence, 16 coefficients are not sufficient for all angles of azimuth. Because the spectral shape of the HRTFs is more complex for elevations below the horizontal plane, it can be assumed that even more cepstral coefficients are required for lower elevations.
The main difference in the method between studies is that Kulkarni and Colburn compared a virtual stimulus with smoothed spectra to a real source, whereas in the present study only virtual stimuli with different degrees of smoothing were compared. It could have been that a comparison of real and virtual stimuli is easier for subjects (e.g. due to slight head movements of the subjects' head between stimuli). However, the results are similar in both studies and, hence, it can be concluded that the method did not influence

the results.

Asano et al. (1990) analyzed the localization performance as a function of spectral smoothing. It was shown that 20 filter coefficients of an auto regressive, moving average (ARMA) filter model, (i.e. 10 for the FIR and 10 for the IIR part of the digital filter) were sufficient to allow the subjects to localize correctly in an absolute localization measurement. Among other things, the authors concluded that information contained in the frequency range below 4 kHz helps to resolve front/back confusions. This finding supports the results from Blauert (1969) that directional information is also contained in the low frequencies area. However, the contribution of the spectral cues in the low frequency region to the directional perception seems to be small because the result of the present study show that an elimination of the low frequency cues was not detectable for the subjects. A similar result was found by Langendijk and Bronkhorst (2001). They eliminated spectral cues in the low frequency region by setting the level to 0 dB and found that eliminating the cues in the frequency range from 4-6 kHz did not reduce the localization performance significantly in an absolute localization task.

In another study of Langendijk and Bronkhorst (2000) the sensitivity of human listeners to interpolated HRTFs was investigated. A discrimination task was used to obtain the detection threshold. The distance measure describing the acoustical differences between target and reference HRTFs was calculated in the following way: The level differences between the target and reference HRTFs of the right ear were computed in 1/3 octave bands. The maximal difference across frequency bands was used as distance measure. Thresholds of 1.5 dB to 2.5 dB were calculated for the detection of a change in stimulus timbre and above 2.5 dB for a change in spatial position.

To compare the results of the present study to the thresholds obtained by Langendijk and Bronkhorst the distance measure given above was applied to the perceptual data obtained here (see Appendix A.2, distance measure D8). However, instead using 1/3 octave bands a Gammtone filter bank was applied. The correlation between this distance measure and the perceptual data of the present study was lower in comparison to the distance measures D_{mon} ($\bar{r} = 0.8$ to $\bar{r} = 0.82$) and D_{bin} ($\bar{r} = 0.63$ to $\bar{r} = 0.72$) . Furthermore, the threshold was set by Langendijk and Bronkhorst to 75% correct responses, whereas in the present study it was set to 65% correct responses.

However, even by calculating the thresholds in the same as proposed by Langendijk and Bronkhorst different values are obtained. By this measure spectral timbre variations are detectable for spectral deviation in one frequency band of approx. 1.3 dB (1.5-2.5 dB is given by Langendijk and Bronkhorst). Substantially higher thresholds are obtained for the detection of a spatial displacement (approx. 5 dB in the present study and $>$ 2.5 dB in the study of Langendijk and Bronkhorst). These differences may be related to the different physical deviations of the localization cues that are introduced in both studies. The target stimuli in the study of Langendijk and Bronkhorst were interpolated HRTFs and in the present study subjects had to detect HRTFs with smoothed spectra. It could

be that the cues introduced by applying interpolated HRTFs to the stimuli are better detectable for the subjects because more relevant spatial information is distorted.

4.5 Experiment II: Spectral morphing

In this section the sensitivity of subjects to a spectral transformation is investigated which also shifts the center frequencies of the peaks and notches of the individual HRTFs. Therefore, this transformation is more destructive to the individual information of the HRTF spectra compared to the spectral detail reduction transformation.
The manipulation applied in this experiment transforms the spectral shape of the individual HRTF spectra to the spectral shape of dummy head HRTF spectra ('spectral morphing'). The measurement paradigm as described in Section 4.2 was used to obtain detection rates for the target stimuli.

4.5.1 Stimuli

The manipulated HRTF spectrum $\hat{H}_\alpha$ of the target stimulus was computed by transforming the individual HRTF H by

$$|\hat{H}_\alpha| = (1-\alpha)|H| + \alpha|H|\frac{|D_{MS}|}{|H_{MS}|} \tag{4.4}$$

where D_{MS} is the macroscopic structure of the corresponding dummy head HRTF spectrum and H_{MS} is the macroscopic structure of the individual HRTF both obtained by smoothing the spectra in sixth octave bands. The right term of the sum in Equation 4.4 represents the absolute transfer function spectrum where the macroscopic shape of the individual HRTF spectrum has been completely transformed to the dummy head shape. By stepwise increasing the factor α from zero to one, the ratio of the individual macroscopic structure and the dummy head structure is varied.
Throughout the study, this process is called 'spectral morphing'. Because 'spectral morphing' is done independently for the left and right ear spectra, the effect of the procedure on the HRTFs can be observed best in the interaural transfer function. As an example the interaural transfer function $ITF_\alpha = H_{\alpha R}/H_{\alpha L}$ of one subject for $\phi = 90°$ is calculated with $\alpha = 0, 0.1, 0.3, 0.5, 0.7, 0.9, 1$ and plotted in Figure 4.7. It can be observed, that a peak around 5.4 kHz shifts in its center frequency to 7 kHz as α is increased. Furthermore, a notch around 9 kHz is broadened and the overall level in the frequency region above 10 kHz varies as a function of alpha. The fine structure, however, remains constant.
The morphed HRTFs obtained in this way were applied to a white noise stimulus which was randomly level roved in the spectrum ($\pm 5dB$ range in 1/6 octave bands). The stimulus sequence consisted of three reference stimuli ($\alpha = 0$) and one target stimulus with

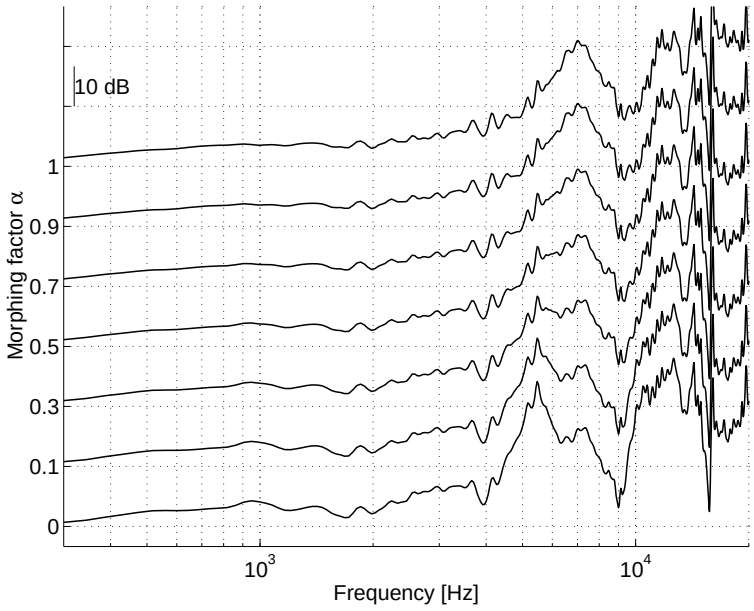


Figure 4.7: The morphed interaural transfer function spectrum of one subject at $\phi = 90°$ azimuth and $\vartheta = 0°$ elevation is shown as a function of the morphing factor α.

a randomly chosen α= 0.1, 0.3, 0.5, 0.7, 0.9 . The stimulus sequence was randomly presented from one of five different azimuth positions. The number of stimulus repetitions and measurement sessions is listed in Table 4.1.
The subject was instructed to identify the interval in which both stimuli differ more with respect to their spatial impression. This instruction was necessary because the random changes of the stimulus spectrum introduced a change in the perceived location in addition to the changes produced by the manipulated HRTFs.

4.5.2 Results

The results for the 'spectral morphing' manipulation are summarized in Figure 4.8. The figure is organized similar to Figure 4.2 and shows the percent correct responses averaged across six subjects as a function of the morphing factor α.
The subjects obviously are very sensitive to the manipulation introduced. The comparatively high standard deviation across subjects is probably due to the heterogeneous stimuli produced by the morphing process that introduces different cues for each subject. The same linear dependency of performance on morphing factor is observed for all stimulus positions, except for $\phi = 90°$ where the function slightly deviates from the linear behavior.
The highest sensitivity is observed for sound incidence out of the median plane and the sensitivity decreases as the source location moves to the side.

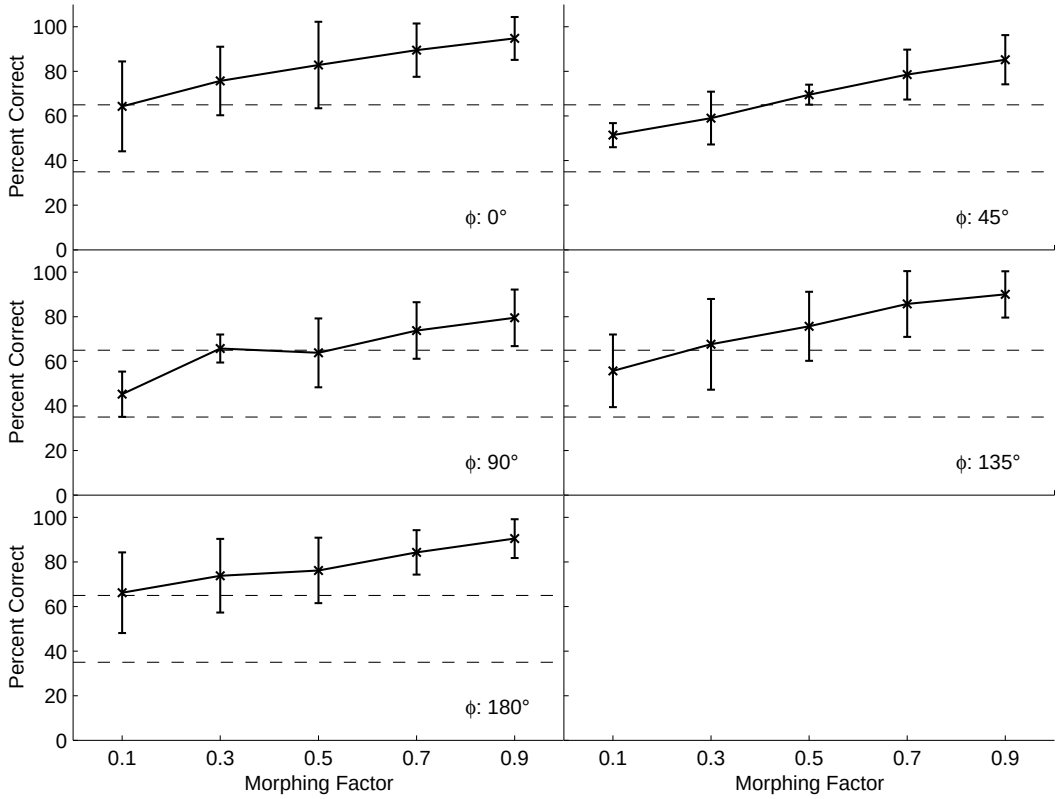


Figure 4.8: Percentage of correct responses are shown as a function of the morphing factor α. Organization of the figure is similar to Figure 4.2.

Relation to physical stimulus parameters

In order to assess the cues that subjects may have used for the discrimination task, Figure 4.9 gives the level differences between the morphed and original interaural transfer functions for one subject for $\phi = 45°$. The ILD deviation was computed from the interaural transfer function (ITF), derived both from the original HRTFs and the morphed HRTFs. The signed level differences between these two ITFs were calculated.
In contrast to HRTF smoothing, only small changes can be observed for lower frequencies. This reflects the physical differences between the dummy head and the individual's head that are only significant on a cm scale (e.g. differences in exact body and pinna geometry) and hence relate to frequencies above approx. 1 kHz, whereas the lower frequencies are influenced by the dummy head and a real listener in roughly the same way. The effect of morphing the transfer functions on the ILD is illustrated in Figure 4.10. Level deviations were computed in the same way as for Figure 4.9 and averaged across two frequency bands below 4 kHz (4.10(a)) and above 4 kHz (4.10(b)). At low frequencies , the mean level difference between the morphed and the original interaural transfer function (ITF) is between 0.1 dB ($\alpha = 0.1$, $\phi = 90°$) and 1.5 dB ($\alpha = 0.9$, $\phi = 45°$). The least difference between the dummy head ITF and the individual ITF at low frequencies is at 90° and the lowest concordance can be observed for $\phi = 45$. For frequencies above

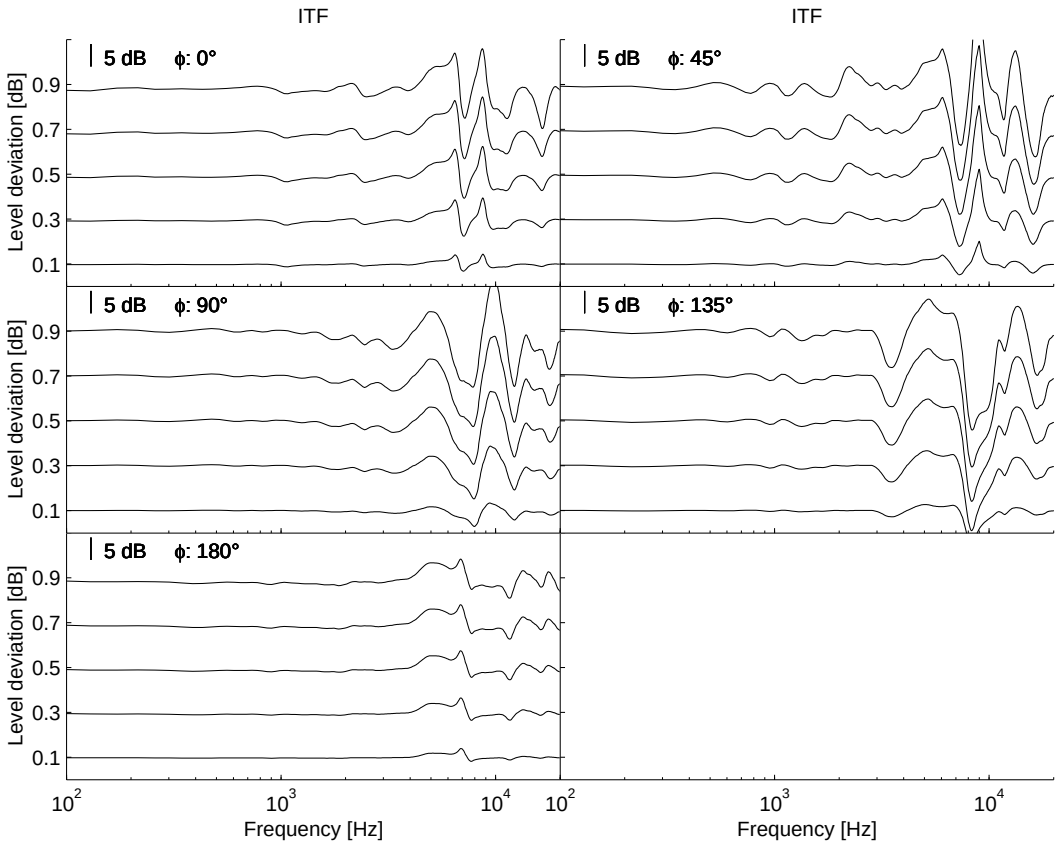
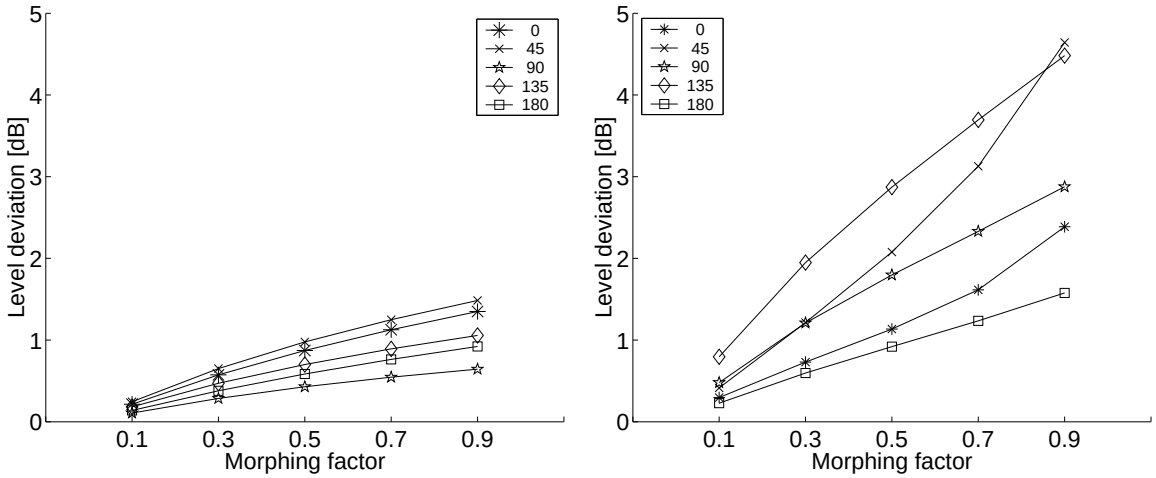


Figure 4.9: Level differences between reference and 'morphed' interaural transfer functions for one subject. In each subplot a different angle of sound incidence is depicted.

4 kHz the dummy head ITFs are deviating more from the individual ITF (minimum: 0.2 dB for $\alpha = 0.1$, $\phi = 180°$; maximum: 4.8 dB for $\alpha = 0.9$, $\phi = 45°$).

In a correlation analysis correlation coefficients between the perceptual data and different distance measures were computed. The distance measure D_{bin} used for the condition 'SS III' in experiment I shows the best correlation (see Appendix A.2). Correlation values are lowest for stimuli at the median plane (r $\approx$ 0.67). At lateral angles correlations are in the range of 0.74 to 0.79. In Figure 4.11 percent correct responses are plotted as a function of D_{bin}. A regression function is plotted as a solid line for each subplot. In each subplot data for a different azimuth of the stimulus is presented. The dashed lines mark the 95% confidence interval for deviations of the detection rates from chance performance.

From the data in Figure 4.11 detection thresholds in dB can be obtained in the same way as described in experiment I. For sound incidence out of the median plane the detection thresholds are influenced by the ceiling effect caused by high detection rates. If the threshold is corrected for this effect, the threshold is approx. 0.6 dB. The thresholds for sound incidence from the side are increased to about 1.2 dB.



(a) $f < 4kHz$ (b) $f > 4kHz$

Figure 4.10: Level deviations between morphed and original ILD in two frequency bands.

4.5.3 Discussion

The 'spectral morphing' manipulation was intended to affect the spatially relevant information of the HRTF spectra. This was done by transforming the individual HRTF spectra to the macroscopic shape of dummy head HRTF spectra. The dummy head used ('Oldenburg dummy head') differs in several aspects from individual heads. First, the geometry of the head and the pinna is only suitable for an average subject and an approximation of an individual head. Second, the pinnae are symmetric with respect to the median plane. Third, the dummy head has no torso and shoulders. Fourth, the dummy head has no ear canal. Furthermore, no hair is attached to the skin. Therefore, the dummy head HRTFs differs in two general criteria from individual HRTFs. First, the different geometry scales properties common to all heads in frequency. For instance, if the individual head is smaller than the dummy head, interference effects are located at higher frequencies for the smaller head. Second, due to the lack of physical structures that generate spatial information (e.g. the dummy head has no shoulders) less information, at least at low frequencies is provided to the individual listeners by listening to dummy head HRTFs. Therefore, transforming the individual HRTF spectra into dummy head spectra transforms information to different frequency areas and eliminates spatial information contained in the individual HRTFs.

As expected, the results show that subjects are very sensitive to the HRTF transformation. Subjects reported that one detection cue was the occurrence of front/back confusions for the targets, mainly for source positions in the median plane. These confusions were perceived even for low values of α. A transformation like 'spectral morphing' destroys the individual spectral cues that are resolving the front/back confusions. There-

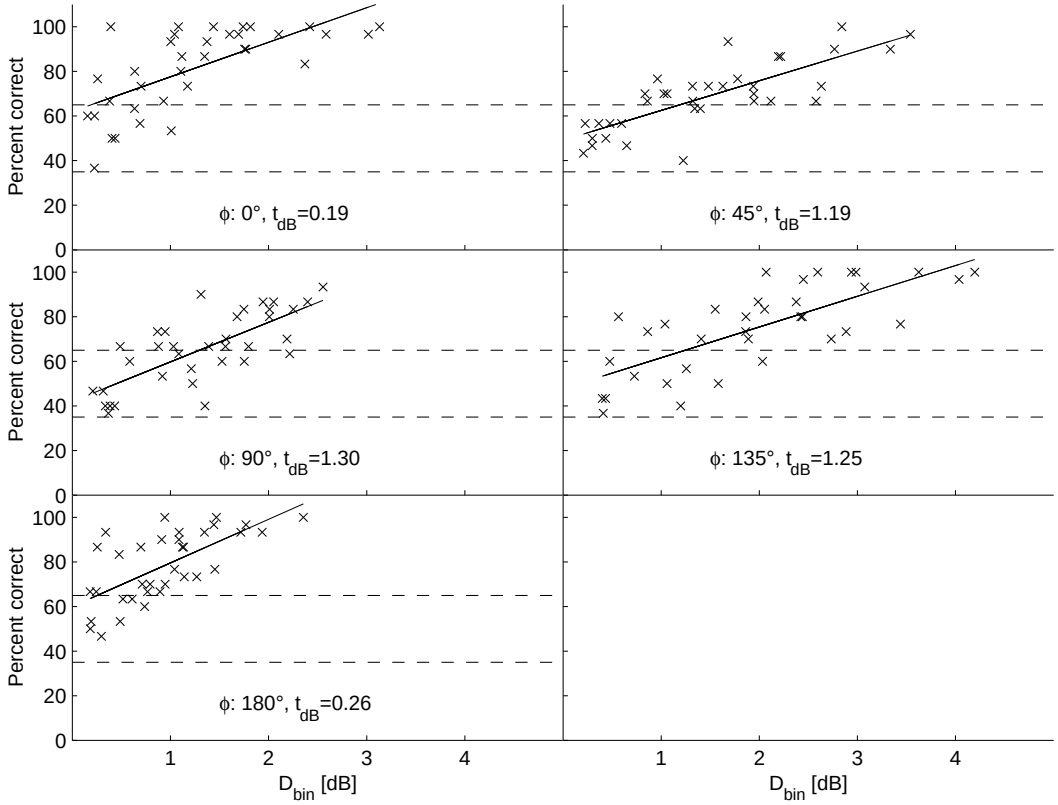


Figure 4.11: Percent correct responses measured for the 'spectral morphing' condition as a function of the distance measure D_{bin}. Data for all subjects averaged across sessions is presented. In each subplot the mean detection thresholds t_{dB} are given. They are computed by calculating the level deviations for which the regression functions intersect the significance threshold.

fore, it is likely that front/back confusions are introduced by 'spectral morphing'. Front/back confusions are a very obvious cue producing a high spatial distance between target and reference and may reduce the threshold. This could explain the high detection rates even for low values of α.

The perceptual data was related to the physical differences introduced by 'spectral morphing' by calculating correlation coefficients between the percentage of correct responses and the distance measure D_{bin} that describes the mean deviation of the ILD in different frequency bands. The correlation values are approx 0.67 for frontal sound incidence and 0.79 for sound incidence from the sides. Higher correlation values for frontal sound incidence can be expected, if the ceiling effect of subjects' response would be removed.

The same distance measure was used for the scrambled white noise condition ('SS III') in the former experiment. The detection thresholds extracted from the perceptual data are comparable (slightly lower for frontal sound incidence) but the manipulation of the spectra were different. Therefore, it can be assumed that the thresholds given by the distance measure D_{bin} are appropriate for serving as a common measure for deviations

of individual HRTF spectra that are irrelevant for the spatial perception.

4.6 Experiment III: ITD variation

The aim of this investigation was to investigate in which way the spectral shape of a broadband noise stimulus affects the ITD JND. The task of the subjects was to detect spatial displacements of virtual stimuli introduced by manipulations of the ITD in two conditions. The two conditions differ in the amount of spatial information in the shape of the stimulus spectra. In the first condition ('plausible ILD'), individual HRTFs were used to filter the white noise stimuli. In the second condition ('constant ILD'), the spectral shape that is introduced by using individual HRTFs was eliminated and set to a constant factor across frequency. However, the mean ILD across frequency was equal under both measurement conditions.
Two hypothesis were given in the introduction that predict the differences in the detection rates between the two conditions in opposed ways. The first hypothesis suggests that the detection rates are lower for the 'plausible ILD' condition because the additional spatial information in the spectra produce a more focused perception of the stimulus and, hence, spatial displacement are easier to detect. The second hypothesis predicts lower detection rates for the 'plausible ILD' condition because the additional spatial information contained in the spectral shape stabilizes the perception of the spatial object. A comparison of the detection rates in both conditions can reveal which of both hypothesis holds true.
The same measurement paradigm as in the experiments I and II was used and only the type of manipulation applied to the HRTFs was changed. The number of subjects participating in the experiments and the number of stimulus repetitions for each condition is given in Table 4.1.

4.6.1 Stimuli

The stimuli for the 'plausible ILD' condition were generated in the following way: The same sample of frozen white noise as in the former experiments was used as a sound source. For creating the reference stimuli white noise was convolved with the individual minimum phase HRTFs of the left and right ear. The target stimulus was generated by shifting the individual HRIR of the left ear in time by $\Delta\tau = \pm 22.7, \pm 68.0$ and $\pm 113.4 \mu s$. This corresponds to the discrete time shift imposed by the sampling frequency.
The reference stimuli for the second condition ('constant ILD') were created by applying the ITD, obtained from the individual HRTFs of the first experiment, to the white noise stimulus. Subsequently, the white noise stimuli for the left and right ear were scaled to

the same RMS level difference as obtained from the white noise stimuli convolved with individual HRTFs. The target stimuli were generated by shifting the stimulus that is presented to the left ear by $\Delta\tau$ =+22.7, +68.0 and +113.4μs. Note, that only positive variations of the ITD were applied to the flat spectrum stimuli.
For both conditions the stimulus sequence was presented randomly from one of five different azimuth positions. Each condition was conducted in a separate measurement session.

4.6.2 Results and Discussion

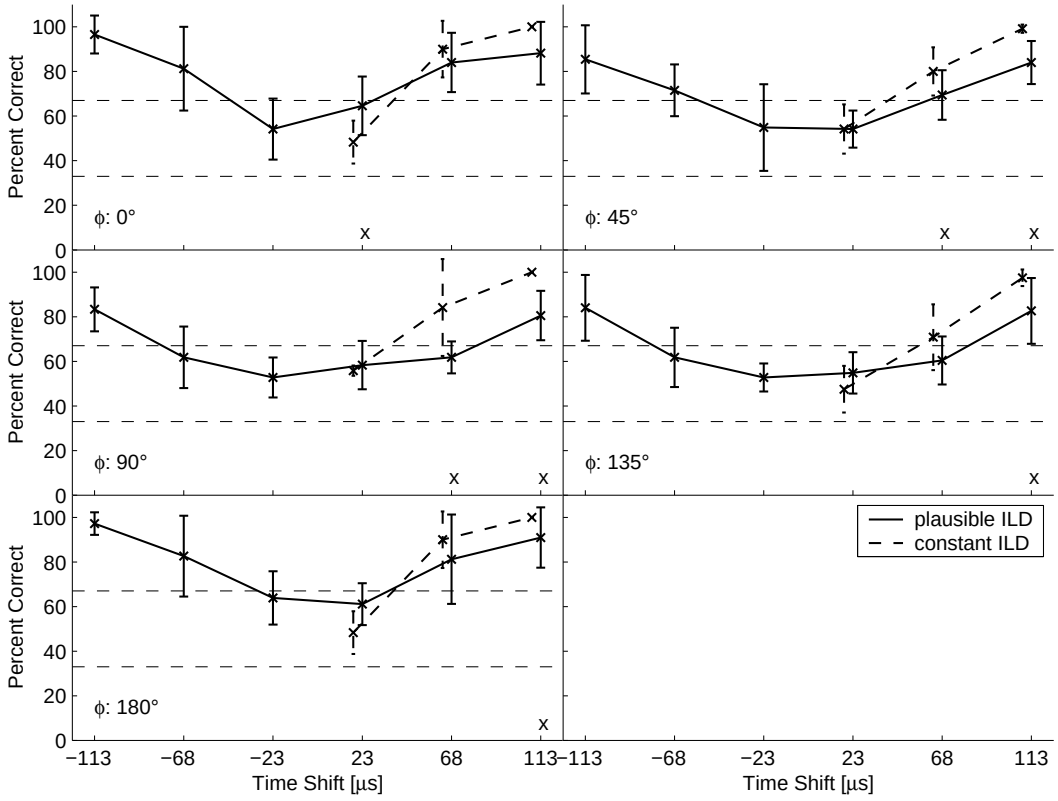


Figure 4.12: Percent correct responses averaged across subjects as a function of ITD variation for HRTF stimuli are depicted (solid lines, condition 'plausible ILD'). Dashed lines give percent correct responses for flat spectrum stimuli (condition 'constant ILD'). The dashed horizontal lines mark the thresholds for deviation from chance performance and the error bar indicate the inter-individual standard deviation.

Percentage correct identifications of manipulated (i.e. interaural time shifted) stimuli averaged across subjects are depicted in Figure 4.12. The percentage of correct responses is plotted as a function of the ITD variation for both measurement conditions. The solid lines connect data for the 'plausible ILD' condition and the dashed lines connects data for the 'constant ILD' condition. Significant differences ($p<0.05$) between the detection

rates in both conditions are marked by a small 'x' at the bottom of each sub-plot. The horizontal dashed line mark the thresholds for deviation from chance performance. Each subplot shows data from a different source positions in azimuth.
In the 'plausible ILD' condition the detection rate for $\Delta\tau = \pm 23\mu s$ is below the significance threshold for any angle of source incidence. However, for $\phi = 0°$ and $\phi = 180°$ percent correct responses are very near to the threshold. A delay of $\pm 68\mu s$ (3 samples) introduced to the lagging ear can be detected at $\phi = 0°, 45°, 180\circ$ but not significantly at $\phi = 90°, 135°$. If the delay is further increased (i.e. $\pm 113\mu s$ delay) it is significantly detected, independent from the angle of sound incidence. It should be noted, that the error bars for a delay of the lagging ear of $\pm 68\mu s$ show, that the sensitivity to the introduced delay varies considerably across subjects. Especially at $\phi = 0°$ and $\phi = 180°$ some subjects are below the significance threshold even for $\Delta\tau = 66\mu s$ and some are above the threshold for $\Delta\tau = 22\mu s$.
The detection rates for the flat spectrum stimuli (condition 'constant ILD') are higher for all ITD variations than the detection rates for the empirical stimuli if the detection rates deviate from chance performance. However, the detection rates are not significantly different for each condition, as can be seen in Figure 4.12 where only statistically significant differences are marked by the 'x' at the bottom of each plot.
The detection performance for $\Delta\tau = 23\mu s$ is always at chance level. In contrast, ITD variations of $68\mu s$ and $113\mu s$ were detectable for subjects independent from the source azimuth.

An estimate of the average ITD JNDs in both conditions was calculated from the

Azimuth [°]	0°	45°	90°	135°	180°
plausible ILD $\Delta\tau > 0$	23	63	78	81	30
plausible ILD $\Delta\tau < 0$	43	56	78	82	36
constant ILD $\Delta\tau > 0$	43	45	41	61	43

Table 4.3: ITD JNDs in microseconds obtained by calculating the intersection of the psychometric function with the detection threshold. The two different rows in the table indicate, if the target ITD was greater or smaller than the reference ITD.

curves given in Figure 4.12 by computing the intersection of the psychometric function with the detection threshold for deviation from chance performance. Since the psychometric functions were symmetrically measured around the reference ITD in the 'plausible ILD' condition ($\pm 1, 3, 5$ samples delay of the lagging ear), two thresholds for each angle of sound incidence are listed in Table 4.3.
Highest sensitivity to ITD changes can be observed for frontal sound in the 'plausible ILD' condition. The values for 0° azimuth are asymmetric, showing higher values for $\Delta\tau > 0$. An ANOVA on the individual data revealed that this effect is not significant (p=0.7).

The absolute size of the JND increases in the 'plausible ILD' condition by a factor of 3-4 as the angle of sound incidence is increased. An ANOVA was performed to assess the significance of the differences across angles. It shows, that the JNDs in the median plane are significantly different from the JNDs at lateral angles in the 'plausible ILD' condition ($p < 0.05$). No significant differences of the ITD JNDs were obtained for angles within the median plane (p=0.63). Furthermore, the differences in ITD JND at lateral angles are not significant. The tendency that the JND at 135° is higher than the JND at 45° is, therefore, also not significant (p=0.15).
In the 'constant ILD' condition the ITD JND at 135° deviates significantly from the ITD JND at the other azimuth positions (p<0.05).
In general, the detection rates in the 'plausible ILD' condition are smaller compared to the detection rates in the 'constant ILD' condition. Thus, two conclusions can be drawn from this finding. First, the ITD JND is influenced by the spectral shape of the stimuli and second, the ITD JND is decreased if additional spatial information is provided in the stimulus spectra. Hence, the results presented here support the hypothesis that the additional spatial information in the spectrum of the stimulus stabilizes the perceived location of the virtual object. In contrast, the alternative hypothesis that the additional spatial information generates a virtual object that is more focused in its spaciousness and that, therefore, spatial displacements are easier to detect, is not supported by the results of this experiment.

Comparison to the literature

To enable an comparison of the ITD JNDs obtained in the present study with data from the literature the ITD and ILD values of the reference stimuli are listed in Table 4.4. In the last row of the table mean values averaged across subjects are shown. For lateral source positions the ITD is increased to the range of approx. $500\mu s$ to $800\mu s$, whereas the reference ILD is increased to the range of approx 10 dB to 14 dB.
ITD JNDs were intensively investigated in the literature (s. (Durlach and Colburn, 1979) for a review). The JND of a 500 Hz tone was found to be around $10\mu s$ (Hershkowitz and Durlach, 1969) and for Gaussian noise between $12.3\mu s$ and $62.2\mu s$ depending on subjects (Mossop and Culling, 1995; Kinkel, 1990). To the knowledge of the authors, the lowest ITD JND was found for noise bursts to be approx. $6\mu s$ (Tobias and Zerlin, 1959).
In a study of Kinkel (1990) the ITD JND was measured for 1/3 octave bandpass noises with center frequencies of 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz as a function of the reference ITD and ILD. For zero ITD and ILD the obtained ITD JND is within the range of $20\mu s - 60\mu s$. An increase of the ITD and ILD reference values to $600\mu s$ and 15 dB, respectively, raised the ITD JND for most stimuli to the range of approx. $60\mu s$ to $160\mu s$. Only an increase of the reference ILD for the 500 Hz and 1 kHz stimuli did not result in a substantial increase of the ITD JND. Thus, on the average the ITD JND

is increased by a factor of approx. 3-4. This increase is consistent with the results of the present study. However, in this study a broadband noise was used. Therefore, a detailed comparison of the obtained ITD JNDs seems to be unappropriate.
The method used in this study limited the ITD JND to a minimum of $22\mu s$ due to the sampling rate of 44.1 kHz. The results from the literature show that lower JNDs have been found (e.g. (Domnitz, 1968; Hershkowitz and Durlach, 1969)). Therefore, the method was not appropriate for capturing the whole range of possible ITD JNDs. However, the results show that subjects were at or below the threshold in the most sensitive case. Hence, it can be assumed that for the stimuli used in this study, the obtained thresholds are representing the actual binaural temporal resolution.
In spite of differences in the methods, the ITD JNDs obtained in the current study are within expectations, both for the empirical and the flat spectrum stimuli.

Subjects \ Azimuth	ITD 0° ILD	ITD 45° ILD	ITD 90° ILD	ITD 135° ILD	ITD 180° ILD
RH	0 | 0.1	560| 11.2	800 |12.9	580| 7.8	0 | 0.7
IB	0 | 0.4	620| 13.0	840 |11.1	560| 6.6	0 | 0.6
HR	0 | 0.4	580| 12.8	820 |13.4	660| 10.5	0 | 0.5
HK	-60 | 3.3	460| 11.3	740 |14.0	640| 11.8	60| 0.0
JO	40 | 0.0	580| 11.1	780 |14.5	560| 5.5	-40| 3.1
MK	-40 | 0.3	500| 10.3	780 |16.0	620| 10.0	40| 0.5
Ø	23.3| 0.8	550| 11.6	793.3|13.7	603| 8.7	10| 0.9

Table 4.4: ITD and ILD of the individual reference stimuli. Dimensions of ITD and ILD are μs and dB, respectively.

4.7 Summary and general discussion

The general aim of the present study was to assess the sensitivity of subjects to deviations of the individual physical localization cues (described by HRTFs). Therefore, detection rates for manipulations of the individual HRTFs of 10 subjects were measured. Two kinds of spectral manipulations were applied to the HRTFs. The first manipulation (spectral detail reduction) is based on the work of Kulkarni and Colburn (1998). The findings of the present study are generally consistent with their results. A high amount of spectral detail can be removed from the HRTF spectra without affecting the perceived stimulus positions of a virtual stimulus (16 cestral coefficients were sufficient in the study of Kulkarni and Colburn and 16-32 were needed in the current study). For this amount of smoothing the frequency variation in the low frequency region is almost completely smoothed out. Hence, low frequency components seem to have a small contribution to the spatial perception because subjects were not able to detect a spatial displacement.

This finding is in contrast to the result from Asano et al. (1990) that the spatial information in the low frequencies (f<2 kHz) aid to resolve front-back confusions. It could be that due to the presentation of virtual stimuli already front-back confusion occurred and that subjects, therefore, were not able to detect changes in the stimuli. However, as pointed out before, the results of this study are consistent with the findings of Kulkarni and Colburn. In their study virtual stimuli were compared to a real sound source and, hence, front-back confusions introduced by smoothing would have been detected easily. Hence, it can be concluded that the lack of spectral information in the low frequency range does not affect the spatial perception at least for broadband stimuli.

The ILD deviations caused by cepstral smoothing at discrimination threshold are well correlated to the perceptual data, i.e. approximately the same ILD deviation in different situations (that roughly coincides with the ILD JND values from the literature) corresponds to the detected change in HRTF. Thus, the ILD deviation was used as a binaural distance measure for the differences of the physical localization cues. The results indicate that a mean ILD deviation (averaged across frequency channels of a Gammatone filter bank) of approx. 0.8 dB is detectable for sound incidence out of the median plane. This detection threshold is increased to 1.2 dB for sound incidence from lateral source positions.

The results given above are based on spatial displacements of the virtual stimuli because the source spectrum of the white noise was roved spectrally for each stimulus. The investigation was extended to non-spatial cues, like timbre, by presenting an unscrambled white noise and a click train stimulus to the subjects. Subjects were able to detect the manipulated HRTFs with higher detection rates compared to the scrambled white noise conditions for both, the white noise and the click train stimuli. Furthermore, the detection rates for click train stimuli were below the rates for white noise stimuli. This result is surprising because the spectral variation is the same for both stimuli. One hypothetical explanation is that in the click train condition subjects' attention was focused on the stimulus pitch of the click train which is introduced by the repetition rate of the clicks. This pitch is not changed by smoothing and, hence, subjects were less able to use the spectral variations as a detection cue. It can be concluded that for more complex stimuli than white noise (for instance, for music) the HRTF spectra can be smoothed by a higher degree without affecting the spectral timbre.

The detection thresholds for timbre variations, computed from the mean level differences of the smoothed and original HRTF spectra of the right ear, showed that for unscrambled white noise the monaural HRTF spectra may not deviate by more than 0.5 dB for sound incidence from the sides. An even lower threshold was computed for frontal sound incidence. For the click train condition a threshold could only be computed for 45° sound incidence (1 dB). In comparison to the threshold for the white noise stimulus it is increased by a factor of approx. 2.

In experiment II the sensitivity to a more complex spectral transformation was inves-

tigated that also shift the peaks and notches of the HRTF spectra. This was done by transforming the macroscopic spectral shape of the individual HRTF spectra to the shape of dummy head HRTF spectra ('spectral morphing'). As expected, subjects were very sensitive to the introduced manipulations. Again, the ILD deviation that is introduced by the 'spectral morphing' procedure served as a distance measure because a correlation analysis showed that the perceptual data is well correlated to this measure. The detection thresholds obtained from this distant measure are basically the same as for the spectral smoothing condition, if the scrambled white noise was presented to the subjects. For frontal sound incidence the thresholds are slightly lower than for the other directions. It can be assumed that the very obvious cue of front-back confusions aid to detect the manipulated HRTFs. This is likely because the 'spectral morphing' transformation distorts the spectral information (i.e. the center frequencies and the amplitudes of the peaks and notches) that is responsible for resolving the front-back confusions. In contrast, front-back confusions were not reported by the subjects if only the spectral detail of the HRTFs was reduced. However, although the front-back confusions introduce another detection cue, the thresholds described by the binaural distance measure (i.e. the deviation of the ILD that is introduced by 'spectral morphing') are almost the same for spectral detail reduction (for the scrambled white noise stimulus) and 'spectral morphing'. It can be concluded that the average ILD deviation across critical bands provides an appropriate measure for spatially relevant changes of the HRTF spectra.
In the last experiment presented in this study, the sensitivity to changes of the ITD was investigated. To investigate if the ITD JND is affected by the plausibility or consistency of the localization cues, two conditions were tested: First, detection rates for ITD variations of white noise stimuli convolved with individual HRIRs were measured. In a further condition, ITD JNDs were measured for white noise stimuli that exhibited the same ITD but had a constant ILD across frequency which is matched to the mean ILD (averaged across frequencies) of the individual HRTFs. Two hypotheses were tested to predict the differences of the detection rates. The first assumes that detection rates are higher for individual HRTFs because the virtual object is more focused in its spaciousness. The second predicts lower detection rates for the flat spectrum stimuli because the localization cues are less consistent and the virtual object is, therefore, less robust against distortions of one localization cue. The results showed that detection rates were *higher* for the less focused flat spectrum stimuli. Hence, more consistent localization cues seem to stabilize the virtual perception of a spatial acoustical object. This is a remarkable outcome since in traditional psychoacoustics it is assumed that interaural time discrimination is largely independent from object properties (such as, e.g. 'spatial diffusiveness'). For both conditions, the ITD JNDs calculated from the detection rates were within the expectations given by results found in the literature.

HRTFs for virtual auditory displays

The 'spectral morphing' procedure of the second experiment is of further interest. Using this method, perceptual relevant distances of individual HRTFs from different subjects can be described quantitatively. This can be done by calculating the morphing factor α, for which the ILD differences given by the distance measure D_{bin}, is above the appropriate detection threshold. For perceptually distant HRTFs low values of α are expected, whereas α is expected to be near to 1 for HRTFs that provide a similar spatial perception. Therefore, HRTFs can be grouped in perceptually similar HRTFs by using α as predictor for the perceptual distance.
However, the value α describes only perceptual distances of the ILD. The ITD is not taken into account by this measure. The results of the study show that thresholds for ITD deviations of non-individual HRTFs are well described by the threshold obtained in the literature. The ITD JND for empirical HRTFs is increased by additional localization cues and, therefore, the results for the ITD JND that can be found in the literature provide an lower limit for the ITD JND.

Chapter 5

Lead discrimination suppression in reverberant environments

Abstract

In reverberant environments the position of a sound source is dominated by the direct sound (the lead), whereas the spatial information of the reflections (the lag) is suppressed. Little attention has been paid in the literature to discrimination suppression of the direct sound in presences of a reflection and the results are not consistent. Thus, discrimination experiments were conducted to find out if in a natural listening environment the evaluation of the spatial information in the direct sound is processed in the same way as in a non-reverberant environment. The task of the subjects was to detect manipulations in the spatial information of the direct sound under reverberant and non-reverberant conditions. A 500 ms white noise stimulus was convolved with individual head related impulse responses (HRIRs) under the non-reverberant condition. In the reverberant condition binaural impulse responses of a seminar room (excluding the direct sound) were added to the HRIRs. Three manipulations were applied to the HRIRs: I) spectral smoothing, II) transformation of the macroscopic spectral shape ('spectral morphing') and III) ITD variations. The results show that for all three experiments the detection rate of the manipulations of the direct sound are significantly reduced under the reverberant condition. Thus, in a reverberant environment the contribution of the direct sound to the spatial perception is reduced. It is hypothesized that the lead discrimination suppression is due to further localization cues in the reflections that stabilize the perceived localization of the stimulus and make it more robust to changes in the direct sound. Due to the discrimination suppression in the spatial information of the lead less individual information is, therefore, needed in the direct sound in reverberant environments.

5.1 Introduction

One of the most important phenomenons of auditory localization for our daily life is the ability to localize the position of a sound source in a reverberant environment. The acoustical reflections produced by the environment are delayed, transformed (e.g. by absorption) copies of the original signal that are added to the direct sound in the ear canal. The sound originating from the source position is always leading the sequence of signals reaching the ear. This fact is used by the auditory system to localize the sound source position by giving precedence to the first wave front (see e.g. (Wallach *et al.*, 1949; Blauert, 1971)) and suppressing the spatial information of the lagging sounds. This effect is, therefore, called precedence effect and most often investigated by two stimuli measurement paradigms in which the reduction of spatial information in the second stimulus (the lag) in presence of the first stimulus (the lead) is investigated (e. g. (Wallach *et al.*, 1949; Perrott *et al.*, 1989; Haas, 1949; Zurek, 1980; Shinn-Cunningham *et al.*, 1993)). A comprehensive review of the precedence effect is given by Litovsky et al. (1999). From these investigations it is known, that the lag contributes only little to the perceived azimuth position of the sound but affects non-spatial cues like loudness and stimulus timbre (Blauert, 1974; Freyman *et al.*, 1998).

Nevertheless, the spatial perception is enhanced in reverberant environments. The energy ratio between the direct sound and the reflections is used by the auditory system as a cue for distance perception (e.g. (Bronkhorst and Houtgast, 1999)). Thus, the localization cue provided by reverberation differs considerably from those provided by the direct sound only. In a non-reverberant environment the direct sound recorded at the eardrum is equivalent to the head related impulse responses (HRIRs) convolved with the sound emanated from the source position. The HRIRs (or their frequency domain representations, the head related transfer functions (HRTFs)) contain all spatial information that can be used by the auditory system to estimate the source position. HRTFs describe binaural (interaural time difference, ITD and interaural level difference, ILD) and monaural (spectral filtering due to interference effects and pinna filtering) localization cues that can be exploited to calculate an estimate of the source direction. However, distance cues are only rudimentarily inherent in the head related transfer transfer functions (HRTFs) for source positions at a distance below 1 m especially in the median plane (Brungart and Rabinowitz, 1995). It is likely that the position of a sound source is determined by integrating over all available localization cues. This implies that redundant or additional localization cues increase the robustness of the spatial perception against distortions in one localization cue. Therefore, the sensitivity to variations of the localization cues is expected to be decreased in reverberant environments.

Furthermore, if the sensitivity to differences of individual HRTFs is reduced by adding reverberation, the amount of individual information needed in the direct sound is expected to be decreased. Thus, by incorporating reverberation to virtual auditory displays

not only the distance perception is enhanced but also the need for individual HRTFs to generate the direct sound is expected to be decreased. This would save costs and effort for the development of individual virtual environment generators.
The localization cues which can be extracted from HRTFs are ILD, ITD and monaural spectral filtering. Thus, to test the hypotheses given above three different discrimination experiments are conducted in which the detection performance of subjects to changes of the localization cues in the direct sound is compared under reverberant and non-reverberant conditions. The assumption is that if reverberation stabilizes the perceived position of an acoustical object higher variations of the localization cues can be introduced in the reverberant condition compared to the non-reverberant condition, without affecting the spatial perception.
To incorporate a test of the second assumption that less individual information is needed in the direct sound in a reverberant environment, two spectral manipulations of the HRTF spectra of the direct sound were chosen that reduce the amount of individual spectral information. In the first experiment cepstral smoothing was applied to the HRTF spectra to reduce the spectral detail. The investigation on the inter-individual standard deviation of the HRTF spectra across subjects given in Chapter 3 shows that the individual information in the HRTF spectra is reduced by cepstral smoothing. Hence, if a higher amount of spectral detail can be reduced in the reverberant condition without affecting the spatial perception, less individual spectral information is needed.
The second manipulation transforms the individual spectral shape of the HRTF spectra to the shape of dummy head HRTF spectra ('spectral morphing') which deviate strongly from individual ones (see Chapter 3). Again, if less individual information is needed in the direct sound in a reverberant conditions it is expected that more non-individual spatial information can be introduced to the stimuli without causing a spatial displacement.
In a further experiment the sensitivity to variations of the ITD is investigated under reverberant and non-reverberant conditions. In Chapter 3 it is shown that the inter-individual differences of ITDs obtained from different subjects averaged across source locations in the horizontal plane is approx. $40\mu s$. This values is within the range of the ITD JND (e.g. (Koehnke *et al.*, 1995)). If the ITD JND is further increased in reverberant environments it can be assumed that in this case individual ITD information is not needed for creating perceptually accurate virtual acoustic stimuli.

Related studies

Investigations that are related to the current study compare the absolute localization performance in reverberant and non-reverberant conditions or investigate discrimination suppression of the lead in presence of a lag.
In a study of Hartmann (1983) it was found that the absolute localization accuracy of a 500 Hz tone is *not* affected by changing the amount of reflections of a concert hall from

an absorbing condition to a reflecting condition. On the other hand, Begault (1992) observed that the localization acuity to speech stimuli created with non-individualized HRTFs *is* reduced if synthetic reverberation is added to the stimuli but the distance perception was enhanced. Thus, from absolute localization experiments conducted in the literature a clear picture concerning the differences in the localization accuracy under reverberant and non-reverberant conditions can not be extracted.
Although discrimination tasks are more sensitive to changes of the stimuli than absolute localization tasks, studies conducting discrimination experiments also do not show consistent results. In a study by Litovsky and Macmillan (1994) the change of the minimum audible angle (MAA) of the lead with and without the presence of the lag was investigated. No significant influence of the lag on the MAA of the lead was found. In a later study (Litovsky, 1997) a slight reduction of the MAA of the lead in the presence of the lag was observed. In this study longer stimuli were used and different groups of subjects with respect to their age.
In a study of Tollin et al. the ITD JND was measured for click stimuli with and without the presence of a lag. It was shown, that the ITD JND of the lead is increased by a factor of two if a lag was present. However, in a reverberant environment multiple reflections are following the direct sound. For distance perception it is likely that the auditory system averages across the first 6 ms (Bronkhorst and Houtgast, 1999). Therefore, it can be assumed that the decrease of the ITD JND in presence of a lag is higher for multiple reflections.

5.2 Methods

The general task of the subjects was to identify spatial displacements of virtual stimuli that were created by manipulated HRIRs convolved with a white noise stimulus. A two interval-two alternative forced choice (2I-2AFC) measurement paradigm was used for the experiments I-III. A stimulus sequence of four stimuli grouped in two intervals was presented to the subjects. The task of the subjects was to identify the interval in which one of the two stimuli deviated with respect to its spatial position.
In each experiment both reverberant and non-reverberant stimuli were presented in separate measurement sessions. The non-reverberant stimuli were created by applying individual HRTFs measured in an anechoic room to a reproducible scrambled white noise stimulus (see Chapter 3 for a description of the HRTF measurements).
For the generation of the reverberant stimuli non-individual binaural room impulse responses of an asymmetric seminar room were added to HRIRs by exchanging the direct sound of the room impulse responses with the HRIRs. The non-individual room impulse responses were measured for one selected listener (subject 'JO'). It can be assumed that the non-individual room impulse responses do not restrict the generality of the results

for the following reason: The reflections are spectrally filtered copies of the direct sound radiated from directions that primarily depend on the room and the orientation of the source relative to the listener. If non-individual room impulse responses are added to the direct sound, the reflections are filtered with non-individual HRIRs, which deviate in ILD and ITD in comparison to the corresponding values obtained from individual HRTFs. However, it is known from investigations on the precedence effect that ILD and ITD JNDs are increased for the lagging sound indicating that the sensitivity to individual cues in the reverberation process is greatly reduced. For instance, the ITD JND is decreased by a factor of 4-5 for reflections within the first five milliseconds (Tollin and Henning, 1998) and the MAA of the lag is increased by a factor of 2-6 depending on the time delay between lead and lag (Perrott *et al.*, 1989). Hence, it can be concluded that the differences in the localization cues introduced by using non-individual room impulse responses should not affect the results.

5.2.1 Subjects

A total number of six subjects (1 female and five male) participated in the experiments. The subjects were aged from 27 to 34 years and had normal hearing. The number of subjects participating in each of the three experiments is listed in the second row of Table 5.1. All subjects were members of the Physics and Psychology department of the University of Oldenburg and had extensive experience in psychoacoustic tasks. Each subject participated in both reverberant and non-reverberant experimental conditions. The author participated in all measurements.

	Exp I	Exp II	Exp III
Trials p. Cond.	40	30	24
Subjects	6	5	5
Sessions	4	6	6

Table 5.1: Number of trials per stimulus condition and measurement situation (row I), number of subjects per measurement condition (row II) and number of sessions (row III).

5.2.2 Stimuli

The same frozen white noise stimulus was used for all experiments. The noise sample had a duration of 500 ms. The on- and offsets were ramped by 5 ms squared cosine ramps. In the experiments I and II the spectrum of the white noise was scrambled randomly before it was convolved with the target or reference HRTFs. Scrambling was performed

in 1/6 octave bands by up to ± 5 dB. In experiment III the noise spectrum was left unchanged.
For each of the experiments I-III, anechoic and reverberant virtual stimuli were prepared. The first group consisted of the white noise sample convolved with manipulated HRTFs without reverberation. Under the reverberant condition reverberation was added to the manipulated HRTFs of the first group. After preparation of the target and reference HRTFs, they were convolved with a white noise sample.

5.2.2.1 Non-reverberant stimuli

Individual HRTFs were measured for each subject (see Chapter 3). Three different kinds of manipulation were applied to the HRTFs.

Experiment I: Reduction of the spectral HRTF details. The spectral detail of the HRTF spectra is reduced by cepstral smoothing. To smooth out the HRTF spectra the logarithm of the absolute HRTF spectra is reconstructed by a Fourier Series

$$log(|\hat{H}(k)|) = \sum_{n=0}^{M} \tilde{C}(n) \cos \frac{2\pi nk}{N} \tag{5.1}$$

where $\tilde{C}(n)$ can be obtained from the cepstrum $C(n)$ of the HRTF spectrum $H(k)$

$$C(n) = \sum_{k=0}^{N-1} \log |H(k)| e^{\frac{i2\pi kn}{N}} \tag{5.2}$$

$$\tilde{C}(n) = \begin{cases} \frac{(C(1)+C^*(1))}{2} & : \quad n = 0 \\ (C(n) + C^*(n)) & : \quad 1 <= n <= N/2 \end{cases}$$

The upper limit M of the series defines how many cosine terms are used for a reconstruction of the spectrum. If M equals $N/2$ (N is the length of the corresponding impulse response) no smoothing occurs. For $M < N/2$ cosine terms representing amplitude fluctuations of higher orders are neglected. Therefore, the spectrum is smoothed out by decreasing M.
The reference stimulus was created by using $M = 128$ coefficients to reconstruct the HRTF spectrum. Target stimuli had HRTF spectra with $M = 8, 16, 32, 64$ terms of the Fourier Series. The phase of each HRTF was calculated from $\hat{H}(k)$ as minimum phase plus a frequency independent group delay to incorporate the ITD.

Experiment II: Transformation of the macroscopic spectral shape ('spectral morphing'). The macroscopic shape of the target HRTF spectra was manipulated by transforming the individual HRTF spectra to the corresponding HRTF spectra of dummy head HRTFs. A description of the dummy head is given by (Trampe, 1988)).

This process is called *'spectral morphing'* throughout the study. It replaces the individual macroscopic spectral HRTF shape by the structure obtained from the HRTF of a dummy head. By Equation 5.3 the absolute spectrum of the individual HRTF $|H|$ is transformed into $|\hat{H}_\alpha|$. The parameter α describes the degree of morphing. $|H_{MS}|$ and $|D_{MS}|$ are representing the macroscopic spectral shape of the individual and the dummy head HRTF, obtained by 6^{th} octave smoothing. By increasing α from zero to one the proportion of the macroscopic dummy head spectra is increased. For $\alpha = 0$ $|H|$ equals $|\hat{H}_\alpha|$ and for $\alpha = 1$ the individual macroscopic shape is completely replaced by the dummy head shape.

$$|\hat{H}_\alpha| = (1-\alpha)|H| + \alpha|H|\frac{|D_{MS}|}{|H_{MS}|} \tag{5.3}$$

The reference HRTF was created by $\alpha = 0$ and the targets were calculated by setting α to 0.1-0.9 with $\Delta\alpha = 0.2$. The phase of the HRTFs is calculated from $|\hat{H}_\alpha|$ as minimum phase plus a frequency independent group delay.

Experiement III: ITD variation. In this experiment the interaural time delay between the left and right ear HRTFs was manipulated. The ITD of the reference stimuli were given by the ITDs of the empirically measured HRTFs. Targets were created by shifting the impulses responses of the lagging ear (left) by $\pm 1, 3, 5$ samples. Due to the sampling frequency of 44.1 kHz ITD variations of approx. $\pm 22\mu s$, $67\mu s$ and $110\mu s$ were introduced.

5.2.2.2 Reverberant stimuli

In each of the experiments I-III non-reverberant stimuli and reverberant stimuli were presented in separate sessions. The non-reverberant stimuli were noise samples convolved with the target or reference HRIRs as described before. Under the reverberant condition reflections were added to the HRTFs and then convolved with the noise sample. To illustrate the time pattern of the room reflections the envelope of the room impulse responses measured by microphones in the ear canals of subject 'JO' is shown in Figure 5.2. Each panel shows the first 40 ms of the impulse response measured in the left (thin lines, shifted in amplitude for visibility) and right ear canals (thick lines) for the source azimuth given in the panel (see Figure 5.1 for a sketch of azimuth positions in the room). It can be seen from this figure that the direct sound is clearly separated from the early reflections. The direct sound is located at approx. 6 ms at the right ear and shifted by the ITD at the left ear. At approx. 11 ms two first reflections separated by approx 1 ms, can be identified. Because the time delay between direct sound and the first two reflections is independent of azimuth, it is likely that these are reflections from

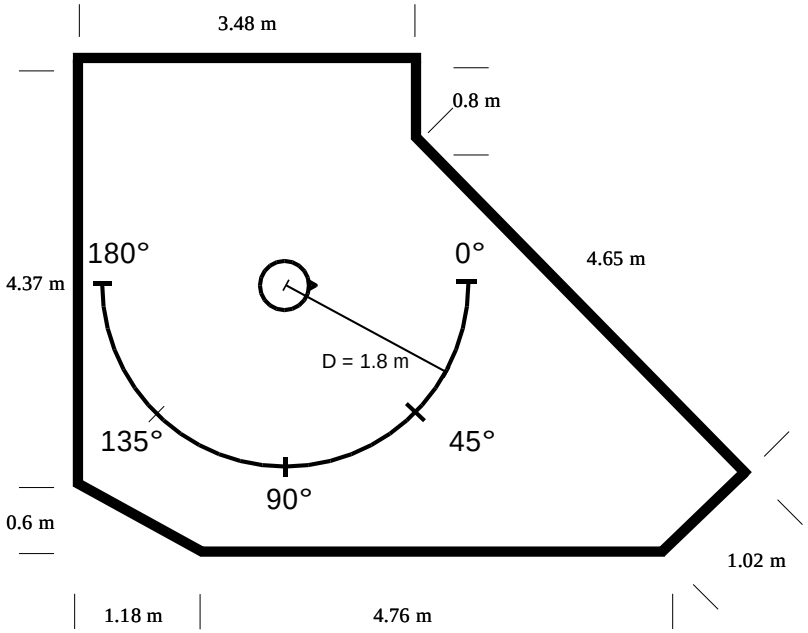


Figure 5.1: Floor plan of the room in which impulse responses were measured. The position of the center of the head was chosen by the restriction that a half circle with a radius of 1.8 m can be installed in the right hemisphere. Impulse responses were measured at the positions marked on the half circle.

the floor and the ceiling. Various reflections from different azimuths are succeeding the first reflections in intervals of 3 ms to 10 ms. For lateral angles, a prominent reflection at the left ear at approx. 12 ms can be identified. From Figure 5.1 it can be seen that this reflection is originated from the wall on the left side of the dummy head. After 40 ms late reflections evolve into a 'noisy' part of the impulse response (not shown here).
Target and reference stimuli in the reverberant condition were created by replacing the direct sound of the room impulse responses with the target and reference HRIRs, respectively. To give an example, the complete process of creating the reverberant stimuli in experiment I is described. First, HRTFs from all relevant azimuthal positions were measured for each subject individually in the anechoic room (see Chapter 3). Smoothed versions of the HRIRs were calculated from the HRIRs by applying cepstral smoothing to the HRTF spectra (see Equation 4.2 in Section 4). Then, room impulse responses were measured from a selected subject ('JO'). The speaker for obtaining the room impulse responses was located at the same azimuths as it was for the HRTF measurements. Subsequently, the direct sound of the room impulse responses was replaced by the previously obtained HRIRs. The reflections were scaled in amplitude in a way that the direct sound and the HRIRs of the right ear have the same RMS values. Preparing the stimuli in this way ensured that the HRIRs in the non-reverberant measurement condition and the direct sound in the reverberant condition were the same.

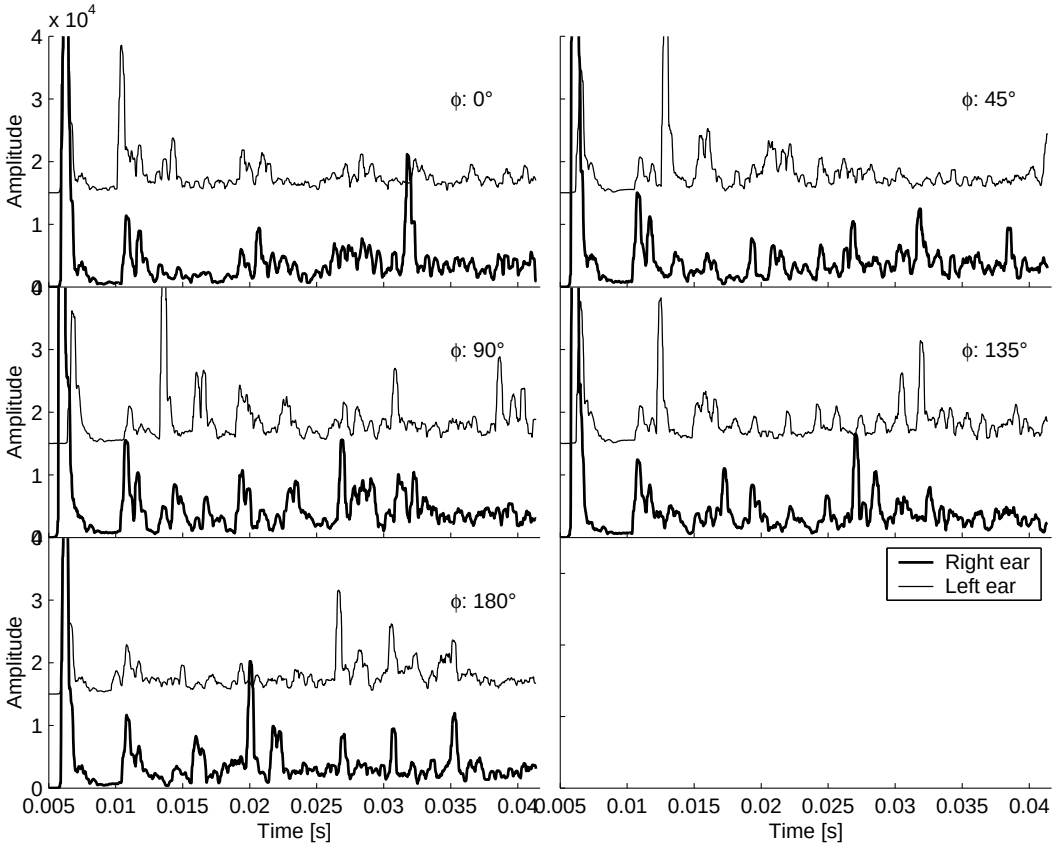


Figure 5.2: RMS values (averaged across 313μs time frames) of room impulse responses measured in the right (thick lines) and left (thin lines, shifted in amplitude for better visibility) ear canal. The sound source was positioned at the azimuth positions in the environment shown in Figure 5.1

5.2.3 Procedure

Subjects were seated in a sound isolated booth (IAC, Model No. 405A) in front of a window. The monitor of the computer controlling the experiments by running a MatLab script was located behind the window. Stimuli were presented to the subjects over a headphone (AKG 501) which was plugged into the output of a sound card (Soundblaster 128). The presentation level was set to a comfortable level for the subjects (approx. 70 dB A measured at the right ear of a dummy head for frontal sound incidence.).
For each trial a stimulus sequence consisting of four stimuli within two intervals was presented. One of the two intervals consisted of one reference and one target stimulus and the other interval consisted of two reference stimuli. The task of the subject was to identify the interval containing the target stimulus. The keyboard of the computer was used as input device. The position of the target within the stimulus sequence was chosen at random. Intervals were separated by 300 ms pauses and stimuli within intervals by 100 ms delays.

For each trial the horizontal location of the stimulus sequence was randomly chosen out of five different azimuth positions ($\phi = 0°, 45°, 90°, 135°, 180°$). The stimulus positions were the same for all three experiments. The number of stimulus repetitions per stimulus condition is listed in the first row of Table 5.1. In the second row the number of subjects participating in each experiment is shown and in the last row the number of sessions that each subject had to attend. Two experimental conditions were conducted in each experiment. In the first condition non-reverberant stimuli were used and in the second condition reverberant stimuli were presented to the subjects.

5.3 Results

In Figures 5.3-5.5 the results of the three discrimination experiments are shown for both stimulus conditions. The organization of the plots is the same for each of the three experiments. In each subplot the percentage of correct responses is shown as a function of the manipulation parameter for a different azimuth angle. Data for non-reverberant stimuli are represented by crosses and for the reverberant stimuli by open rhombi. In all conditions mean values across subjects are shown. The horizontal dashed lines indicate the 95% significance threshold for deviation from chance performance.
No standard deviations or error bars are shown in order to simplify the plots. To analyze the significance of the differences between the reverberant and non-reverberant conditions, a non-parametric ANOVA (Kruskal-Wallis) was computed. If the differences are significant ($p < 0.05$) a box plotted by dashed lines is enclosing the corresponding data points. For high significance ($p < 0.01$) the box is plotted by solid lines.

5.3.1 Experiment I: HRTF smoothing

In Figure 5.3 the results for detecting the target stimuli with smoothed HRTF spectra are shown for the reverberant and non-reverberant conditions. Percentage of correct responses are plotted as a function of the number of smoothing coefficients.
The figure illustrates that 16 ($\phi = 0°, 90°, 180°$) to 32 ($\phi = 45°, 135°$) cepstral coefficients are sufficient for providing all spatial information in the non-reverberant condition. Significant reductions of the detection rates occur for all angles of sound incidence in the reverberant condition. The detection rates are not significantly different from chance performance for all angles of azimuth, except for 135°. For this azimuth the detection rates are above the threshold for 8 cepstral coefficients.
The differences in the detection rates between the non-reverberant and the reverberant condition are significant for 8 cepstral coefficients for all angles of sound incidence. The largest differences occur for 135° azimuth were they are highly significant for 8 to 32 smoothing coefficients. To quantify the detection differences in the reverberant and non-

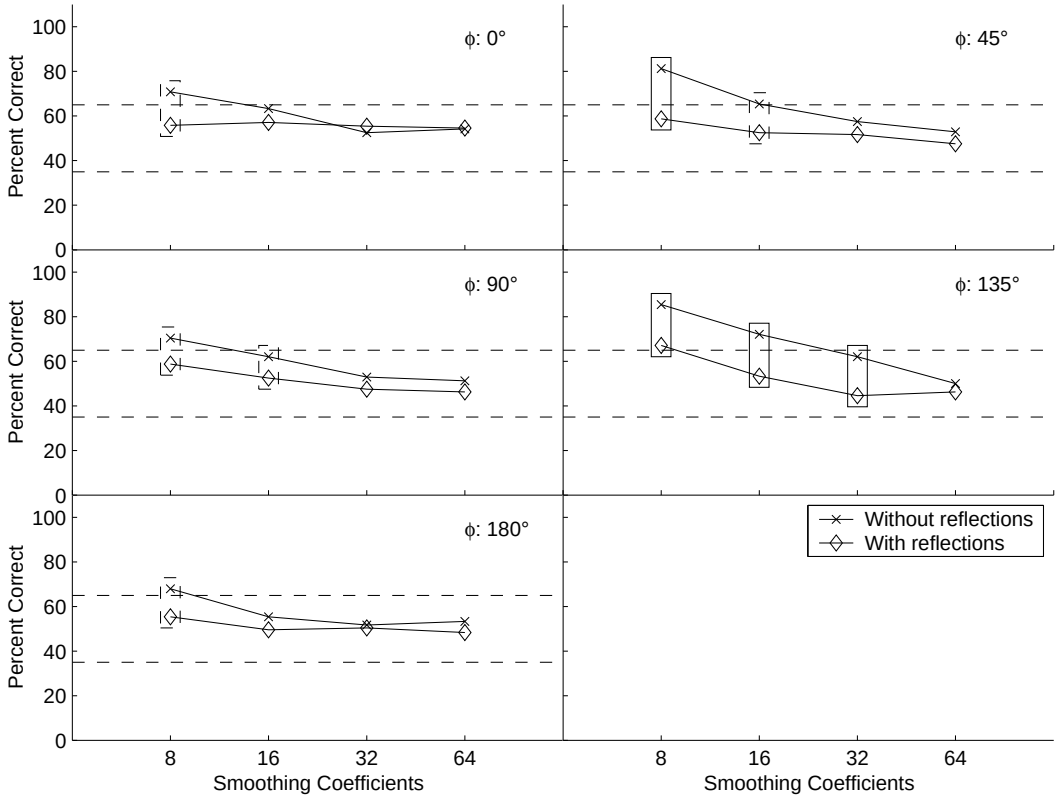


Figure 5.3: Detection rates for stimuli with smoothed spectra in reverberant (open rhombi) and non-reverberant (crosses) conditions. If the symbols are enclosed by a box the differences in the detection rates are significant (i.e. $p<0.05$) as indicated by dashed lines and highly significant (i.e. $p<0.01$) as indicated by solid lines.

reverberant condition, detection thresholds were computed and are listed in Table 5.2 for the reverberant (R) and non-reverberant (NR) condition. The thresholds are given in terms of the ILD deviation between the ILDs of the reference and target HRTFs. To calculate the thresholds, the psychometric functions were plotted as a function of the ILD deviation (averaged across frequency). The threshold was defined to be the ILD deviation for which the linear interpolation of the detection rates as a function of the corresponding ILD deviation intersects the 95% significance threshold for deviation from chance performance.

The ILD deviations were obtained by computing the ILDs of the target and reference

Condition \ Azimuth	0°	45°	90°	135°	180°
ILD deviation, R[dB]	>1.1	>2.1	>1.6	3.2	> 0.87
ILD deviation, NR[dB]	0.9	1.5	1.44	1.4	0.85

Table 5.2: ILD deviation thresholds for the detection of smoothed HRTFs in reverberant and non-reverberant conditions. If the detection rate was below threshold even for the strongest cue the thresholds are marked by a $'>'$ sign.

HRTFs in each filter bank channel of a Gammatone filter bank. ILD differences between target and reference stimuli were computed in each filter bank channel and averaged across frequency. This threshold was computed because the outcome of a correlation analysis was that the ILD deviation calculated in this way shows the highest correlation to the perceptual data in the non-reverberant condition (see Section 4.4).
If the detection rate is below the threshold for eight cepstral coefficients, the ILD deviation for this degree of smoothing is listed and marked by a $'>'$ sign to indicate that the threshold is above the listed value. Only for 135° of azimuth the detection rate is above threshold in the reverberant condition for eight cepstral smoothing coefficients. The threshold for this source direction is raised by a factor of two in this case. For the other angles of sound incidence it can be speculated that similar thresholds reductions occur.

5.3.2 Experiment II: Spectral morphing

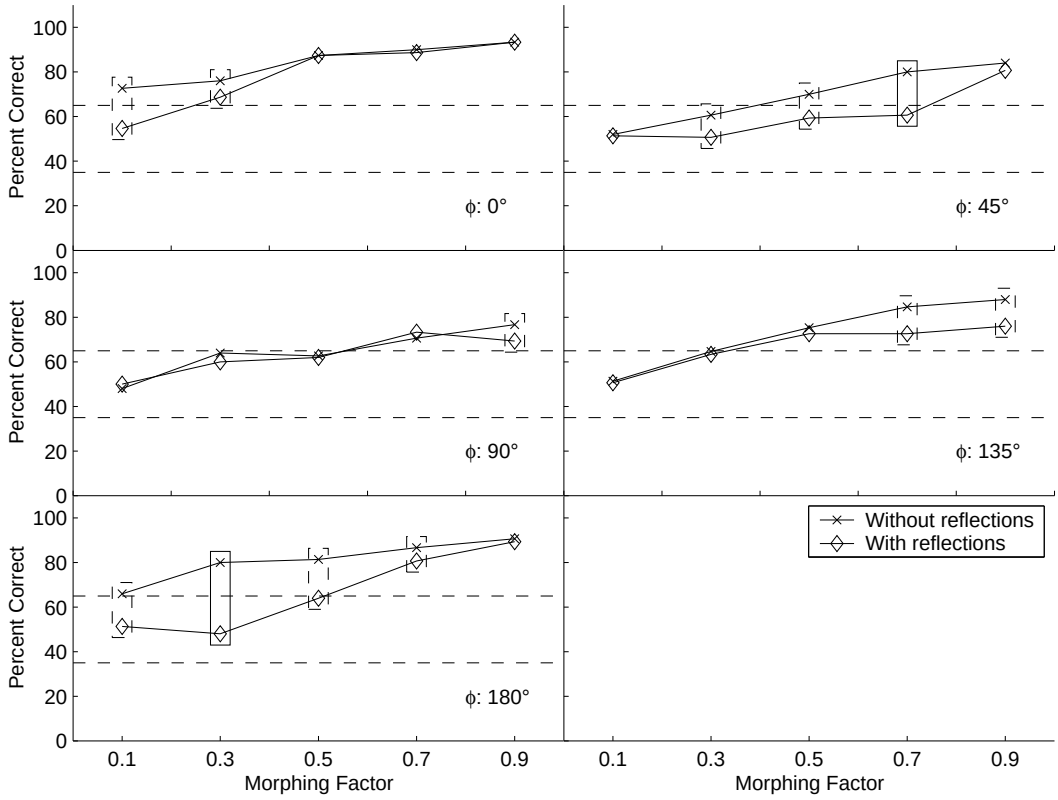


Figure 5.4: Detection rates for stimuli created with spectrally morphed HRTFs in reverberant (open rhombi) and non-reverberant (crosses) conditions.

The results of the 'spectral morphing' experiment are presented in Figure 5.4. The percentage of correct responses in the reverberant (rhombus symbol) and non-reverberant (crosses) condition are plotted as a function of the morphing factor α.

Condition \ Azimuth	0°	45°	90°	135°	180°
ILD deviation, R[dB]	0.75	2.2	1.51	1.63	1
ILD deviation, NR[dB]	<0.41	1.3	1.46	1.34	<0.32

Table 5.3: ILD deviation thresholds for the detection of spectrally morphed HRTFs in reverberant and non-reverberant conditions. If the detection rate was above the threshold even for the smallest cue the thresholds are marked by a $'<'$ sign.

It can be seen that subjects are highly sensitive to the 'spectral morphing' manipulation in the non-reverberant condition for sound incidence out the median plane (i.e., $\phi = 0°$ and $\phi = 180°$). The detection rates deviate from chance performance even for $\alpha = 0.1$. For lateral angles the sensitivity to the manipulation is reduced being lowest at 90° azimuth.
In the reverberant condition the pattern of the sensitivity as a function of source direction is changed. The lowest sensitivity can be observed at 45° azimuth and the highest for 0° and 135°.
Significant reduction of the detection rates (in comparison to the non-reverberant condition) can be seen for all angles of azimuth in the reverberant condition. The highest differences occur for $0°, 45°, 180°$ azimuth. However, the detection rates for 90° and 135° azimuth are nearly identical (i.e., at chance level for low values of α). Only at higher values of α significant differences can be seen. The thresholds listed in Table 5.3 were computed in the same way as in the HRTF smoothing experiment (s. Section 5.3.1). If the detection rate is not below the significance threshold (for instance at zero degree azimuth, non-reverberant condition) the ILD deviation for the lowest value of α is presented and marked by a $'<'$ sign. It can be seen that the detection thresholds are decreased by a factor of approx 1.7 for $\phi = 0°$ and 45°. For 180° of azimuth even stronger reduction of the sensitivity to the manipulation can be observed (> 3). As pointed out before no significant threshold differences between the reverberant and non-reverberant condition can be seen for $\phi = 90°$ and $\phi = 135°$.

5.3.3 Experiment III: ITD variation

In Figure 5.5 the results for the ITD variation experiment are shown. The number of correct responses in percent is plotted as a function of the ITD variation $\Delta\tau$. The crosses represent the non-reverberant condition and the rhombi represent the reverberant condition.
In general, the sensitivity to the ITD variation is reduced in the reverberant condition. For sound incidence out of the median plane the shape of the psychometric function is maintained, but the percent correct score is decreased by a nearly constant factor for all $\Delta\tau$. The differences between the percentage of correct responses in the reverberant and

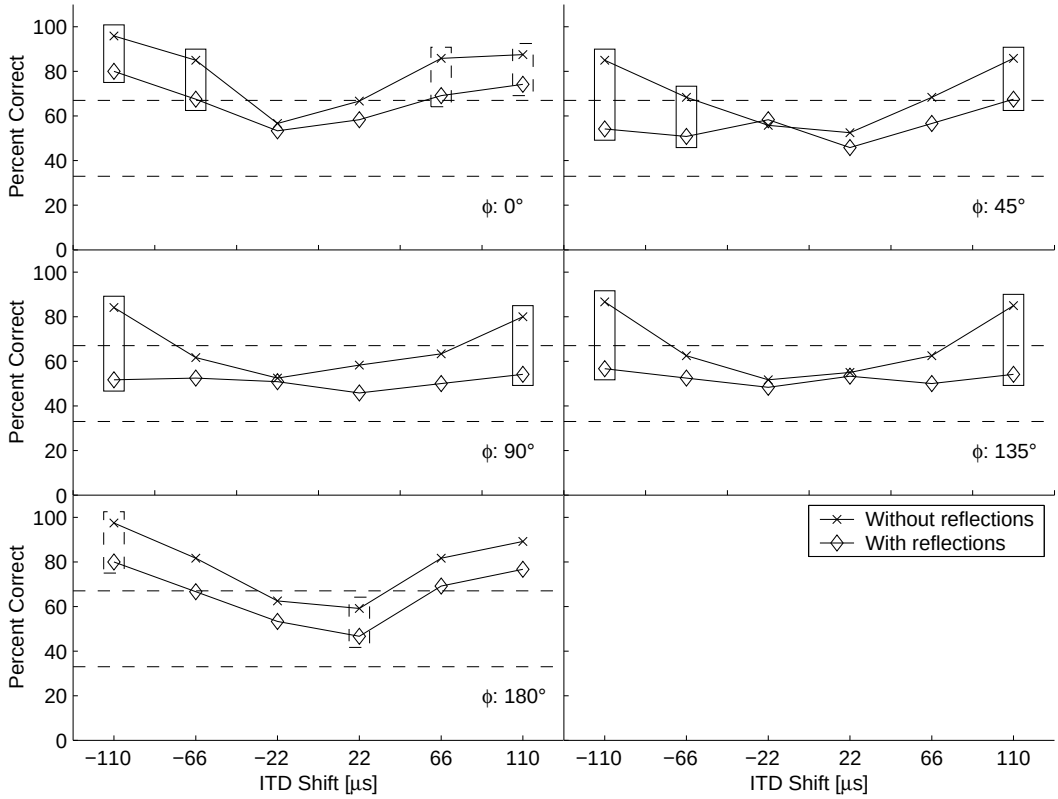


Figure 5.5: Detection rates for stimuli with shifted ITDs in reverberant (open rhombi) and non-reverberant (crosses) conditions.

non-reverberant condition at 0° of azimuth are significant for $\Delta\tau >= 66\ \mu s$. The reduction in sensitivity is similar at 180° compared to the frontal hemisphere and is significant for two introduced ITD variations. For lateral sound incidence ($\phi = 45°, 90°, 135°$) the sensitivity to the ITD variation in the reverberant condition is decreased below the significance threshold for all $\Delta\tau$. Only at 45° the average detection rate is slightly above the threshold for $\Delta\tau = 110\ \mu s$. The differences in percent correct responses in the non-reverberant and reverberant condition are highly significant for $\Delta\tau = 110\mu s$ at all lateral source positions. Detection thresholds were computed by calculating the intersection of the psychometric functions with the detection thresholds marked by the horizontal dashed lines. For the reverberant condition (R), this procedure was only applicable for source locations in the median plane. At lateral positions the detection rate is below the detection threshold. Threshold were calculated where possible and summarized in Table 5.4. Thresholds for source locations in the median plane are increased by a factor of approx. two in the reverberant condition.

Condition/Azimuth	0°	45°	90°	135°	180°
NR: $\Delta\tau < 0$	44	55	79	79	30
NR: $\Delta\tau > 0$	28	60	81	82	36
R: $\Delta\tau < 0$	66	> 110	79	> 110	69
R: $\Delta\tau > 0$	58	111	30	> 110	64

Table 5.4: Average detection thresholds of ITD variation manipulation are computed from the intersection of the psychometric function in Figure 5.5 with the 95% confidence level *NR* indicates the non-reverberant condition and *R* the reverberant case. Thresholds are given in μs.

.

5.4 Discussion

The general aim of this study was to investigate if the localization cues contained in the HRTFs (of the direct sound) are evaluated differently in a reverberant environment in comparison to a non-reverberant condition. Therefore, the sensitivity of three different types of manipulations were measured in both conditions. In the first experiment the spectral details of the HRTFs were reduced. The results show that the detection performance was reduced in the reverberant condition in comparison to the non-reverberant condition for all angles of sound incidence. Only at 135° azimuth the detection performance was above chance level in the reverberant condition for eight cepstral coefficients. For the other source positions no significant detection could be observed in the reverberant condition. For these source positions 4 or even 2 cepstral coefficients could be sufficient for providing all spatial information in the reverberant condition. The thresholds computed for 135° of azimuth showed, that the ILD deviation of the target to the reference ILD can be two times higher in the reverberant condition than in the non-reverberant condition.

The 'spectral morphing' experiment was intended to disturb the individual information in the macroscopic structure of the HRTFs. This was done by a stepwise transformation of the individual HRTF spectrum to the spectrum of a dummy head. The manipulation distorts the individual spatial information in the center frequencies of the peaks and notches of the HRTFs. The results of this transformation show, that for source positions in the median plane and for 45° azimuth the sensitivity to the manipulation is reduced in the reverberant condition. For 90° and 135° azimuth no reduction in threshold can be observed in the reverberant condition. Hence, the sensitivity of the auditory system to ILD changes (or changes to the spectral composition of the ILD) is not reduced due to reflections by the same amount for each angle of sound incidence. This result differs from the smoothing experiment, where the sensitivity was reduced for all angles of sound incidence.

In the third experiment the sensitivity to ITD variations was compared in reverberant and non-reverberant conditions. The sensitivity to the manipulation was found to be reduced significantly for all angles of azimuth. For source positions in the median plane the ITD JND is increased by a factor of two. If the sound is emanated from lateral positions, the detection rates for the target stimulus does not deviate from chance performance. However, the ITD JNDs are at least increased to $110\mu s$. This corresponds to an increase of the ITD JND by a factor of 1.4 for lateral positions. From these results it can be concluded that the detection thresholds are elevated by a factor of approximately two, both for the spectral variations and the variation of the ITD.

Relations to the precedence effect

To conclude, the sensitivity to changes in the binaural localization cues contained in the HRTF (the direct sound) is reduced in a reverberant environment. This can be explained by the following simple assumption. If the precedence effect would operate 'correctly' than the spatial information contained in the reflections would not disturb the localization perception. The spatial information would only be taken from the direct sound of the stimulus and, hence, the detection performance for reverberant stimuli would be the same as for non-reverberant stimuli. Because the sensitivity to the manipulations in the reverberant condition *is* reduced, it can be concluded that the precedence effect failes to operate 'correctly'. This means, that the early reflections influence the spatial perception of the direct sound to a certain degree.
In a study of Litovsky et al. (1994) no significant reduction of the minimum audible angle (MAA) of the direct sound (the lead) in presence of a reflection (the lag) has been found. MAA were measured in a single burst condition (without the presentation of the lag) and compared to the MAA of the lead in a lead-lag condition. Although slight differences occurred in the MAA, they were not significant. In this study very short noise stimuli (6 ms) were used. In a later study (Litovsky, 1997) the same stimulus conditions were compared to each other using longer stimuli and different groups of subjects (children and adults). Significant differences between the single burst and lead discrimination task were found for 25 ms noise bursts. In this case the MAA increased from 0.78° to 1.15°. This increase was even higher for children (1.55° to 4.4° for five year old children and 5.65° to 23.05° for 18 month old children). Hence, this investigation is consistent with the results of our study, showing that in a reverberant environment the sensitivity to changes in the direct sound is reduced.
In a MAA experiment both the ILD and the ITD of the target signal are varied. Hence, the reduction in sensitivity could be caused by a reduction of the ILD or ITD JND. To the knowledge of the author, the effect of the ILD JND increase in the direct sound in the presence of reflections has not been investigated yet by other studies. However, in a study of Tollin et al. (1998) the ITD JND of the lead was measured for precedence

effect stimuli (two clicks) and for single clicks. Threshold elevation factors (TEFs) were computed by calculating ratios of the ITD JND of the lead to the single click ITD JND. It was found that the TEFs increase as a function of the inter-click interval (ICI) separating the lead from the lag. For ICIs of 0.8 ms to 10 ms the TEF is approx. two. This indicates, that the ITD JND of the lead is two times higher in presence of the lag than without. This finding is confirmed by the results of experiment III, where the ITD is also increased by a factor of two in the reverberant condition (for sound incidence out of the median plane). It is remarkable that although realistic reverberation was used here instead of a single click in the study by Tollin et al. (1998) the reduction in sensitivity is the same in both studies. It can be concluded, that the effect is caused by early reflections of the room impulse responses and that the later reflections play a minor role.

Hypothesis: Perceptual stabilization by reflections

In a non-reverberant environment the localization acuity is determined by the spatial information contained in the HRTFs and the ability of the auditory system to extract the spatial cues from the HRTFs. In a reverberant environment the direct sound, (which is identical to the HRIR for ideal click stimuli) also determines the spatial perception of the source location. However, the results of this study show, that due to reflections of the sound by the environment, the sensitivity to HRTFs manipulations is reduced. This does not mean a priori that the localization performance in an absolute localization task is reduced by reflections.

Two different hypothesis can be supposed: First, it can be assumed that subjects were less sensitive because the spatial perception of the stimulus is in a way stabilized by the reflections. This means, that the reflections add information that can be used by the localization process to build the spatial object. The relative contribution of the HRTF information in the direct sound would be decreased and, therefore, the JNDs to HRTF manipulations would be increased.

The second hypothesis assumes that reflections with different azimuth position than the azimuth position of the direct sound confuse the auditory system and lead to a more fuzzy perception of the acoustic object. Manipulations to the HRTFs cues of the direct sound are then less detectable to the subjects in the reverberant condition, because the stimulus is less concentrated in its spaciousness. However, the results of the current study show, that the sensitivity to HRTF manipulations is approx. reduced by a factor of two. Translated to absolute localization experiments this would result in a localization blur being two times higher in the reverberant condition. To the knowledge of the author this has not yet been found by localization experiments. Hence, this hypothesis seems less plausible than the first hypothesis.

The first hypothesis is further supported by the anecdotal report of the subjects that in the reverberant condition front/back confusion in the 'spectral morphing' experiment

III were less often observed than in the non-reverberant condition. In order to further analyze the reason for the stabilization of the spatial percept due to reflections, the physical cues provided by the reflections have to be considered further. It is known from the literature that reflections add distance information to the stimulus (Békésy, 1938; Mershon and King, 1975; Sheeline, 1983) which is only basically inherent in the HRTFs in non-reverberant environments (Brungart and Rabinowitz, 1995). Furthermore, the first reflections can have the same source azimuth as the direct sound. This holds especially for reflections from the floor and a low ceiling as was pointed out before by Hartmann (1983). In this case, the reflections reinforce the information of the direct sound. Thus, manipulations in the localization cues of the direct sound are in a way corrected by the reflections. Hartman (1983) found that localization performance of a rectangular gated tone (500 Hz) was decreased for a higher ceiling compared to a lower ceiling. It was concluded that the reflections from the ceiling have the same source azimuth as the source and therefore facilitate the localization of the sound.
The azimuth direction of the first reflections reaching the ear of the listener, however, is depending on the geometry of the room and the orientation of the source and listener position within this room.
More information about the source position in the stimulus (provided by reflections) increase the robustness against distortions in the binaural cues of the direct sound. However, this argument would only hold true if the localization acuity is not dramatically reduced in a reverberant environment. In the study of Hartmann (1983) it was found that the localization accuracy of a 500 Hz tone was independent of the reverberation time of a room with variable acoustics. It was concluded, that for the different degrees of reverberation and absorption (7 dB difference in the level of the reflections) the localizations performance did not change. On the other hand, it was shown in a study of Begault (1992) that for speech stimuli with synthetic reverberation, the localization acuity was reduced compared to the non-reverberant condition, while distance perception was enhanced. Hence, although the stability argument seems to be plausible, no compelling data can be found in the literature that support this view.

Consequences for the use of HRTFs in reverberant environments

Independent from the hypothesis of how the reduction in sensitivity is caused by the reflections it can be concluded from the results of this study that less spatial information in the direct sound is needed in reverberant environments. Even for only eight cepstral smoothing coefficients the detection rate is below the threshold in the reverberant condition. From Chapter 3 of this thesis it can be seen that for eight cepstral coefficients the inter-individual standard deviation of the HRTF spectra across subjects is reduced. It can be concluded, that less individual information in the HRTFs is needed, if reverberation is added to the HRTFs of the direct sound. This is supported by the results of

experiment II. For three of five source positions the sensitivity to the 'spectral morphing' procedure is reduced in the reverberant condition. However, for higher values of the morphing factor α, the detection rate is above the threshold. Therefore, although the sensitivity of subjects to individual information is reduced, the reduction is not sufficient for using dummy head HRTFs without a change in the spatial perception.
The investigations of the HRTFs presented in Chapter 3 show that the spectral difference between subjects are smaller than the differences between a subject and the dummy head. Therefore, it can be assumed that the reduction in sensitivity to the spectral cues is sufficient for using non-individualized HRTFs. This can be investigated by using non-individualized HRTFs for the spectral morphing' procedure rather than dummy head HRTFs. If the detection performance is below the threshold even for higher values of α this would indicate that in reverberant conditions individual spectral information is not needed. However, this has to be investigated.
The sensitivity to ITD variations is also reduced in reverberant environments. From the investigation in Chapter 3 on the ITD standard deviation across subjects ($\bar{\sigma} = 40.1\ \mu s$) it can be seen that the differences between subjects are within the dimension of the ITD JND estimated in experiment III. Hence, in reverberant conditions the need for individual ITD information in the direct sound is reduced.

5.5 General conclusion

The results of this study can be summarized as follows.

- Sensitivity to manipulations of the ILD and ITD is reduced by a factor of approx. two in reverberant conditions
- Therefore, less individual spatial information is needed in the direct sound of a stimulus compared to non-reverberant environments
- The reduction in sensitivity could be caused by additional localization cues provided by the reflections that enhance the robustness against distortions of the spatial information in the direct sound.

Chapter 6

Spatial elevation perception of a spectral source cue

Abstract

Spectral scrambling is applied to the spectrum of the stimulus in localization experiments to prevent the subject from using spectral timbre variations as a cue. It has not been investigated yet, if the spectral scrambling introduces a localization cue that affects the apparent stimulus position. The spectral scrambling of the source spectrum could introduce a monaural cue that influences the elevation perception. Therefore, in the experiment presented here, the influence of a spectral cue in the source spectrum on the perceived elevation was studied for a noise stimulus (500 ms length) that is projected to the horizontal plane by using virtual acoustics. The spectrum of the source sound contains a monaural spectral cue that points to an elevation in the range of $-40°$ *to* $60°$*. The task of the subject was to judge the perceived elevation in an absolute localization paradigm as a function of the spectral cue in the source spectrum. The results show that the spectral cue in the source spectrum significantly influences the perceived elevation with a maximum effect of* $20°$*. Hence, there is a need for developing scrambling methods that only change the perceived timbre but not the perceived localization of a given sound.*

6.1 Introduction

The localization performance of the auditory system of human subjects is normally measured by presenting a sound source at a certain stimulus position and asking the subject to report the perceived source location (see Chapter 2.3). To estimate the source position of the stimulus the subjects can used binaural cues (interaural time differences, ITD and interaural level differences, ILD) as well as monaural spectral cues that are introduced by interference effects and pinna filtering. However, the monaural cues can

serve as a spectral timbre cue that could be used by subjects to learn the timbre that corresponds to a certain stimulus position. To prevent the subjects from using timbre cues for the identification of the stimulus position, the spectrum of the sound source is often randomly scrambled in a certain level range before the stimulus is presented to the subjects.

Spectral scrambling has been used in a variety of localization studies. For instance, Wightman and Kistler varied the spectral amplitude of the source stimulus in critical bands by up to 20 dB in absolute localization experiments (Wightman and Kistler, 1989b; Wightman and Kistler, 1992; Wightman and Kistler, 1997). The same manipulation of the source spectrum was used by Wenzel et al. (1993). Kulkarni et al. scrambled the source spectrum in 1/3 octave bands by up to $\pm 5\ dB$ to prevent the subject from using non-spatial cues in a real/virtual source discrimination task. Langendijk and Bronkhorst varied the stimulus spectrum in 1/3 octave bands in order to investigate if subjects are able to virtual stimuli generated with interpolated HRTFs (Langendijk and Bronkhorst, 2000). In the experiments described in Sections 4 and 5 the stimulus spectrum was scrambled in 1/6 octave bands by up to $\pm 5\ dB$.

However, the spectral shape, that is introduced to the source spectrum by scrambling, could contain spatial information that could be processed by the auditory system. Therefore, the perceived stimulus position could change depending on the scrambled source spectrum. Hence, to quantify the bias that could be introduced by spectral scrambling, the affect of a monaural spectral cue in the spectrum of a broadband stimulus is investigated in the current study by performing an absolute localization experiment.

It is well known that narrow band stimuli can cause confusions to the auditory system. The apparent source position can be determined by the center frequency of the narrow band sound independent of the actual source position. For instance, Blauert (1969) found that the perceived source position of 1/3-octave band noise signals, emanated from locations in the median plane, is only determined by the center frequency of the stimuli. Butler and Helwig (1983) varied the center frequency of 1 kHz wide noises and showed that the perceived spatial position in the median plane goes from front to back, as the center frequency is increased from 4 kHz to 12 kHz. Musicant and Butler (1985) showed that the center frequencies of 1 kHz wide noise also determine the perceived localization in the horizontal plane, where binaural cues are usable for the subjects.

The physical properties that relate to the judged locations of the narrow band stimuli can be found in the head related transfer functions (HRTFs). They describe the directional dependent transformation of a sound from its source location to a point in the ear canal. The HRTFs for the judged locations have peaks at frequencies that correspond to the center frequencies of the narrow band signals. Depending on studies these peaks are called 'boosted bands' (Blauert, 1969), 'covert peaks' (Flannery and Butler, 1981; Musicant and Butler, 1984) or 'proximal stimulus spectra' (Middlebrooks, 1992). Thus, salient peaks in the spectra of a stimulus bias the apparent stimulus location to the po-

sition for which the HRTF spectra have a peak in the corresponding frequency range. It is likely that the spectrum of scrambled broadband stimuli has prominent peaks in some frequency bands. Therefore, the auditory system could relate the peaks introduced in the source spectrum by spectral scrambling to spectral filtering by the HRTFs. As a result the apparent position of the stimuli would vary randomly for each scrambled stimulus spectrum. Although this explanation seems to be plausible, it has not been investigated in a systematic way if monaural source cues in a broadband stimulus spectrum affect the spatial perception.

Hence, in the study presented here it is investigated if the elevation perception of a broad band stimulus is affected by a monaural cue in the source spectrum. A noise stimulus is randomly presented from one of five different azimuth positions in the horizontal plane by using the methods of virtual acoustics (e.g. (Wightman and Kistler, 1989a; Hammershoi, 1995)). Before the virtual stimulus is generated the spectrum of the noise is multiplied with the spectrum of a HRTF of one ear measured at the same azimuth position. However, the elevation of the HRTF spectrum was chosen from $\theta = -40°, -20°, 0°, 20°, 40°, 60°$. Hence, a broad band stimulus is projected to the horizontal plane that contains a monaural source cue that points to a different elevation at the same azimuth. The task of the subjects was to judge the source location as a function of the monaural source cue in an absolute localization task. If the perception of the stimulus is independent from its spectrum then the subjects should localize each stimulus in the horizontal plane. If, however, the monaural source cue influences the perception of the stimulus, the judged elevation should increase as the elevation of the monaural source cue is increased. Two experiments were conducted. In the first experiment the monaural spectrum of the left ear HRTF was applied to the white noise source and in the second experiment the monaural spectrum of the right ear HRTF was applied to the stimulus spectrum.

6.2 Method

Subjects

Six subjects (four male and two female) aged from 28-35 participated voluntarily in the localization experiment. All had normal hearing and extensive experience in psychoacoustic tasks. They were members of the physics and psychology departments of the University of Oldenburg.

Stimuli

A catalogue of individual HRTFs measured at a high number of source positions was recorded in a separate session (see Chapter 3 for a description of the measurements). All

HRTFs needed to generate the source stimuli were taken from this database.
A frozen white noise sample (500 ms length, ramped with 5 ms squared cosines) was used as a source stimulus. The spectrum of the noise was multiplied with the spectrum of the individual HRTF of one ear at azimuth ϕ and elevation θ. The spectrally transformed noise sample was then convolved with the left and right ear HRTFs measured at the same azimuth ϕ but the elevation was set to the horizontal plane. Hence, the monaural information in the source spectrum points to a different elevation than the HRTFs used to project the noise sample to the horizontal plane. The azimuth positions were chosen from $\phi = 0°, 45°, 90°, 135°, 180°$. The HRTF spectra for the monaural source cue were obtained from the elevations $\theta = -40° - +60°, \Delta\theta = 20°$ at each of the listed azimuths. Two sets of stimuli were created. In the first set, the monaural elevation cue of the left ear was used to transform the source spectrum (condition ML) and in the second set the monaural cue of the right ear (condition MR) was applied. No headphone correction was performed. To illustrate the spectra multiplied to the spectrum of the sound source

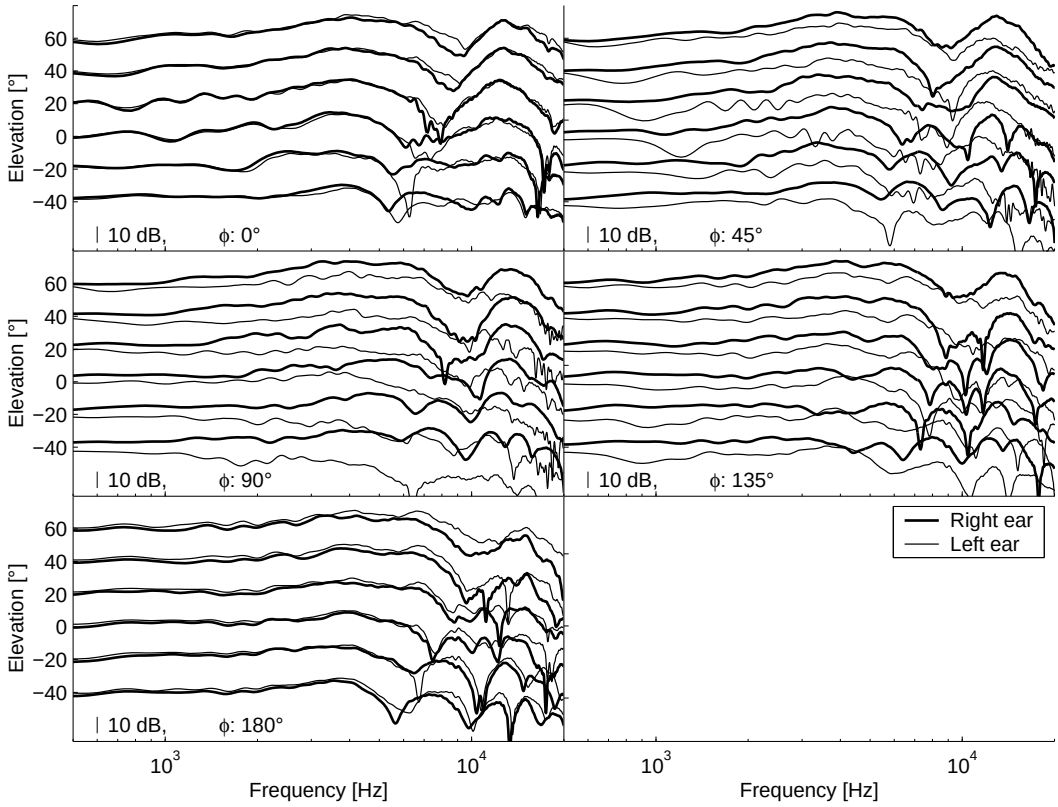


Figure 6.1: HRTF spectra of one subject used for the transformation of the source spectrum for this subject. The thick solid lines show spectra for the right ear and the thin solid lines are representing the left ear.

(in order to generate the monaural source cue), in Figure 6.1 the HRTF spectra of one subject are shown. In each subplot HRTFs at azimuth ϕ for the elevations $\theta = -40° - +60°, \Delta\theta = 20°$ are presented. Right ear HRTFs are plotted by thick solid lines

and left ear HRTFs by thin solid lines. The spectra show the typical shape of HRTFs as a function of azimuth and elevation.

Procedure

The localization experiments were conducted in a sound isolated booth (IAC 405A). The subjects were seated on a small chair in front of a window. The IBM compatible computer that controlled the experiments was located outside the room behind the window. A WinShell batch program controlled the stimulus presentation and the recording of the subjective data.
The stimuli were computed off-line and stored on the harddisc of the computer. An AKG 501 headphone served to present the stimuli to the subject. It was directly plugged to the output of a SoundBlaster 128 sound card. The presentation level was approx. 70 dB(A) (measured at the right ear of a dummy head for frontal sound incidence). After stimulus presentation the subjects recorded the perceived stimulus location by using the GELP technique (see Chapter 2 for a comprehensive description of the GELP technique). The recording device consisted of a sphere with a diameter of 30 cm located in front of the subject. The subject had to point to a location on the sphere that corresponds to the perceived stimulus position, as if the subject would be sitting in the center of the sphere. A Polhemus Inside Track was used to capture the position of the input device (a receiver of the Inside Track) on the spherical surface. The computer recorded the position of the receiver on the surface when it was placed there for at least one second. To acknowledge the recording, a brief signal was presented to the subject by headphone.
The azimuth position as well as the elevation of the monaural source cue was chosen at random for each trial. Each stimulus condition was repeated 10 times.
Subjects conducted two separate sessions. In the first session the source spectrum was transformed by the left ear HRTF spectrum (condition ML) and in the second session the source spectrum was transformed by the right ear spectrum (condition MR).

6.3 Results

An inspection of the individual responses revealed that the subjective judgments for each azimuth show an individual bias in elevation, either to lower or higher elevations, depending on the subject. This bias might be due to the monaural elevation cue in the source spectrum that confuses the auditory system and leads to an inaccurate elevation perception. In order to eliminate the bias, it was individually subtracted from the localization data before averaging across subjects. Note, that the elimination of the bias does not affect the shape of the curves presented in Figure 6.2 but the absolute position of the curves is shifted vertically. Furthermore, the inter-individual standard deviation is

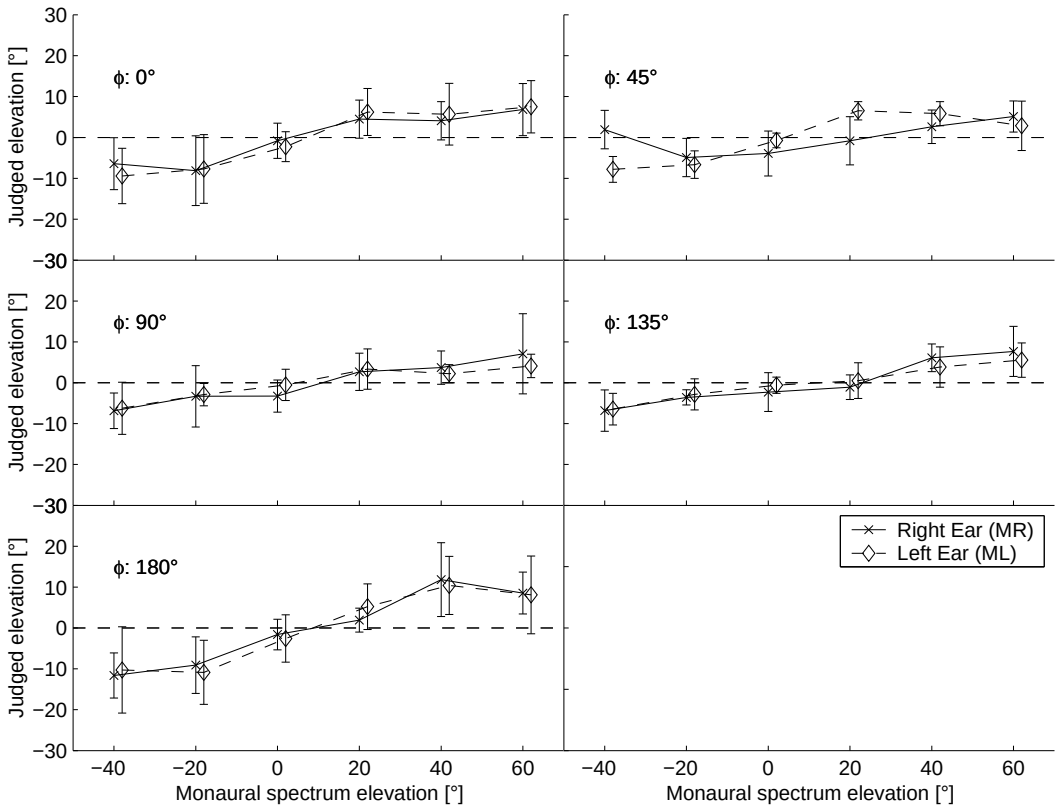


Figure 6.2: Perceived elevation as a function of the monaural source cue for different angles of azimuth averaged across subjects. Elevation judgements for monaural source cues taken from the right ear are connected by solid lines and by dashed lines for monaural cues from the left ear HRTF. The error bars indicate the inter-individual standard deviation.

reduced. The amount of this bias is listed in Table 6.1 for each subject individually. The bias was eliminated because the absolute localization error in elevation was less relevant for the present study than the change in elevation perception for stimuli with different monaural source cues.

The results of the localization experiment are summarized in Figure 6.2. Each subplot shows the mean perceived elevation averaged across subjects after removing the bias as a function of the source cue elevation. The solid line shows data for the MR condition (right ear HRTF spectrum) and the dashed line shows data for the ML condition (left ear HRTF spectrum).

The shape of the curve is nearly identical for all azimuth positions. The lowest perceived elevation angle is approx. $-10°$, increasing as the source cue elevation is increased from $-20°$ to $60°$. At $180°$ azimuth no further increase of the perceived source elevation can be observed for a source cue elevation increasing from $40°$ to $60°$. The maximum range of the perceived elevation is approx. $20°$, with a trend to a smaller range at lateral azimuth positions compared to positions in the median plane.

Subjects\Condition	ML					MR				
Azimuth	0°	45°	90°	135°	180°	0°	45°	90°	135°	180°
HK	23	43	16	-13	-23	24	27	4	-4	-13
JD	-3	2	-1	-1	3	4	-2	-3	2	3
HR	4	30	3	1	19	0	15	-7	-8	5
RH	3	4	-1	7	9	10	-2	-8	6	27
MK	10	19	8	-4	6	43	28	-5	5	50
IB	-1	9	3	-1	-7	5	3	-6	-10	-1

Table 6.1: The bias in degree (i.e. the mean localization error in elevation) is listed for each subject and each angle of azimuth for both measurement conditions.

The shape of the curve for the MR and ML condition is nearly identical. Only at 45° of azimuth the left ear source cue and the right ear source cue provide a different perception of the stimulus elevation.
The significance of the perceived differences in stimulus elevation as a function of the source cue elevation was analyzed by a non-parametric ANOVA (Kruskall-Wallis). The null hypothesis was that the judged elevation for a monaural source cue from $-40°$ is equal to the judged elevation for a monaural source cue, obtained from higher elevations. The analysis reveals that the elevation judgements for a monaural source cue from 20° elevation are significantly different from the $-40°$ elevation for all angles of azimuth ($p < 0.05$), except for 135° azimuth. However, at this azimuth the mean judgements between $-40°$ and 40° are significantly different ($p < 0.05$).
The differences in the judged elevations can be related to differences in the monaural source cue. This is illustrated in Figure 6.3. Here, the absolute level differences between the HRTFs spectra at $-40°$ and higher elevations are shown separately for the left (thin solid lines) and right (thick solid lines) ear. Each subplot shows level differences in dB averaged across subjects as a function of frequency for one azimuth.
The variation of the monaural spectra as a function of elevation is concentrated on the frequency range around 6-8 kHz for azimuth locations in the median plane ($\phi = 0°, 180°$). The level differences in this frequency band increase steadily for increasing elevations. Only minor effects can be seen in other frequency areas. Due to the symmetry of the head with respect to the median plane, similar differences for both ears can be observed. At 45° of azimuth an increase of the source elevation raises the level in the 6-8 kHz area of the left ear. Smaller changes in the same frequency region can be observed for the right ear. Compared to the other azimuth angles the variation of the right ear HRTF as a function of source elevation is rather small and can, thus, explain the deviating elevation perception for this direction (Figure 6.2, upper right panel, solid line). For azimuth positions at 90° and 135° the main effects are a general increase in the level of the left ear. Furthermore, an increase of the level in the frequency range around 9 kHz

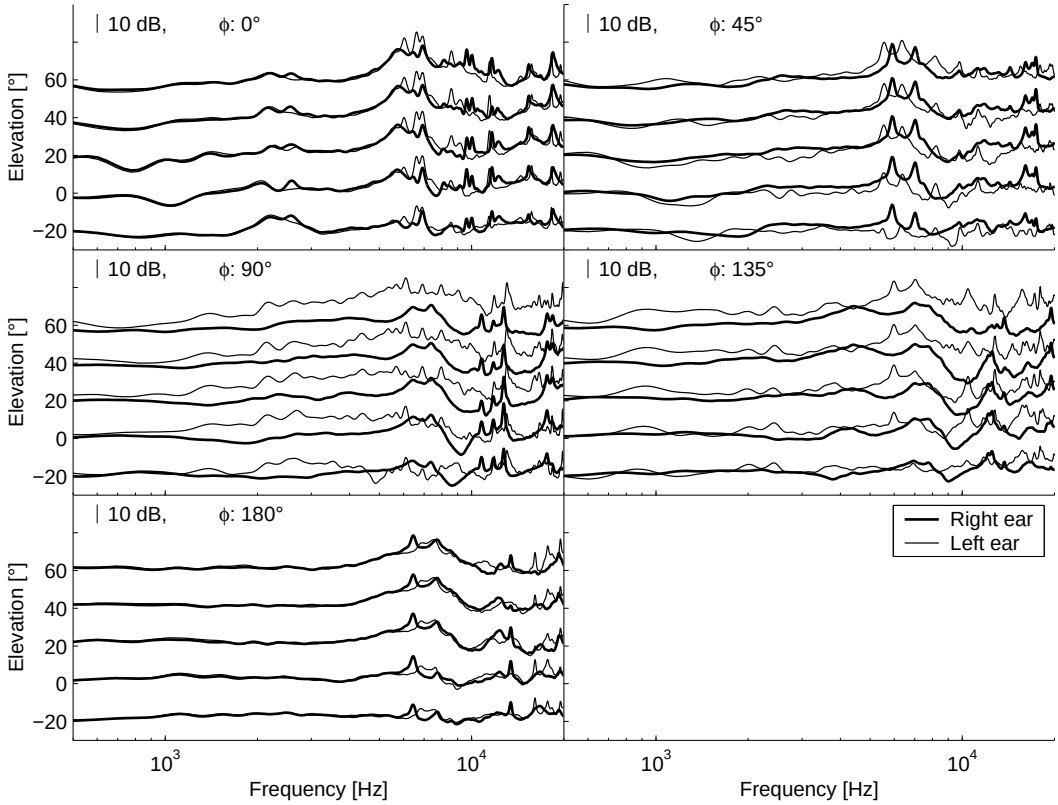


Figure 6.3: Level differences between the HRTF spectra at −40° elevation and higher elevations for different angles of azimuth are shown. Level differences for the right ear are plotted by thick solid lines, and for the left ear by thin solid lines.

at the right ear is introduced by higher source elevations.

6.4 Discussion

The results of the localization experiment show that a monaural cue in the spectrum of the source can significantly influence the elevation perception of a broad band stimulus. The monaural cues taken from the left and right ear HRTFs of sources at −40° to 60° elevation, modify the perceived elevation in a range of approximately 20° with a tendency of a smaller range for lateral source positions.

The analysis of the physical cues that caused the different elevation judgements shows that for most azimuth angles of sound incidence level changes occurred primarely in the frequency band from 6-8 kHz. The range of the level differences in theses frequency bands is approx. 10 dB. Level differences in a broader frequency area can only be observed in the HRTF spectra of the contralateral ear at 90° and 135°. The frequency band around 8 kHz was specified by Blauert as a boosted band for the above direction (Blauert, 1969). The results of the analysis of the physical cues confirms the finding that an increase in the level in this band increases the perceived stimulus elevation. However,

a general level increase in a broader frequency range, as observed for the left ear at 90° and 135°, also seems to cause an increased elevation perception.

The general intention of this study was to asses the effect of spectral scrambling on the elevation perception of a stimulus. Most methods for scrambling the source spectrum rove the spectrum in a range ($> 10\ dB$) that seems to be sufficient for producing peaks in the source spectrum yielding a variation of the perceived elevation. Furthermore, the spectral source cue in the current study only introduced level differences in the 'above' band. It is likely that the random level variation in the source spectrum introduced by spectral scrambling also introduces peaks and notches in other directional bands. Therefore, it can be assumed that spectral scrambling causes further variations of the perceived stimulus location, for instance, in the front-back dimension.

The use of virtual acoustics for the presentation of stimuli does not necessarily limit the generality of the results. It can be assumed that an equivalent experiment in the free-field would show similar results. However, it is possible that even small rotations of the head during stimulus presentation could lead to a breakdown of the effect. If the head position is changed during the stimulus presentation, the auditory system could estimate the source spectrum by computing the amplitude ratio of the stimulus for two different head positions. This would enable the auditory system to distinguish between the source spectrum and the spectral transformation of the HRTFs.

Therefore, it would be of interest to investigate if the change in the perceived elevation as a function of the monaural source cue can also be observed for stimulus presentations in the free-field. However, the head should be fixed by using a bite bar to provide an experimental condition that is equivalent to the virtual stimulus presentation.

The results of this study have consequences for absolute localization experiments as well as for localization discrimination tasks. If the source spectrum is scrambled before each stimulus presentation in an absolute localization experiment, the perceived source location may subjectively change as a function of source scrambling. An increased variabiliy of the results of localization experiments are in this case not only introduced by the localization uncertainty but also by the spectral scrambling of the source sound. Hence, the localization uncertainty measured by scrambled sound stimuli will never be smaller than the spatial variation of the sound source as a function of scrambling.

In spatial discrimination tasks subjects are asked to identify a target stimulus with a spatial displacement relative to a reference stimulus. To avoid the subjects from using non-spatial cues, both the target and reference stimulus are scrambled in their spectral content. Again, scrambling could affect the spatial position of both stimuli. Thus, only spatial displacements greater than the spatial variation as a function of source scrambling can be significantly detected, because both stimuli are changing their position in each trial.

The investigation presented in this paper shows that the source spectrum affects the perceived source position even if all binaural and monaural information is provided by

the HRTFs. Therefore, further research is needed to develop scrambling methods that introduce no spatial cues but only change the timbre of the stimulus. Without such a procedure, the sensitivity of localization measurements to the spatial resolution is always restricted to the spatial variation caused by spectral scrambling. The effect of random scrambling on the perceived source position is hardly predictable, because variations in a high dimensional parameter space are performed. Therefore, an empirical investigation of the effect of spectral scrambling on the apparent source position might be appropriate. This could be done by an empirical procedure outlined as follows: A pair of noise signals could be presented via virtual acoustics to an arbitrary angle of azimuth. In each trial the spectrum of one of the stimuli is randomly scrambled in a way similar as described in the present study. The task of the subject would be to indicate if both stimuli are perceived at the same position. If the amplitudes of the spectral variation in the frequency bands of the scrambled noise stimulus are recorded for each trail, the spectrally roved stimuli for which the source position did not change could be extracted. This information could be used to generate a scrambling procedure that does not affect the spatial stimulus position.

6.5 Conclusions

From the investigation presented here the following conclusions can be drawn:

- Appropriate spectral cues in the source spectrum can bias the perceived localization of a broadband sound by up to 20° in elevation in virtual acoustic localization tasks.
- The physical cues in the source spectrum that produce these elevation are consistent with the directional bands for the 'upward' direction found by Blauert (1969).
- New spectral scrambling methods should be developed in the future where only a change in the perceived timbre, but no change in perceived elevation or localization, respectively, is produced.

Chapter 7

Summary and Conclusion

The general aim of this thesis was to assess the perceptual robustness of the auditory system to variations of the respective physical localization cues and (closely related) to characterize the amount of individual information that is needed in head related transfer functions (HRTFs) to achieve an accurate perception of virtual acoustic stimuli.
To measure individual HRTFs a mechanical setup (the TASP system) was constructed that allows for a rapid and accurate positioning of a physical sound source on a sphere of possible source locations (see Chapter 2). The usability and reliability of this measurement system was investigated by conducting two free-field localization experiments. The results were similar to comparable experiments from the literature even though the measurement setups differed in several aspects. Therefore, it can be concluded that the experimental setup used here does not substantially influence the measured localization accuracy. However, it seems necessary to re-establish the subjects head position at the center of the TASP system after each stimulus presentation. Since this was not the case in the present experiments, the data data did not show the increased localization accuracy for frontal sound incidence usually obtained with a fixed head position.
The individual subjects localization decision was recorded using the Gods eye view localization pointing (GELP) technique which was evaluated in different experimental series with respect to its accuracy. Even though the subjects were less accurate in handling the GELP technique in a darkened room than in a lighted room, the recorded localization accuracy of acoustical targets is obviously not restricted by using the tactile sense to handle the GELP technique.
In order to analyze physical differences between HRTFs of different subjects, the TASP system was used to measure individual HRTFs from 11 subjects and one dummy head (see Chapter 3). The results show high inter-individual differences in the monaural and binaural cues. Hence, dummy head HRTFs can in principle not replace individual HRTFs since the former would lead to an inaccurate spatial perception for the majority of subjects. Furthermore, the inter-individual differences between individual HRTFs are

smaller than the differences between individual and dummy head HRTFs. Hence, if individual HRTFs are not available for building virtual acoustic displays, non-individual HRTFs of a selected listener (showing a high localization performance in free-field localization experiments) instead of dummy head HRTFs are recommended for the best expected localization performance under theses circumstances.

In the process of realizing HRTFs as digital filters, the HRTF spectra are often smoothed to reduce the computational effort to process these filters. In the second section of Chapter 3 the effect of two different smoothing procedures (cepstral smoothing and $1/N$ octave smoothing) on the localization cues was investigated. It is shown that both the monaural and binaural localization cues are affected by smoothing. A comparison of the effect of cepstral smoothing and $1/N$ octave smoothing on the ILD revealed that the logarithmic $1/N$ octave smoothing is less appropriate for reducing the spectral detail of HRTF spectra because comparably high ILD deviations are introduced. Furthermore, the reduction of the filter length is less effective with respect to $1/N$ octave smoothing in comparison to cepstral smoothing.

In Chapter 4 the investigation is expanded to perceptual consequences of spectral and temporal HRTF manipulations. Discrimination experiments were conducted to obtain thresholds for perceptually just noticeable deviations of the respective physical localization cues. The results indicate that a high amount of spectral detail can be eliminated from the HRTF spectra without affecting the *spatial* perception. However, if the subjects were able to use spectral timbre cues, the detection thresholds are substantially decreased. Furthermore, subjects were very sensitive to variations of the macroscopic structure of the HRTF spectra which were varied by introducing non-individual spectral information of dummy head HRTFs, especially for source positions in the median plane. It can be concluded that subjects are more sensitive to distortions of the individual localization cues if the center frequencies of the perceptually relevant peaks and notches of the HRTF spectra are affected.

A correlation analysis showed that the average ILD deviation that is introduced by manipulating the HRTF spectra correlates well to the perceived spatial displacement. Hence, the ILD deviation was used as a distance measure for obtaining thresholds for perceptually just noticeable variations of the HRTF spectra. Thus, an important conclusion is that the ILD deviation can be used for predicting the perceptual effect of spectral manipulations.

The introduced 'spectral morphing' method could be used to quantify perceptually relevant distances of individual HRTFs by calculating the 'morphing' factor for which the corresponding ILD deviation exceeds the appropriate thresholds which were obtained in the psychoacoustic experiments. In this way, perceptually relevant distances of HRTF spectra from different subjects could be quantified.

It was furthermore shown, that the ITD JND is increased if the stimuli have a plausible ILD that is determined from individually measured HRTFs. It can be assumed that the

additional spatial information provided by the plausible frequency composition of the ILD enhances the perceptual robustness to ITD variations of the virtual acoustic object.
In Chapter 5 it is investigated if reflections and reverberation affect the evaluation of the localization cues in the direct sound (which is assumed to be the primary factor in localization). To do so, the sensitivity of subjects to manipulations of the localization cues in the direct sound is compared under reverberant and non-reverberant conditions. The results show that the detection thresholds increase by a factor of approx. 2 in the reverberant condition for spectral and temporal manipulations. Hence, reflections do influence the evaluation of the localization cues in the direct sound. It can be assumed that the increased thresholds for manipulations of the localization cues in the direct sound are not caused by an increased localization blur but by an increased perceptual robustness of the acoustical object due to additional spatial information provided by the reflections.
Hence, both the results of the investigation on the ITD JND and the results obtained in Chapter 5 suggest that additional spatial information provided to the auditory system (which is available in real acoustics, but not necessarily in virtual acoustics) stabilizes the perception of virtual acoustic stimuli and enhance the robustness to distortions or deviations of the localization cues.
The physical differences between individual and non-individual HRTFs can be interpreted as distortions of the individual localization cues. Therefore, the results of this thesis show that the need for individual HRTFs is substantially reduced if additional localization cues (e.g. distance information provided by reflections or dynamic cues due to head rotation) can be used by the auditory system.
In some of the experiments presented above, the perceptual dimension was restricted to spatial cues by scrambling the source spectrum of the stimulus. Subjects anecdotally reported that not only the target stimulus was spatially displaced in the discrimination experiments but also the reference stimuli were moving depending on the spectral timbre variation that was introduced by spectral scrambling. Hence, in Chapter 6 the effect of a source spectrum variation on the perceived stimulus location was investigated. The results of an absolute localization experiment show that a spectral cue in the source spectrum can vary the perceived stimulus elevation in a range of 20°. This effect is mainly caused by an increase in level in a comparatively small frequency band that corresponds to the directional 'above' band specified by Blauert (1969). It can be concluded that spectral scrambling can vary the perceived source location in the estimated range.
This outcome has consequences for the interpretation of the results obtained in Chapters 4 and 5. If spectral scrambling was applied to the stimuli, random variations of the perceived stimulus position could have been perceived by the subjects. It is likely that if an ideal scrambling procedure which only varies the spectral timbre of the stimuli would have been used, subjects would have been able to detect smaller spatial changes of the stimulus. Hence, it can be concluded that in this case the ILD deviation thresholds for

perceptually just noticeable distortions of the HRTF spectra would have been slightly decreased.

Taken together, the current thesis has provided new methods for characterizing the human ability to localize sounds both in virtual and real acoustical environments. It has also shown the limits of the physical representation of sound stimuli that have to be obeyed when simulating an acoustical scene. A particular noteworthy aspect is the stabilization of spatial perception produced by the "first wavefront" by adding natural acoustical features, such as a natural ILD pattern across frequency and reflections/reverberations. It can be expected that the current results will be useful not only for gaining more insights about the nature of human sound processing, but also for the design and verification of systems employing virtual acoustics, for example in telecommunication, remote control and home entertainment.

Appendix A

A.1 Free field localization experiments in the literature

The studies of Wightman and Kistler (1989a), Gilkey et al. (1995) and Makous & Middlebrooks (1990) are summarized briefly. The results of these studies are used for a comparison of the results of the current study in Section 4.

Wightman and Kistler compared the localization performance under free-field and virtual conditions. In the free-field condition, noise stimuli were presented over six speakers, mounted on a semicircular steel arc. The ends of the arc were mounted above and below the subject. Source positions in elevation were chosen by using one of six speakers. The stimulus spectrum (a pulsed train of gaussian noise, eight 250 ms pulses separated by 300 ms) was roved in level (20 dB range in critical bands) to prevent the subject from learning the stimulus timbre. The stimulus locations were spaced by 10° in azimuth and by 18° in elevation ranging from −36° to 54°. Before data collection, the subjects were intensively trained to learn the verbal report technique of azimuth and elevation angles.

In order to introduce and validate the GELP technique, Gilkey et al. conducted two experiments. In the first one, click train stimuli (300 ms length) were presented by one of a pool of 239 loudspeakers, distributed on the surface of a sphere with a diameter of 4.3 meters. The speaker resolution on the surface was between 8° and 15°c surrounding the subject in azimuth and covering the elevation range from −45° to 90°. The GELP technique was used to record the subjectively perceived source location. In a second experiment, the source locations were verbally conveyed by the experimenter. In both experiments the head of the subjects was centered in the localization dome by a bite bar. A comparison of experiment I and II was used to validate the GELP technique. In order to compare their work with the literature, Gilkey et al. re-analyzed data from Wightman & Kistler.

In the study from Makous and Middlebrooks the stimuli were presented via one of 36 loudspeakers (10° spacing) mounted on a circular hoop with a diameter of 2.4 meters.

The rotation axis lies within the interaural axis. The source locations covered all azimuths (10° resolution) and elevation within −45° to 55° with respect to a double pole system of coordinates. The stimuli had a flat spectrum within 1.6 kHz to 16 kHz and pseudo random phase. The subjectively perceived source locations were recorded by monitoring the subjects head position. The subjects were asked to point to the source location with their noses. Between measurement trials an acoustical reference stimulus was presented in front of the subject to allow for a realignment of the subjects head. Two measurement conditions were used to measure the individual localization performance (Open Loop condition) and the inherent motor response (Closed loop condition). In the former condition the stimulus duration was 150ms and in the latter the stimulus lasted until the subjective data was recorded. However, it should be noted that in all previous mentioned studies the stimuli were not only distributed in the horizontal and median plane as in the current study.

A.2 Correlations between distance measures and percent correct responses for Chapter 4

Correlation coefficients between the perceptual data (obtained in Chapter 4) and different distance measures are given in Tables A.2-A.2. The distance measures are described below. Note that the distance measures $D10$, $D11$ are D_{mon} and D_{bin}, respectively.

Description of the distance measures

1. D1: Absolute difference between ILDs of reference and target HRTFs averaged across frequency bins on a linear scale.

2. D2: Absolute difference between right ear HRTF spectra of target and references HRTFs averaged across frequency bins on a linear scale.

3. D3: Absolute difference between left ear HRTF spectra of target and references HRTFs averaged across frequency bins on a linear scale.

4. D4: Maximum absolute difference between ILDs of reference and target HRTFs across frequency bins on a linear scale.

5. D5: Maximum absolute difference between right ear HRTF spectra of target and references HRTFs across frequency bins on a linear scale.

6. D6: Maximum absolute difference between right ear HRTF spectra of target and references HRTFs across frequency bins on a linear scale.

7. D7: Maximum absolute difference between ILDs of reference and target HRTFs across frequency channels of a Gammatone filter bank.

8. D8: Maximum absolute difference between right ear HRTF spectra of reference and target HRTFs across frequency channels of a Gammatone filter bank.

9. D9: Maximum absolute difference between left ear HRTF spectra of reference and target HRTFs across frequency channels of a Gammatone filter bank.

10. D10: Mean absolute difference between ILDs of reference and target HRTFs averaged across frequency channels of a Gammatone filter bank. This distance measure is called D_{bin} throughout the study.

11. D11: Mean absolute difference between right ear HRTF spectra of reference and target HRTFs averaged across frequency channels of a Gammatone filter bank. This distance measure is called D_{mon} throughout the study.

12. D12: Mean absolute difference between left ear HRTF spectra of reference and target HRTFs averaged across frequency channels of a Gammatone filter bank.

SS I

Distance measure / Azimuth	0°	45°	90°	135°	180°	Ø
D1	0,2	0,47	0,74	0,8	0,57	0,7
D2	0,21	0,55	0,78	0,81	0,75	0,78
D3	0,26	0,41	0,65	0,77	0,64	0,69
D4	0,18	0,41	0,66	0,75	0,42	0,61
D5	0,04	0,57	0,68	0,66	0,64	0,66
D6	0,26	0,41	0,65	0,77	0,64	0,69
D7	0,12	0,42	0,59	0,85	0,38	0,61
D8	0,02	0,37	0,78	0,86	0,76	0,8
D9	0,25	0,45	0,46	0,84	0,35	0,55
D10	0,12	0,43	0,75	0,79	0,65	0,73
D11	0,14	0,42	0,81	0,85	0,8	0,82
D12	0,33	0,48	0,69	0,83	0,76	0,76

Table A.1: Correlation coefficients between percentage of correct responses in condition 'SS I' and twelve different distance measures are shown. Mean values averaged across source azimuths from $90° - 180°$ are given in the last column.

SS II

Distance measure / Azimuth	0°	45°	90°	135°	180°	Ø
D1	0,22	0,58	0,37	0,44	0,13	0,31
D2	0,31	0,72	0,14	0,23	0,26	0,21
D3	0,19	0,42	0,48	0,53	0,22	0,41
D4	0,19	0,65	0,41	0,47	-0,01	0,29
D5	0,15	0,72	0,08	0,14	0,16	0,13
D6	0,19	0,42	0,48	0,53	0,22	0,41
D7	-0,05	0,58	0,35	0,45	0,37	0,39
D8	0,06	0,58	0,15	0,27	0,29	0,24
D9	0,32	0,48	0,27	0,43	0,39	0,36
D10	0,11	0,54	0,23	0,46	0,49	0,39
D11	0,18	0,63	0,19	0,29	0,41	0,3
D12	0,31	0,52	0,38	0,49	0,45	0,44

Table A.2: Correlation coefficients between percentage of correct responses in condition 'SS II' and twelve different distance measures are shown. Mean values averaged across source azimuths from $0° - 180°$ are given in the last column.

SS III

Distance measure / Azimuth	0°	45°	90°	135°	180°	Ø
D1	0,7	0,87	0,68	0,66	0,45	0,67
D2	0,72	0,81	0,68	0,63	0,56	0,68
D3	0,75	0,8	0,72	0,64	0,49	0,68
D4	0,73	0,77	0,67	0,53	0,36	0,61
D5	0,71	0,78	0,67	0,74	0,4	0,66
D6	0,75	0,8	0,72	0,64	0,49	0,68
D7	0,71	0,83	0,69	0,69	0,59	0,7
D8	0,57	0,67	0,67	0,66	0,56	0,63
D9	0,51	0,77	0,66	0,57	0,45	0,59
D10	0,66	0,88	0,71	0,75	0,59	0,72
D11	0,72	0,74	0,67	0,65	0,57	0,67
D12	0,63	0,86	0,71	0,75	0,48	0,69

Table A.3: Correlation coefficients between percentage of correct responses in condition 'SS III' and twelve different distance measures are shown. Mean values averaged across source azimuths from $0° - 180°$ are given in the last column.

Spectral morphing

Distance measure / Azimuth	0°	45°	90°	135°	180°	Ø
D1	0,62	0,85	0,7	0,74	0,62	0,71
D2	0,7	0,81	0,79	0,63	0,48	0,68
D3	0,55	0,77	0,68	0,73	0,41	0,63
D4	0,66	0,8	0,67	0,76	0,57	0,69
D5	0,78	0,7	0,68	0,53	0,55	0,65
D6	0,55	0,77	0,68	0,73	0,41	0,63
D7	0,64	0,81	0,64	0,81	0,59	0,7
D8	0,74	0,72	0,77	0,58	0,5	0,66
D9	0,55	0,75	0,72	0,74	0,24	0,6
D10	0,68	0,79	0,74	0,75	0,67	0,73
D11	0,69	0,82	0,77	0,59	0,43	0,66
D12	0,54	0,78	0,76	0,75	0,23	0,61

Table A.4: Correlation coefficients between percentage of correct responses in condition 'spectral morphing' and twelve different distance measures are shown. Mean values averaged across source azimuths from $0° - 180°$ are given in the last column.

References

Alrutz, Herbert (**1983**). *Über die Anwendung von Pseudorauschfolgen zur Messung an linearen Übertragungssystemen*, (Phd-thesis). Ernst-August-Universität Göttingen.

Asano, F. (**1990**). Role of spectral cues in median plane localization. *J. Acoust. Soc. Am.*, **88**(1):159–168.

Begault, D. R. (**1992**). Perceptual effects of synthetic reverberation on three-dimensional audio systems. *J. Aud. Eng. Soc.*, **40**(11):895–904.

Békésy, G. v. (**1938**). Über die Entstehung der Entfernungsempfindung beim Hören. *Akustische Zeitschrift*, **3**:21–31.

Blauert, J. (**1969**). Sound localization in the median plane. *Acustica*, **22**:205–213.

Blauert, J. (**1971**). Localization and the law of the first wav front in the median plane. *J. Acoust. Soc. Am.*, **50**:466–470.

Blauert, J. (**1974**). *Räumliches Hören.* S. Hirzel Verlag.

Blauert, J. (**1998**). *verbal communication.*

Borish, J. and J. B. Angell (**1983**). An efficient algorithm for measuring the impulse response using pseudorandom noise. *J. Audio Eng. Soc.*, **31**(7):478.

Bronkhorst, A. W. (**1995**). Localization of real and virtual sources. *J. Acoust. Soc. Am.*, **98**:2542–2553.

Bronkhorst, A. W. and T. Houtgast (**1999**). Auditory distance perception in rooms. *Nature*, **397**(6719):517.

Brungart, D. S. and W. M. Rabinowitz (**1995**). Auditory localization of nearby sources. head related transfer functions. *J. Acoust. Soc. Am.*, **16**:331–353.

Butler, R. A. and C. C. Helwig (**1983**). The spatial attributes of stimulus frequency in the median sattigal plane and their role in sound localization. *Am. J. Otolaryngol.*, **4**:165–173.

Butler, R.A. and K. Belendiuk (**1977**). Spectral cues utilized in the localization of sound in the median sattigal plane. *J. Acoust. Soc. Am.*, **61**:1264–1267.

D., Musicant A. and R.A. Butler (**1985**). Influence of monaural spectral cues on binaural localization. *J. Acoust. Soc. Am.*, **77**:202–208.

Domnitz, R. (**1968**). The interaural time JND as a simultaneous function of interaural time and interaural amplitude. *J. Acoust. Soc. Am.*, **50**:1549–1552.

Durlach, N. I. and H.S. Colburn (**1979**). *Binaural phenomena.*, Chp. 10, pp. 365–466. Academic Press.

Fisher, I.N., T. Lewis and B.J. Embleton (**1987**). *Statistical analysis of spherical data.* Cambrigde: Cambridge University Press.

Flannery, R. and R. A. Butler (**1981**). Spectral cues provided by the pinna for monaural localization. *Percept. Psychophys.*, **29**:438–444.

Freyman, R. L., D. D. McCall and R. K. Clifton (**1998**). Intensity discrimination for precedence effect stimuli. *J. Acoust. Soc. Am.*, **103**(4):2031–2041.

Gilkey, R.H., M.D. Good, M. A. Ericson, J. Brinkman and J.M. Steward (**1995**). A pointing technique for rapidly collecting responses in auditory research. *Behav. Res. Meth. Instr. and Comp.*, **27**:1–11.

Good, M. and R. H. Gilkey (**1996**). Sound localization in noise: The effect of signal to noise ratio. *J. Acoust. Soc. Am.*, **99**:1108–1117.

Haas, H. (**1949**). The influence of a single echo on the audibility of speech. *J. Audiol. Eng. Soc.*, **20**:145–159.

Hammershoi, D. (**1995**). *Binaural technique - a method of true 3D sound reproduction.*, (Phd-Thesis). Aalborg University Press: Aalborg University.

Hammershoi, D. and H. Møller (**1996**). Sound transmission to and along the ear canal. *J. Acoust. Soc. Am.*, **100**:408–427.

Hartmann, W. M. (**1983**). Localization of sound in rooms. *J. Acoust. Soc. Am.*, **74**:1380–1391.

Hebrank, J. and D. Wight (**1974**). Spectral cues used in the localization of sound sources in the median plane. *J. Acoust. Soc. Am.*, **56**:1829–1834.

Hershkowitz, R. M. and N. I. Durlach (**1969**). Interaural time and amplitude JND's for a 500-hz tone. *J. Acoust. Soc. Am.*, **46**:1464–1467.

Kinkel, M. (**1990**). *Zusammenhang verschiedener Parameter des binauralen Hörens bei Normal und-Schwerhöringen.*, (Phd-Thesis). Georg-August-Universität zu Göttingen.

Koehnke, J., C. P. Culotta, Hawley M. L. and H. S. Colburn (**1995**). Effects of reference interaural time and intensity differences on binaural performance in listeners with normal and impaired hearing. *Ear and Hearing*, **16**:331–353.

Kuhn, G.F. (**1977**). Model for the interaural time differences in the azimuthal plane. *J. Acoust. Soc. Am.*, **62**(1):157–167.

Kulkarni, A. and H.S. Colburn (**1998**). Role of spectral detail in sound localization. *Nature*, **396**:747–749.

Kulkarni, A., Isabelle S. K. and H. S. Colburn (**1999**). Sensitivity to head-related transfer function phase spectra. *J. Acoust. Soc. Am.*, **105**(5):2821–2840.

Langendijk, E. H. A. and A. W. Bronkhorst (**1997**). Collecting localization responses with a virtual acoustic pointer. *J. Acoust. Soc. Am.*, **101**:3106.

Langendijk, E. H. A. and A. W. Bronkhorst (**2001**). *Contribution of spectral cues to human sound localization*. Submitted to JASA.

Langendijk, E. H. A. and A.W. Bronkhorst (**2000**). Fidelity of three-dimensional sound reproduction using a virtual auditory display. *J. Acoust. Soc. Am.*, **107**:528–537.

Litovsky, R. Y. (**1997**). Developmental changes in the precedence effect: Estimates of minimum audible angle. *J. Acoust. Soc. Am.*, **102**:1739–1745.

Litovsky, R. Y., H. S. Colburn, W. A. Yost and S. J. Guzman (**1999**). The precedence effect. *J. Acoust. Soc. Am.*, **106**(4):1633–1654.

Litovsky, R. Y. and N. A. Macmillan (**1994**). Sound localization under conditions of the precedence effect: Effects of azimuth and standard stimuli. *J. Acoust. Soc. Am.*, **96**:752–758.

Lorenzi, C., S. Gatehouse and C. Lever (**1999**). Sound localization in noise in normal hearing listeners. *J. Acoust. Soc. Am.*, **105**(3):1810–1820.

Makous, J. C. and J. C. Middlebrooks (**1990**). Two-dimensional sound localization by human listeners. *J. Acoust. Soc. Am.*, **87**:2188–2200.

Mehrgardt, S. and V. Mellert (**1977**). Richtungshören in der Medianebene und Schallbeugung am Kopf. *J. Acoust. Soc. Am.*, **61**:1567–1576.

Mershon, D. H. and E. King (**1975**). Intensity and reverberation as factors in the auditory perception of egocentric distance. *Perception and Psychophysics*, **18**:409–415.

Middlebrooks, J.C. (**1992**). Narrow-band sound localization related to external ear acoustics. *J. Acoust. Soc. Am.*, **92**:2607–2624.

Middlebrooks, J.C. (**1999a**). Individual differences in external-ear transfer functions reduced by scaling in frequency. *J. Acoust. Soc. Am.*, **106**:1480–1492.

Middlebrooks, J.C. (**1999b**). Virtual localization improved by scaling non-individualized external-ear transfer functions in frequency. *J. Acoust. Soc. Am.*, **106**:1493–1510.

Middlebrooks, J.C. and D.M. Green (**1991**). Sound localization by human listeners. *Annu. Rev. Psych.*, **42**:135–159.

Middlebrooks, J.C., E. A. Macpherson and Z. A. Onsan (**2000**). Psychophysical customization of directional transfer functions for virtual sound localization. *J. Acoust. Soc. Am.*, **108**(6):3088–3091.

Møller, H., M. F. Sørensen, D. Hammershøi and C. B. Jensen (**1995**). Head-related transfer functions of human subjects. *J. Audio. Eng. Soc.*, **43**(5):300–321.

Moore, B. C. J., R. W. Peters and Glasberg B. R. (**1990**). Auditory filter shapes at low center frequencies. *J. Acoust. Soc. Am.*, **88**:132–140.

Morimoto, M. and H. Aokata (**1984**). Localization cues of sound sources in the upper hemisphere. *J. Acoust. Soc. Jpn*, **5**(3):165–173.

Mossop, J. E. and J. F. Culling (**1995**). Lateralization of large interaural delays. *J. Acoust. Soc. Am.*, **16**:331–353.

Musicant, A. D. and R. A. Butler (**1984**). The psychophysical basis of monaural localization. *Hear. Res.*, **14**:185–190.

Oldfield, S.R. and S. A. Parker (**1984a**). Acuity of sound localization: a topography of auditory space I. Normal hearing condition. *Perception*, **13**:581–600.

Oldfield, S.R. and S. A. Parker (**1984b**). Acuity of sound localization: a topography of auditory space II. Pinna cues absent. *Perception*, **13**:601–617.

Oppenheim, A.V. and R.W. Schafer (**1975**). *Digital Signal Processing*. Prentice-Hall.

Otten, J. (**1997**). *Psychoakustische Messungen zur Lokalisationsfähigkeit beim Menschen.*, (Diplom-Thesis). University of Oldenburg.

Perrott, D. R., K. Marlborough and P. Merrill (**1989**). Minimum audible angle thresholds obtained under conditions in which the precedence effect is assumed to operate. *J. Acoust. Soc. Am.*, **85**(1):282–288.

Pösselt, C., J. Schröter, M. Opitz, P. L. Diverny and J. Blauert (**1986**). Gerneration of binaural signals for research and home entertainment. *Proc.* 12^{th} *Int. Cong. on Acoustics (Toronto).*

Rayleigh, Lord (**1907**). On our perception of sound direction. *Philos. Ma.*, **13**:214–232.

Rife, Douglas D. and J. Vanderkooy (**1993**). Transfer-function measurement with maximum-length sequences. *J. Audio Eng. Soc.*, **37**(6):419.

Roffler, S. K. and R. A. Butler (**1968**). Factors that influence the localization of sound in the vertical plane. *J. Acoust. Soc. Am.*, **43**:1255–1259.

Shaw, E. A. G. (**1974**). Transformation of the sound pressure level from the free-field to the eardrum in the horizontal plane. *J. Acoust. Soc. Am.*, **56**:1848–1861.

Shaw, E. A. G. (**1997**). *Binaural and Spatial Hearing in Real and Virtual Environmentes*, Chp. 2, pp. 25–47. Lawrence Erlbaum Associates.

Shaw, E.A.G. and R. Teranishi (**1968**). Sound pressure generated in an external-ear replica and real human ears by a nearby point source. *J. Acoust. Soc. Am.*, **44**(1):240–249.

Sheeline, C. W. (**1983**). *An investigation of the effects of direct and reverberant signal interaction on auditory distance perception*, (Phd Thesis). Stanford University.

Shinn-Cunningham, B. G., P. M. Zurek and N. I. Durlach (**1993**). Adjustment and discrimination measurements of the precedence effect. *J. Acoust. Soc. Am.*, **93**(5):2923–2932.

Teranishi, R. and E.A.G. Shaw (**1968**). External ear acoustics models with simple geometry. *J. Acoust. Soc. Am.*, **44**:257–263.

Tobias, J. V. and S. Zerlin (**1959**). Lateralization threshold as a function of stimulus duration. *J. Acoust. Soc. Am.*, **31**:1591–1594.

Tollin, D. J. and G. B. Henning (**1998**). Some aspects of the lateralization of echoed sound in man. I. the classical interaural-delay based precedence effect. *J. Acoust. Soc. Am.*, **104**:3030–3038.

Trampe, Ulrich (**1988**). *Akustische Übertragung des durchschnittlichen antropomorphen Außenohrs*, (Diplom Arbeit). Unversität Oldenburg.

Wallach, H., E. B. Newman and M. R. Rosenzweig (**1949**). The precedence effect in sound localization. *Am. J. Psychol.*, **LXII**:315–336.

Wenzel, E.M., M. Arruda, D.J. Kistler and F.L. Wightman (**1993**). Localization using non-individualized transfer functions. *J. Acoust. Soc. Am.*, **94**(1):111–123.

Wiener, F.M. and D.A. Ross (**1946**). The pressure distribution in the auditory canal in a progressive sound field. *J. Acoust. Soc. Am.*, **18**:401–498.

Wightman, F. I. and D. Kistler (**1989a**). Headphone simulation of free-field listening: I Stimulus synthesis. *J. Acoust. Soc. Am.*, **85**:858–867.

Wightman, F. I. and D. Kistler (**1989b**). Headphone simulation of free-field listening: II Psychoacoustic validation. *J. Acoust. Soc. Am.*, **85**:858–867.

Wightman, F. I. and D. Kistler (**1997**). Monaural sound localization revisited. *J. Acoust. Soc. Am.*, **101**:1050–1063.

Wightman, F.L. and D.J. Kistler (**1992**). The dominant role of low-frewuency interaural time differences in sound localization. *J. Acoust. Soc. Am.*, **91**:1648–1661.

Woodworth, R.S. (**1954**). *Experimental psychology.* Rinehart & Winston.

Zurek, P. M. (**1980**). The precedence effect and its possible role in the avoidance of interaural ambiguities. *J. Acoust. Soc. Am.*, **67**(3):952–964.

Zwicker, E. and H. Fastl (**1990**). *Psychoacoustics: Facts and Models.* Heidelberg, Germany: Springer-Verlag.

Erklärung

Hiermit erkläre ich, daß ich die vorliegende Arbeit selbständig verfaßt und nur die angegebenen Quellen und Hilfsmittel verwendet habe.

Oldenburg, den 13. Juli 2001 Jörn Otten

Danksagung

Mein herzlichster Dank gilt all den Menschen, die mir geholfen haben, diese Arbeit zu verrichten und erfolgreich zu beenden. Insbesondere bedanke ich mich bei

Prof. Dr. Dr. Birger Kollmeier für die Ermöglichung dieser Arbeit und für die Schaffung der hervorragenden (menschlichen wie materiellen) Arbeitsbedingungen in der AG Medizinische Physik. Seine hilfreichen und führenden Kommentare insbesondere zum Schluß dieser Arbeit haben wesentlich zum Gesamtbild beigetragen.

Prof. Dr. Volker Mellert für die freundliche Übernahme des Koreferats.

Prof. Dr. Steven Colburn für anregende Diskussionen über das Richtungshören und den Beitrages der HRTFs zum selbigen.

Dr. Adelbert Bronkhorst für seine konstruktiven Kritiken und sein Interesse an dieser Arbeit.

Michael Kleinschmidt und Rainer Huber dafür, dass sie es mit mir jahrelang in einer Sardinenbüchse ausgehalten haben.

Dr. Thomas Brand, für kleine und große Hilfen in jeder Lebenslage, der bereitwilligen Preisgabe seiner unglaublichen 'Datenbank' an Informationen und der Teilung eines Hobbies, welches langsam 'High-End' wird...

Holle Kirchner, die mit ihrem klarem Blick und kritischen Bemerkungen in den Anfängen dieser Arbeit eine wesentliche Unterstützung war.

den Mitgliedern der AG Medizinische Physik, die durch das freundliche Miteinander eine entspannte und weiträumige Arbeits-Atmosphäre geschaffen haben.

den Versuchspersonen Helmut Riedel, Michael Kleinschmidt, Rainer Huber, Ingo Baumann, Holle Kirchner, Karin Troidl, Dirk Junius, Dr. Carsten Reckhardt und Jörg Damaschke.

Jonas und Julian, die alle Probleme vergessen lassen.

meiner Frau Sandra. Ohne ihre selbstlose Unterstützung wäre diese Arbeit nicht möglich gewesen.

Lebenslauf

Am 22.06.1970 wurde ich, Jörn Otten, in Leer/Ostfriesland als drittes Kind von Imbke Otten, geb. Hörmann, und Otto Otten geboren. In dem Zeitraum von 1976 - 1980 besuchte ich die Ludgeri Grundschule in Leer und wechselte nach der Orientierungstufe (1980 - 1982) auf das Ubbo-Emnius Gymnasium, welches ich ab 1982 besuchte. Die schulische Ausbildung konnte ich 1990 mit dem Abschluß der Allgemeinen Hochschulreife beenden. Den Zivildienst beim Paritätischen Wohlfahrtsverband absolvierte ich in dem Zeitraum von Mai 1990 bis Juli 1991. Das Studium der Physik wurde an der Universität Oldenburg im Oktober 1991 begonnen und mit dem Abschluß als Diplom Physiker in Juni 1997 beendet. Die vorliegende Dissertation wurde seit Juli 1997 als Stipendiat des Graduiertenkollegs 'Psychoakoustik' unter der Leitung von Prof. Dr. Dr. Birger Kollmeier angefertigt. Seit dem 01.05.2001 bin ich als wissenschaftlicher Mitarbeiter am Hörzentrum Oldenburg tätig.

Factors influencing acoustical localization

Vom Fachbereich Physik der Universität Oldenburg
zur Erlangung des Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
angenommene Dissertation

Tag der Disputation: 21.09.2001

Erstreferent: Prof. Dr. Dr. B. Kollmeier

Korreferent: Prof. Dr. V. Mellert

Jörn Otten
geb. am 22. Juni 1970
in Leer / Ostfriesland

Contents

Chapter 1

General introduction

The ability of the auditory system to determine the spatial position of a sound source is essential for the orientation in our daily life environment. Due to a comprehensive analysis of the sound field generated by the source, human listeners are able to assess the direction, the distance and the spaciousness of a sound source. In contrast to the visual system, this capability is not restricted to a limited spatial range and, thus, the auditory sense does not only extend the perception of the environment to the acoustical modality but also extends the range of spatial cognition to the whole range of spatial directions. This extension allows us to be completely enveloped in the environment and it is, therefore, not surprising that we often close our eyes (for instance, in a music concert or even on a silent meadow) if we do not want to focus our attention to the spatially restricted range provided by the eyes.

The spatial information that is used by the auditory system to localize a sound source in a non-reverberant environment is captured by head related transfer function (HRTFs). They describe the transformation of the sound from its source location in the free-field to the microphone in the left or right ear canal. HRTFs can be measured by recording a sound emanating from a speaker at a certain location in space by small probe microphones in the ear canal of a subject. The auditory system uses two different kinds of cues that can be extracted from the HRTFs to estimate the source location. The *binaural* cues are calculated from a comparison of the HRTFs of the left and right ear. The interaural level difference (ILD) is caused by head shadowing and interference effects and describes the differences in level at the left and right ear as a function of frequency. The interaural time difference (ITD) reflects the differences in the path length (for lateral source positions) from the sound source to the left and right ear, respectively. The ITD and ILD are proposed by Lord Rayleigh (1907) to be the localization cues that characterize the spatial position of a sound source in the horizontal plane. However, there is no unique relation between the binaural cues and the position of a sound source in space because a whole cone of source positions can

be specified for which the ILD and ITD are almost constant (see (Woodworth, 1954) for a description of the 'cone of confusions'). In the 70^{th} the role of the pinnae (the outer ear) in sound localization began to emerge (see Blauert (1974) for a review). Shaw and Teranishi (1968) were able to show that the pinna cavities have a variety of resonance modes at characteristic frequencies. The amplitudes and the frequencies for which the resonances occur depend on the direction of sound incidence. Hence, the spectrum of the sound source is transformed by the resonances of the pinnae in a way that is characteristic for the source position of the sound. The spectral cue is denoted as 'monaural' since it is introduced independently at each ear. In addition to the binaural cues it represents the second group of spatial information. This 'spectral fingerprint' generated by the spectral transformation is different for sound incidence from each point on a 'cone of confusion' and, hence, monaural spectral cues are used to estimate the sound elevation as well as to decide if the sound is coming out of the frontal or rear hemisphere ((Hebrank and Wight, 1974; Butler and Belendiuk, 1977; Morimoto and Aokata, 1984; Asano, 1990)).

Since all spatial information that can be used by the auditory system to estimate the position of a sound in a non-reverberant environment is given by HRTFs, they provide the capability to simulate a free-field presentation of a sound. By presenting the signal convolved with head related impulses responses (HRIRs, that are the time domain representations of the HRTFs) of the left and right ear over headphones, a perception similar to a free-field condition can be achieved. This technique is called 'virtual acoustics' and allows to present externalized sound sources over headphones with an localization accuracy that is comparable to the acuity for real free-field presentations (e.g. (Wightman and Kistler, 1989a; Wightman and Kistler, 1989b; Hammershoi, 1995; Otten, 1997)). Virtual acoustics can be used to build computer controlled virtual auditory displays (VADs), that are capable of projecting sounds to any desired location in space, for instance, as a component of a virtual environment generator.

Two major problems emerge for VADs. First, the source positions could be distributed on a whole sphere of possible source locations and, hence, HRTFs are needed from a high number of source locations. Therefore, a measurement setup is needed that allows for flexible positioning of a physical sound source on a spherical surface. To reduce the measurement effort, fast and accurate positioning is required and the procedure to measure the HRTFs should introduces only small time delays.

Furthermore, it is not sufficient to measure a comprehensive set of head related transfer functions for only one selected listener. To achieve the same perceptual impression for each subject, individual HRTFs have to be used in VADs. If non-individual HRTFs are used to generate virtual sounds, the main problems that occur are an increased localization blur for the elevation perception and an increased occurrence of front-back confusions (i.e. the source position is perceived on a point on the appropriate cone of confusion that is opposite to the source location where the HRTFs were measured from.)

(Wenzel *et al.*, 1993). Both kinds of localization errors are introduced by deviations between the HRTF spectra of the listener and the HRTF spectra provided by the VAD. Because of the need for individual HRTFs, VADs are very costly to implement and, therefore, they are far away from being applicable for the common run of mankind. However, there are lots of potentialities for VADs to improve communication in our daily life, for instance for man-machine communication (especially for blind people) or for each application for which the distribution of information in a 3D space could be useful (for instance, telephone conferences or to improve communication in aircrafts). Thus, further research is needed to understand which aspects of individual HRTFs (providing the most basic localization cues) are needed for an accurate spatial perception.

This thesis deals with both the experimental needs for measuring HRTFs and the need for individual information in the localization cues to achieve an accurate perception of spatially localized objects.
Thus, the thesis is structured as follows: In Chapter 1 a mechanical setup is introduced (TASP, Two Arc Source Positioning) that allows for a rapid and accurate positioning of physical sound sources to almost any point on a spherical surface. The usability of the TASP system is investigated by free-field localization experiments and the results are validated by a comparison to data from the literature. The GELP technique (God's eye view Localization Pointing) introduced by Gilkey et al. (1995) is used to collect the subjective responses and it is investigated, furthermore, in which way the use of the GELP technique affects the recorded localization data.
In order to analyze inter-individual differences between HRTFs the TASP system is used to measure HRTFs from 11 subjects and one dummy head. This investigation is presented in Chapter 3 of this thesis. The HRTFs are described in terms of individual binaural and monaural localization cues and differences between HRTFs. For virtual acoustics, HRTFs are often realized as digital minimum phase finite impulse response (FIR) filters with smoothed spectra. The effects of spectral detail reduction on minimum phase HRTFs is investigated in the second section of Chapter 3.
While in Chapter 3 the investigation is focused on the effect of smoothing on the physical localization cues, in Chapter 4 the scope of the study is extended to perceptual consequences of HRTFs manipulations. By conducting discrimination experiments perceptual thresholds for spectral and temporal (variations of the ITD) manipulations of the individual physical localization cues are obtained to assess deviations of the individual localization cues that are not noticeable for human subjects.
In reverberant environments the direct sound emanated by the sound source is followed by reflections from objects surrounding the listener. The auditory system suppresses the spatial information in the reflections and estimates the source position mainly by means of the spatial information in direct sound. This effect is called 'precedence effect' because the auditory system gives precedence to the spatial information in the direct

sound. However, it can be assumed that the evaluation of the localization cues in the direct sound is influenced by reflections. To test this hypothesis it is investigated by discrimination experiments in Chapter 5 if the perception of changes in the localization cues of the direct sound differs under reverberant and non-reverberant conditions.
A common method to restrict the perceptual dimension in localization experiments to spatial cues is to rove the source spectrum ('spectral scrambling') before filtering the stimulus with HRTFs. Without this technique, subjects would be able to use the stimulus timbre as a cue. The scrambling procedure is also applied to stimuli in the parts of the experiments in Chapters 4 and 5. However, it could be that spectral scrambling introduces spatial information to the virtual stimuli that affects the localization of the stimuli. Thus, in Chapter 6 it is investigated by using an absolute localization paradigm, if spectral scrambling can vary the perceived stimulus positions.

Chapter 2

Influence of procedural factors on localization in the free-field using a two-arc loudspeaker system

Abstract

A computer controlled mechanical loudspeaker positioning system (TASP, two arc source positioning) is presented. It allows for continuous sampling of source positions in azimuth and elevation. To validate the system, free-field localization measurements in the horizontal plane ($\phi = 0° - 180°$, $\Delta\phi = 15°$) and in the median plane ($\theta = -40,° -60°, \Delta\theta = 20°$) were conducted. The stimulus was a 300 ms click train. A comparison to localization measurements from the literature revealed that consistent results are achieved even though the setup presented here deviates in several aspects from those described in the literature. However, to capture the improved localization performance for frontal sound incidence a head monitoring technique to center the head seems to be necessary. The GELP technique (Gilkey et al., 1995) was used to collect the localization data. To validate the use of the GELP technique in a darkened room the free-field localization performance was compared to data obtained from three control experiments with non-acoustical localization tasks. In the first control experiment, numerical values of the target azimuth and elevation were presented. In the second and third experiment, visual stimuli were presented in a lighted or darkened room. A comparison of the control experiments with the acoustical free-field localization experiment showed that the localization accuracy in the free-field setup employed here is not restricted by using the GELP technique in a darkened room.

2.1 Introduction

Study of localization ability has gained considerable interest in recent years (e. g. (Oldfield and Parker, 1984a; Makous and Middlebrooks, 1990; Good and Gilkey, 1996; Lorenzi *et al.*, 1999)), even though a variety of studies in this area have been conducted since the beginning of the 20^{th} century (see Blauert 1974 for a review). For measuring the localization ability in an anechoic chamber (free-field condition), either a fixed array of speakers has been employed or one or two speakers that can mechanically be positioned at certain locations. However, the mechanical setups for positioning the sound sources were not able to cover the whole range of spatially relevant source positions with high resolution. Since this has not yet been achieved in a satisfactory way, this contribution presents and evaluates a new setup that overcomes some of the restrictions of the systems known from the literature.

Different approaches for positioning a sound source on a virtual spherical surface of source locations with high resolution were used in the recent literature. Gilkey et al. (1995) used a static sphere of 272 loudspeakers evenly distributed on a surface of a sphere with a diameter of about 4.3 m. This construction allows for a rapid collection of localization data because there is no need to move a sound source between stimulus intervals. The main disadvantages are the fixed resolution of possible source locations and the considerable amount of reflecting surfaces of the metal construction and the speakers itself. Another possibility is to use only one arc of speakers that is rotating around a fixed axis (e.g (Wightman and Kistler, 1989a; Makous and Middlebrooks, 1990)). This concept reduces the reflecting surface by a considerable amount compared to the localization dome and increases the maximal resolution in at least one dimension (azimuth or elevation). If the rotation axis lies within the interaural axis, the construction is optimal for a double pole system of coordinates (Morimoto and Aokata, 1984), whereas a vertical rotation axis through the center of the head prefers a single pole system of coordinates. However, in both cases the resolution is restricted by the fixed location of the speakers either in elevation or in azimuth. A disadvantage of this approach compared to a fixed array of speakers lies in the time delay needed for a rotation of the arc. A setup similar to the one presented here was realized by Bronkhorst (1995). The subject is seated in the center of a rotatable arc (diameter 1.4 m). The rotation axis coincides with the interaural axis. However, the leverage of the arc at 0° elevations makes it difficult to control and effectively limits the diameter of the arc.

The measurement setup introduced in the current paper is termed TASP: Two Arc Source Positioning system. It is capable of positioning one of two speakers at nearly every point on a spherical surface with a diameter of approx. 4 m. The TASP system consists of two rotating hemi-arcs, with a vertical rotation axis going through the center of the head of the subject. The angle of azimuth of the sound source is adjusted by a

rotation of the arcs. Two sledges moving along the arcs position the elevation of the sledges (see Figure 2.1). The usability of the TASP system is verified by conducting a free-field localization experiment in the horizontal and median plane. The results of this experiment are compared to data from the literature (Section 2.3).

A prerequisite for the collection of localization data is that the subject can transform the subjective acoustical perception into an objective recordable variable. In the literature, a variety of different methods was used to collect localization data. For instance, Wightman and Kistler (1989b) asked their subjects to make a verbal report of the source location in terms of azimuth and elevation angle. The subjects had to train the report intensively before data collection began. In a study of Makous and Middlebrooks (1990) the subjects had to point to the stimulus location with their nose. The data was collected by monitoring the head orientation. Oldfield and Parker (1984a) used a pistol-like input device and asked the subject to 'shoot' the stimulus position. Langendijk used a virtual acoustic pointer controlled by a joystick-like input device (Langendijk and Bronkhorst, 1997). It was shown that the virtual pointer technique is more accurate than the verbal report technique. The GELP (God's Eye View Localization Pointing technique [1](Gilkey *et al.*, 1995)) uses a little globe in front of the subject that represents the sphere of stimulus locations surrounding the subject. The subject has to point to the location on the sphere that corresponds to the stimulus location on the sphere of possible source locations (see Section 2.4.1 for a detailed description). The study of Gilkey et al. showed that the input accuracy was as accurate as for the verbal report technique but not as accurate as the technique used by Makous and Middelbrooks. However, data collection was much faster by using the GELP technique (16-20 trials per minute) compared to the verbal report (2-3 trials per minute) and the 'nose pointing' (3-4 trials per minute) technique.

In the current study the GELP technique is used to record the subjective localization data because it is easy to implement and allows for a rapid collection of data. Furthermore, it turned out that the subjects did not need any training.

In the experiments conducted by Gilkey et al. to validate the GELP technique, the subjects were able to see the surface of the GELP globe. In contrast, the localization experiments presented in this study had to be performed in a darkened room to prevent the subject from seeing the moving parts of the TASP system. Therefore, the subject has to use his/her tactile instead of visual sense to point to the correct input location. It can be assumed that the capability of the subject to handle the GELP technique is reduced if only the tactile sense can be used.

Consequently, two different experiments are described in this paper to validate the usability of the TASP system in combination with the GELP technique. The suitability of the TASP system for serving as a positioning system in localization experiments is

[1]The GELP technique is similar to a technique developed by Blauert et al., called 'Bochumer Kugel', (Blauert, 1998)

examined in the first section of this study (Section 2.3). Since an investigation of the localization ability for each possible location on a sphere is beyond the scope of this study, the source locations were distributed only in the horizontal and median plane to reduce the overall measurement time and the size of the data set.
In the second section (Section 2.4), it is investigated if the GELP technique can also be used in a darkened room where the subjects were not able to see the spherical surface of the GELP technique. Three control experiments were conducted in which non-acoustical stimuli were presented to the subjects. In the first experiment, stimulus locations were presented numerically on a screen in terms of azimuth and elevation ('numeric' condition). In a second experiment the subject had to estimate the position of one of the TASP sledges in a lighted room (visual I). The third measurement was conducted in the darkened anechoic chamber. A little diode fixed in the center of the loudspeaker served as a target (visual II). The general assumption is that if the input performance (i.e. the capability of the subjects to point to the desired locations on the spherical surface in the non-acoustical stimulus conditions) is higher in the control conditions than in the free-field localization experiment, the localization accuracy is not restricted by using the GELP technique in a darkened room.

2.2 Technical description of TASP

Figure 2.1: The TASP (Two Arc Source Positioning) system inside the anechoic room.

The apparatus presented here was constructed under the constrains of maximum resolution in azimuth and elevation and as little positioning time delay and amount of surface reflections as possible. Furthermore, the setup can only be installed temporarily in the

anechoic room so that the mechanical installation procedure is required to be as short as possible. Consequently, the construction was chosen to consist of a fixed part at the ceiling and a removable part being attached to it. The mechanical installation procedure of the removable part takes about two and a half hours.
Figure 2.1 presents a photograph of the TASP (Two Arc Source Positioning) system within the anechoic room of the University of Oldenburg. Figure 2.2 depicts a scheme of the main functional parts. The removable part consists of two opposed hemi-arcs with a moveable loudspeaker sled attached to it. A sound source is positioned to the desired location by dragging the sledge into the correct elevation and turning the arc to the desired azimuth. The two opposed arcs divide the sphere of possible source locations into two hemispheres. Hence, the frontal and the rear hemisphere are covered by the two hemi-arcs.
The dimensions of the Oldenburg anechoic room hosting the TASP system (Figure 2.1) is 8,5m x 5m x 4m (width, depth, height) with a 1.3 m absorber depth and a lower cutoff frequency of 50 Hz. The TASP system itself is mounted at the ceiling by a double cross consisting of four iron double T profiles (1). A metal plate in the center of the double cross carries the main rotational axis (2) and the stepping motor (3, Positec VRDM31122) is responsible for the azimuthal rotation. The rotating system itself (4) is constructed as an open circle with a dihedral angle of 90°.

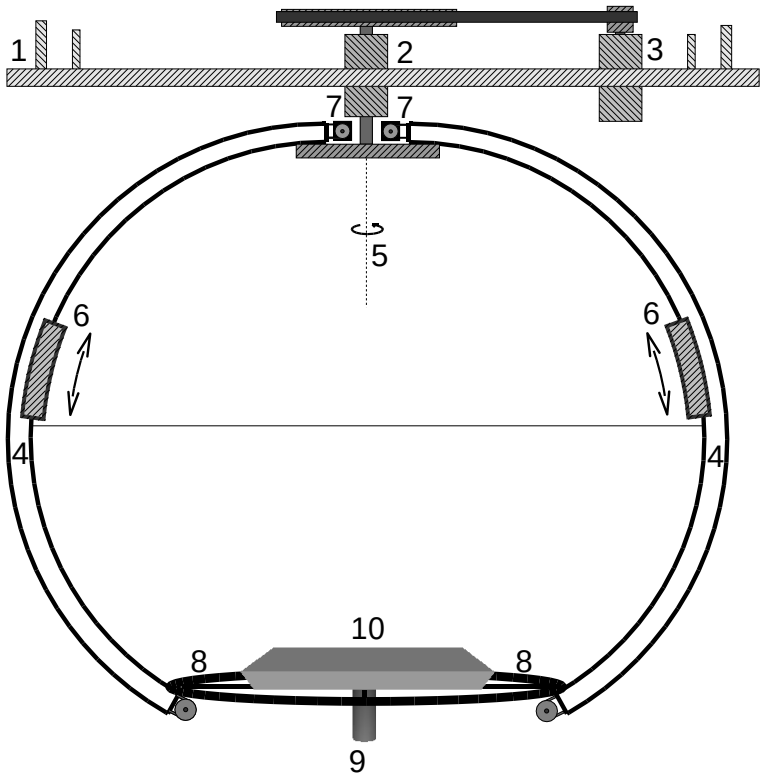


Figure 2.2: Scheme of the TASP system. See text for a detailed description of the numbered parts.

The rotation axis (5) corresponds to the axis of symmetry of the open circle. Two little sledges (6) using the inner part of the double T profile of the arc as tracks, serve as transports for the sound sources. Two stepping motors (7, Positec VRDM 3913 LWC), one for each sledge, are mounted directly at the rotation axis of the arc. They allow for an independent movement of the sledges on both hemi-arcs. A toothed belt is affixed to the sledge. Driven by the gear wheel of the respective stepping motor, the sledge is dragged into the desired direction. In this way the elevation of the sound source is adjusted. To prevent the hemi-arcs from oscillating around the rotation axis, their lower ends are connected via a metal ring (8) which is pivoted at its center point by a solid cylinder (9). This cylinder also serves as a stand for the platform (10) which carries the subjects chair or the dummy head to be positioned in the center of the sphere.
Figure 2.3 depicts the connection of the controlling software to the stepping motors. An IBM compatible 486 PC controlled by the WinShell[2] command line is connected via the serial port to a programmable stepping motor control device (Positec WPM 311). Two power devices (Positec WD3-004 and WD3-008) drive the stepping motors for positioning of the source.

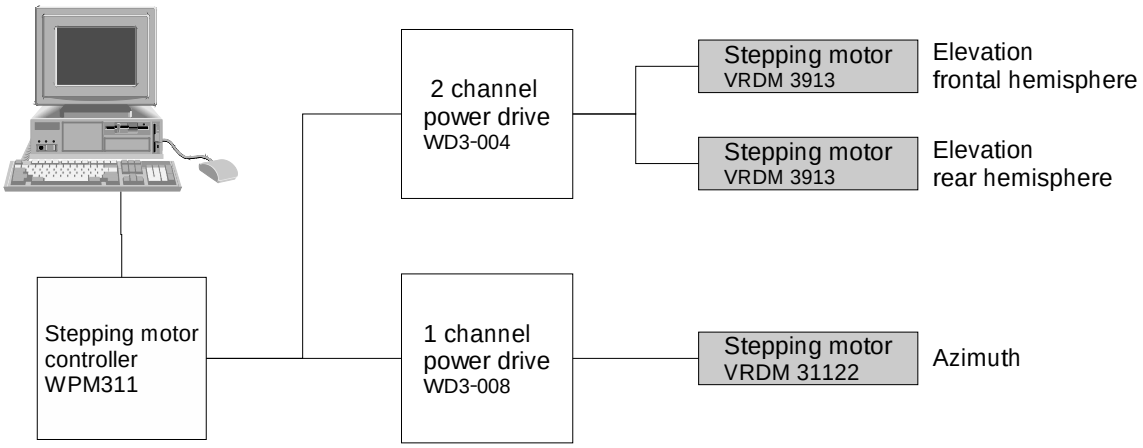


Figure 2.3: Controlling of the stepping motors.

2.2.0.1 Performance

The performance of the TASP system can be described by a) the time delay between the presentation of stimuli at different source locations b) the overall range of possible source locations c) the maximum resolution in azimuth and elevation d) the amount of reflecting surface disturbing measurements in anechoic conditions e) cues of the source position generated by TASP system itself.
With respect to the positioning time delay it turned out that the positioning in azimuth

[2]The WinShell is a command line experiment control system, developed by members of the work group 'Medizinische Physik' at the Universität Oldenburg which is capable of linking libraries providing control commands for hardware devices.

is much more critical than the movement in elevation. A non-continuous alteration of the rotation velocity causes the arc to oscillate around its axis of rotation. Therefore, onset and offset ramps have to be used to allow for a smooth movement and to limit the angular momentum to be applied. These ramps slow down the process of positioning the arc to the correct azimuth. To prevent the subject from using the delay as a cue for the relative distance between subsequent stimuli, a variable angular velocity of the rotation was introduced in a way that the delay is nearly independent of the relative distance of two successive stimuli. Hence, a fixed time delay can be specified with 6 s for the azimuth positioning and 2 s for the elevation positioning.

The mechanical constrains do not restrict the range of azimuth but limit the elevation to the range between $-40°$ and $+80°$. This should be sufficient for all kinds of investigations that are related to directional hearing. The minimal distance between two source positions is nearly arbitrarily small. It was limited by software to one degree in azimuth and elevation.

Although the reflecting surface of the TASP technique is quite small compared to other setups (like a localization dome, for instance) the environment of the subject is not without reflections. The sound wave generated by the speaker of one hemi-arc is reflected by the opposite hemi-arc and its speaker as well as the whole construction under the ceiling that carries the rotating part.

The rotation of the hemi-arc around the z-axis provides no hint to the speaker location because the driving motor is mounted overhead and the movement of the hemi-arc in the air is very silent. However, the sliding of the sledges along the arc is not noiseless. If one sledge is moved to a certain elevation, the toothed belt driving the sledge grates along the surface of the arc. The originating noise is not correlated to the speaker elevation but allows the subject to identify the azimuth of the arc. Hence, in measurement conditions were stimuli positions are distributed in azimuth and elevation, the noise from the sledge movement has to be masked by an external sound source. Another possibility would be to first position the elevation and afterwards the azimuth.

The localization measurements described in this study used only movements with a fixed azimuth or elevation providing no mechanical localization cue. Therefore, no masking noise was needed.

2.3 Free-field localization

2.3.1 Method

2.3.1.1 Subjects

Eight normal hearing subjects, six male and two female aged from 27 to 34 participated voluntarily in the free-field localization task. All subjects were members of the faculty and had extensive experience in psychoacoustic tasks but none of them was involved in localization experiments before. However, subject 'JO' is one of the authors.

2.3.1.2 Stimuli

The stimuli used for the presentation were click trains with a duration of 300 ms presented at a level of approx 60 dB(A). Clicks were repeated at a rate of 100 Hz. The onsets and offsets were gated by 25ms squared cosine ramps. The stimuli were equalized by the transfer function of the speaker within the frequency range of 100Hz to 14 kHz. After positioning the speaker to the desired position the stimulus was presented only once. The subject had no limitation in time to convey the perceived source location to the computer using the GELP technique (s. Section 2.4 for a comprehensive description of the GELP technique implementation).

2.3.1.3 Procedure

The localization performance in the horizontal plane and in the median sattigal plane was measured in two separate sessions. The measurements were conducted in the darkened anechoic room using the TASP technique for positioning the sound source. The subject was seated on a chair and adjusted in height so that the interaural axis lies within the horizontal plane. The head was not fixed by a chin rest or an equivalent method. Instead, the subject was told to focus the straight forward position (where the speaker was located during the instruction to the subject when the light were still on) and to re-establish this position after the input of the localization perception to the GELP technique. Before the beginning of each measurement the room was darkened and the speakers were moved at random three times without emanating a stimulus. Because the movement of the speaker in the horizontal plane does not give any cue for the detection of the speaker location in the darkened room, the subjects reliably lost the speaker location after the threefold positioning. Three seconds after the last movement the stimulus was presented. After recording the localization data by the computer, a 200 ms gated sine wave was presented from a speaker mounted under the subject's chair platform to acknowledge the recording. This signal was normally localized inside the

head and should not influence the localization task.
Each subject conducted two sessions. In the first session the source location was randomly chosen from 24 positions in the horizontal plane at 0° elevation (15° spacing). The subjects were not informed about the discrete distribution of the possible stimulus locations. Each position was measured three times resulting in 72 trials per session.
In a second session, source elevations in the median plane were distributed from −30° to +60° in the frontal and rear half-plane with a constant distance between locations of 10 degrees. The measurement routine was the same as the previous one except for the different source locations.
Data collection began with the first presentation of the stimulus. Thus, the subject were untrained and had only experience in using the GELP technique by participating in the validating experiments described in Section 2.4, which were conducted with each subject before the free-field localization measurements.

2.3.1.4 Localization data analysis

To compare the outcome of the free-field localization experiment to data from the literature, the analytical methods used by Wightman and Kistler (1989b) and Gilkey et al. (1995) were adapted.
The judgement centroid describes the mean judgment of subjects response to a stimulus from one certain location. It is calculated by summing up the normalized vectors from the center of the GELP sphere to the position on the surface indicated by the subject.
The mean absolute error is calculated by computing the absolute difference between the target and the judgement angle either in azimuth or elevation. The angle of error is computed for each response individually and then averaged across source locations and subjects.
The spread of responses for one target location is described by the parameter κ^{-1}. The concept of κ^{-1} was adapted from the statistics of spherical distributions and is similar to the standard deviation (Fisher *et al.*, 1987). A detailed description how to calculate κ^{-1} for localization data is given by (Wightman and Kistler, 1989b) and (Gilkey *et al.*, 1995).
A special problem in the investigation of localization data is the appearance of front to back reversals (e.g. (Makous and Middlebrooks, 1990; Wightman and Kistler, 1989b; Gilkey *et al.*, 1995)). The binaural localization cues are only capable to determine the source position on a 'cone of confusion' (Woodworth, 1954) for which the binaural parameters are constant. The position within each cone of confusions is resolved by utilizing monaural spectral cues. Hence, applying the former methods to the raw localization data would result in large azimuth errors which do not reflect the binaural localization accuracy. One way to resolve the confusion is to mirror the judgement to the (front-back) hemisphere where the distance between target location and judged location is smallest.

This concept can introduce errors for target locations near 90° of azimuth because it is possible that judgments are 'resolved' which were genuine localization errors. It is assumed that the number of errors introduced is small compared to the benefit of the mirroring procedure which avoids an overestimation of the errors due to front-back confusions.
In addition, a linear regression function was calculated for the localization data and the correlation coefficient between the presented and the judged locations was computed.

2.3.2 Results

Azimuth

The results for eight subjects participating in the localization experiments are shown in Figure 2.4. The judgment centroids are plotted as a function of the target angles in azimuth. The dotted line marks the ideal performance of correct responses. A linear regression function is plotted as a solid line in each diagram. To provide more information on the spread of data, the centroids are stretched along the judgement dimension, if κ^{-1} for the actual angle is greater than the mean value of κ^{-1} averaged across all azimuthal angles for this subject. In this case, the stretching is proportional to κ^{-1}. If κ^{-1} is less than the mean, the diameter of the centroid is set to a lower limit. Therefore, the ellipses mark an increased variability of the judgements relative to the mean spread of data. The centers of the ellipses still coincide with the original centroid.
In each sub-plot of Figure 2.4 additional information on the inter-individual differences in localization performance is provided. The bars in the lower right quadrant of each sub-plot give the individual performance normalized by the mean performance averaged across subjects. The dark gray bar in the left half of the surrounding box shows the mean error angle and the light gray bar on the right side reflects κ^{-1} for each individual subject. The bar heights were calculated by the same general procedure for the mean angle of error and κ^{-1}. The top of the surrounding box is the maximum value across all subjects and the dotted vertical line represents the mean value across subjects. In this way the diagram shows the individual performance expressed by the individual mean angle of error and κ^{-1} relative to the mean performance across subjects.
Although the localization performance is quite high, the subjects show the same pattern in those localization errors that still occur. The localization acuity is near optimum for frontal (0°) and rear (±180°) sound source incidence. If the source is positioned at more lateral angles ($\varphi < 90°$), the subjects tend to project the source to the side. However, the effect is small for subjects MK and JO. The localization uncertainty marked by κ^{-1} indicates that stimuli coming from angles between ±130° and ±180° are more difficult to localize than sounds in the frontal hemisphere. Again, this effect is quite small and not shown by each subject.

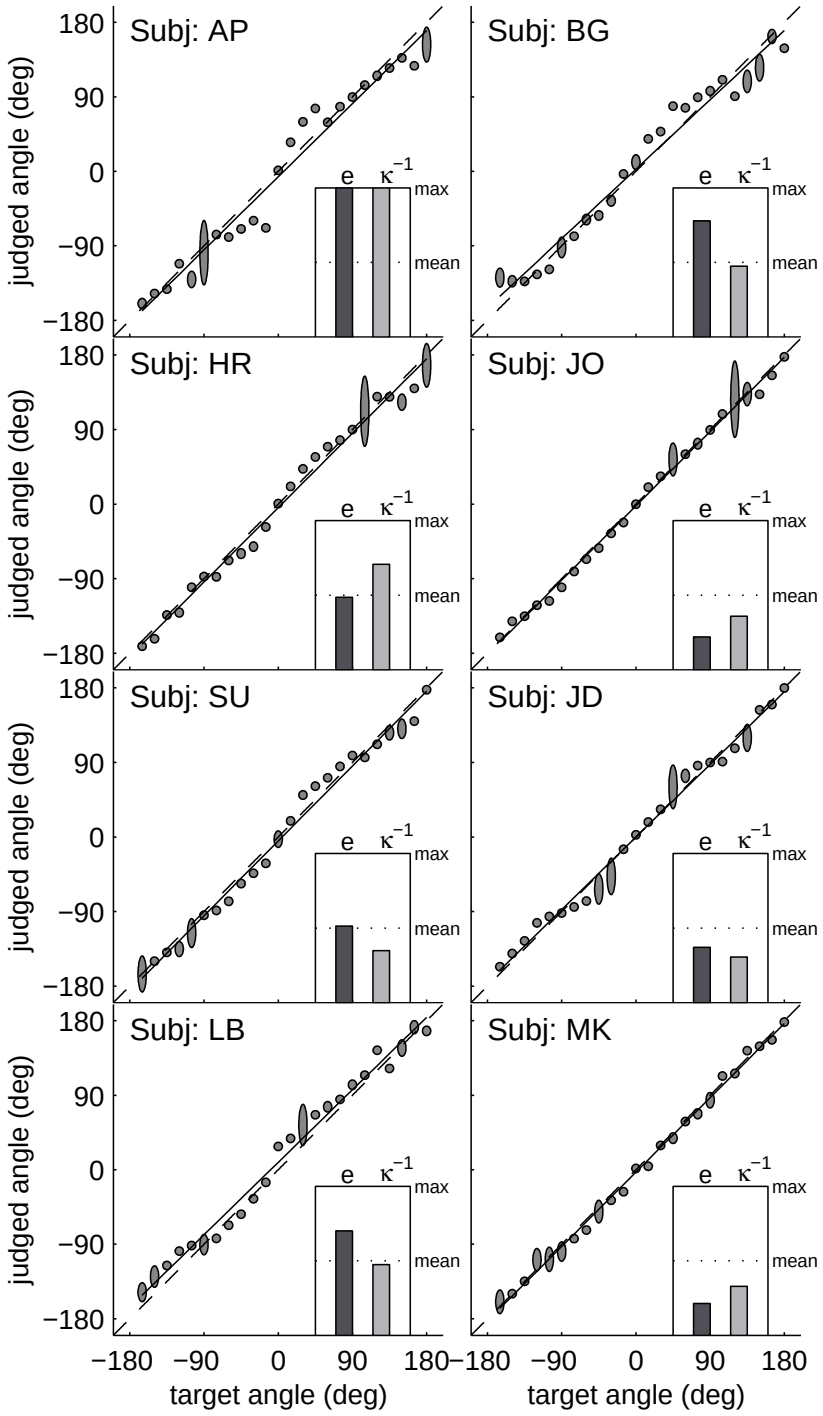


Figure 2.4: Extended centroid diagrams of the localization performance in azimuth. The stretched centroids identify a spread of the data that is higher than the mean across all locations for that subject. The bar diagrams in each plot represent the inter-individual differences in localization performance. The left, dark gray bar shows the mean absolute error $\bar{e}$ for the presented subject relative to the mean across all subjects (dotted horizontal line). The top of the box represents the maximum values across subjects. On the right side the light gray bar represents the same for κ^{-1}.

Subject	m	b	r	$\bar{e}$	κ^{-1}	F/B [%]
AP	0.981	-6.15	0.982	19.69(13.37)	0.065(0.036)	23
BG	0.931	2.84	0.985	16.11(13.05)	0.023(0.017)	17
HR	0.993	-4.05	0.994	11.38(10.78)	0.041(0.013)	7
JO	0.996	-2.11	0.998	07.08(06.42)	0.015(0.004)	0
SU	1.000	-4.71	0.994	11.83(11.34)	0.014(0.009)	7
JD	0.974	-0.05	0.995	09.52(09.42)	0.011(0.004)	6
LB	0.973	9.01	0.992	14.85(12.67)	0.024(0.011)	7
MK	0.998	-2.83	0.998	06.99(07.92)	0.012(0.007)	4
Ø	0.980	-1.00	0.992	12.18(10.62)	0.026(0.013)	9

Table 2.1: Results from the localization measurement in azimuth. Listed are the slope m and intercept b of the linear regression function, the correlation coefficient r, the mean absolute angle of error $\bar{e}$ (median values are shown in parenthesis), κ^{-1} and the percentage of front-back confusions.

The bar diagrams show that the angle of error and the spread of the input (κ^{-1}) are positively correlated ($r = 0.74$). This indicates, that under the present conditions the absolute localization uncertainty is well described by one of the measures. Subject 'AP' shows the poorest localization performance with the greatest angle of error and κ^{-1}.
It should be noted that subject 'JO', being one of the authors, shows a better than normal performance. It is likely that the lower errors are due to the a priori knowledge of the author that the source positions are discretely distributed in azimuth. This would allow for a substantial decrease in localization error because the absolute localization task changes to a identification task across different locations. However, the accuracy is still restricted by the accuracy of the GELP technique (see Section 2.4).
Table 2.1 summarizes the quantitative parameters of the localization results. Presented are the slope and intercept of the linear regression function (m, b), the correlation coefficient r between the target and judged angle, the mean absolute error $\bar{e}$, κ^{-1} and the number of front-back confusions in percent. Mean values across all subjects are presented in the last row of the table. These values will be used to compare the result of the current study to data from the literature.
Figure 2.5 shows the mean absolute error (solid line) and the signed error angle (dash-dotted line) averaged across subjects as a function of the source azimuth. The absolute error is a measure of the general localization uncertainty and the signed error reflects the bias to a certain direction. The absolute error varies slightly around the average of 12.3° with minima at 0° and ±90°. A prominent maximum can be seen at 45°. The positive values of the signed error for azimuthal source positions less than 90° indicate that the subjects tend to overestimate the angle in the frontal hemisphere. The opposite is true for the rear hemisphere. The negative values for target angles greater then 90° show an

underestimation of the azimuth position. It can be concluded, that subjects have a bias towards the more extreme lateral positions.

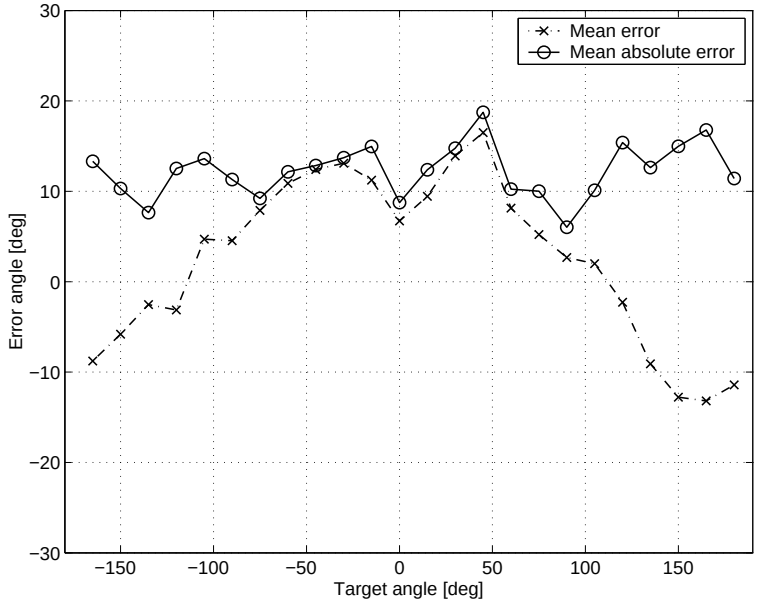


Figure 2.5: Mean absolute error (solid line) and mean error (dashed-dotted line), averaged across subjects, plotted as a function of azimuth.

Elevation

Figure 2.6 shows the localization data for source positions in the median plane. The spread of the distribution for each elevation is marked by stretching the centroid proportional to κ^{-1}. Data for source positions in the frontal hemisphere (0° azimuth) are plotted as light gray centroids and the corresponding centroids in the rear hemisphere (180° azimuth) are dark grey. Linear regression functions have been computed independently for the two hemispheres and are plotted within each subplot (frontal hemisphere: solid line, rear hemisphere dash-dotted line). The bar diagram is calculated in the same way as for the azimuth condition. However, the values were separately calculated for frontal and rear sound incidence and then averaged across hemispheres.
In contrast to the azimuth condition, localization performance varies considerably across subjects. In general, elevations greater than 20° are overestimated. This effect is more prominent for rear sound incidence. Targets at elevations lower than 0° are well localized by nearly all subjects. Only subject SU shows greater deviations from the target angles in this situation. Subject AP shows a large localization uncertainty with high errors and a wide spread of data. Higher elevations are strongly overestimated and only frontal locations near the horizontal plane are correctly localized. The subject stated that she had a high uncertainty on the stimulus position and felt like she was only guessing most locations. The data from subject BG for rear elevations is also remarkable.

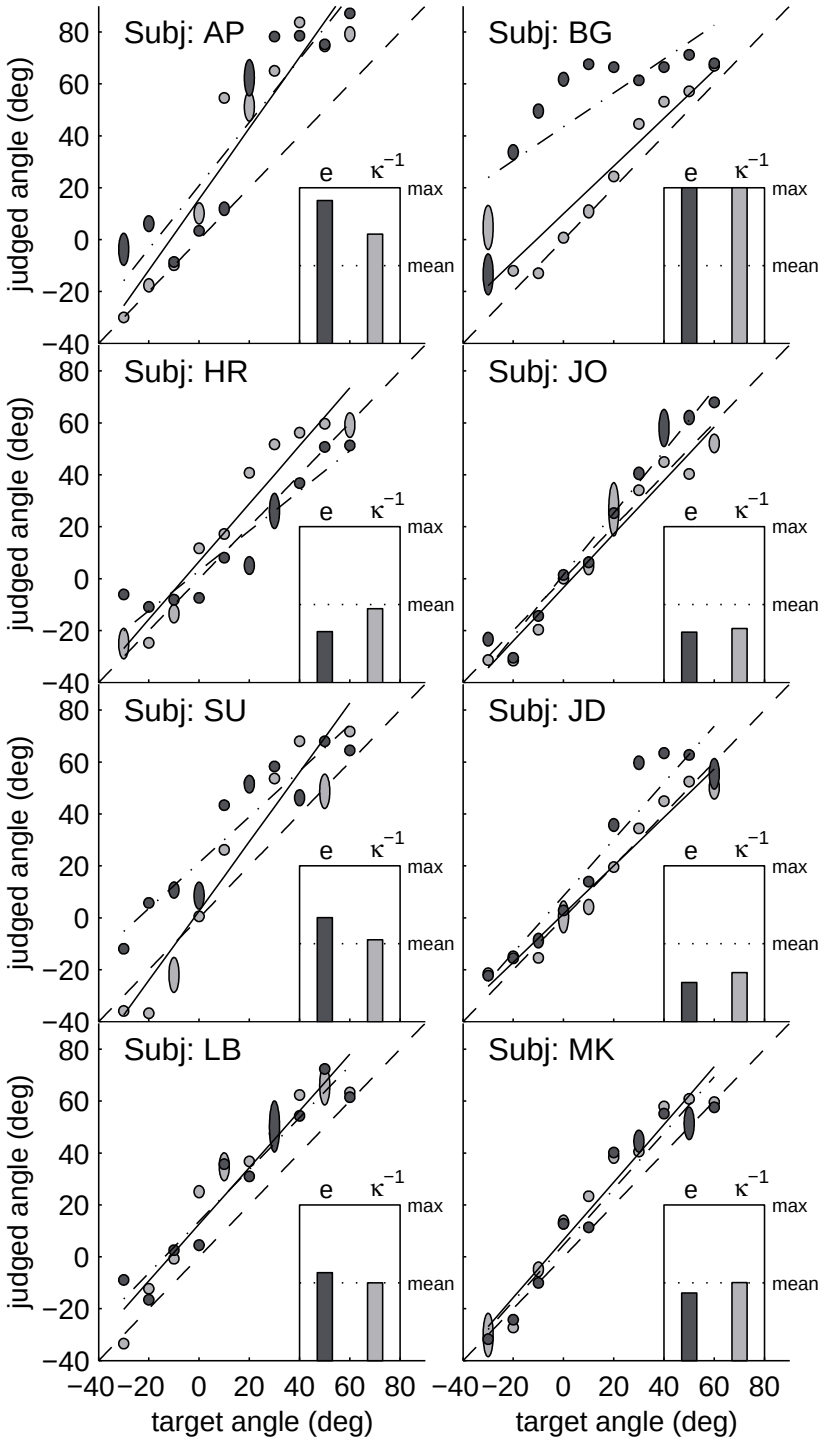


Figure 2.6: Localization performance for source locations in the median plane. The light gray centroids represent source positions for frontal sound incidence and the dark gray source positions for rear sound incidence. Regression functions are calculated for both hemispheres separately (solid line: frontal hemisphere, dash-dotted line: rear hemisphere). The bar diagrams show the individual localization ability relative to the mean across all subjects.

Although the perception of the stimulus location is very accurate for frontal sound incidence, the rear elevations are highly overestimated. However, the judged elevation is limited to 60° elevation. The subject reported that she knew the limitation of target locations to 60° and, therefore, did not judge higher elevations. If she had not known this, she would have judged higher elevations resulting in a more linear behavior of the localization data at higher elevations.
The results from a quantitative investigation of the localization data are summarized in Table 2.2. Each parameter was computed independently for the frontal and rear hemisphere. Mean values across subjects are presented in the last row of that table. It can be seen from the data, that the localization accuracy in the rear hemisphere is reduced compared to the frontal hemisphere. The inter-individual differences in localization performance represented by the bar diagrams in Figure 2.6 in each sub-plot are similar to the diagrams for the horizontal plane. Subjects AP and BG show a poorer localization performance than the mean and subject JO (one of the authors) a better than normal. However, the localization accuracy of the subjects JD, MK, HR is comparable to JO 's data, indicating that the a priori knowledge of the possible source positions and their discrete distribution is not as important as in the azimuth condition.
The mean absolute error (dark gray bar) and κ^{-1} are highly correlated (r=0.92).

2.3.3 Comparison with data from the literature

The results of the localization experiments described in the former sections are compared to data from the literature in Table 2.3 [3,4]. The derived parameters 'm' (slope of regression line), 'b' (intercept of y-axis), 'r' (correlation between target and judged locations) that are listed in Table 2.3, are separately computed for azimuth and elevation, whereas '$\bar{e}$' (mean absolute error), 'κ^{-1}' (spread of judgement) and 'fb' (number of front-back confusions in percent) are averaged across both dimensions.

[3]A brief description of the free-field localization experiments that were used for a comparison of the localization accuracy is given in Appendix A.

[4]The data from the literature was obtained as follows: If data was given for each subject, the mean across subjects was computed. The row 'Gilkey' represents data of experiment I in (Gilkey *et al.*, 1995). The next row 'Gilkey (W & K) ' shows data from Wightman & Kistler (1989a), subject SDO and SDE re-analyzed by Gilkey et al.. The row 'W & K' represents native data from (Wightman and Kistler, 1989a) taken from their Table II (correlation and reversals) and Table III (mean angle of error and κ^{-1} averaged across 0° and 18° for source positions in the azimuth and across all elevations in the frontal and rear quadrant for the comparison in the elevation). The last row shows the mean values from the current study taken from Tables 2.1 and 2.2. The average angle of error and the mean of κ^{-1} were computed differently by Gilkey and Wightman & Kistler. In the former study median values were calculated, whereas the latter presented mean values. To account for these deviations, both median and mean values were computed for these parameters in the current study. Median values are listed in parenthesis.

Subject	m_f	m_r	b_f	b_r	r_f	r_r	$\bar{e}_f[°]$	$\bar{e}_r[°]$	κ_f^{-1}	κ_r^{-1}	f/b [%]
AP	1.37	1.22	15.6	20.8	0.951	0.921	22.5(21.86)	27.6(27.01)	0.034(0.027)	0.153(0.036)	23.3
BG	0.92	0.65	9.93	43.49	0.938	0.761	16.24(14.29)	38.35(38.93)	0.094(0.070)	0.046(0.029)	08.3
HR	1.11	0.77	6.58	3.02	0.967	0.943	11.51(08.76)	10.96(9.30)	0.012(0.006)	0.036(0.029)	11.7
JO	1.03	1.19	-3.41	-1.46	0.978	0.985	07.24(09.35)	08.65(08.58)	0.012(0.003)	0.018(0.017)	00.0
SU	1.33	0.88	2.62	21.23	0.946	0.939	16.75(16.47)	20.89(19.43)	0.021(0.010)	0.041(0.032)	18.3
JD	0.93	1.09	1.47	8.35	0.982	0.955	05.72(05.30)	11.92(08.83)	0.008(0.005)	0.013(0.012)	13.3
LB	1.09	1.00	12.52	13.69	0.964	0.962	16.67(17.08)	15.70(13.44)	0.045(0.018)	0.033(0.021)	10.0
MK	1.11	1.08	6.51	4.48	0.973	0.967	11.73(12.09)	11.46(11.61)	0.020(0.014)	0.031(0.025)	00.0
Ø	1.11	0.95	6.55	14.25	0.962	0.929	13.55(13.15)	18.19(17.14)	0.030(0.019)	0.046(0.025)	10.0

Table 2.2: Localization performance in the median plane. The analyzing parameters are the same as in Table 2.1 but separately calculated for the frontal and rear hemisphere. Values in parenthesis are median values. All other values are mean values across source locations. The indices indicate if the parameters were calculated for frontal (f) or rear (r) sound incidence.

Paper	m_a	m_e	b_a	b_e	r_a	r_e	$\bar{e}$	κ^{-1}	fb
Gilkey	0.97	0.703	-2.47	6.87	0.996	0.889	(18.2)	(0.035)	
Gilkey (W&K)	1.01	0.77	0.85	8.45	0.995	0.829	(20.95)	(0.047)	
W & K					0.982	0.903	21.04	0.052	6
Otten	0.98	1.03	-1.00	10.4	0.992	0.945	14.64(13.64)	0.034(0.019)	9.5

Table 2.3: Comparison of parameters from this study with data from the literature (slope m, y-axis intersection b, correlation coefficient r, mean absolute error $\bar{e}$, mean spread κ^{-1}) and front-back confusions in percent fb. Indices denote those values that are computed separately for azimuth and elevation. Values in parenthesis are median values[4].

A detailed comparison of the values listed in Table 2.3 is not appropriate because of differences in the methods between studies and the inter-individual differences between subjects. However, it is obvious that similar results are obtained in the current study compared to the data from the literature. Although the subjects of this study were completely untrained, there is a tendency for higher localization accuracy in the current study, marked by a comparatively low mean angle of error $\bar{e}$ and spread of the data as indicated by κ^{-1}. This can be related to the restricted range of source positions in the current study because the highest localization uncertainty occurs at higher elevations for rear sound incidence (e.g. (Oldfield and Parker, 1984a)). In the current study only few source positions are located in this region. As expected, the localization uncertainty is increased in this region but the mean localization performance is dominated by the higher accuracy at the remaining source positions. Furthermore, in the study of Wightman and Kistler scrambled white noise stimuli were used to prevent the subject from using monaural cues, whereas unscrambled stimuli were used in the current study. Hence, the subjects were also able to use monaural spectral cues for estimation of the source location. It is likely that the acuity is increased by additional spatial information provided by monaural cues.

A comparison across studies of the mean absolute localization error in azimuth is given in Figure 2.7 [5]. The overall shape of the error as a function of azimuth is very similar across studies. There is a trend towards smaller errors for frontal source positions in the data of the cited literature that can not be observed in the results of the current study. This could be caused by the lack of some kind of head fixation (e.g. a bite bar (Gilkey et al.) or an acoustical reference location (Makous and Middlebrooks)). Subjects were allowed to move their head between trials and had no reference point to re-establish the head orientation before the next stimulus was presented.

[5]The data for Gilkey et al. and Makous & Middlebrooks was obtained from Table 2 in (Gilkey *et al.*, 1995) by averaging across $\pm 5°$ elevation. The data from this study is a re-plot of the data from Figure 2.5 collapsed over the left/right hemispheres.

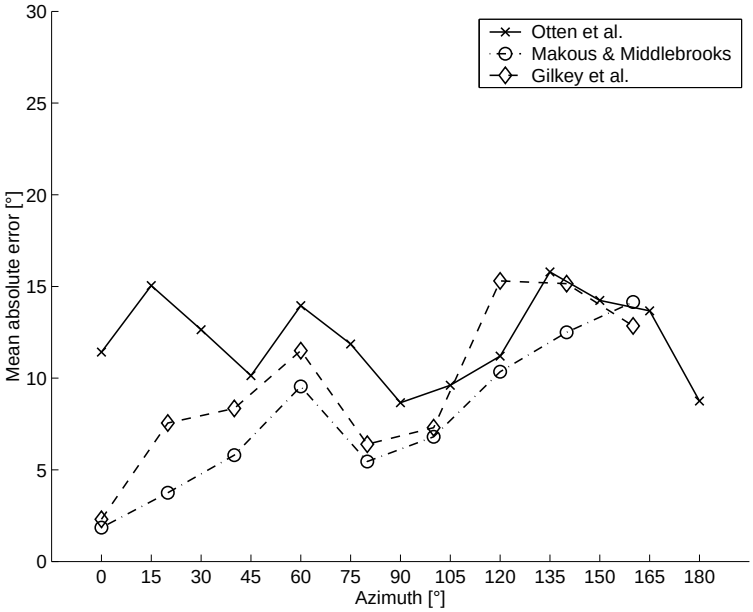


Figure 2.7: Comparison of the mean absolute error measured by Makous & Middlebrooks and Gilkey et al. with the current study.

Hence, the increased spatial resolution for frontal sound incidence could be concealed by changes of the orientation of the listeners head between stimulus trials.

2.4 Validation of the GELP technique

The experiments presented in this section were conducted to validate the different implementation of the GELP technique and its use in a darkened room.

2.4.1 Method

2.4.1.1 Implementation of the GELP technique

The general idea behind the GELP technique is that the spherical surface of possible source locations surrounding the subject is mapped to a globe with a much smaller diameter in front of the subject. This mapping is done by projecting the center of the subjects' head to the center of the sphere in front of the subject. The subject has to point to the corresponding point on the globe as if the subject was sitting inside.
The globe employed here has a diameter of 30 cm and consists of polystyrene. To facilitate the orientation on the sphere in a darkened room, the horizontal plane, the median plane and the planes with a constant elevation of $-30°$, $30°$ and $60°$ were carved into the surface. The sphere is placed on a wooden stand at height of 80 cm, which

makes it comfortable for the subject to reach any point on the sphere. The subject was seated on a chair that could be adjusted in height. To measure the position indicated by the subject, a Polhemus inside track pointer was used. The emitter (Model 3A06906) is mounted at the stand of the sphere and a normal receiver (Model 4A0332) was used to point to the source locations[6]. If the receiver had a distance greater than 1 cm from the surface of the sphere recording of data was not possible. The position of the pointer was recorded by the computer if the receiver was placed on the surface for one second. A short tone, transmitted by a loudspeaker mounted under the subjects' chair, acknowledged the recording of the data. The inside track was controlled by the computer that was also responsible for the movement of the TASP system.

2.4.1.2 Subjects

A total of 15 subjects participated voluntarily in the experiments. At least seven subjects participated in each experiment. The subjects were aged from 27 to 34 years and had normal hearing. All subjects were members of the physics and psychology department of the University of Oldenburg. Except for subject 'JO', none of the subjects received any training or had pre-knowledge about the measurements.

2.4.1.3 Procedure

Three control experiments where conducted in separated sessions. Two numerical values, representing azimuth and elevation coordinates, were displayed on a monitor screen in front of the subject ('numeric' condition). The task of the subject was to point to the corresponding location on the spherical surface of the GELP technique. In this experiment subjects were sitting in a normal reverberant room because no acoustical stimulus was presented. After recoding the response of the subject, feedback in terms of the judged azimuth and elevation angles was given. The stimuli locations were equally distributed in azimuth (15° spacing). However, for each azimuth a different angle of elevation was randomly chosen from $-30°$ to $60°$ in steps of $10°$. The positions were presented three times in random order. Note that only one randomly distributed elevation at each azimuth was chosen.
The general measurement procedure in the conditions 'visual I+II' was equivalent to the free-field localization measurement (see above). The task of the subject in the 'visual I' condition was to judge the location of one sledge of the TASP system in the lighted anechoic room. A different set of source positions but distributed in the same way as in the 'numeric' condition was chosen. To identify which of the two speakers of the TASP system was the target, a short click train was emitted.

[6]Although it has no nib as the Stylus (compare Gilkey et al. (1995)) we felt that direct contact with the surface of the sphere facilitates the use of the pointer.

To examine if subjects are able to handle the GELP technique in a darkened room, they had to judge the location of a little diode, mounted in the center of the speaker ('visual II'). It was not possible for the subjects to see the globe of the GELP technique. Hence, the subjects had to use their tactile sense to find the desired location on the spherical surface.

2.4.2 Results

In Figure 2.8 results from the control experiments and the free-field localization experiment are shown for two representative subjects. In the left column data for subject 'JO' is shown and in the right column the results for subject 'MK' are given for source positions in the horizontal plane. The centroids were stretched proportionally to κ^{-1}, if κ^{-1} for the current azimuth is greater than the mean across all azimuth positions for that subject. A linear regression function is plotted in each panel.
In the first row data obtained in the 'numeric' condition is plotted. The judgements are near to the optimum performance for every angle of azimuth. Both subjects are able to position the pointer of the GELP technique very accurately.
In the 'visual I' condition more spread can be seen in the response pattern. Furthermore, the centroids are slightly more distant from the optimum performance. This tendency remains for the two other experiments and the highest error can be seen in the free-field localization task. This qualitative description is quantified in Figure 2.9. Here, the mean absolute error (averaged across the left and right hemisphere) as a function of the stimulus azimuth is plotted for the four different experimental conditions. In addition, data from 'experiment II' from Gilkey et al. (1995), averaged across $\pm 5°$ elevation is shown (dashed lines). This experiment is very similar to the 'numeric' condition in the current study. It deviates only in the presentation of the azimuth and elevation angles, which were reported verbally to the subjects.
The absolute error is lowest in the 'numeric' experiment and highest for the acoustical free-field presentation. The input performance in the 'visual II' condition (darkened room) is substantially reduced in comparison to the 'visual I' condition (lighted room). Hence, the handling of the GELP technique in the darkened room seems to be complicated. However, the main constraint given in the introduction was that the localization error in the free-field localization experiment is higher in comparison to the error in the control experiments. A non-parametric ANOVA (Kruskal-Wallis) performed on the mean localization errors for the 'visual II' condition and the free-field localization experiment shows that mean localization error in the acoustical localization experiment is still higher ($p < 0.01$).

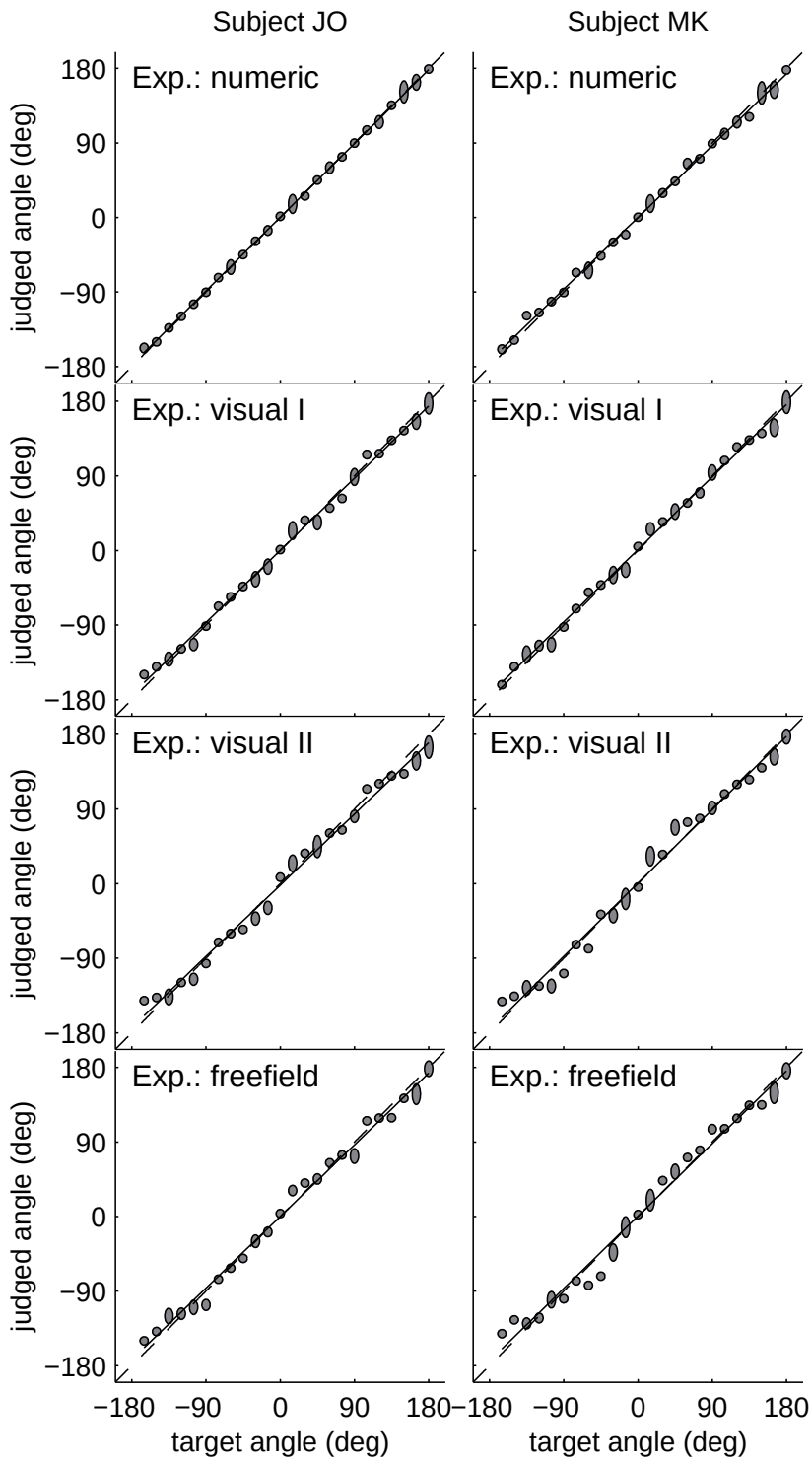


Figure 2.8: Validation of the GELP technique: Judged azimuthal angles for various conditions ('numeric', 'visual I': lighted room, 'visual II': darkened room, acoustical) as a function of the azimuthal target angle. Results for two representative subjects (left and right side) are shown for three control conditions (row 1-3) and for the free-field localization experiment.

Minima can be seen in the regions around 0°, 180° and 90° for each condition. These regions were marked by carves on the surface of the GELP sphere and this seems to ease the handling of the technique. A comparison of the results from the 'numeric' condition to the data from Gilkey et al. reveals that the performance is comparable in the frontal hemisphere. An increase of the mean absolute angle of error for increased azimuths can be observed for the data from Gilkey et al. This might be due to fixation of the subject's head by a bite bar that makes it more difficult to point to rear positions on the surface of the GELP sphere. The smaller error in the current study could also be caused by the greater size of the sphere (30 cm compared to 20 cm in the study of Gilkey et al.) because a small displacement of the pointer on the surface of the globe generates smaller errors if the diameter of the sphere is increased.

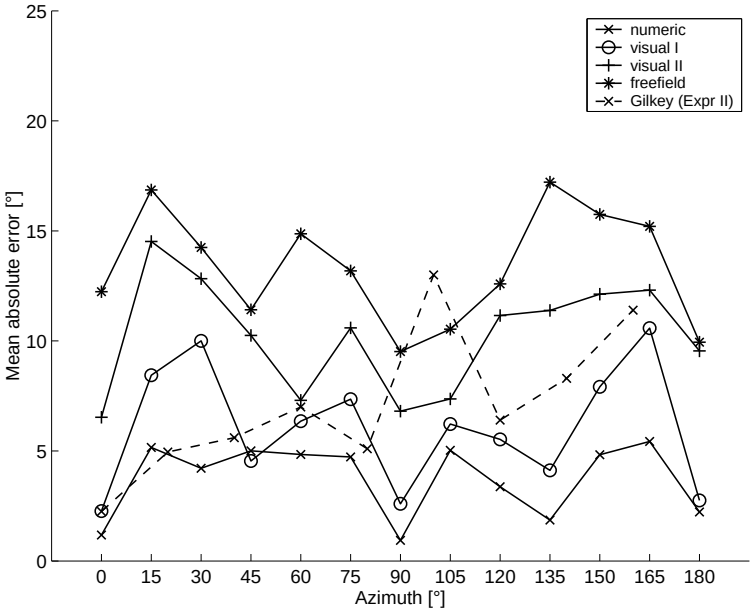


Figure 2.9: Mean absolute error averaged across subjects and left/right hemispheres under four conditions are shown as solid lines (see legend). The dashed line shows data of the verbal presentation experiment II from Gilkey et al. (1995).

2.5 Discussion

2.5.1 TASP and free-field localization

In the current study a method for positioning a sound source on a spherical surface was presented. The TASP system allows almost continuous sampling of the virtual sphere of source positions. The upper and lower limits of the elevation angle are $-40°$ and $+80°$ and the whole azimuth range is covered. The average time interval between two stimulus presentations is about approx. 6 s.

In a free-field localization experiment eight subjects were requested to localize a click train stimulus presented from positions out of the horizontal and median plane. The localization performance for positions in the horizontal plane were very accurate for most subjects. However, two subjects showed a considerably lower localization performance in azimuth as well as in elevation. Inter-individual differences in the localization performance have also been found in the literature. For instance, subject 'SDE' in the study from Wightman and Kistler (1989b) showed an accuracy that was considerably lower than the average. The lower localization performance was mainly found for judgements of the elevation and is also expressed by the number of front-back confusions. Wightman and Kistler related the lower localization performance to the physical spectral cues provided by the head related transfer functions (HRTFs). They showed that for subject 'SDE' less spectral information was contained in the spectral cues compared to other subjects. Hence, the decreased localization accuracy for subject 'SDE' might have been caused by less spatial information provided by the HRTFs. In the current study, the lower localization performance of the two subjects also occurred in the horizontal domain. It is unlikely that the HRTFs for the subjects with a decreased acuity provide less *binaural* information. Hence, it can be assumed that the low localization performance is not only caused by a lack of spatial information contained in the HRTFs but by a decreased utilization of the physical cues available to the subjects.

In the current study no method was used to center the head of the subject to the center of the hemi-arcs of the TASP system. Hence, it was possible that the position of the head was changed slightly between stimulus presentations. An analysis of the mean absolute error in azimuth showed that the error was higher for frontal sound incidence in comparison to studies in which a head fixation (Gilkey *et al.*, 1995) or a reference position given by an acoustical stimulus from 0° azimuth and elevation (Makous and Middlebrooks, 1990) was used. Therefore, to be able to measure the higher localization accuracy for frontal sound incidence, the head of the subject has to be centered to the middle of the sphere of possible source locations before each stimulus presentation. However, a fixation of the head reduces the flexibility of the subjects and a stimulus from a reference location could change the absolute localization task to a discrimination task for the reference location. Hence, it seems to be suitable to center the head by a head monitoring technique that gives verbal or visual instructions to the subject to center the head. Such a technique has been used, for instance, by Kulkarni and Colburn (1998) to center the head for measuring head related transfer functions.

A comparison of the mean localization performance (averaged across subjects and source positions) revealed that despite of the differences in the methods (reduced set of source positions, click train stimulus, recording technique and untrained subjects) the acuity is comparable across studies. It can be concluded that the use of the TASP system for positioning the sound source did not influence the localization data.

GELP in a darkened room

The GELP technique was used to collect the localization data in the free-field localization experiment. Although the technique was already validated by Gilkey et al., a re-examination was necessary because the ability of subjects to handle the sphere in a darkened room was unclear. Therefore, three different control experiments were conducted with non-acoustic spatial stimuli. In the first experiment the coordinates of the stimulus position were given in terms of azimuth and elevation angles to the subject. The subjects were able to point to the corresponding positions of the surface of the GELP sphere in the lighted room very accurately (mean angle of error: approx. 4°). In two further control experiments subjects' capability to handle the GELP technique in a lighted and a darkened room was investigated. In the 'visual I' condition subjects had to judge the position of a sledge of the TASP system (mean angle of error: approx. 6°) and in the 'visual II' condition a little diode in the center of the sledge served as a target in the darkened room (mean angle of error: approx. 9.5°). The differences between the mean angle of error in these two conditions can be related to two properties: First, the geometry of the anechoic room and the visual cue of any reference direction could be used by the subjects in the 'visual I' condition. The absence of this aid in the 'visual II' condition could complicate the allocation of source positions to positions on the GELP globe. Second, the subjects were not able to see the surface of the GELP sphere in the darkened room. This also seems to increase the input uncertainty. However, a comparison of the 'visual II' condition to the free-field experiment shows that the mean absolute error for the presentation of an acoustical stimulus in the free-field is still above the error obtained in the 'visual II' condition.

In order to validate the GELP technique, Gilkey et al. conducted an experiment which was similar to the 'numeric' condition in the current study. A comparison of the data from both experiments showed that the subjects in the current study were able to handle the technique with a higher accuracy. This can be related to the bigger size of the GELP sphere and the lack of a head fixation. Although the head fixation increases the localization accuracy for frontal sound incidence, it seems to reduce the input accuracy for positions on the rear surface of the GELP sphere. Therefore, it can be concluded that an adjustment of the head position by emanating a stimulus from a reference position (as used by Makous and Middlebrooks) or by monitoring the head position should increase the localization accuracy for frontal sound incidence. These techniques should be preferred because they do not reduce the flexibility of the subjects to handle the GELP technique.

The influence of using the GELP technique in the dark could be further investigated by conducting the 'numeric' condition in a darkened room and presenting the azimuth and elevation coordinates by a verbal report. A comparison of the input accuracy in the lighted and darkened room could directly show the error that is introduced by using

only the tactile sense for handling the GELP technique.
A main advantage of the GELP system is that it enables to collect localization data at a high rate. The handling of the GELP technique in the dark substantially lowers the collection rate. Gilkey et al. stated that they were able to measure 16-20 source positions per minute (by using a static loudspeaker array). This rate can not be achieved if the subject can only use the tactile sense for handling the GELP technique. Although the collection rate was not measured explicitly, it can be specified with 3-5 stimulus positions per minute for the measurement setup presented here.
However, the GELP technique seems to be a suitable method for collecting localization data. Its implementation is less expansive than the head monitoring technique (Makous and Middlebrooks, 1990), at the same time it is as accurate as the verbal report (Wightman and Kistler, 1989b) and even without any training a high accuracy can be accomplished by subjects.
In general, it can be concluded that the combination of the GELP technique with the TASP system is a suitable setup for measuring the localization accuracy. This method can be enhanced by using a head monitoring technique to re-establish the position of the subjects head before the localization stimulus is presented.

Chapter 3

Head related transfer functions and the effect of spectral smoothing on individual localization cues

Abstract

Head related transfer functions (HRTFs) were measured from 11 subjects and one dummy head with high resolution in azimuth and elevation. The head related impulses responses (HRIRs) were obtained at the blocked ear canal entrance by using maximum length sequences (MLS). Binaural and monaural localization cues are calculated from the HRTFs and presented for selected source positions. The inter-individual differences of the localization cues are investigated by their standard deviations across subjects as a function of azimuth and elevation. Furthermore, the individual HRTFs are compared to the HRTFs from the dummy head. The results show, that both the binaural and monaural localization cues of the HRTFs strongly vary across subjects at low elevations and are less individual at high elevations. A comparison between the individual HRTFs and the dummy head HRTFs revealed, that the dummy head can not serve as an average listener, if spatially correct perception is needed. In order to reduce the amount of data required for an individual spatial auralization, the effect of cepstral and $1/N$ octave spectral smoothing is investigated on I) the inter-individual standard deviation of the spectra across subjects, II) the interaural level difference (ILD), III) the interaural time difference (ITD) and IV) the length of the HRIRs. $1/N$ octave smoothing introduces high ILD deviations to the smoothed HRTFs and is, therefore, not recommended for spectral HRTF smoothing. Cepstral smoothing with 16 coefficients, on the other hand, introduces only perceptually irrelevant changes to the binaural and monaural localization cues. Note, that this is only true if the ITD of the minimum phase HRTFs are computed from low-pass filtered impulses responses. An further advantage of cepstral smoothing is that is reduces the length

of the impulse responses more effectively than $1/N$ *octave smoothing.*

3.1 Introduction

The physical properties that are exploited by the auditory system to estimate the position of a sound source are captured by head related transfer functions (HRTFs). They described the directional dependent transformation of a sound from its source location to a point within the ear canal. The cues that are provided by HRTFs and which are characteristic for each source position can be divided in two groups. The binaural localization cues (interaural level difference, ILD and interaural time difference, ITD) are obtained from a comparison of the left and right ear HRTFs, whereas the spectral filtering of the source spectrum due to interferences effects and pinna filtering is introduced at each ear individually and is, therefore, called monaural cue (see (Blauert, 1974; Middlebrooks and Green, 1991) for comprehensive reviews of localization cues). If the HRTFs of the left and right ear for a certain source direction are known, they can be used to introduce the localization cues for that spatial direction to an arbitrary sound source by convolving it with the head related impulse responses (HRIRs), which are the corresponding time domain representations of HRTFs. Thus, a set of HRTFs, sampled from the whole spatial range of directional perception provides the possibility to project a sound source by headphones to any of the sampled locations. This technique is called virtual acoustics. It has been shown, that a virtual source presentation, based on individual measured HRTFs, is capable of producing an acoustical perception with an accuracy that is near to the free-field condition (Wightman and Kistler, 1989a; Wightman and Kistler, 1989b; Hammershoi, 1995; Otten, 1997; Kulkarni and Colburn, 1998).

Although HRTFs from different subjects have similar shapes in ITD, ILD and spectral filtering, the details of each cue are highly individual (e.g. (Møller *et al.*, 1995)). Therefore, it is not sufficient to use non-individualized HRTFs to yield a localization performance that is comparable to the free-field acuity (Wenzel *et al.*, 1993). However, it is a major effort to measure individual HRTFs for a number of source positions covering the whole range of spatial directions. The use of dummy heads that provide the localization cues of an average subject would facilitate the generation of virtual displays. Therefore, a comparison of dummy head HRTFs and individual HRTFs provides valuable information about the needs for creating dummy heads and appropriate auralization methods for virtual acoustic environments.

The aim of the first section in this investigation is to describe HRTFs measured from 11 subjects and one dummy head. These HRTFs are used in the subsequent chapters of this thesis to create individual virtual stimuli. The HRTFs are described by extracting the monaural and binaural localization cues and by presenting standard deviations across

subjects. Furthermore, the capability of the dummy head to serve as an average subject is investigated by comparing dummy head HRTFs to individual HRTFs.
Virtual acoustic displays are created by realizing HRTFs as digital filters. To reduce the computational effort of the digital filters, the filter order is often reduced by roughly approximating the HRTF spectra. Furthermore, if finite impulse response (FIR) filter are used, the filter length is reduced by applying minimum phases to the HRTFs because they have a minimal energy delay (Oppenheim and Schafer, 1975). However, HRTFs that are approximated by digital filters have to provide the same directional properties as the original HRTFs. That implies, that each manipulation (e.g. all pass filtering, smoothing) may not alter perceptually relevant localization cues.
To gain further insight into the effects that smoothing has on minimum phase HRTFs, the effect of smoothing is analyzed on I) the inter-individual standard deviation of the HRTF spectra, II) the interaural level differences and III) the interaural time differences. Furthermore, to asses the computation time that is saved by smoothing, the length of the impulse responses is analyzed as a function of spectral detail reduction.

3.2 HRTF measurements

HRTFs were described by a variety of studies (see (Møller *et al.*, 1995) for a summary). The main difference in the method between studies is the type of microphone and its location within the ear canal. The position of the microphone within the ear canal or the cavum concha and its influence on the HRTFs was investigated by several researchers (e.g. (Wiener and Ross, 1946; Mehrgardt and Mellert, 1977; Hammershoi and Møller, 1996)). The investigations show that, within the frequency range of interest, all spatial information is present at any point within the ear canal because the transition of the sound from the entrance of the ear canal to the eardrum does not add spatial information. Hence, it is not necessary to place a probe tube near to the ear drum. A recording location at the entrance is sufficient for capturing all spatially relevant information. In the present study, the position of the microphones was several millimeters inside the ear canal, which was blocked by the microphones.
Impulse responses can be obtained from a variety of measurement techniques (e.g. single clicks, sweeps, noises, etc.). Because a high number of impulse responses have to be measured, the method has to allow for accurate measurements in a short time. By using maximum length sequences (MLS) a high signal to noise ratio is obtained by only few measurement repetitions (Alrutz, 1983; Rife and Vanderkooy, 1993).

3.2.1 Theory

In the following it is described in which way the HRTF $A(\omega, \phi, \theta)$ can be obtained by using a MLS stimulus. The angles ϕ and θ denote the position of the sound source in spherical coordinates.
The MLS stimulus is acoustically radiated by a loudspeaker and recorded by microphones in the ear canal of the subject. The sound pressure $y(t, \phi, \theta)$ recorded in the ear canal consists of the MLS stimulus $m(t)$ convolved with the impulse response of the complete electroacoustical transducer system $h(t, \phi, \theta)$.

$$y(t, \phi, \theta) = h(t, \phi, \theta) * m(t) \tag{3.1}$$

One important feature of the MLS sequence is, that its auto-correlation function is a delta impulse with a small DC offset, which is inverse proportional to the length of the sequence. The DC offset is neglected in the present case because it is assumed that the sequence is sufficiently long. Thus, by a cross correlation of $y(t, \phi, \theta)$ with the MLS sequence the impulse response of the complete system including the HRIRs can be extracted[1].

$$m(t) \otimes y(t, \phi, \theta) = m(t) \otimes m(t) * h(t, \phi, \theta) = \delta(t) * h(t, \phi, \theta) \tag{3.2}$$

The impulse response $h(t, \phi, \theta)$ can be split into in the impulse response of the electroacoustical system $e(t)$ (that is directionally independent) and the HRIR $a(t, \phi, \theta)$.

$$h(t, \phi, \theta) = e(t) * a(t, \phi, \theta) \tag{3.3}$$

The impulse response of the measurement system $e(t)$ can be obtained by recoding the MLS stimulus at the position corresponding to the center of the head with the head absent. If $e(t)$ is known it can be used to extract the HRIR by deconvolving $h(t, \phi, \theta)$ with $e(t)$. The easiest way to accomplish this is to transform Equation 3.3 into the time domain and to solve for $A(\omega, \phi, \theta)$:

$$A(\omega, \phi, \theta) = \frac{H(\omega, \phi, \theta)}{E(\omega)}. \tag{3.4}$$

The inverse transfer function $E^{-1}(\omega)$ is only stable if it is minimum phase (see (Oppenheim and Schafer, 1975)). If it is not minimum phase the calculation given by Equation 3.4 can be performed on the absolute spectra and a minimum phase phase or a linear phase can be applied to the HRTFs. In this case the absolute spectrum of $E(\omega)$ may not contain zeros. The construction of minimum phase HRTFs is discussed in Section 3.3.4.

[1]The computational effort to calculate a cross correlation is proportional to the square of n (n denotes the length of both correlation sequences). To reduce the computation time a Fast-Hadamard transformation was used. In principal, it applies a butterfly algorithm to the correlation process that reduces the computation time to $n \times \log(n)$.

3.2.2 Methods

3.2.2.1 Subjects

Eleven subjects aged from 27 to 34 served as subjects. In addition, the HRTFs of a dummy head (Trampe, 1988) were measured. The dummy head has a rubber like surface with a shape of a normal head without hair. The outer ears have been modelled from the actual ear impression of an 'average' person and were constructed by a computer controlled cutter. At the entrance of each ear canal a B&K microphone (1/2 inch, 4165 capsule) records the sound pressure. The head has no shoulders and torso.

3.2.2.2 Experimental setup

A DSP board (AT&T DSP32C) hosted in a 486-IBM compatible PC was used for the output of the MLS stimulus. The signal was transmitted through an amplifier (Alesis RA 100) to the TASP system (Two Arc Source Positioning, see Chapter 2) for a description of the TASP system). The TASP system was used to position the electro - acoustical transducer (Manger MSW in a self constructed closed enclosure) to the desired location. Two microphones (Sennheiser KE4-211) were used to record the stimulus several millimeters inside the blocked ear canal. The recorded signal was amplified by a microphone amplifier (Unides Design, Model MPA10D) and directly fed into the AD converter entrance of the DSP Board. Stimuli were averaged by summing them up in the DSP memory.

3.2.2.3 Stimuli

Maximum length sequences with a length of 4095 samples were used. The sampling frequency was 50 kHz for subjects and 100 kHz for the dummy head. The stimuli were calculated off-line and stored in the DSP memory. Each position was measured five times (ten times for the dummy head) and averaged in the time domain. The whole range of azimuth positions with a resolution of 5° degree for subjects and 1° for the dummy head was measured. For each azimuth the HRTFs at elevations from $-40°$ to $+70°$ (dummy head: $-30°$ to $+60°$) were recorded with a resolution of 5°.

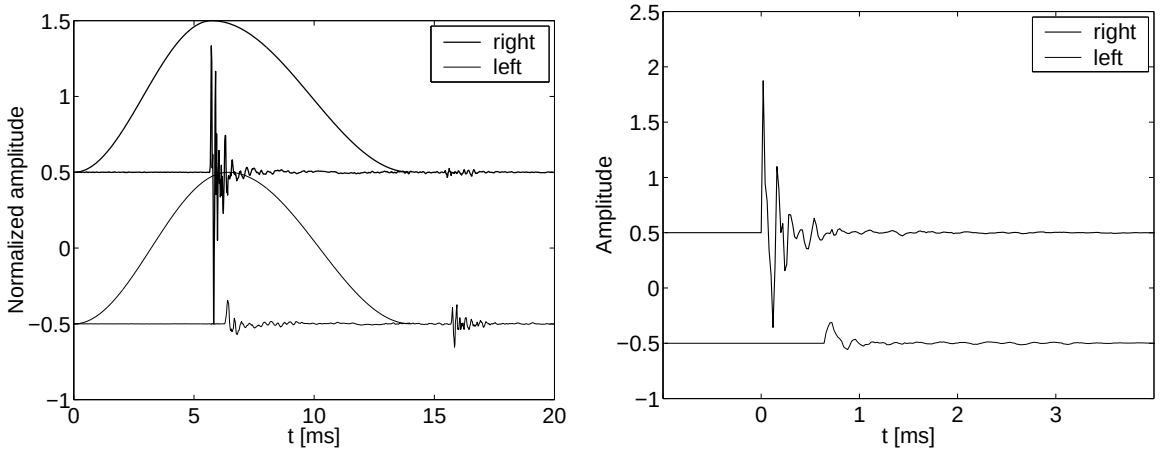
3.2.2.4 Procedure

The subject was seated in a modified bureau chair. A chin rest was fixed to the chair providing a comfortable deposit for the subjects head. The rod of the chin rest was led near to the chest of the subject to prevent it from disturbing the stimulus sound field. A simple method was used to adjust the subject's head to the center of the two

rotating arcs of the TASP system. The two speakers of the TASP setup were initially positioned in the horizontal plane at 0° and 180° azimuth, respectively. A long cord connecting the enclosures of the speaker at 0° elevation was stretched from the rear to the frontal speaker. A knot in the middle of the cord marked the center of the TASP system. The head of the subject was adjusted in a way, that the entrance of the right ear canal was exactly next to the knot of the cord. Before this procedure was performed, it was assured that the median sagittal plane of the subject coincided with the median plane of the TASP system. This was already determined by the geometrical position of the chair and the head rest in the center of the TASP system. The subjects head position was finally checked by eye. After the head had been centered, the cord was removed and the subject was told to remain as still as possible.
The dummy head, which was mounted on a stand, was adjusted in the same way. To remove specular reflections from the platform, the ground was covered with foam.
A measurement at 60° azimuth was taken to avoid an overload of the AD converter on the DSP board. To reduce the delay time between two measurements, the elevation of both the frontal and the rear sledge were adjusted at the same time. For each orientation of the arc in azimuth, first the measurement with the source in the frontal hemisphere and then the measurement in the rear hemisphere was performed. All measurements in elevation were conducted before the arc was moved to a new position in azimuth. The HRTFs were recorded in four separate sessions with overlapping azimuths at −40°, 0°, and +40°. After each session a reference measurement was performed with the microphones located at the center of the TASP system (resp. the center of the subjects head).

3.2.2.5 Data manipulation

The raw microphone signal, consisting of the MLS sequence convolved with the impulse response of the complete electro-acoustical setup was recorded. To obtain the impulse response, a Fast-Hadamard-Transformation was computed (Alrutz, 1983; Borish and Angell, 1983). The HRTFs were extracted by deconvolving the impulse responses with the reference transfer function measured at the center of the head with the head absent. The absolute level differences between sessions were resolved by comparing the levels of the double recordings of the overlapping regions in azimuth. For instance, all elevations for −40° azimuth were measured in the first and second recording session. The overall level of the impulses in the second session was adjusted by the level of the measurements in the first session for the same source locations. Figure 3.1 shows the windowing of the impulse responses $h(t)$ that was used to eliminate reflections from the TASP system before the HRTFs are extracted. The window was constructed by two squared cosine ramps. The sound source was located at 90° azimuth and 0° elevation and the impulse responses are shifted in amplitude for visibility. Reflections can be identified in the left panel of Figure 3.1 at 16 ms. This corresponds to a distance from one ear to the



(a) Measured impulse responses.

(b) Head related impulse responses.

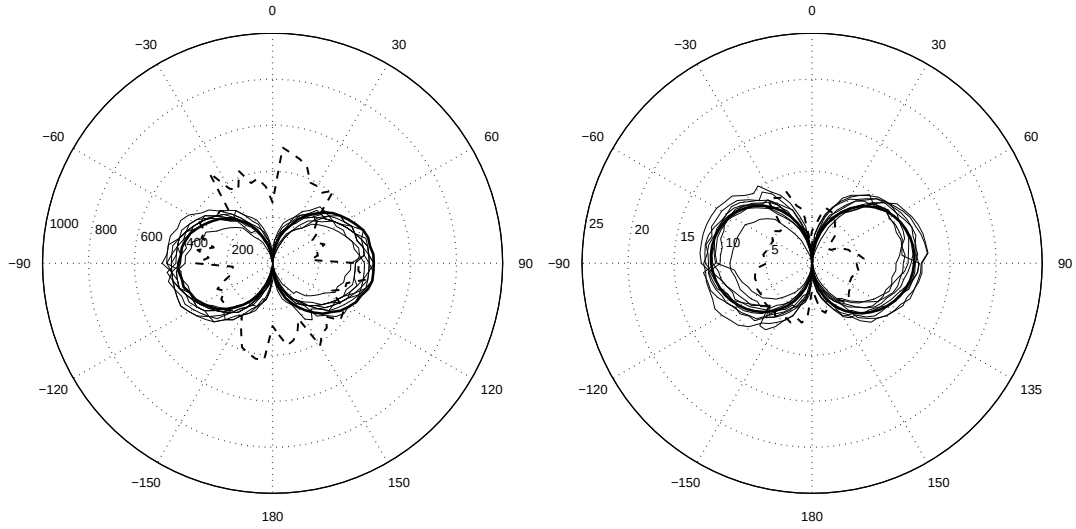
Figure 3.1: The window used for an elimination of reflections is depicted in the left panel. It eliminates reflections at approx. 16 ms. In the right panel the corresponding head related impulse responses are shown. The HRIRs were recorded for a sound source located at 90° azimuth and 0° elevation. In both figures the impulses responses are shifted in amplitude for visibility.

reflecting surface of about 3.44 m, being approximately the diameter of the arc of the TASP system. In the right panel of Figure 3.1 the corresponding HRIRs $a_{r,l}(t)$ for the left and right are shown that were extracted from $h_{r,l}(t)$.

3.2.3 Results and Discussion

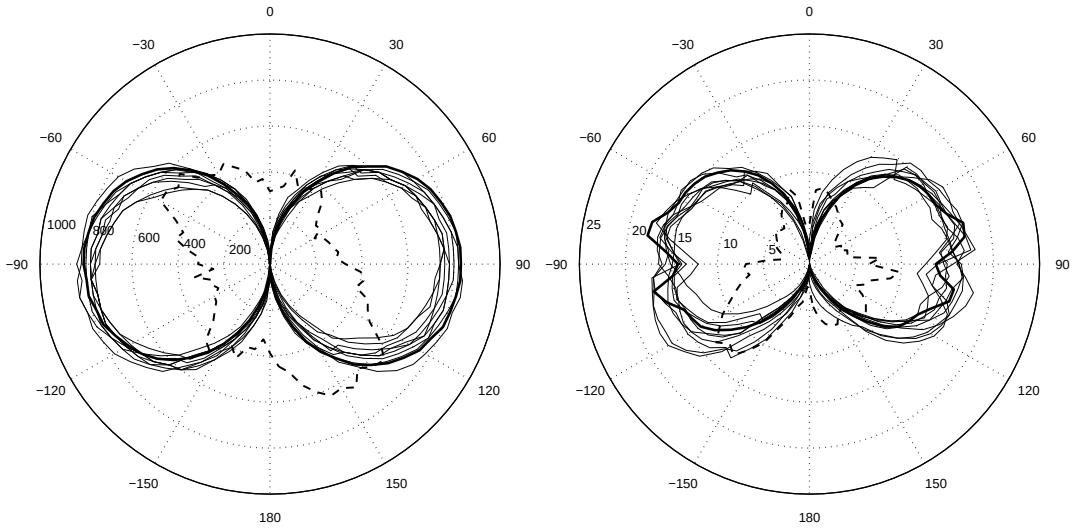
3.2.3.1 ITD and ILD

Interaural time and interaural level differences calculated from HRIRs of nine subjects and one dummy head are shown in Figures 3.2 and 3.3. The ITD is computed as the time shift of the maximum of the cross correlation function of the left and right ear impulse responses. The impulses were low-pass filtered at 500 Hz edge frequency before cross correlation. The ILD is computed as the absolute level difference between the unfiltered left and right ear HRIRs. This is equivalent to a calculation of the *signed* level differences of the HRTF spectra averaged across frequency.
Each polar plot shows absolute values of the interaural parameters as a function of azimuth. The thin dark lines show data for the subjects and the thick lines data for the dummy head. In addition, the standard deviation σ across all subjects, but not the dummy head, is plotted as a dashed line.



(a) ITD at 60° elevation. $\bar{\sigma} = 32.4\mu s$

(b) ILD at 60° elevation. $\bar{\sigma} = 1.06dB$

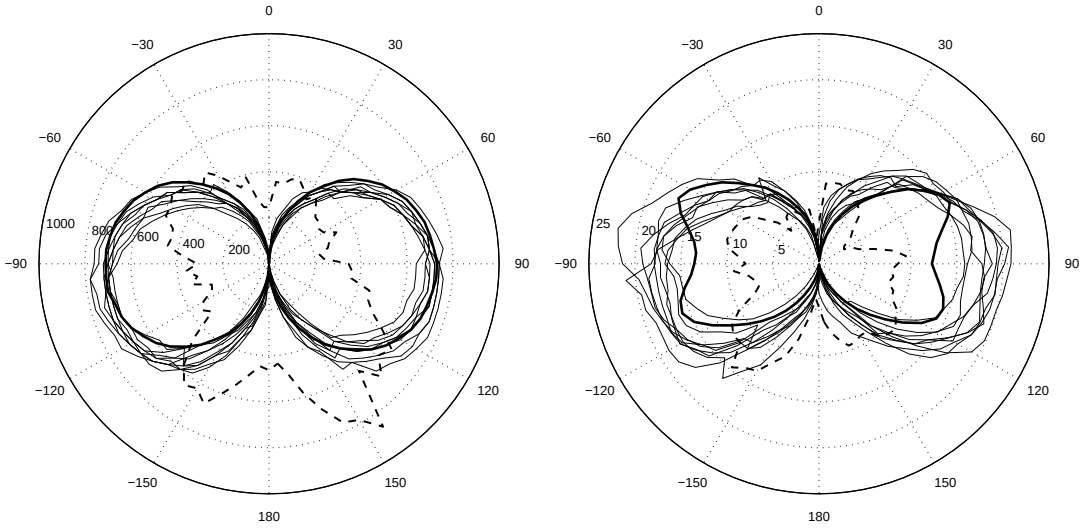


(c) ITD at 0° elevation. $\bar{\sigma} = 40.6\mu s$

(d) ILD at 0° elevation. $\bar{\sigma} = 1.29dB$

Figure 3.2: Polar plots of ILD (left column) and ITD (right column) calculated from HRIRs of 9 subjects (thin lines) and one dummy head (thick lines). Standard deviations of ILDs and ITDs are plotted as dashed lines and zoomed by a factor of 10 for ITDs and 5 for ILDs. Mean standard deviations across azimuth are given in the caption of each figure.

To make it visible within the axis range, the standard deviation was scaled by a factor of 10 for the ITD and a factor of 5 for the ILD. In the left half of Figures 3.2 and 3.3 the ITDs for the elevations of 60°, 0° and −30° are shown, whereas in the right column the ILDs for the same elevations are depicted.



(a) ITD at$-30°$ elevation. $\bar{\sigma} = 43.6\mu s$

(b) ILD at $-30°$ elevation. $\bar{\sigma} = 1.62dB$

Figure 3.3: Same as Figure 3.2 but at $-30°$ elevation.

The maximum ITD value across locations can be observed at extreme lateral source positions at 0° elevation (3.2(c)). Below and above the horizontal plane the maximum ITD decreases as the absolute distance between the source position and the median plane decreases (3.2(a) and 3.3(a)). The circle like structure of the ITD reflects the sinusoidal behavior of the ITD as a function of azimuth. A prediction of the ITD, obtained from theoretical considerations for low frequency components, is given by $\tau = \frac{3a}{c} sin(\varphi)$ (Kuhn, 1977) where a is the radius of the head and c is the velocity of sound in air. This theoretical ITD would result in two circles in the polar plot, one for each hemisphere. The standard deviation across subjects ($\bar{\sigma}$) averaged across all azimuths shows, that for higher elevations the inter-individual differences are significantly smaller than for low elevations ($p < 0.01$). The highest values of σ can be observed, especially at low elevations (Figure 3.3(a)), for azimuthal angles around $\pm 30°$ and $\pm 150°$.
The ILD pattern is similar to the ITD. For higher elevations the ILD values are smaller (3.2(b)), increasing for elevations near to the horizontal plane (3.2(d)). In contrast to the ITD, the ILD for $-30°$ elevation (3.3(b)) are partly larger than in the horizontal plane. The more complex interference pattern of the left and right ear HRTFs, induced by the torso and the shoulders, accounts for this effect. Because the interference is very sensitive to different shapes of the head and the torso, the inter-individual standard deviation is higher for low elevations ($\bar{\sigma}$ = 1.06dB at 60° elevation and $\bar{\sigma}$ = 1.62dB at 0° elevation, p<0.01). The maximum standard deviation across subjects can be seen around $\pm 30°$ and $\pm 150°$ azimuth.

The dummy head ITDs and ILDs are represented by the thick line in each plot of Figures 3.2 and 3.3. Although the range of the dummy head ILDs and ITDs is within the one for individual listeners, substantial differences between subjects and the dummy head can be observed. Furthermore, the deviation pattern varies across elevations. For instance, the dummy head ILD at 60° and 0° elevation is within the individual range, but is substantially smaller at −30° elevation. These differences may occur due to the lacking shoulders and torso of the dummy head. It can be seen, furthermore, that the dummy head cues are much more symmetrical with respect to the median plane and the interaural axis, resulting from the symmetrical geometrical design of the dummy head.

3.2.3.2 Spectral cues: Azimuth

In the following section the logarithmic power spectra of the HRTFs are presented for selected azimuths (0°, 45°, 90°, 135°) at constant elevations (−30°, 0°, 60°). Each subfigure of the Figures 3.4 - 3.6 presents the spectra for five subjects (thin lines) and the dummy head (thick lines). The spectra of the left and right ear HRTFs are plotted in the left and right columns, respectively. Additionally, the standard deviation across subjects is plotted as a thin solid lines at the bottom of each panel with the corresponding axis at the right side.

The HRTF spectra of the subjects and the dummy head are characterized by means of the following properties: a) spectral shape as a function of the source position b) differences between subjects (described by the standard deviation) and c) differences between the dummy head and the individual spectra.

Normally, two prominent resonances can be observed in the HRTF spectra (Shaw, 1997). The first one at approx. 2-3 kHz corresponds to the ear canal resonance. However, this resonance can only be seen if the HRTFs were measured in the open ear canal. The blocked meatus method used in our study, only shows the second prominent resonance at around 4-5 kHz, which belongs to the first mode of a concha resonance (Teranishi and Shaw, 1968). It is exited at both ears from nearly all directions.

The level in the frequency range above 10 kHz is influenced by the head shadow effect. Hence, for lateral sound incidence the level is decreased in the high frequencies (e.g. Figure 3.5, $\varphi = 90°$). The complexity of the HRTF spectra is highest at the contralateral ear for low source elevations and lowest for high elevations at the ipsilateral ear (comp. Figure 3.4 and 3.6, $\varphi = 90°$).

The level in the frequency range below 500 Hz varies only slightly as a function of the source position and shows little variability across subjects. Because the wavelength of this frequency range is about 70 cm, the level variation can not be assigned to the head shadow effect. However, it is likely that reflections from the chair and the legs of the subject cause the low level variations.

The standard deviation of the HRTF spectra is plotted at the bottom of each panel.

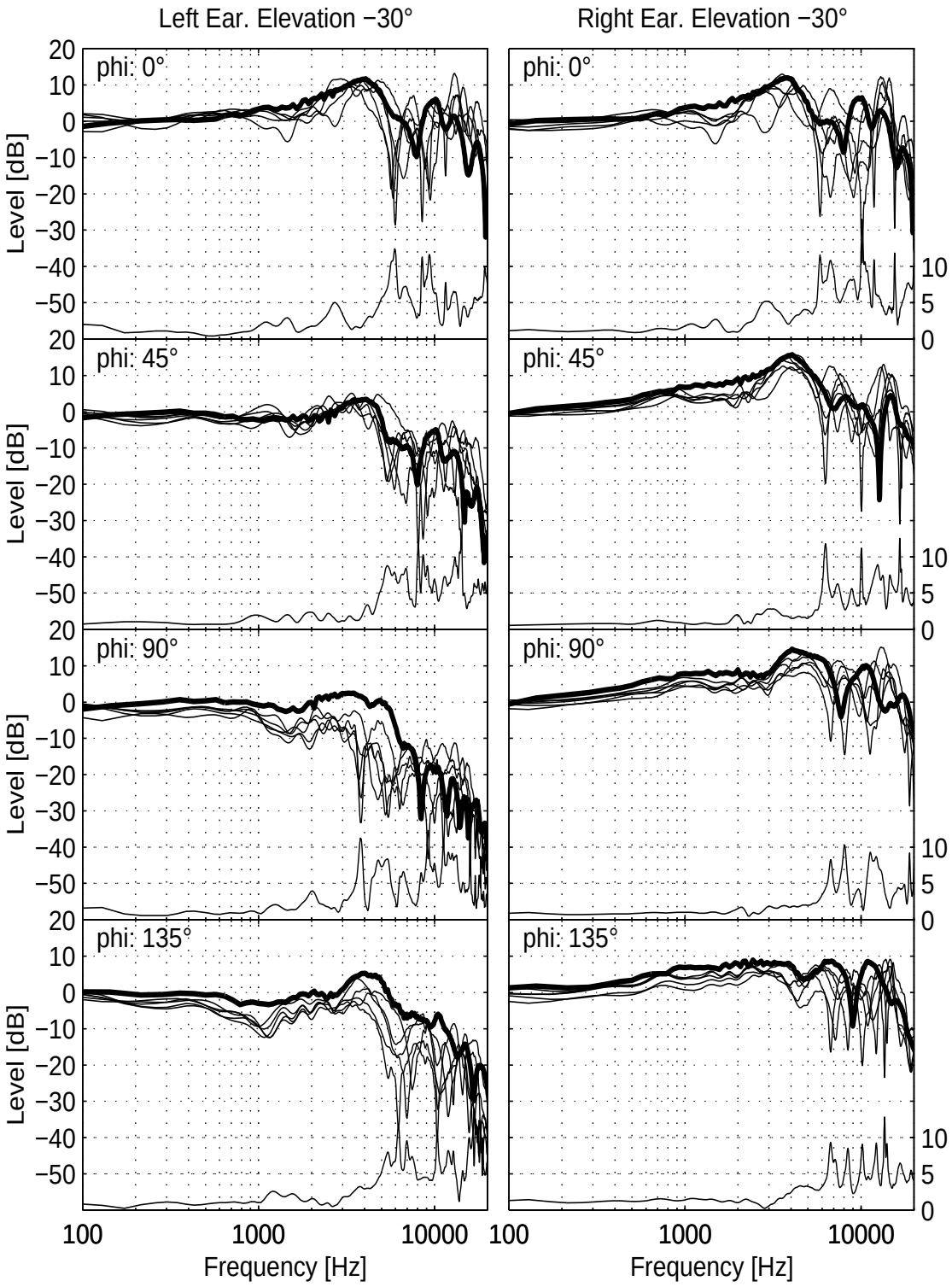


Figure 3.4: HRTFs of the left and right ear recorded from five subjects (thin lines) and one dummy head (thick line) at different azimuths and −30° elevation. The standard deviation of the individual HRTF spectra is plotted as a thin solid line at the bottom of each sub-plot.

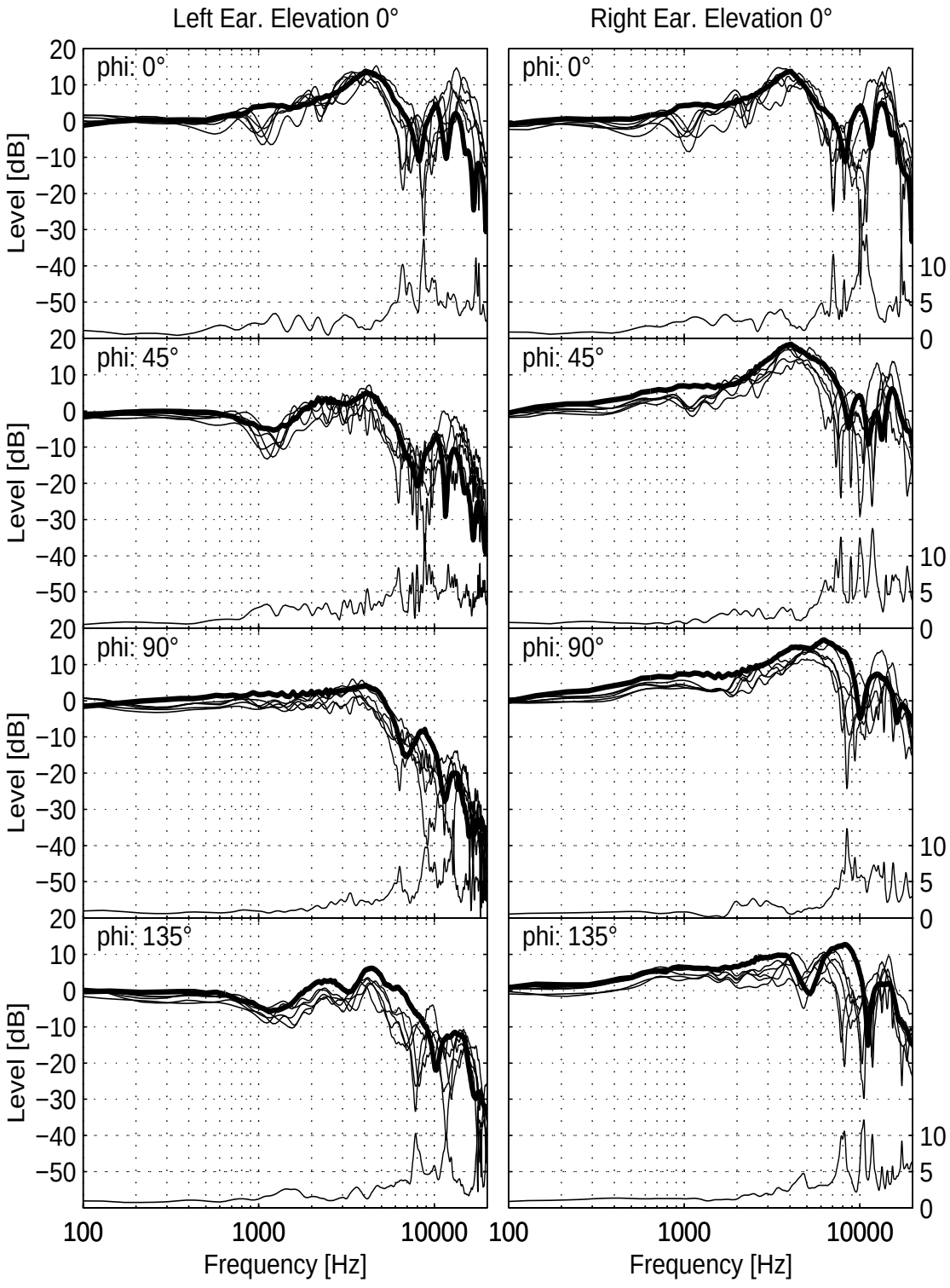


Figure 3.5: Same as Figure 3.4 at 0° elevation.

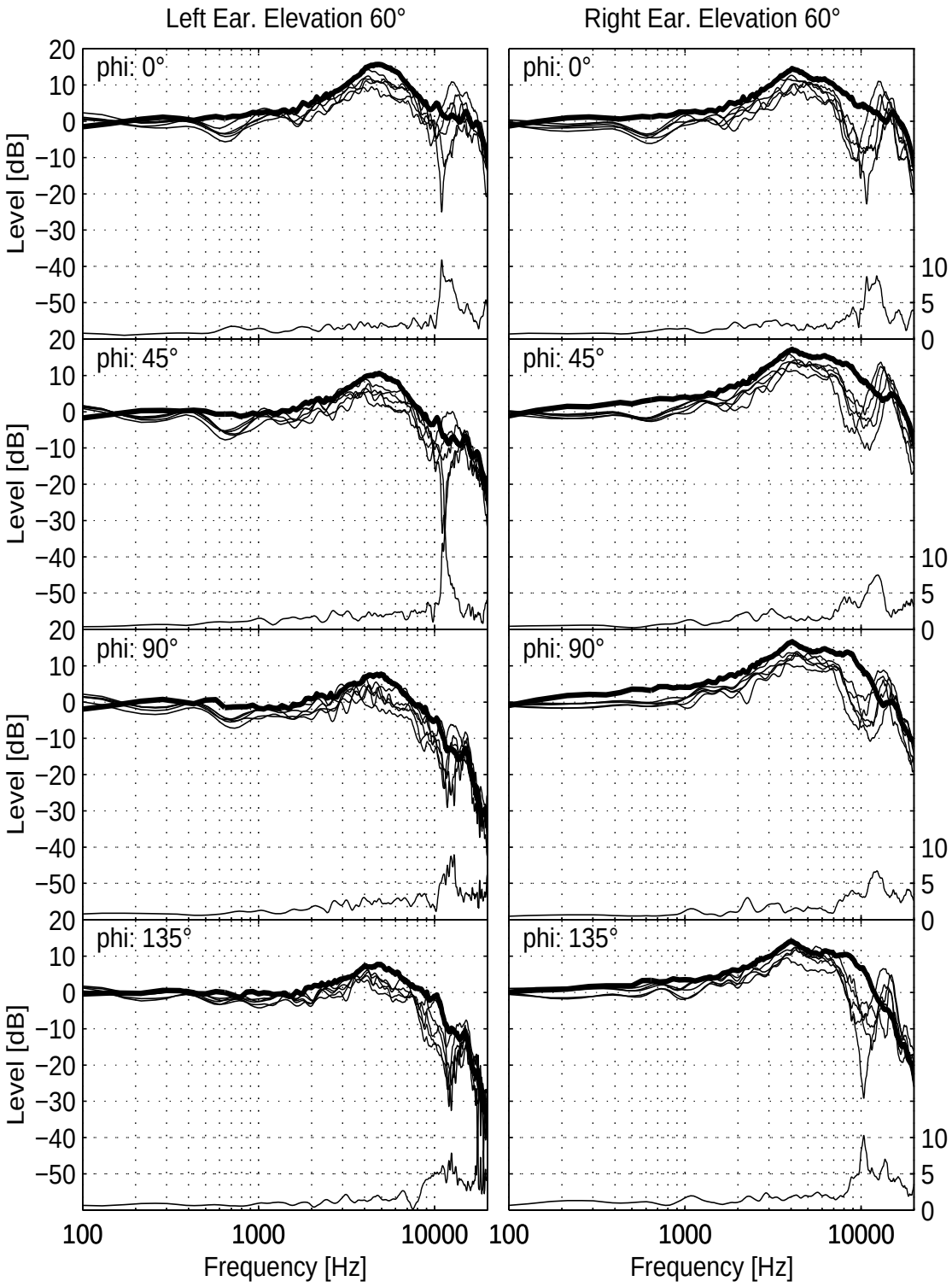


Figure 3.6: Same as Figure 3.4 at 60° elevation.

It is less than 3 dB for frequencies below 5 kHz. If the wavelength is within the dimension of the outer ear ($f > 6$ kHz), the standard deviation has peaky maxima due to spectral notches in this region that differ with respect to their center frequency. The inter-individual differences of the HRTF spectra are at maximum at approx. 10 kHz. The low standard deviation across all subjects at higher elevations is caused by the relatively smooth transfer functions (Figure 3.6). At lower source elevations, the peak around 10 kHz is broadened with a maximum standard deviation of about 10 dB.
The thick line in each panel of Figures 3.4 - 3.6 represents the dummy head HRTF spectra. The differences between dummy head spectra and individual spectra vary across source positions and frequencies. Although the dummy head was positioned on a stand and has no torso or shoulder, it follows the individual dependence for frequencies below 3 kHz within ±3dB but shows large deviations at higher frequencies.

3.2.3.3 Spectral cues: Elevation

In Figures 3.7 and 3.8 the spectral variation as a function of the source elevation (−30°, 0°, 30°, 60°) at constant azimuth (0° and 90°) is presented in the same way as given above.
The spectra of the left and right ear are symmetric up to about 5 kHz at low source elevations (Figure 3.7). The notch frequencies deviate between both ears at higher frequencies. The most prominent feature of the spectra are the concha resonance at 4 kHz and a notch in the area of 8-10 kHz. It can be seen, that the concha resonance is stable in its center frequency for all elevations. The notch shifts to higher frequencies as the source is elevated and is lowered in depth. At 90° azimuth and −30° elevation (Figure 3.8) the spectra of the contralateral HRTF are dominated by sharp notch resonances in a broad frequency range. With increasing source elevation the spectra are increasingly symmetrical across both ears.
The inter-individual differences between the HRTF spectra are represented by the standard deviation at the bottom of each panel of Figures 3.7 and 3.8. At low elevations, the standard deviation for frequencies up to 4 kHz increases for higher frequencies up to 10 dB. If the source is elevated the inter-individual differences are decreasing.
The dummy head spectra (thick line) show less variation as a function of frequency than the individual HRTFs. Especially in the frequency range above 8 kHz less interference effects can be observed. The concha resonance and the notch at 8 kHz are clearly identifiable and clarifies the behavior of the individual spectra. However, only the variation of the dummy head HRTF spectra below 4 kHz are comparable in level to the subjects HRTF. For higher frequencies, the dummy head spectra deviate strongly from the individual ones. Furthermore, the level of the high frequency range above 8 kHz is overestimated by the dummy head HRTFs.

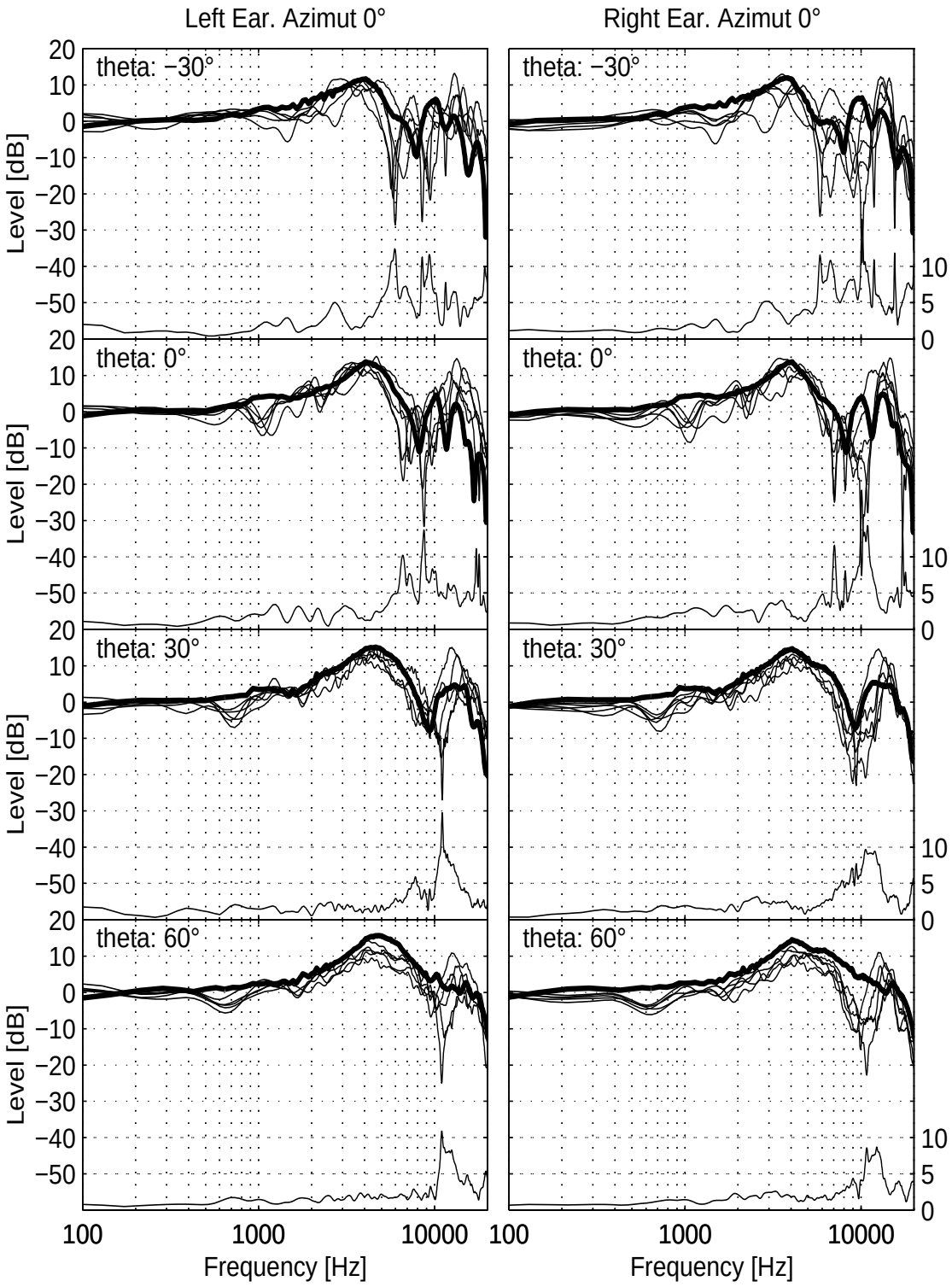


Figure 3.7: HRTF spectra of the left and right ear recorded from five subjects (thin lines) and one dummy head (thick lines) at 0° azimuth for different elevations. Additionally, the standard deviation of the individual HRTFs is plotted as a thin solid line at the bottom of each sub-plot.

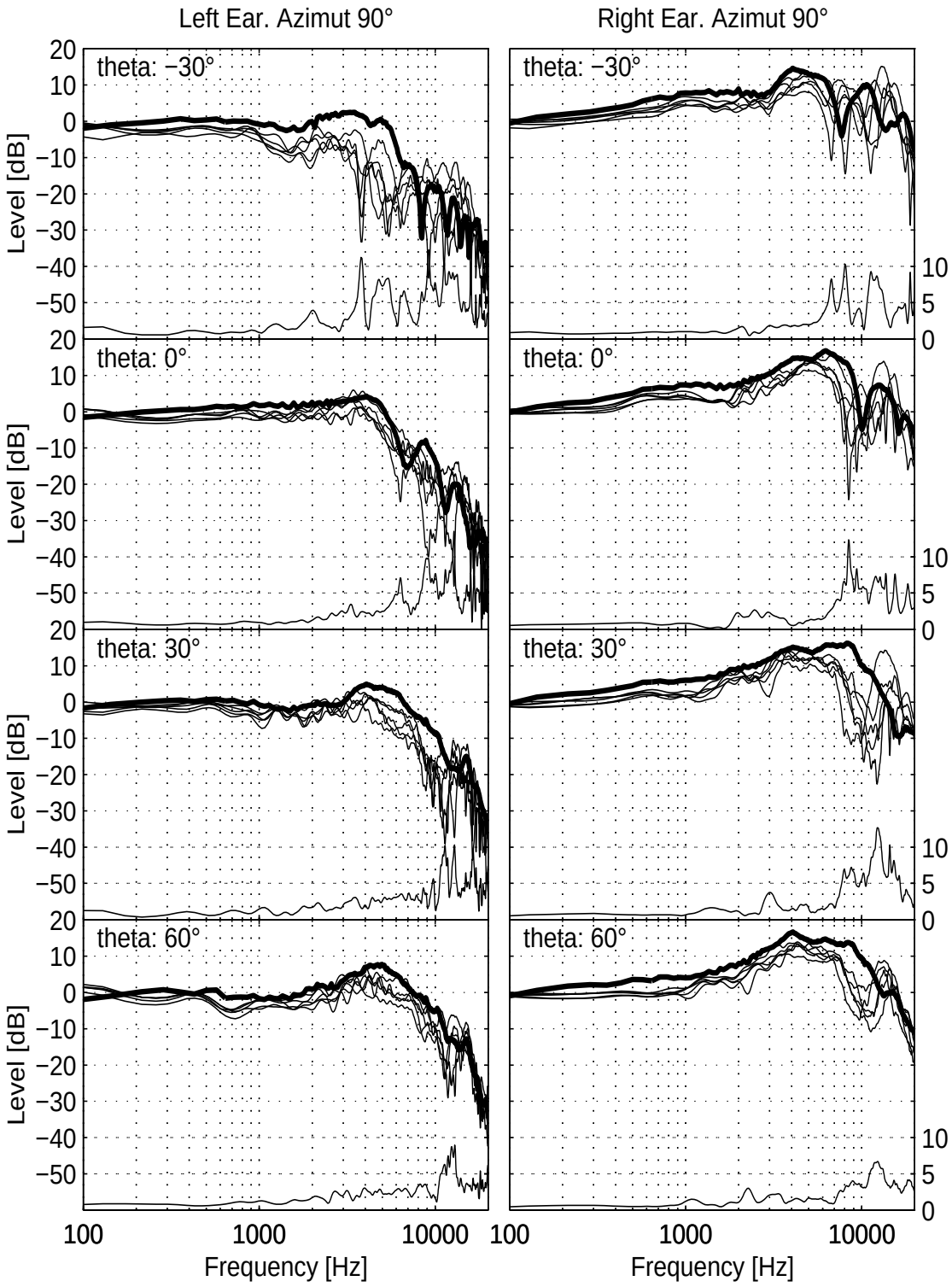
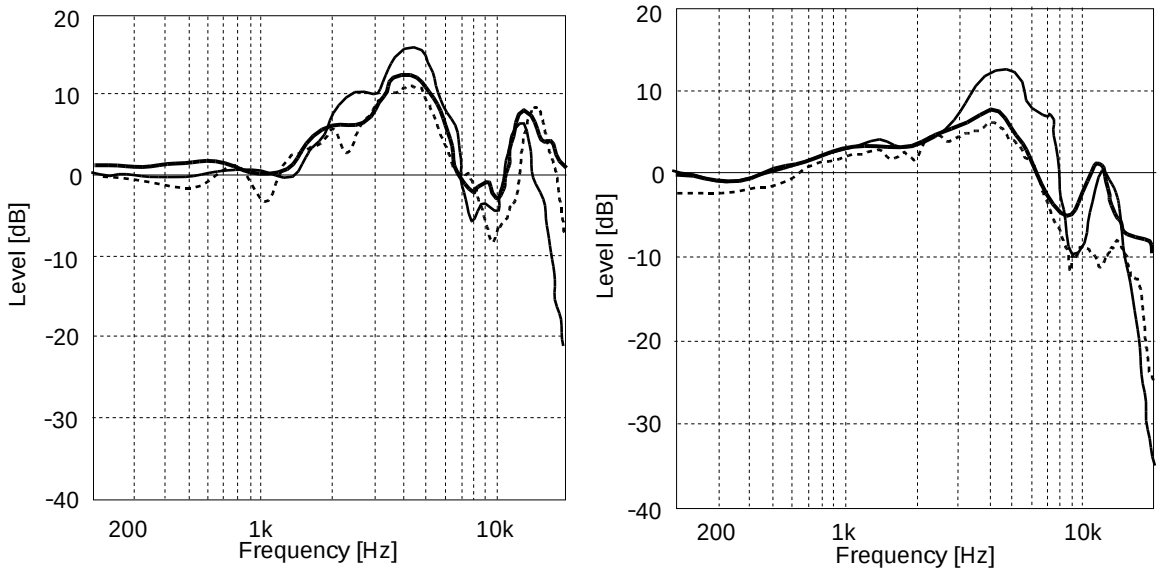


Figure 3.8: Same as Figure 3.4 at 90° azimuth.

3.2.4 Comparison of mean HRTFs

In the high frequency area the spectral variation as a function of frequency is highly variable across subjects, especially at low elevations. The spectral variation in the frequency range below 8 kHz provides less spatial information because the two resonances (i.e. the ear canal resonance and the resonance of the cavum conchae) are stimulated from nearly all spatial positions. Therefore, to compare HRTFs measured in different studies, mean HRTFs averaged across subjects are computed even though the suitability of averaged transfer functions for the presentation of spatial cues in the HRTFs spectra is doubtful: The inter-individual differences in the high frequency area are high and, therefore, by averaging HRTFs most of the spatially relevant details of the HRTFs are eliminated. However, comprehensive comparisons of mean HRTFs have been presented by Shaw (1974) and Møller et al. (1995). In the study of Møller et al. HRTFs were measured at the blocked ear canal and in the open ear canal. The results from both measurements techniques mainly differ in the frequency range below 10 kHz. HRTFs measured at the blocked meatus do not show the ear canal resonance, which is prominent in the open ear canal HRTFs. However, it was concluded by Møller et al. that the blocked meatus measurements still capture all spatially relevant information. Hence, only this kind of data are considered here.



(a) Mean HRTFs at 0° azimuth.

(b) Mean HRTFs at 180° azimuth.

Figure 3.9: Comparison of mean HRTFs across literature. The thick solid lines show mean HRTFs from Møller et al. (1995) averaged across 40 subjects and the thin solid lines represent data from Pösselt et al. (1986) averaged across 11 subjects. Mean HRTFs of the present study are given by thin dashed lines (10 subjects).

In Figure 3.9 mean HRTFs from the studies of Pösselt et al. (1986) (11 subject, thin solid lines) and Møller et al. (1995) (40 subjects, thick solid lines) are shown and compared to mean HRTFs computed from the data of the present study (10 subjects, thin dotted lines). In the left panel mean HRTFs for 0° azimuth are shown and in the right panel the source was positioned at 180° azimuth.
The general shape of the HRTF spectra is consistent across studies, although there are differences in the details. For frontal sound incidence there is a good agreement between the HRTFs of the present study and the HRTFs measured by Møller et al. The data from Pösselt et al. deviate in the amplitude of the peak in the frequency region at 2-5 kHz. Even higher deviations can be see for 180° azimuth in a slightly higher frequency area, whereas the data from Møller and the present study are very similar. However, the data obtained in the present study deviates for rear sound incidence from the cited data at frequencies above 10 kHz. A peak that is present in the data from Møller et al. and Pösselt et al. can not be seen in the mean HRTFs of the present study.
The differences of the mean HRTFs could be due to different groups of subjects and different positions of the microphones in the ear canal. In both, the study of Pösselt et al. and the present study subjects were sitting, whereas in the study of Møller et al. subjects were standing. Influences of the position of the torso on the HRTF spectra should only occur in the low frequency region. However, there is no better agreement between the studies where the subjects were sitting. Hence, the orientation of the torso does not consistently influence the shape of the low frequency HRTF spectra.

3.3 Influences of spectral smoothing on HRTFs

In order to assess the effect of reducing spectral information included in the HRTF by smoothing, the following have to be observed.

1. The standard deviation across spectra of individual HRTFs is a measure of the individual information contained in the spectra. Therefore, by investigating the standard deviation as a function of smoothing the amount of individual information in the HRTFs can be assessed. This is presented in Section 3.3.2.

2. Spectral smoothing is performed independently for the left and right ear HRTF. Therefore, also the ILD is affected by monaural smoothing. The relation between monaural smoothing and the variation of the ILD is considered in Section 3.3.3.

3. The effect of smoothing the HRTF spectra on the interaural time difference (ITD) is not easy to assess. It is a common practice to smooth the absolute spectrum and to model the HRTF phase as minimum phase plus a frequency independent group

delay τ_{Emp}, which has to be obtained from the empirical impulse responses. The minimum phase of the HRTF model is calculated from the logarithm of the absolute spectrum (s. (Oppenheim and Schafer, 1975)). Therefore, different degrees of smoothing also result in different phase spectra and different impulse responses. The ITD, however, is calculated from the impulse responses and can, therefore, vary as a function of smoothing. This is investigated in Section 3.3.4.

4. The last point concerns the length of the impulse responses as a function of smoothing which is analyzed in Section 3.3.5.

3.3.1 Smoothing methods

Two different types of smoothing are applied to the individual HRTF spectra: cepstral smoothing (see Section 4.4) and $1/N$ octave smoothing. The parameter M of the cepstral procedure describes the number of cosine terms used for the Fourier reconstruction of the spectra.
While cepstral smoothing is a linear approach with respect to the frequency axis, the humans ears' frequency resolution might be represented better by using a logarithmic approach with respect to the frequency axis. Hence, the amplitude spectra in $1/N$ octave bands are averaged by a moving average

$$H_M S(2\pi\nu) = \frac{1}{\nu_2 - \nu_1} \sum_{k=\nu_1}^{\nu_2} H(2\pi k\nu) \tag{3.5}$$

where $\nu_1 = \nu * 2^{-(\frac{1}{2N})}$ and $\nu_2 = \nu * 2^{+(\frac{1}{2N})}$ are the edges of the averaging band.
In all investigations presented below cepstral and $1/N$ octave smoothing were applied to the HRTF spectra. However, data for both procedures is only shown separately, if the results deviate between both procedures. Otherwise, results for cepstral smoothing are presented.

3.3.2 Smoothing and inter-individual differences

In Figure 3.10 the frequency dependent standard deviation of individual HRTF spectra across 10 subjects recorded at two positions in the horizontal plane (45° (top row) and 90° of azimuth (bottom row)) is plotted as a function of cepstral smoothing.
It can be seen that an increasing amount of smoothing flattens the standard deviation in the high frequency region. For eight cepstral coefficients the standard deviation is nearly constant across frequency. Hence, the inter-individual differences of the HRTF spectra are reduced to different amounts of energy in broad frequency bands.
In order to investigate the influence of the elevation on the inter-individual standard deviation of the left and right ear HRTF spectra, respectively, the mean standard deviation

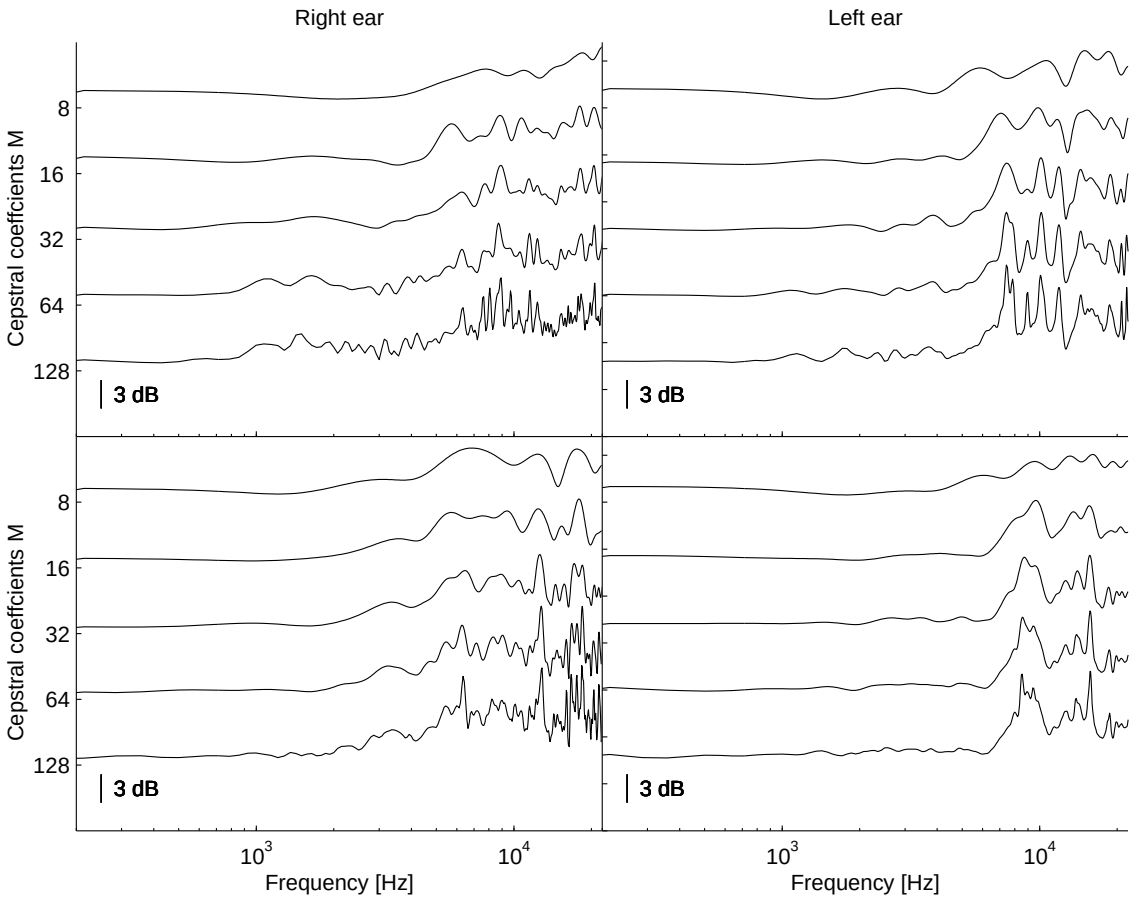
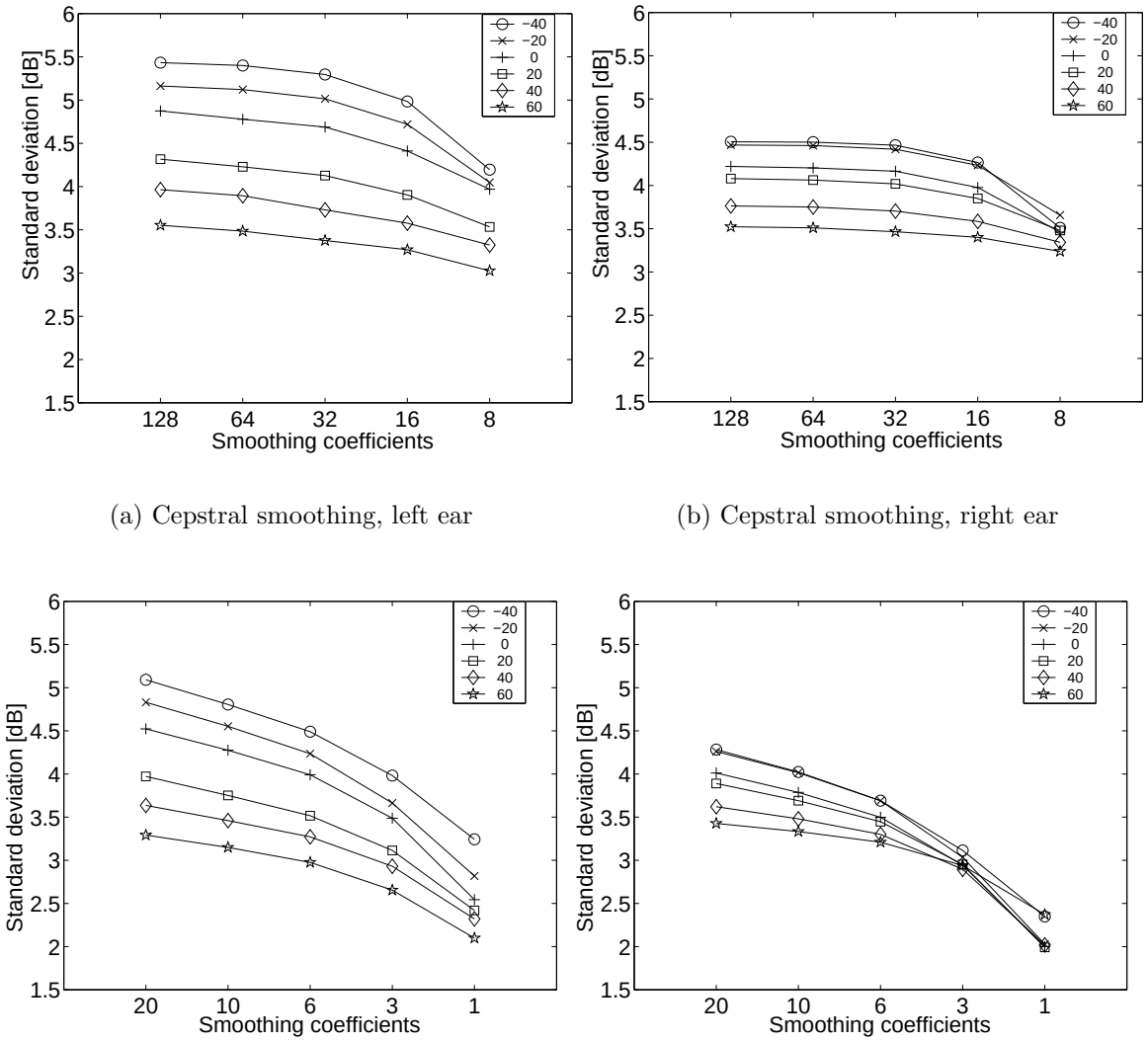


Figure 3.10: Standard deviation of the HRTFs spectra of 10 subjects for a sound source located in the horizontal plane at 45° azimuth (top row) and 90° (bottom row) azimuth for a variable number of cepstral smoothing coefficients (M) are shown for both ears separately.

averaged across frequency is plotted in Figure 3.11. Mean values for 10 subjects were computed across azimuth ($\phi = 0° - 180°, \Delta\phi = 15°$) for different elevations ($\Delta\theta = 20°$) and are given as a function of smoothing coefficients. In the top row cepstral smoothing was used, whereas data for $1/N$ octave smoothing is depicted in the bottom row.
The highest standard deviation can be observed for low elevations and the smallest standard deviation for high elevations. This tendency is more distinct at the contralateral than at the ipsilateral ear. Logarithmic smoothing, depicted in the bottom row of Figure 3.11 reduces the standard deviation more than the linear cepstral approach. This can be explained by the fact, that the linear smoothing conserves the more individual information in the high frequency region, whereas the logarithmic algorithm smoothes it out.



(a) Cepstral smoothing, left ear

(b) Cepstral smoothing, right ear

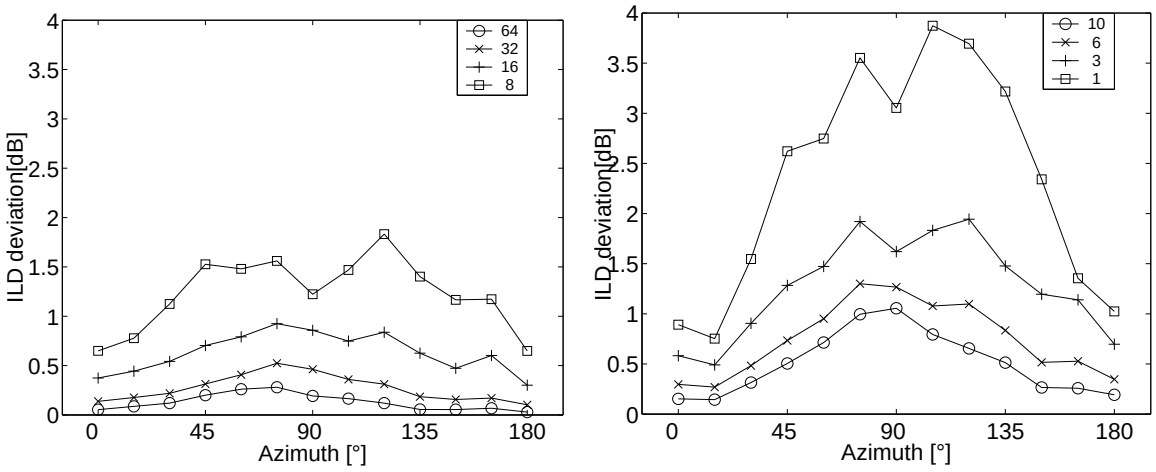
(c) Nth octave smoothing, left ear

(d) Nth octave smoothing, right ear

Figure 3.11: Standard deviation of HRTF spectra of 10 subjects averaged across azimuths plotted as a function of smoothing for different elevations. Cepstral smoothing was used in panels a) and b) and logarithmic smoothing in panels c) and d).

3.3.3 ILD deviations of smoothed transfer functions

In Figure 3.12 the ILD deviation that is introduced by smoothing the HRTF spectra is shown. The ILD deviation is calculated as the absolute level difference between the frequency dependent ILDs computed from the empirical HRTFs and the smoothed HRTFs, averaged across frequency and 10 subjects. In Figure 3.12 the ILD deviation is plotted as a function of azimuth at 0° elevation for different degrees of smoothing. In the left panel (Figure 3.12(a)) the HRTF spectra were smoothed by cepstral smoothing and in



(a) Cepstral smoothing

(b) $1/N$ octave smoothing

Figure 3.12: Deviations between original and smoothed ILDs are plotted as a function of azimuth and smoothing parameters.

the right panel (Figure 3.12(b)) by $1/N$ octave smoothing.
For 32 cepstral coefficients the influence of smoothing on the ILD is rather small (< 0.5 dB). If 16 coefficients are used the ILD deviations are up to 1 dB for sound incidence from the side. The ILD deviation increases to approx. 2 dB at lateral source positions if only eight cepstral coefficients are used.
Higher deviations can be observed in Figure 3.12(b) for $1/N$ octave smoothing. For 1/10 octave smoothing the ILD deviation is up to 1 dB at the side. If the averaging bandwidth is increased, the ILD deviation increases to up to 4 dB.
The head shadowing effect that causes the ILD, vanishes for frequencies for which the diameter of the head is small compared to the wavelength of the sound. Hence, only small ILDs can be observed in the low frequency range and higher ILDs occur mainly in the high frequencies. Furthermore, sharp notches can be observed in the HRTF spectra in the high frequencies that are introduced by interference effects and pinna filtering. Since the averaging bandwidth is increased for $1/N$ octave smoothing in the high frequencies the notches are smoothed out and even the macroscopic spectral shape is affected. The smaller averaging bandwidth in the low frequencies provides no advantage because the ILD is nearly zero in this region.
In contrast, cepstral smoothing reduces the same amount of spectral detail in each frequency band and, hence, the spectral detail in the high frequencies is better conserved by cepstral smoothing. Therefore, cepstral smoothing produces a smaller ILD variation than $1/N$ octave smoothing does.
In auditory models the frequency channels are approximatively separated by 1/3 octaves. The results of this investigation show, that this kind of auditory processing is not ap-

propriate for compressing the information available in HRTFs, since it can be assumed that the ILD deviation of approx. 2 dB is detectable for subjects.

3.3.4 ITD deviations of smoothed transfer functions

If the transfer function $E(\omega, \phi, \theta)$ in Equation 3.4 is not minimum phase the inverse transfer function $E^{-1}(\omega, \phi, \theta)$ can be unstable. In this case the calculation of $A(\omega, \phi, \theta)$ should be restricted to the absolute spectrum of the HRTF and an appropriated phase can be applied to the spectrum.
It has been shown by Mehrgardt and Mellert (1977) and Kulkarni et al. (1999) that the empirical HRTF phase is almost minimum phase plus an frequency independent group delay.
A minimum phase can be obtained from the absolute HRTF spectrum by

$$P_{min}(\omega, \phi, \theta) = \Xi(-ln(|A(\omega, \phi, \theta)|)) \tag{3.6}$$

where Ξ is the Hilbert transform. The complex HRTF is then given by

$$A(\omega, \phi, \theta) = |A(\omega, \phi, \theta)| \times e^{-iP_{min}(\omega,\phi,\theta)}. \tag{3.7}$$

An important property of minimum phase transfer functions is that they have a minimal energy delay ((Oppenheim and Schafer, 1975)). As a consequence, the group delay of the minimum phase impulse response is always nearly zero. Therefore, if both the left and the right ear HRTFs are minimum phase, the ITD is nearly zero independent on source location. To apply an appropriate ITD an frequency independent group delay is introduced to one ear that matches the ITD obtained from the empirical impulse responses. However, the ITD of the pure minimum phase HRTFs is not equal to zero for all source positions. Therefore, the frequency independent group delay that is applied to the minimum phase HRTFs has to be corrected for the inherent time delay of the minimum phase HRTFs. This correction term has to be subtracted from the ITD that is introduced to the minimum phase impulse responses.
Only the low frequency range of the ITD is perceptually relevant, because for high frequencies the phase differences at the two ears are ambiguous. Thus, it is important that the low frequency ITD of the minimum phase plus frequency independent delay HRTFs is consistent with the empirical ITD. Hence, in this study the group delay is calculated in a way that the low frequency ITDs of the minimum phase plus delay HRTFs and the empirical HRTFs are equal.
As pointed out in Section 3.3 the ITD of minimum phase plus delay HRIRs is directly related to the spectrum. Therefore, the smoothed HRTF spectra have to be taken into account when the group delay, that is introduced to the minimum phase HRTFs, is calculated.

Taken together, three different ITDs have to be calculated for introducing an low frequency ITD to minimum phase impulse responses that matches the low frequency ITD of the empirical HRTFs. They are obtained as follows:

$$\begin{aligned} \tau_{Emp} &= argmax(\Gamma(h_l, f) \otimes \Gamma(h_r, f) \\ \tau_{Corr} &= argmax(\Gamma(h_{l,min,S}, f) \otimes \Gamma(h_{r,min,S}, f) \\ \tau_{Min} &= \tau_{Emp} - \tau_{Corr} \end{aligned} \tag{3.8}$$

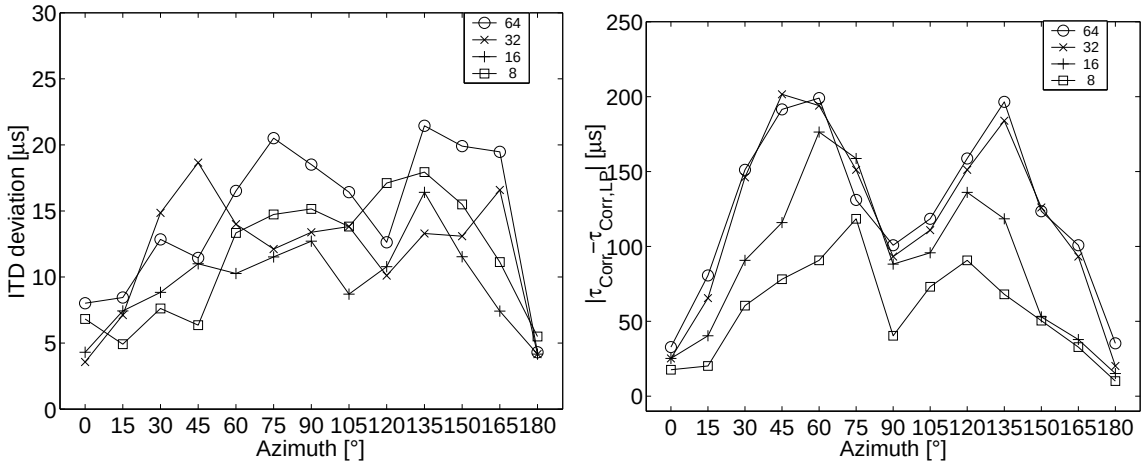
τ_{Emp} is the ITD of the empirical HRTFs calculated as the time shift of the maximum of the cross-correlation function (marked by the symbol $\otimes$) of the left (h_l) and right (h_r) ear HRIRs. '$argmax$' denotes the time shift of the maximum of the cross correlation function. The function $\Gamma(h, f)$ is the low-pass filtered impulse response $h(t)$ with edge frequency f. It is applied to extract the low frequency ITD ($f = 500\ Hz$). The correction term τ_{Corr} is calculated from the minimum phase HRIRs with smoothed spectra ($h_{r/l,min,S}$). The index S denotes the degree of smoothing either for $1/N$ octave or cepstral smoothing. Then the frequency-independent group delay introduced to the minimum phase impulse responses is given by τ_{Min}.
Based on this calculation, the low frequency ITD of the minimum phase plus frequency independent group delay HRIRs should match the empirical low frequency ITD τ_{Emp}, independent of spectral smoothing. To verify this, the ITD τ_{ReCalc} is re-calculated from the minimum phase plus frequency independent group delay HRIRs by

$$\tau_{ReCalc} = max(\Gamma(h_{l,min,S}(t + \tau_{Min}), f) \otimes \Gamma(h_{r,min,S}(t), f)) \tag{3.9}$$

The ITD error between the minimum phase plus frequency independent group delay HRIRs and the empirical HRIRs is then given by

$$\tau_{Err} = \tau_{Emp} - \tau_{ReCalc} \tag{3.10}$$

In Figure 3.13(a) the ITD error τ_{Err} (averaged across 10 subjects) is plotted as a function of azimuth for four different degrees of smoothing. The error is small for sound incidence out of the median plane ($\simeq 5 - 8\mu s$) and increases at lateral angles. Furthermore, the ITD error is not independent from smoothing and is varying in a range of approx. $10\mu s$. The results for $1/N$ octave smoothing are comparable and not shown here. It is important to note, that the perceptually relevant low frequency ITD τ_{ReCalc} only matches the empirical low frequency ITD τ_{Emp} if the minimum phase correction term τ_{Min} is computed from low pass filtered HRIRs. In a study of Kulkarni et al. (where the correction term τ_{Corr} was introduced) the sensitivity of subjects to HRTFs phases was investigated by discrimination experiments. It was shown, that minimum phase plus frequency independent group delay HRTFs were distinguishable from empirical HRTFs at lateral source positions. At these positions the low frequency ITD of the minimum phase plus frequency independent group delay HRTFs deviated from the empirical ITD. It was concluded that these deviations served as a cues for the subjects.



(a) $|\tau_{emp} - \tau_{ReCalc}|$ (b) $|\tau_{Corr} - \tau_{Corr,LP}|$

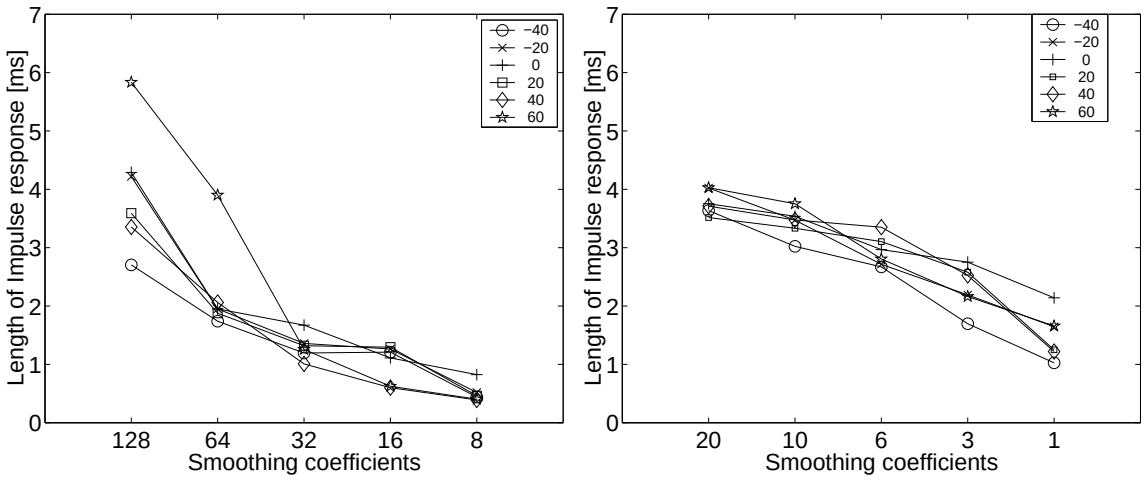
Figure 3.13: Left Panel: The differences between the low frequency ITDs of the empirical HRTFs (τ_{Emp}) and the minimum phase plus frequency independent group delay models (τ_{ReCalc}) are shown for different degrees of cepstral smoothing as a function of azimuth. Right panel: The difference of the correction term τ_{min} calculated from unfiltered and low-pass filtered minimum phase impulse responses for different degrees of smoothing are shown for source positions in azimuth. The data is averaged across 10 subjects for both plots.

However, in their study the ITDs were computed from unfiltered impulses responses (i.e. by removing Γ in equation 3.8). ITDs calculated in this way represent the group delay of the broadband signal. If the correction term τ_{Corr} is also computed from broadband impulses, the low frequency ITD deviates from the empirical low frequency ITD. This is illustrated in Figure 3.13(b). The absolute differences between the correction term τ_{Corr} computed from unfiltered and low-pass filtered minimum phase HRIRs are plotted as a function of azimuth for different degrees of cepstral smoothing. It can be seen from this figure that the low frequency group delay of the minimum phase HRIRs is clearly different from the overall group delay of the unfiltered minimum phase impulse responses. The range of the differences strongly depend on spectral smoothing, whereas the general shape of the curve is conserved.

The differences between the low frequency ITD of the empirical HRTFs and the ITD obtained from minimum phase plus frequency independent group delay HRIRs shown in Figure 3.13(a) are below the detection threshold (Durlach and Colburn, 1979; Kinkel, 1990) and are, therefore, perceptually irrelevant. However, the differences of the minimum phase correction term shown in Figure 3.13(b) are above the detection threshold for lateral sound incidence, both the absolute differences between the empirical and the minimum phase plus delay ITDs and the differences between different degrees of

smoothing.
It can be concluded from this investigation, that only if τ_{Corr} is calculated from low pass filtered minimum phase impulse responses the smoothed minimum phase plus frequency independent group delay HRTFs are perceptually indistinguishable from empirical HRTFs.

3.3.5 Impulse response shortening by spectral smoothing



(a) Cepstral smoothing.

(b) $1/N$ octave smoothing.

Figure 3.14: The length of the HRIRs are plotted as a function of cepstral smoothing coefficients (left panel) and $1/N$ octave smoothing for different elevations averaged across subjects and azimuth.

Smoothing effectively reduces the frequency resolution of the HRTF spectra. The frequency resolution is directly related to the length of the impulse response: An impulse response of length τ can be considered as an impulse response of infinite length that is multiplied with a rectangular function of length τ that is one for $0 < t < \tau$ and zero outside this interval. Therefore, if the impulse response with length τ is Fourier transformed, the spectrum of the finite impulse response can be interpreted as the Fourier transform of the spectrum of the impulse response convolved with the Fourier transform of the rectangular function, which is a 'sinc' function $(sin(x)/x)$. The convolution of the spectrum of the impulse response with the 'sinc' function can be regarded as a weighted moving average of the frequency spectrum. The width of the first maximum of the 'sinc' function is proportional to $1/\tau$ and, therefore, shorter impulse responses have lower frequency resolution. Hence, the frequency resolution is directly related to the length of the impulse response. An analytical derivation of the length of the HRIRs as a function of cepstral and $1/N$ octave smoothing might be very complicated and is beyond the

scope of the study. Therefore, the length of the impulse response is directly measured for different degrees of smoothing.
In Figure 3.14 the effect of spectral smoothing on the length of the impulse responses is investigated. The length of the impulse responses was calculated by considering the squared impulse responses of the right ear HRTFs within time frames of 110 ms. The end of the HRIR was then defined as the point where the energy has decreased to 1.5 times the energy estimated from the noise floor.
The HRIR length of the right ear was averaged across azimuth ($\phi = 0° - 180°$) and 10 subjects and plotted as a function of elevation for different degrees of smoothing. In Figure 3.14(a) cepstral smoothing and in Figure 3.14(b) $1/N$ octave smoothing was applied . It can be seen that for cepstral smoothing the length of the impulse response is reduced by a factor of up to 6 at high elevations, if M is reduced from 128 to 8 cepstral coefficients. On the average, the length is reduced by a factor of approx. 3. Similarly, for $1/N$ octave smoothing a reduction of the impulse response length by a factor of approx. 3 can be observed, for an increase of the averaging bandwidth from 1/20 to 1/1 octave. If the HRTF is realized as a FIR filter, the filter coefficients are identical to the HRIR. The computational effort to process the filter depends linearly on the number of filter coefficients. Hence, by smoothing the HRTFs the computational effort is reduced by a factor of approx. 3.

3.4 Summary and general discussion

HRTF measurements

In the first section of this chapter, HRTF measurements from 11 subjects and one dummy head were presented. The HRTFs were sampled from a high number of source positions (5° resolution for individual and 1° for the dummy head) using the TASP system (see Chapter 2) . The monaural and binaural localization cues of the HRTFs show the typical spatial dependencies that were found in the literature (e.g. (Mehrgardt and Mellert, 1977; Shaw, 1974; Wightman and Kistler, 1989a; Møller *et al.*, 1995)) and, hence, the mean HRTFs obtained here are in good agreement with results from Møller et al. (1995) and PÖsselt et al. (1986).
The standard deviations of the monaural and binaural cues across subjects are highest for low elevations and decrease as the source elevation is raised. The standard deviation of the monaural spectral cues can be separated into two frequency regions. In the low frequency region, the standard deviation across listeners is small (typically below 2 dB). In the high frequency region the standard deviation increases to up to 10 dB. The cross over frequency between both frequency bands is approx. 4 kHz for low elevations and increases to approx. 8 kHz for high elevations. Thus, HRTFs from low elevations

show more individual properties than HRTFs from high elevations. In the literature standard deviations across subjects are only shown for selected source positions mostly in the horizontal plane (Wightman and Kistler, 1989a) or for HRTFs collapsed over a wide range of spatial positions (Møller *et al.*, 1995). Although the general behavior of the standard deviation shown in the literature is consistent with the results of this study, the reduction of the standard deviation for elevated source positions has not been explicitly shown so far.

Dummy heads are intended to represent an average head of an individual subject. The HRTFs of the dummy head employed here ('Oldenburg dummy head') show that the binaural cues are fairly within the range of individual cues for higher elevations. However, due to the lack of a torso and shoulder the binaural cues at low elevations strongly deviate from individual cues. The spread of the binaural cues across individuals clearly limit the use of dummy head HRTFs as a replacement for individual recordings. Since the deviations of the dummy head cues from the individual binaural cues vary considerable across source locations, the dummy head cues would only be suitable for some source positions. Furthermore, the deviations of the monaural cues provided by the dummy head from the ones originating from the subjects' own ears are even larger than in the binaural case. It is known from the literature, that the spectral filtering of the pinna in the high frequencies is responsible for resolving front-back confusions and to estimate the source elevation (e.g. (Oldfield and Parker, 1984a; Oldfield and Parker, 1984b)). However, the monaural dummy head cues strongly deviate from the individual cues especially at high frequencies. These differences depend on frequency and source position, and an extraction of a systematic pattern of these differences is difficult. Therefore, the results of this study suggest that the 'Oldenburg dummy head' can not be used as an average head of a listener if spatially correct perceptual representations of virtual stimuli are required. If no possibility is given to measure HRTFs of an individual listener at least the HRTFs from a different listener should be employed because in this case at least the low frequency spectra are well matched. However, Wenzel et al. (1993) showed that localization performance is reduced by using non-individualized HRTFs. On possibility to overcome this problem is to scale the non-individual HRTF spectra in frequency to match the individual center frequencies of the peaks and notches (Middlebrooks, 1999a; Middlebrooks, 1999b). The scale factor can be obtained by performing a psychoacoustic task that lasts about one hour (Middlebrooks *et al.*, 2000)).

Spectral smoothing

In the second part of this investigation the effect of spectral smoothing on HRTFs was investigated. Spectral smoothing obviously reduces individual information in the high frequencies. If less than 16 cepstral smoothing coefficients are used, individual information in the high frequency region is reduced to level differences in relatively broad

frequency bands. However, the small peaks and notches code individual spatial information and should, therefore, possibly left unchanged in the smoothing process. This consideration is supported by an investigation of Kulkarni et al. and the results of Section 4. Both studies show that differences between original and smoothed HRTFs with regard to localization can be detected, if less than 16 cepstral smoothing coefficients are used.

This smoothing limit is also supported by the analysis of the ILD deviations as a function of smoothing. For 8 cepstral coefficients the broadband ILD deviation exceeds 1 dB. This suggest that the ILD deviation is detectable by subjects (Durlach and Colburn, 1979). Furthermore, the results of this investigation show, that $1/N$ octave smoothing is not appropriate for smoothing HRTF spectra. The increasing amount of smoothing for high frequencies result in ILD deviations that are above the detection threshold even for 1/3 octave smoothing.

Smoothing does not only affect the ILD but also the ITD. However, it is shown in the present study that the ITD deviation is assumed to be perceptually irrelevant if the ITD that is incorporated to the minimum phase impulse responses is calculated from low pass filtered versions of the empirical HRIRs. Furthermore, the correction term that eliminates inherent ITDs of the minimum phase HRIRs has also to be calculated from low pass filter impulse responses. If, however, this correction term is computed from unfiltered impulse responses it can be assumed that the low frequency ITD of the minimum phase plus frequency independent group delay HRTFs deviates from the ITD of the empirical HRTFs in a perceptually relevant range.

This result is consistent with findings of Kulkarni et al. (1999). In their study, the group delay of the minimum phase HRTFs and the minimum phase correction term were computed from unfiltered HRTFs. The results of a discrimination experiment showed, that subjects were able to distinguish minimum phase plus frequency independent group delay HRTFs from empirical HRTFs for sound incidence from the sides. On the basis of the considerations presented in the current study it can be supposed that the subjects would not have been able to detect the minimum phase plus frequency independent group delay stimuli if the incorporated ITD would have been correctly matched in the low frequency range.

In the last investigation presented in this study the length of the HRIRs as a function of smoothing was calculated. For both cepstral and $1/N$ octave smoothing the length of the impulse responses is substantially reduced. The length of the original impulse responses averaged across azimuth positions ranged from 4.5 ms to 6 ms depending on elevation. In a study of Kulkarni and Colburn (1998) it was shown that 16 cepstral coefficients were sufficient for providing all spatially relevant information of the HRTF spectra. The results of this study show that for this amount of smoothing the length of the impulse responses is below 1.5 ms. From the investigation of the effect of smoothing on the ILD it can be concluded that 1/6 octave averaging produces perceptually irrelevant deviations

for logarithmic smoothing. The length of the corresponding impulse responses for this amount of smoothing ranges from 2-3 ms. Hence, by applying cepstral smoothing to the HRTF spectra the resulting impulse responses are shorter in comparison to $1/N$ octave smoothing. Therefore, both the effect of smoothing on the ILD (see above) and on the HRIR length lead to the same conclusion that linear smoothing is more appropriate for smoothing HRTFs than $1/N$ octave smoothing.

3.5 Conclusions

The investigations of this study show that

- localization cues show high inter-individual differences in ITD, ILD and monaural spectral cues. The cues are highly individual at low source elevations and less individual at higher elevations.
- dummy head HRTFs and also non-individualized HRTFs can not replace individual HRTFs, if perceptually correct virtual stimuli are needed.
- smoothing the HRTF spectra by cepstral smoothing (16 coefficients) reduces the inter-individual standard deviation of HRTF spectra by approx. 0.5 dB and approx. 1 dB for logarithmic 1/3 octave smoothing.
- the ITD of minimum phase plus frequency independent group delay HRIRs has to be calculated from low pass filtered empirical HRTFs.
- the length of minimum phase HRIRs can be reduced by a factor of 3 by smoothing the corresponding HRTF spectra.
- linear cepstral smoothing is more appropriate for HRTF spectra than $1/N$ octave smoothing because the ILD is less affected.
- linear cepstral smoothing with 16 coefficients seems to be the best compromise between preservation of localization cues and minimization of computational effort.

Chapter 4

Sensitivity of Human Listeners to Manipulations of the Head Related Transfer Functions.

Abstract

The sensitivity to changes of the individual localization cues, described by head related transfer functions, is investigated in three discrimination experiments by a two interval, two alternative forced choice (2I-2AFC) measurement paradigm. In the first two experiments, the sensitivity to manipulations of the HRTF spectra was investigated. In experiment I the spectral detail of the HRTF spectra was reduced by cepstral smoothing. The smoothed HRTFs were applied to white noise, click train and scrambled white noise stimuli (500 ms length). Hence, in two conditions of experiment I subjects could use timbre cues whereas in one condition only spatial cues were provided. Stimuli were presented from five different directions in azimuth ($0° - 180°$*, $45°$ *resolution). The results of experiment I show that the detection of the smoothed HRTFs strongly depends on the source stimulus. 16-32 cepstral coefficients (depending on source position) are sufficient for providing all spatially relevant information to the subjects but more than 64 cepstral coefficients have to be used to leave the stimulus timbre unaffected. In experiment II the individual HRTF spectra were stepwise transformed to spectra of a dummy head ('spectral morphing') and applied to a scrambled white noise stimulus (500 ms) to affect also the center frequencies of the peaks and notches of the HRTF spectra. The results show that subjects were very sensitive to the 'morphing' procedure for frontal sound incidence and less sensitive for source positions at the side. Both for spectral detail reduction (scrambled whited noise) and for 'spectral morphing' the ILD deviation (averaged across frequency bands) introduced by the spectral manipulations is an appropriate measure for the cues that subjects could have used. ILD deviations of approx. 0.5-0.8 dB for frontal sound*

incidence and of approx. 1.2 dB for lateral sound incidence can be detected by subjects. In experiment III it is shown that subjects are less sensitive to changes of the interaural time difference (ITD), if a plausible frequency distribution of the ILD was applied to the stimuli as opposed to literature experiments with a fixed ILD across frequency. It is assumed that the lower sensitivity is caused by additional spatial information in the natural ILD that stabilizes the perception of the virtual objects and makes it less sensitive to distortions of the ITD.

4.1 Introduction

The physical entities that are used by the auditory system to localize a sound source in a non-reverberant environment are captured by *head related transfer functions* (HRTFs) (Mehrgardt and Mellert, 1977; Wightman and Kistler, 1989a; Møller *et al.*, 1995; Hammershoi and Møller, 1996)). They describe the directional dependent transfer functions of an acoustical object from its source location to a point within the ear canal of the left resp. the right ear. The auditory system analyzes the directional transformation by a comparison of the sound pressures at the two ears (binaural cues) and the spectral filtering of the head and pinna at each ear (monaural cues) to estimate the location of the sound source. The general contribution of binaural and monaural cues to the localization process are well known (see (Blauert, 1974; Middlebrooks and Green, 1991) for reviews). Both, the monaural and the binaural localization cues are highly dependent on the individual subject because the shape of the head and especially the complex structure of the pinna differ from subject to subject (cp. Chapter 3, (Mehrgardt and Mellert, 1977; Møller *et al.*, 1995; Middlebrooks, 1999a)).

A natural and accurate perception of an externalized sound can be achieved by headphone listening, if the time domain representation of the individual HRTFs (the head related impulse responses, HRIRs) are convolved with a monophone sound stream (Wightman and Kistler, 1989a; Wightman and Kistler, 1989b; Bronkhorst, 1995; Langendijk and Bronkhorst, 2000). For an accurate perception of virtual stimuli individual HRTFs are needed because non-individualized HRTFs reduce the localization performance (Wenzel *et al.*, 1993).

Hence, the physical differences between individual HRTFs of different subjects are normally above the detection threshold of deviations from the individual localization cues. The primary goal of this study is to provide thresholds for perceptually irrelevant deviations of the individual localization cues described by HRTFs. Therefore, the sensitivity of subjects to manipulations of the individual HRTFs is investigated.

The measurement paradigm needed to analyze the perceptual effect of HRTF manipulations has to capture all possible perceptual changes of the virtual stimuli. An absolute

localization measurement paradigm is only capable of capturing a change of the spatial stimulus position. However, a manipulation of HRTFs may result in a different stimulus timbre or spaciousness while preserving the spatial centroid. Therefore, all perceptual differences introduced by the HRTF manipulations should be taken into account. Hence, a discrimination task was chosen for the psychoacoustical experiments.

Spectral HRTF manipulations

Only few studies in the literature conducted discrimination experiments of HRTF manipulations. In a study of Langendijk and Bronkhorst (2000) the detection performance of stimuli created from interpolated HRTFs was investigated. The authors concluded that subjects were able to detect a change in stimulus timbre if spectral differences of 1.5 dB to 2.5 dB in one 1/3 octave band of the right ear HRTF spectrum occur. Spatial displacement was detected by differences greater than 2.5 dB. However, data for different source positions is only qualitatively described. Hence, the sensitivity of the auditory system to changes in the HRTFs as a function of source position has not been investigated. Presumably, the thresholds for frontal source positions will be lower and those for more lateral angles will be higher than the mean threshold across source positions.
Kulkarni and Colburn (1998) reduced the spectral details of the HRTF spectra by cepstral smoothing. They showed that the subjects were not able to discriminate a virtual from a real sound source, as long as more than sixteen terms of a fourier series expansion were used for a reconstruction of the HRTF spectra. Although psychometric functions were shown for different angles of azimuth, the stimulus spectrum was scrambled and hence, only spatial displacements of the manipulated stimulus could be detected by the subjects. However, for virtual acoustic displays, for instance in a virtual recording studio, also non-spatial cues like timbre should be unaffected by spectral smoothing.
Hence, in experiment I of the present study the results from Kulkarni and Colburn are extended to non-spatial cues (i.e., detection of spectral cues introduced by smoothing).
A different approach with respect to the manipulation of the HRTF spectra was used in experiment II. The peaks and notches of the HRTF spectra are the primary cues for elevation perception and aid to resolve front-back confusion (e.g. (Roffler and Butler, 1968; Oldfield and Parker, 1984a; Oldfield and Parker, 1984b; Middlebrooks, 1992)). Although the general shape of the HRTF spectra is similar across subjects, the center frequencies of the peaks and notches vary between subjects, representing individual information (see Chapter 3). Smoothing changes the amplitude of the notches and peaks but does not change their center frequencies. Therefore, it is possible that subjects are more sensitive to spectral manipulations that also shift the center frequencies of the peaks and notches. In this case the detection thresholds obtained from the spectral detail reduction experiment do not describe the general sensitivity to spectral variations.
Therefore, in the second experiment of this study a transformation was applied to the

HRTF spectra that also varies the center frequencies of the peaks and notches. This was done by transforming the shape of individual HRTF spectra to the shape of dummy head HRTF spectra. The results of the investigation on HRTF spectra in Chapter 3 show that dummy head HRTFs differ from individual HRTFs in amplitude as well as in the center frequencies of peaks and notches in the spectra. Thus, by a stepwise incorporation of the macroscopic dummy head HRTF spectra the frequency distribution of the individual HRTFs is changed. This process is called 'spectral morphing' further on. A comparison of the results obtained from experiment I (spectral detail reduction) and experiment II ('spectral morphing') can reveal if subjects are more sensitive to a transformation that primarily affects the individual information coded in the HRTF spectra.

ITD variations

Both experiments, reduction of spectral detail and 'spectral morphing', focus on the HRTF spectra and the sensitivity of human listeners to its spectral structure. However, it is known from the literature that the ITD plays an important role in sound localization and even dominates the cues based on HRTF spectra (e.g. (Wightman and Kistler, 1992)). The sensitivity of subjects to changes of the ITD (just noticeable differences, JNDs) have been investigated intensively by headphone experiments (review (Durlach and Colburn, 1979; Kinkel, 1990)). Normally, JNDs are obtained for tone stimuli (e.g. (Hershkowitz and Durlach, 1969; Domnitz, 1968)), broadband noise (e.g. (Tobias and Zerlin, 1959; Mossop and Culling, 1995)) or narrowband noise ((Kinkel, 1990)). The ITD JNDs vary considerably across studies. Depending on method and subjects JNDs are in the range of 6 μs to 60μs for a reference with zero ITD. However, it is a common finding that ITD JNDs are increasing by a factor of approx. 2-3 for higher ITD references.
The stimulus ILD also affects the ITD JND, tending to cause higher values if the ILD is increased (e.g. (Koehnke *et al.*, 1995)).
The ILD-ITD combinations presented to subjects in studies from the literature, are often implausible for the auditory system because they do not occur in the daily listening scenario (except for conditions were ILD and ITD are zero). Furthermore, the ILD of broadband stimuli is normally constant across frequency. This is very implausible for the auditory system because the interaural level differences in the low frequencies are close to 0 dB due to the vanishing head shadow effect. The most plausible ITD-ILD combinations are given by stimuli convolved with individual HRIRs. Two hypotheses can be stated that predict the results of measuring the ITD JND with empirical stimuli in opposed ways: The spaciousness of a virtual stimulus depends on the consistency of the localization cues. A smaller spatial extend of the object is expected if all localization cues point to the same direction. If, however, the spatial information deviates across cues (for instance, the ITD is pointing to the left hemisphere and the ILD is pointing to

the right hemisphere), an increased blur of the virtual object can be observed. Hence, it can be assumed that changes of the localization cues of a focused stimulus are easier to detect than changes of the localization cues of an object with increased spaciousness. Based on this hypothesis the ITD JND would be *smaller* for stimuli with natural ITD-ILD combinations than for unnatural combinations.
On the other hand, it can be assumed that plausible combinations of the localization cues provide a more robust perception of the virtual object because all localization cues point into the same direction. Hence, a change of one localization cue is less important for correct spatial perception if consistent redundant information is given by the remaining cues. Based on this hypothesis the ITD JND would be *larger* for stimuli with natural ITD-ILD combinations than for unnatural ones.
To test both alternative hypotheses two experimental conditions were investigated. In the first condition, detection rates of ITD variations for stimuli convolved with individual HRTFs were measured. In the second condition, the stimuli had the same ITD but the ILD was constant across frequency. However, the ILDs of the individual stimuli and the flat spectrum stimuli were matched with respect to their level difference averaged across frequencies. Hence, the individual stimuli provide more consistent localization cues than the flat spectrum stimuli. A comparison of the results obtained in both conditions tests both alternative hypotheses that predict the variation of the ITD JND differently.

Hence, the following experiments are presented in this study to obtain thresholds for differences in the individual localization cues: In the first experiment of this study in Section 4.4 detection rates were measured for stimuli with smoothed HRTF spectra. In the second experiment in Section 4.5 the HRTF spectra were manipulated by transforming the individual spectral shape to the shape of dummy head HRTF spectra (see (Trampe, 1988) for a description of the dummy head). Finally, the sensitivity to ITD manipulations is investigated in Section 4.6.

4.2 General Method

The subject was sitting in an sound isolated booth (IAC, Model No. 405A) with dimensions of 3x3x2m (length, depth, height). An IBM compatible computer was located outside the cabin and controlled the experiments by running a MatLab script. The subject was seated in front of a window with the computer monitor behind it. The stimuli were presented to the subject by an AKG 501 headphone which was plugged to the audio output of a SoundBlaster 128 sound card. The presentation level was set to approx. 70 dB A[1].

The measurement paradigm was a two interval, two alternative forced choice paradigm (2I-2AFC). The stimulus sequence consisted of two intervals, each containing two virtually presented sounds. The HRTFs for creating the virtual stimuli were individually measured in a separated session (see chapter 3 for a description of individual HRTFs). Within each interval, both stimuli were separated by 300 ms silence. A pause of 500 ms separated both intervals within each trial. One of five different directions was chosen at random for each trial (ϕ =0°, 45°, 90°, 135°, 180°). The task of the subject was to select the interval in which the two stimuli were different in the aspects that were defined in the instructions given before. The instructions differed slightly from condition to condition so that the subject should focus the attention to the predefined differences in the stimulus sequence. The keyboard of the PC was used for recording the interval number. The only differences between the experiments I-III are the source stimulus (white noise, scrambled white noise, click train) and the type of manipulation (spectral detail reduction, 'spectral morphing', ITD variation) that was applied to the HRTFs.

4.3 Subjects

A total number of 10 subjects (eight male and two female) aged from 27 to 34 with clinical normal hearing participated voluntarily in the experiments. The number of subjects participating under each condition is listed in Table 4.1. All subjects were members of the physics and psychology department of the University of Oldenburg and had extensive experience in psychoacoustic tasks. The author participated in all measurements.

	SS I	SS II	SS III	'spectral morphing'	ITD variation
Subjects	8	7	7	6	6
Sessions	4	4	4	3	3
Trials p. Cond.	20	20	40	30	24

Table 4.1: Number of subjects per measurement condition (row I), number of sessions (row II) and number of trials per stimulus condition and measurement situation (row III).

4.4 Experiment I: Cepstral smoothing

In this experiment the sensitivity of the subjects to a reduction of the spectral HRTF detail was investigated [2]. The first two conditions ('SS I & II') were intended to estimate

[1]The presentation level was measured by presenting a virtual stimulus from 0° azimuth and elevation to a dummy head (Trampe, 1988) over headphone. The microphones in the ear canal of the dummy head (B&K 1/2", 4165 capsule) were directly plugged to a sound level meter (B&K 2610, fast averaging)

the sensitivity of the listeners to any changes of the stimulus perception introduced by cepstral smoothing. The instruction given to the subject, therefore, was to identify the interval in which stimulus differed by any spatial or non-spatial cue. In a further condition only spatial cues were provided to the subject by scrambling the source spectrum ('SS III'). However, due to spectral scrambling, every stimulus in the sequence changed its spatial position slightly. Therefore, the subject was instructed to select the interval in which the spatial deviation between stimuli was larger.
Separate measurement sessions were performed for each degree of smoothing. The number of stimulus repetitions for each stimulus condition (degree of smoothing and azimuthal angle) is given in the third row of Table 4.1. The measurement paradigm is described in Section 4.2.

4.4.1 Stimuli

To smooth the HRTF spectra the ceptral smoothing procedure used by Kulkarni et al. (1998) was adapted. The cepstrum of a HRTF spectrum $H(k)$ with N frequency components can be computed by applying the inverse Discrete Fourier Transform to the logarithm of the absolute magnitude spectrum $|H|$

$$C(n) = \sum_{k=0}^{N-1} \log |H(k)| e^{\frac{i2\pi kn}{N}} \tag{4.1}$$

[2]In this study, the spectral detail is defined as the amplitude variation of the frequencies that is removed by cepstral smoothing for the following reasons: The concept of auditory filters is well established to represent the frequency resolution of the auditory system. The bark scale (Zwicker and Fastl, 1990) and the ERB (equivalent rectangular band) (Moore *et al.*, 1990) scale are frequency scales which were developed to represent the properties of the auditory system with respect to frequency resolution. An appropriate method for smoothing the HRTF spectra may be to average the energy within each bark or erb band. This is approximately a logarithmic smoothing. The spectral detail in this case can be described as the amplitude variation of the frequencies within each band. This point of view is supported by the investigation of Asano et al. (1990). In this work it is shown, that only the macroscopic structure of the high frequency HRTF spectra contains spatial information and the details can be smoothed out without influencing the localization performance.
However, logarithmic smoothing reduces spectral information more in the high than in the low frequency regions. The head shadow and interference effects ($f > 2$ kHz) and the relevant monaural spectral transformation of the outer ear ($f > 5$ kHz), are located in high frequency areas. It seems, therefore, to be reasonable that the high frequencies should be accurately reproduced because more spatial information is contained in the high frequencies than in the low frequencies. This consideration would recommend a more linear weighting of the frequency scale. For this reason, the spectral detail was smoothed out here by cepstral smoothing, which is a linear approach with respect to the frequency scale. This consideration is consistent with the results obtained in Chapter 3. One outcome of this Chapter is that $1/N$ octave smoothing is less appropriate for spectral detail reduction because comparably high reduction of individual spectral information and deviation of the ILD is introduced.

where n varies from one to $N/2$. In this way, the variation of the absolute magnitude as a function of frequency is analyzed. The real part of this transformation can be used to reconstruct the logarithmic magnitude spectrum by a Fourier Synthesis.

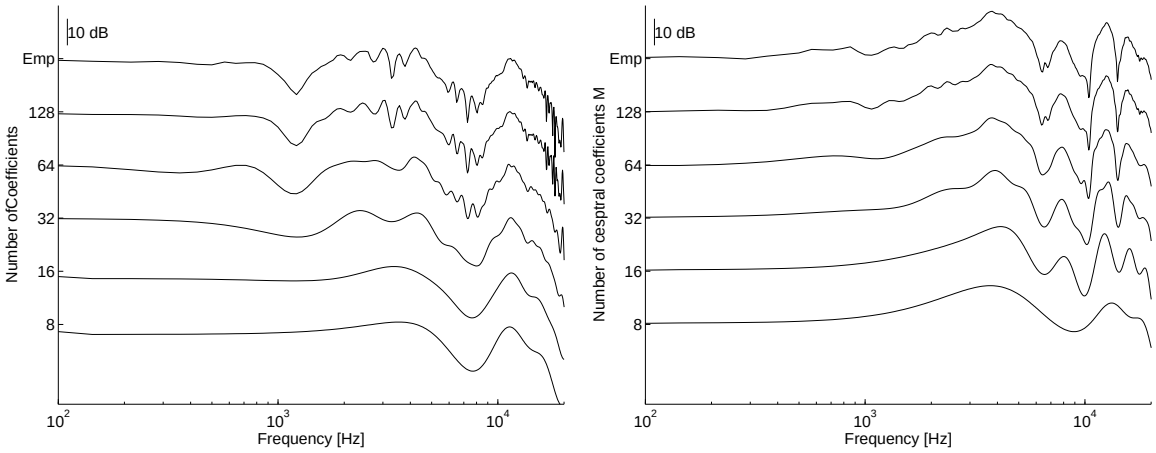
$$log(|\hat{H}(k)|) = \sum_{n=0}^{M} \tilde{C}(n) \cos \frac{2\pi nk}{N} \tag{4.2}$$

where $\tilde{C}(n)$ is defined as

$$\tilde{C}(n) = \begin{cases} \frac{(C(1)+C^*(1))}{2} & : \quad n = 0 \\ (C(n) + C^*(n)) & : \quad 1 <= n <= N/2 \end{cases}$$

For $M = N/2$ the reconstructed spectrum $\tilde{H}(k)$ equals the original one. A smoothed version of $H(k)$ can be obtained for M< $N/2$. In this way the cosine terms representing higher oscillations of the logarithmic magnitude spectrum are not used for a re-synthesis of the spectrum.

In Figure 4.1 smoothed versions of the left and right ear HRTFs (45° azimuth and



(a) Left ear.

(b) Right ear.

Figure 4.1: Smoothed HRTF spectra of one subject at $\phi = 45°$ azimuth and $\vartheta = 0°$ elevation. The empirical HRTF spectrum is plotted at the top of each panel.

0° elevation) of one subject are plotted for $M = 8, 16, 32, 64, 128$. The top line of each diagram represents the empirical HRTF spectrum. The logarithmic scale of the x-axis highlights the effects of linear smoothing on a logarithmic frequency scale. The macroscopic structure in the high frequency area of Figure 4.1(b) is well reconstructed even for $M = 16$. The notch in the mid frequency area around 1.3 kHz of the left ear spectrum (Figure 4.1(a)) is completely smoothed out for M=16. If eight coefficients are used for the synthesis procedure, the macroscopic structure in the high frequency region is only roughly approximated.

The stimulus sequence presented at each measurement trial consisted of three reference stimuli and one target stimulus. The reference stimuli were smoothed HRTFs with M=128. It is known from the work of Kulkarni et al. that this degree of smoothness is not distinguishable from the empirical HRTF (Kulkarni and Colburn, 1998). The targets were smoothed HRTFs with M=16, 32 and 64 in the conditions 'SS I' and 'SS II'. In addition, only eight coefficients were used in the 'SS III' condition.
A minimum phase was estimated from the smoothed HRTF spectra by

$$H_{min}(k) = \Xi(-ln(|\hat{H}(k)|)) \tag{4.3}$$

where Ξ denotes the Hilbert transform. This phase was applied to the smoothed spectrum. After transforming the HRTFs into the time domain, they are convolved with the source stimulus. Three different source sounds were used in separate conditions. In the first condition ('SS I') a 500 ms frozen white noise sample served as a source sound. The on- and offsets were ramped with 5 ms squared cosine ramps. A click train of 500 ms duration was used in the condition 'SS II'. The clicks were repeated at a rate of 100 Hz and onsets and offsets of the train were ramped in the same way as for the white noise. In the third condition the spectrum of the white noise stimulus (from condition 'SS I') was roved in sixth octave bands by up to $\pm$ 5 dB to prevent the subject from using timbre cues. The number of stimulus repetitions for each condition is summarized in the second row of Table 4.1.

4.4.2 Results

Detection rates for smoothed HRTFs in the measurement conditions 'SS I-III' are shown in Figure 4.2. The percentage of correct responses averaged across subjects is plotted as a function of the smoothing parameter M. The dashed lines in each subplot represent the 95% significance level for deviations from chance performance. The subplots are showing data obtained from different source positions in azimuth (denoted by ϕ).
Open squares depict the 'SS I' measurement condition. For azimuth $\phi = 0°$ and $\phi = 45°$ the correct response rate is near 100%. Even with $M = 64$ the smoothed HRTFs stimulus can be discriminated from the reference easily. The detection rates for $M = 32$ are below the threshold only for $\phi = 180°$. If sixteen cepstral coefficients are used for smoothing, the manipulated stimulus is detectable for all sound positions.
The results from the click train condition ('SS II', open diamonds) show much less detectability of the HRTF manipulation than for the white noise situation ('SS I'). Even for frontal sound incidence ($\phi = 0°, 45°$) the detection rates for $M = 32$ and $M = 64$ are near to or below the threshold. If the sound originates from lateral and rear azimuths ($\phi = 90° - 180°$) the smoothing manipulations are not detectable for the subjects, independent from the number of reconstruction coefficients.
In the third measurement condition ('SS III', crosses) spectrally roved white noise was

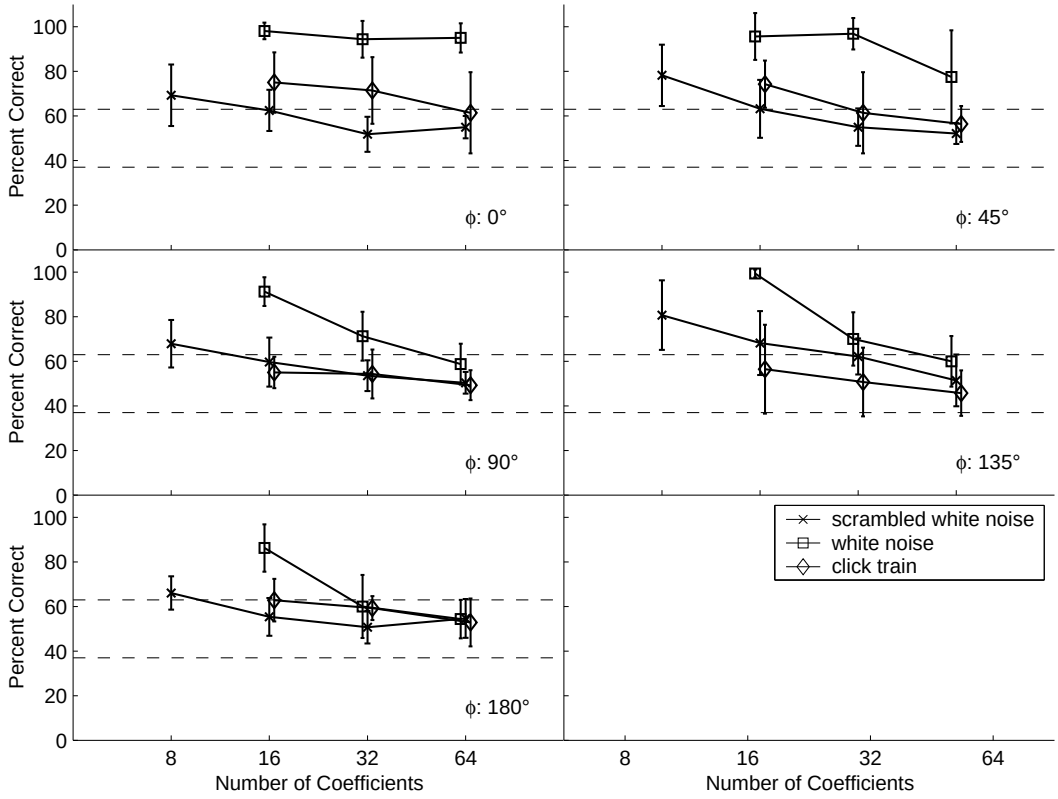


Figure 4.2: Results from the conditions 'SS I - III'. Percent correct responses averaged across subjects are plotted as a function of the number of smoothing coefficients. The error bars represent inter-individual standard deviations. The dashed lines mark the 95% significance threshold for being above chance level. Different angles of sound incidence are depicted in each subplot.

used as a sound source to prevent the subject from using non-spatial cues for the detection task. In general, the detection rates are below the threshold if more than 8 cepstral coefficients are used. Only at 135° of azimuth the detection rate approaches threshold for 32 cepstral coefficients. Except for this azimuth angle the detection rates for the scrambled white noise condition are lowest compared to the other measurement conditions.

Relation to physical stimulus parameters

In order to relate the physical cues that were available to the subjects to their performance, Figures 4.3 and 4.4 give the level differences between the smoothed and the reference HRTFs for the right and left ear, respectively. Each subplot shows the unsigned differences between the HRTF spectra reconstructed with 128 coefficients and the smoothed target spectra with $M = 8, 16, 32, 64$ plotted on a logarithmic frequency scale for one subject and angle of sound incidence ϕ. The subplots differ in the angle of sound incidence. A logarithmic frequency axis is used since it relates better to the perceptual

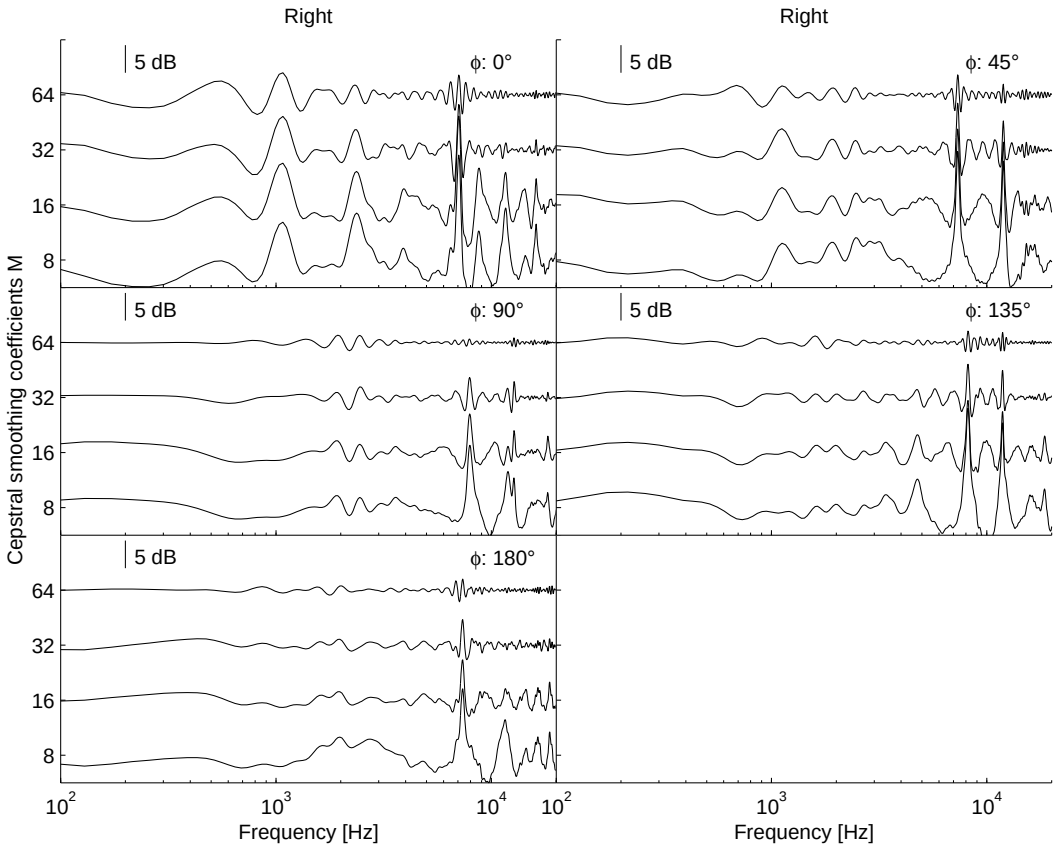


Figure 4.3: Level differences between reference and smoothed HRTF spectra of the right ear for one subject.

cues that can be exploited by the subject.

It can be seen from Figures 4.3 and 4.4 that roughly the same structural differences in spectral shape occur for all degrees of smoothing, while the magnitude of these differences increases with increasing smoothing, predominantly in the high frequency region. The corresponding effect of smoothing on the ILD is given in Figure 4.5. The broad band ILD difference is computed from the absolute level deviation between the smoothed and original interaural transfer function (ITF) averaged across frequencies and subjects. In Figure 4.5(a) level deviations were averaged for frequencies up to 4 kHz and in Figure 4.5(b) for frequencies above 4 kHz. From Figure 4.5(a) it can be seen that the influence of the smoothing process on the low frequency area strongly depends on ϕ. Only small level deviations can be observed for frontal and rear sound incidence ($\Delta ILD < 0.7dB$), but for lateral sources the level deviation reaches values up to 2.6 dB. For frequencies above 4 kHz the ILD deviations depend less on source azimuth. At positions on the cone of confusions ($0°, 180°$ and $45°, 135°$) the ILD deviations are very similar.

To relate the physical cues presented above to the perceptual data, correlation coefficients between percent correct responses and different distance measures of the smoothed and original HRTFs were calculated (see Appendix A.2). Two different distant measures

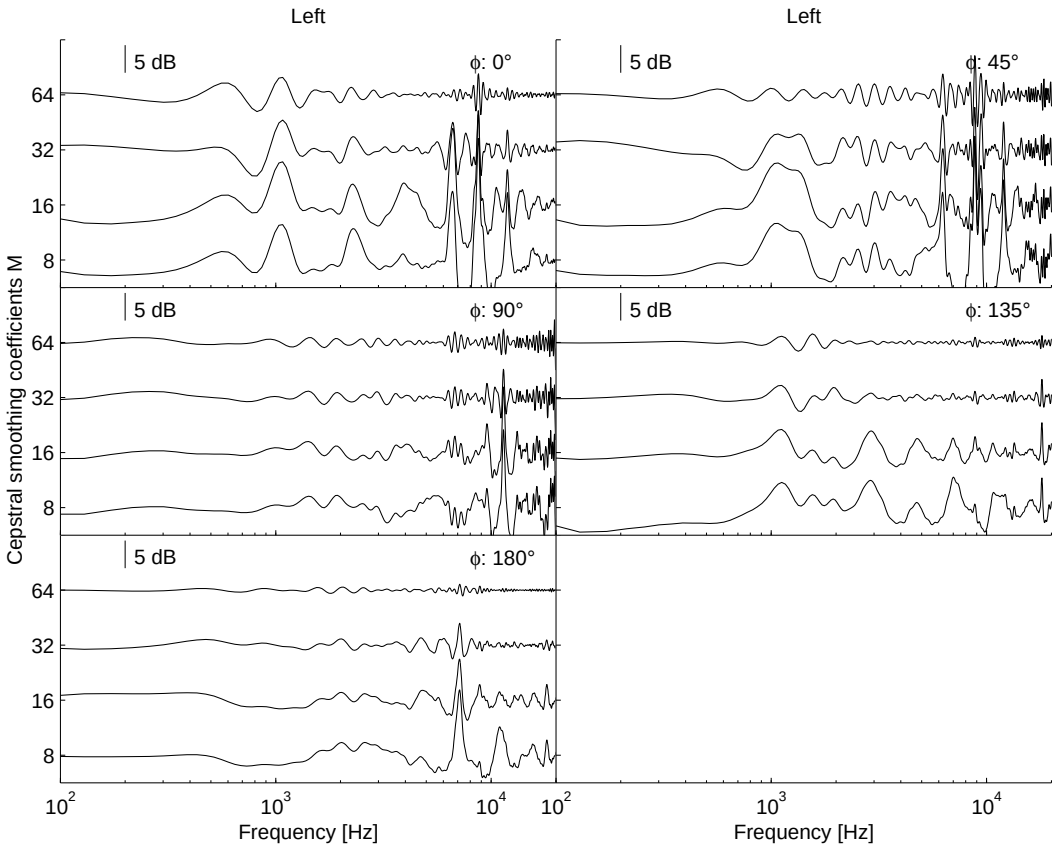
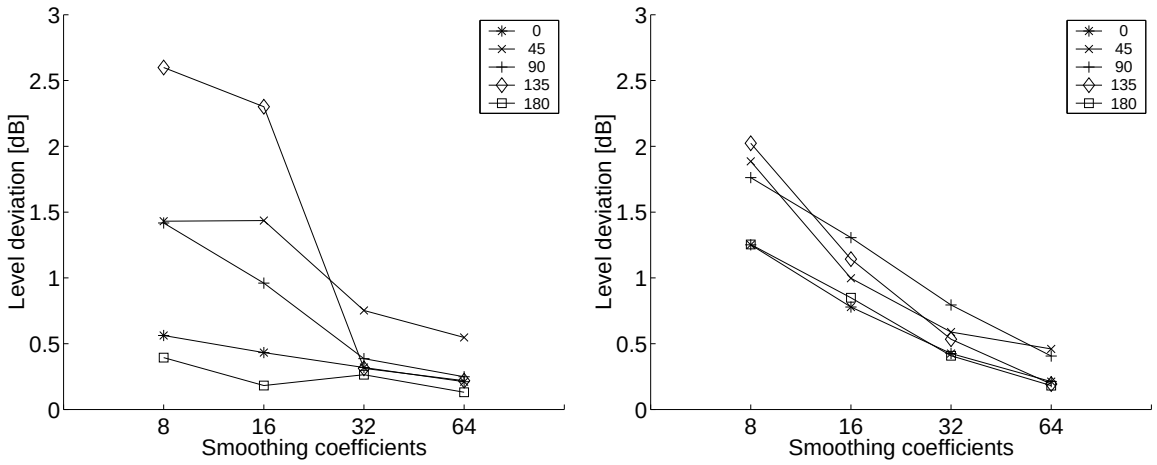


Figure 4.4: Level differences between reference and smoothed HRTF spectra of the left ear for one subject.

that show high correlations for the conditions 'SS I', 'SS II' and 'SS III' are given here. The HRTF spectra were first filtered by a Gammatone filter bank. In the 'SS I' and 'SS II' condition, absolute level differences between the smoothed and original HRTFs of the right ear were calculated for each filter bank channel and averaged across frequency. This distance measure is called D_{mon}. To derive a binaural distance measure D_{bin} for the 'SS III' condition, interaural level differences for each frequency channel were computed both for smoothed and original HRTFs. Then, the level deviations between smoothed and un-smoothed ILDs were calculated in each frequency channel. Finally, the mean across frequencies was computed. Correlation coefficients for the distance measure D_{mon} and the percent correct values in the conditions 'SS I' and 'SS II' are listed in Table 4.2 (see Appendix A.2 for a complete table with correlation coefficients for all distance measures). In the third row the correlation coefficients for D_{bin} and the percent correct responses of the condition 'SS III' are given. The low correlation values for $\phi = 0°$ and $\phi = 45°$ in the 'SS I' condition are due to the ceiling effect of subjects' response. For the other angles of azimuth the correlation coefficients are at least 0.8. Only low correlations can be found for the 'SS II' condition. This can be related to the detection rates that do not deviate significantly from chance performance for $\phi = 90° - 180°$. It



(a) $f < 4kHz$ (b) $f > 4kHz$

Figure 4.5: Level deviation between smoothed and original ILD calculated in two frequency bands.

can be assumed that the correlation rises if the degree of smoothing is increased. The distance measure shows higher correlation to the perceptual data at 45° of azimuth and nearly no correlation to the data for $\phi = 0°$. The correlation analysis shows that some distance measures have higher correlations for 0° azimuth (see Appendix A.2). However, since these correlations are still very low (≤ 0.32) no alternative distance measure for the condition 'SS II' is given here.

Higher correlations can be observed for the 'SS III' condition which takes its maximum value at 45°. In general, lateral source positions show higher correlation values than source positions in the median plane.

Condition	0°	45°	90°	135°	180°
SS I	0.14	0.42	0.81	0.85	0.8
SS II	0.18	0.63	0.19	0.29	0.41
SS III	0.66	0.88	0.71	0.75	0.59

Table 4.2: Correlation values between percent correct responses and the distance measure D_{mon} are listed in the first and second row for the conditions 'SS I' and 'SS II', respectively. In the third row correlation values for the detection rates and the distance measure D_{bin} are shown.

Figure 4.6 displays the number of correct responses in percent as a function of the ILD deviation described by the distance measure D_{bin} for the 'SS III' condition. Each subplot shows data from a different angle of sound incidence. Regression lines are plotted as solid lines for each angle of azimuth. The dashed lines mark the 95% confidence interval for

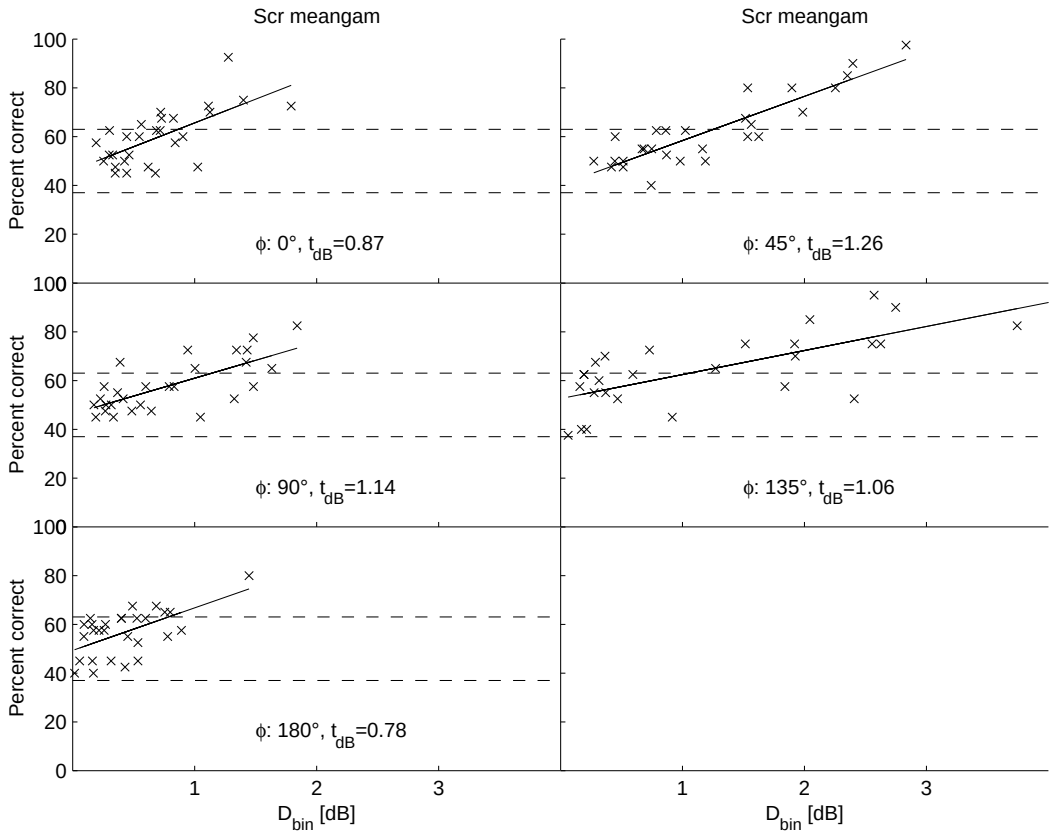


Figure 4.6: Percent correct responses as a function of the acoustical differences between target and reference stimuli. Data for all subjects averaged across sessions is presented. The dashed lines are representing the 95% confidence bounds for chance performance. In each subplot the mean detection thresholds t_{dB} are given. They are computed by calculating the level deviations for which the regression functions intersect the significance threshold.

deviations of the responses from chance performance. From these data, a physical detection threshold can be specified as the level for which the regression function intersects the significance threshold for deviation from chance performance. These thresholds can be interpreted as the average physical value which causes a manipulation in the HRTF to be detectable. The exact thresholds are given in each subplot of Figure 4.6.
The thresholds are approximatively 1 dB with a slight decrease for sound positions in the median plane and an increase for lateral source positions. If the data is plotted for the 'SS I' condition in the same way, it is obvious that subjects were able to detect the target stimulus for level differences greater than 0.5 dB. Because the psychometric functions for $\phi = 0°$ and $\phi = 45°$ are always above threshold, no corresponding physical detection threshold can be presented for these positions. However, the threshold is at least below 0.5 dB.
An estimate of the thresholds for the click train condition can only be given for $\phi = 45°$. A calculation of the threshold from a plot similar to Figure 4.6 shows that it amounts

to about 1 dB.

4.4.3 Discussion

In experiment I the HRTFs of the target stimuli were manipulated by smoothing the spectra. Although the three measurement conditions 'SS I - III' only differed in the source sound (white noise, click train and scrambled white noise), the detection performance of the subject as a function of smoothing differed considerably. Subjects were most sensitive to changes of the HRTFs when a white noise served as a source. The lowest sensitivity was observed for the scrambled white noise stimulus. The detection performance for the click trains is between the other two conditions.
It is remarkable that the detection rate for the white noise stimulus decreases for lateral source positions. The results show that smoothing with 32 and 64 cepstral coefficients was easily detected for sound incidence from 0° and 45° but the detection rate at, for instance, 90° is near to or equal to chance level. Because no spectral scrambling was applied to the stimulus, it is likely that the subjects used timbre variations as a detection cue. To relate this effect to physical cues, monaural level deviations for these positions were computed, averaged across frequency and subjects. The results show that the monaural level deviations are nearly equal for 0° and 90° at the same degree of smoothing. Hence, it seems that subjects were not only using monaural but also some binaural cues for the detection process.
The results show, furthermore, that the detection thresholds for the click train condition ('SS II') are lower than for the white noise condition ('SS I'). Both stimuli contain the same amount of spectral deviation between the reference and target stimulus. However, the click train has, in contrast to the white noise, a tonal component that corresponds to the repetition rate of the clicks (100Hz). This tonal component may dominate the perception of the click train and, hence, reduces the attention of the subject to spectral changes of the stimulus. This consideration could explain the lower rate of percent correct responses for the click train. It can be concluded from this result that HRTFs may be smoothed by a higher degree, if more complex stimuli than white noise are convolved with the smoothed HRTFs. Hence, the detection thresholds for white noise appear to be upper limits.
Two different distant measures of the physical differences of the spectral localization cues introduced by smoothing were correlated to the perceptual data. The correlation analysis revealed that the monaural level differences (described by the distance measure D_{mon}) between the original and the smoothed HRTF spectra correlate well to the perceptual data of the 'SS I' condition. This suggests, that subjects were using mainly monaural cues for the detection process. The thresholds calculated by the distance measure D_{mon} suggest that detectable spectral timbre variations for stimuli at lateral source positions are introduced by cepstral smoothing if D_{mon} exceeds 0.5 dB. For

frontal sound incidence the threshold is even below 0.5 dB.
The binaural distance measure 'D_{bin}' describes the mean ILD deviation introduced by smoothing and correlates well to the perceptual data in the 'SS III' condition. Because timbre cues are excluded in this conditions, subjects had to use spatial displacements of the stimuli for the detection of the manipulated HRTFs. Detection thresholds given by D_{bin} are approx. 0.8 dB for sound incidence from the median plane and approx. 1.2 dB for lateral positions.
An appropriate distance measure that correlates to the perceptual data in the 'SS II' condition have not been found. This is mainly caused by the low detection rates that do not deviate from chance performance for most source positions. The detection threshold is only exceeded for $\phi = 0°$ and 45° azimuth. Consequently, higher detection correlations can be observed at 45°. However, an explanation for the low correlation of the physical cues to the perceptual data for 0° azimuth can not be given here.

Comparison to the literature

The measurement task and the kind of manipulation introduced to the HRTF spectra are based on a study of Kulkarni and Colburn (1998). In their study, the reference stimuli were delivered in the free-field via a loudspeaker and compared to headphone presented virtual stimuli in a discrimination task. For the virtual stimuli, HRTFs with different degrees of spectral smoothing (8, 16, 32, 64, 128, 256, 512 cepstral coefficients) were convolved with a white noise stimulus (80 ms length) which was randomly scrambled in its spectrum (1/3 octave bands, ± 5 dB range). It was found that 16 cepstral smoothing coefficients are sufficient for providing all spatially relevant spectral information for the investigated source directions. This result is consistent with the findings of the corresponding experiment ('SS III') in the current study. However, Kulkarni and Colburn excluded the 135° azimuth position. For this source direction the results of our study show that the detection performance is slightly above chance level even for 32 cepstral coefficients. This is caused by a comparatively high ILD deviation in the frequency region below 4 kHz. Hence, 16 coefficients are not sufficient for all angles of azimuth. Because the spectral shape of the HRTFs is more complex for elevations below the horizontal plane, it can be assumed that even more cepstral coefficients are required for lower elevations.
The main difference in the method between studies is that Kulkarni and Colburn compared a virtual stimulus with smoothed spectra to a real source, whereas in the present study only virtual stimuli with different degrees of smoothing were compared. It could have been that a comparison of real and virtual stimuli is easier for subjects (e.g. due to slight head movements of the subjects' head between stimuli). However, the results are similar in both studies and, hence, it can be concluded that the method did not influence

the results.

Asano et al. (1990) analyzed the localization performance as a function of spectral smoothing. It was shown that 20 filter coefficients of an auto regressive, moving average (ARMA) filter model, (i.e. 10 for the FIR and 10 for the IIR part of the digital filter) were sufficient to allow the subjects to localize correctly in an absolute localization measurement. Among other things, the authors concluded that information contained in the frequency range below 4 kHz helps to resolve front/back confusions. This finding supports the results from Blauert (1969) that directional information is also contained in the low frequencies area. However, the contribution of the spectral cues in the low frequency region to the directional perception seems to be small because the result of the present study show that an elimination of the low frequency cues was not detectable for the subjects. A similar result was found by Langendijk and Bronkhorst (2001). They eliminated spectral cues in the low frequency region by setting the level to 0 dB and found that eliminating the cues in the frequency range from 4-6 kHz did not reduce the localization performance significantly in an absolute localization task.

In another study of Langendijk and Bronkhorst (2000) the sensitivity of human listeners to interpolated HRTFs was investigated. A discrimination task was used to obtain the detection threshold. The distance measure describing the acoustical differences between target and reference HRTFs was calculated in the following way: The level differences between the target and reference HRTFs of the right ear were computed in 1/3 octave bands. The maximal difference across frequency bands was used as distance measure. Thresholds of 1.5 dB to 2.5 dB were calculated for the detection of a change in stimulus timbre and above 2.5 dB for a change in spatial position.

To compare the results of the present study to the thresholds obtained by Langendijk and Bronkhorst the distance measure given above was applied to the perceptual data obtained here (see Appendix A.2, distance measure D8). However, instead using 1/3 octave bands a Gammtone filter bank was applied. The correlation between this distance measure and the perceptual data of the present study was lower in comparison to the distance measures D_{mon} ($\bar{r} = 0.8$ to $\bar{r} = 0.82$) and D_{bin} ($\bar{r} = 0.63$ to $\bar{r} = 0.72$) . Furthermore, the threshold was set by Langendijk and Bronkhorst to 75% correct responses, whereas in the present study it was set to 65% correct responses.

However, even by calculating the thresholds in the same as proposed by Langendijk and Bronkhorst different values are obtained. By this measure spectral timbre variations are detectable for spectral deviation in one frequency band of approx. 1.3 dB (1.5-2.5 dB is given by Langendijk and Bronkhorst). Substantially higher thresholds are obtained for the detection of a spatial displacement (approx. 5 dB in the present study and $>$ 2.5 dB in the study of Langendijk and Bronkhorst). These differences may be related to the different physical deviations of the localization cues that are introduced in both studies. The target stimuli in the study of Langendijk and Bronkhorst were interpolated HRTFs and in the present study subjects had to detect HRTFs with smoothed spectra. It could

be that the cues introduced by applying interpolated HRTFs to the stimuli are better detectable for the subjects because more relevant spatial information is distorted.

4.5 Experiment II: Spectral morphing

In this section the sensitivity of subjects to a spectral transformation is investigated which also shifts the center frequencies of the peaks and notches of the individual HRTFs. Therefore, this transformation is more destructive to the individual information of the HRTF spectra compared to the spectral detail reduction transformation.
The manipulation applied in this experiment transforms the spectral shape of the individual HRTF spectra to the spectral shape of dummy head HRTF spectra ('spectral morphing'). The measurement paradigm as described in Section 4.2 was used to obtain detection rates for the target stimuli.

4.5.1 Stimuli

The manipulated HRTF spectrum $\hat{H}_\alpha$ of the target stimulus was computed by transforming the individual HRTF H by

$$|\hat{H}_\alpha| = (1-\alpha)|H| + \alpha|H|\frac{|D_{MS}|}{|H_{MS}|} \tag{4.4}$$

where D_{MS} is the macroscopic structure of the corresponding dummy head HRTF spectrum and H_{MS} is the macroscopic structure of the individual HRTF both obtained by smoothing the spectra in sixth octave bands. The right term of the sum in Equation 4.4 represents the absolute transfer function spectrum where the macroscopic shape of the individual HRTF spectrum has been completely transformed to the dummy head shape. By stepwise increasing the factor α from zero to one, the ratio of the individual macroscopic structure and the dummy head structure is varied.
Throughout the study, this process is called 'spectral morphing'. Because 'spectral morphing' is done independently for the left and right ear spectra, the effect of the procedure on the HRTFs can be observed best in the interaural transfer function. As an example the interaural transfer function $ITF_\alpha = H_{\alpha R}/H_{\alpha L}$ of one subject for $\phi = 90°$ is calculated with $\alpha = 0, 0.1, 0.3, 0.5, 0.7, 0.9, 1$ and plotted in Figure 4.7. It can be observed, that a peak around 5.4 kHz shifts in its center frequency to 7 kHz as α is increased. Furthermore, a notch around 9 kHz is broadened and the overall level in the frequency region above 10 kHz varies as a function of alpha. The fine structure, however, remains constant.
The morphed HRTFs obtained in this way were applied to a white noise stimulus which was randomly level roved in the spectrum ($\pm 5dB$ range in 1/6 octave bands). The stimulus sequence consisted of three reference stimuli ($\alpha = 0$) and one target stimulus with

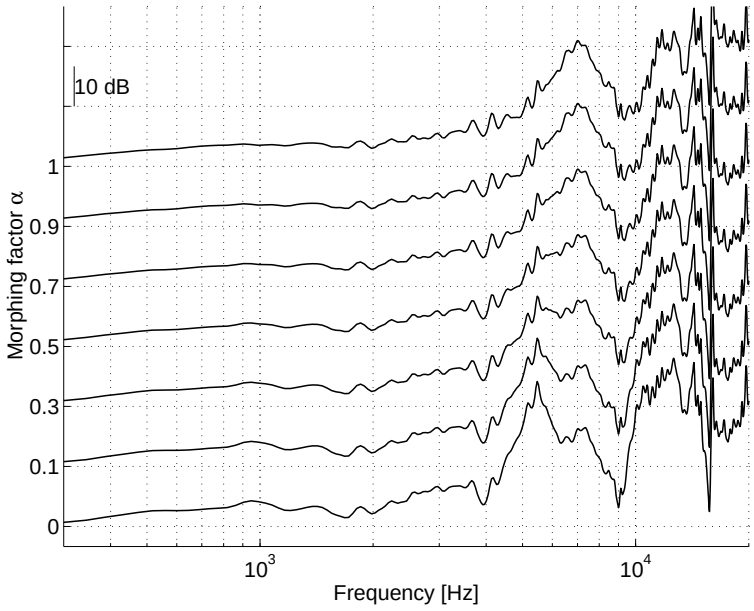


Figure 4.7: The morphed interaural transfer function spectrum of one subject at $\phi = 90°$ azimuth and $\vartheta = 0°$ elevation is shown as a function of the morphing factor α.

a randomly chosen α= 0.1, 0.3, 0.5, 0.7, 0.9 . The stimulus sequence was randomly presented from one of five different azimuth positions. The number of stimulus repetitions and measurement sessions is listed in Table 4.1.
The subject was instructed to identify the interval in which both stimuli differ more with respect to their spatial impression. This instruction was necessary because the random changes of the stimulus spectrum introduced a change in the perceived location in addition to the changes produced by the manipulated HRTFs.

4.5.2 Results

The results for the 'spectral morphing' manipulation are summarized in Figure 4.8. The figure is organized similar to Figure 4.2 and shows the percent correct responses averaged across six subjects as a function of the morphing factor α.
The subjects obviously are very sensitive to the manipulation introduced. The comparatively high standard deviation across subjects is probably due to the heterogeneous stimuli produced by the morphing process that introduces different cues for each subject. The same linear dependency of performance on morphing factor is observed for all stimulus positions, except for $\phi = 90°$ where the function slightly deviates from the linear behavior.
The highest sensitivity is observed for sound incidence out of the median plane and the sensitivity decreases as the source location moves to the side.

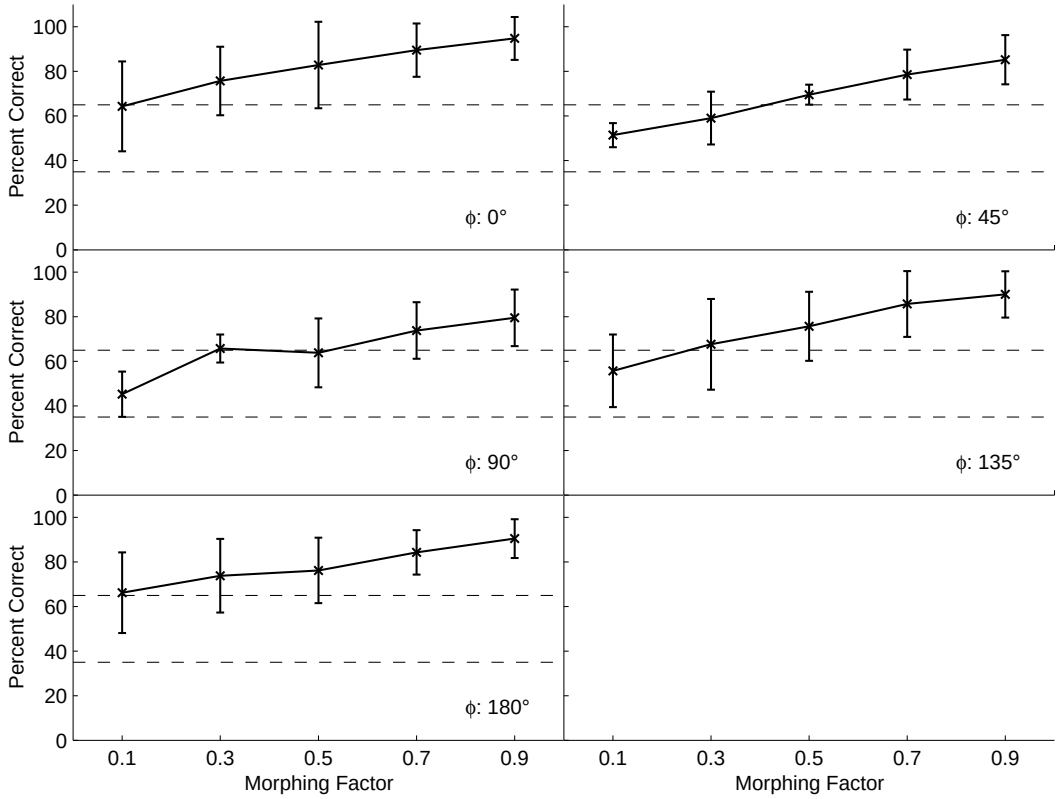


Figure 4.8: Percentage of correct responses are shown as a function of the morphing factor α. Organization of the figure is similar to Figure 4.2.

Relation to physical stimulus parameters

In order to assess the cues that subjects may have used for the discrimination task, Figure 4.9 gives the level differences between the morphed and original interaural transfer functions for one subject for $\phi = 45°$. The ILD deviation was computed from the interaural transfer function (ITF), derived both from the original HRTFs and the morphed HRTFs. The signed level differences between these two ITFs were calculated.
In contrast to HRTF smoothing, only small changes can be observed for lower frequencies. This reflects the physical differences between the dummy head and the individual's head that are only significant on a cm scale (e.g. differences in exact body and pinna geometry) and hence relate to frequencies above approx. 1 kHz, whereas the lower frequencies are influenced by the dummy head and a real listener in roughly the same way. The effect of morphing the transfer functions on the ILD is illustrated in Figure 4.10. Level deviations were computed in the same way as for Figure 4.9 and averaged across two frequency bands below 4 kHz (4.10(a)) and above 4 kHz (4.10(b)). At low frequencies , the mean level difference between the morphed and the original interaural transfer function (ITF) is between 0.1 dB ($\alpha = 0.1$, $\phi = 90°$) and 1.5 dB ($\alpha = 0.9$, $\phi = 45°$). The least difference between the dummy head ITF and the individual ITF at low frequencies is at 90° and the lowest concordance can be observed for $\phi = 45$. For frequencies above

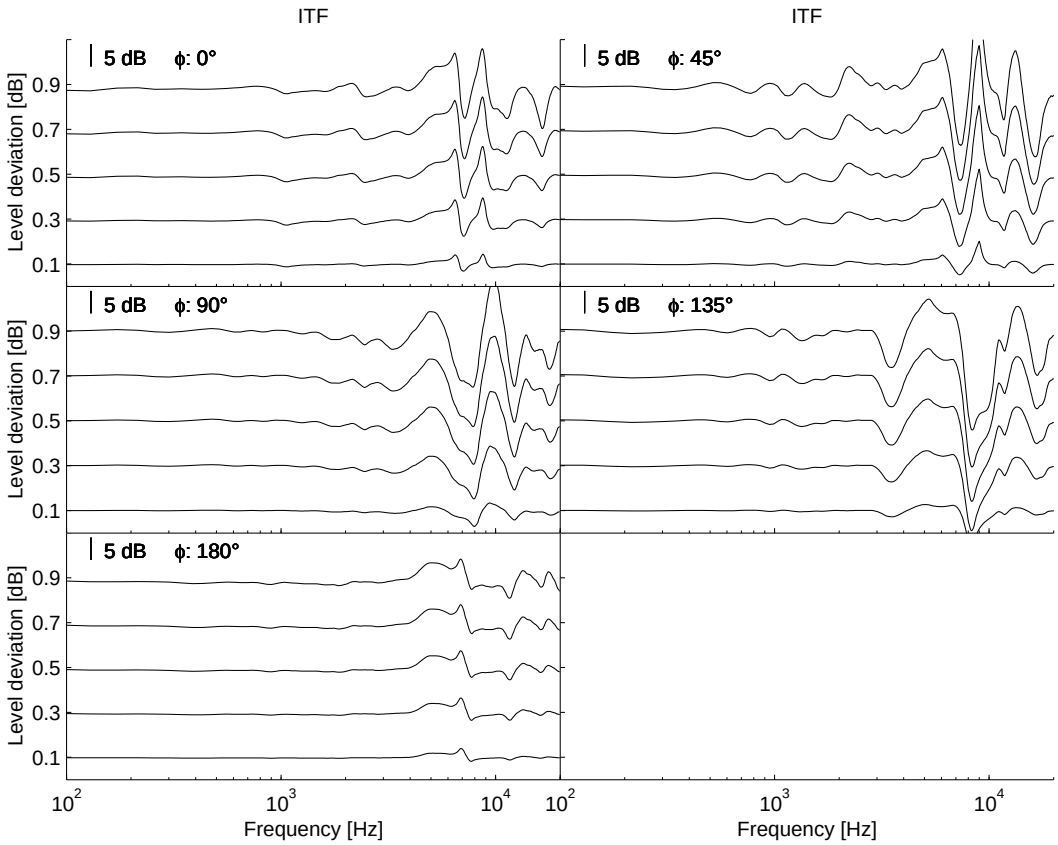
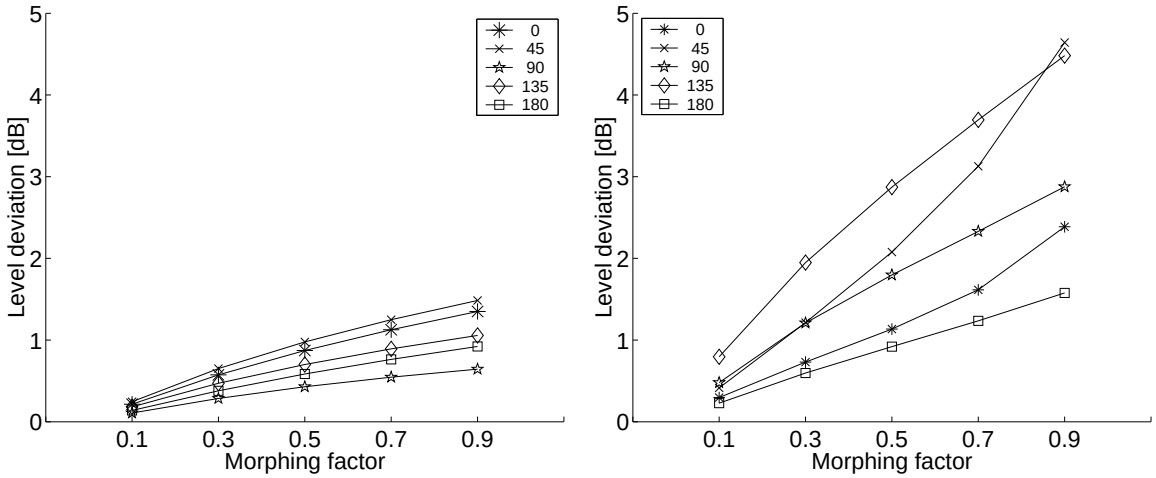


Figure 4.9: Level differences between reference and 'morphed' interaural transfer functions for one subject. In each subplot a different angle of sound incidence is depicted.

4 kHz the dummy head ITFs are deviating more from the individual ITF (minimum: 0.2 dB for $\alpha = 0.1$, $\phi = 180°$; maximum: 4.8 dB for $\alpha = 0.9$, $\phi = 45°$).

In a correlation analysis correlation coefficients between the perceptual data and different distance measures were computed. The distance measure D_{bin} used for the condition 'SS III' in experiment I shows the best correlation (see Appendix A.2). Correlation values are lowest for stimuli at the median plane (r $\approx$ 0.67). At lateral angles correlations are in the range of 0.74 to 0.79. In Figure 4.11 percent correct responses are plotted as a function of D_{bin}. A regression function is plotted as a solid line for each subplot. In each subplot data for a different azimuth of the stimulus is presented. The dashed lines mark the 95% confidence interval for deviations of the detection rates from chance performance.

From the data in Figure 4.11 detection thresholds in dB can be obtained in the same way as described in experiment I. For sound incidence out of the median plane the detection thresholds are influenced by the ceiling effect caused by high detection rates. If the threshold is corrected for this effect, the threshold is approx. 0.6 dB. The thresholds for sound incidence from the side are increased to about 1.2 dB.



(a) $f < 4kHz$ (b) $f > 4kHz$

Figure 4.10: Level deviations between morphed and original ILD in two frequency bands.

4.5.3 Discussion

The 'spectral morphing' manipulation was intended to affect the spatially relevant information of the HRTF spectra. This was done by transforming the individual HRTF spectra to the macroscopic shape of dummy head HRTF spectra. The dummy head used ('Oldenburg dummy head') differs in several aspects from individual heads. First, the geometry of the head and the pinna is only suitable for an average subject and an approximation of an individual head. Second, the pinnae are symmetric with respect to the median plane. Third, the dummy head has no torso and shoulders. Fourth, the dummy head has no ear canal. Furthermore, no hair is attached to the skin. Therefore, the dummy head HRTFs differs in two general criteria from individual HRTFs. First, the different geometry scales properties common to all heads in frequency. For instance, if the individual head is smaller than the dummy head, interference effects are located at higher frequencies for the smaller head. Second, due to the lack of physical structures that generate spatial information (e.g. the dummy head has no shoulders) less information, at least at low frequencies is provided to the individual listeners by listening to dummy head HRTFs. Therefore, transforming the individual HRTF spectra into dummy head spectra transforms information to different frequency areas and eliminates spatial information contained in the individual HRTFs.

As expected, the results show that subjects are very sensitive to the HRTF transformation. Subjects reported that one detection cue was the occurrence of front/back confusions for the targets, mainly for source positions in the median plane. These confusions were perceived even for low values of α. A transformation like 'spectral morphing' destroys the individual spectral cues that are resolving the front/back confusions. There-

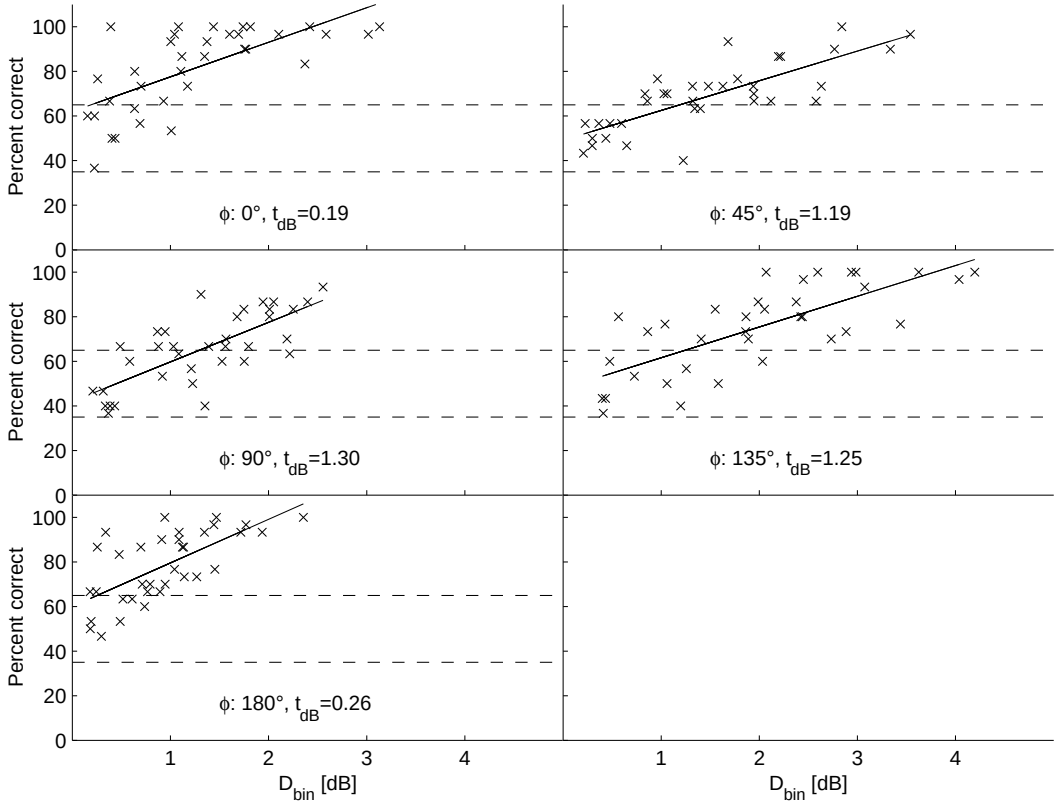


Figure 4.11: Percent correct responses measured for the 'spectral morphing' condition as a function of the distance measure D_{bin}. Data for all subjects averaged across sessions is presented. In each subplot the mean detection thresholds t_{dB} are given. They are computed by calculating the level deviations for which the regression functions intersect the significance threshold.

fore, it is likely that front/back confusions are introduced by 'spectral morphing'. Front/back confusions are a very obvious cue producing a high spatial distance between target and reference and may reduce the threshold. This could explain the high detection rates even for low values of α.

The perceptual data was related to the physical differences introduced by 'spectral morphing' by calculating correlation coefficients between the percentage of correct responses and the distance measure D_{bin} that describes the mean deviation of the ILD in different frequency bands. The correlation values are approx 0.67 for frontal sound incidence and 0.79 for sound incidence from the sides. Higher correlation values for frontal sound incidence can be expected, if the ceiling effect of subjects' response would be removed.

The same distance measure was used for the scrambled white noise condition ('SS III') in the former experiment. The detection thresholds extracted from the perceptual data are comparable (slightly lower for frontal sound incidence) but the manipulation of the spectra were different. Therefore, it can be assumed that the thresholds given by the distance measure D_{bin} are appropriate for serving as a common measure for deviations

of individual HRTF spectra that are irrelevant for the spatial perception.

4.6 Experiment III: ITD variation

The aim of this investigation was to investigate in which way the spectral shape of a broadband noise stimulus affects the ITD JND. The task of the subjects was to detect spatial displacements of virtual stimuli introduced by manipulations of the ITD in two conditions. The two conditions differ in the amount of spatial information in the shape of the stimulus spectra. In the first condition ('plausible ILD'), individual HRTFs were used to filter the white noise stimuli. In the second condition ('constant ILD'), the spectral shape that is introduced by using individual HRTFs was eliminated and set to a constant factor across frequency. However, the mean ILD across frequency was equal under both measurement conditions.
Two hypothesis were given in the introduction that predict the differences in the detection rates between the two conditions in opposed ways. The first hypothesis suggests that the detection rates are lower for the 'plausible ILD' condition because the additional spatial information in the spectra produce a more focused perception of the stimulus and, hence, spatial displacement are easier to detect. The second hypothesis predicts lower detection rates for the 'plausible ILD' condition because the additional spatial information contained in the spectral shape stabilizes the perception of the spatial object. A comparison of the detection rates in both conditions can reveal which of both hypothesis holds true.
The same measurement paradigm as in the experiments I and II was used and only the type of manipulation applied to the HRTFs was changed. The number of subjects participating in the experiments and the number of stimulus repetitions for each condition is given in Table 4.1.

4.6.1 Stimuli

The stimuli for the 'plausible ILD' condition were generated in the following way: The same sample of frozen white noise as in the former experiments was used as a sound source. For creating the reference stimuli white noise was convolved with the individual minimum phase HRTFs of the left and right ear. The target stimulus was generated by shifting the individual HRIR of the left ear in time by $\Delta\tau = \pm 22.7, \pm 68.0$ and $\pm 113.4 \mu s$. This corresponds to the discrete time shift imposed by the sampling frequency.
The reference stimuli for the second condition ('constant ILD') were created by applying the ITD, obtained from the individual HRTFs of the first experiment, to the white noise stimulus. Subsequently, the white noise stimuli for the left and right ear were scaled to

the same RMS level difference as obtained from the white noise stimuli convolved with individual HRTFs. The target stimuli were generated by shifting the stimulus that is presented to the left ear by $\Delta\tau$ =+22.7, +68.0 and +113.4μs. Note, that only positive variations of the ITD were applied to the flat spectrum stimuli.
For both conditions the stimulus sequence was presented randomly from one of five different azimuth positions. Each condition was conducted in a separate measurement session.

4.6.2 Results and Discussion

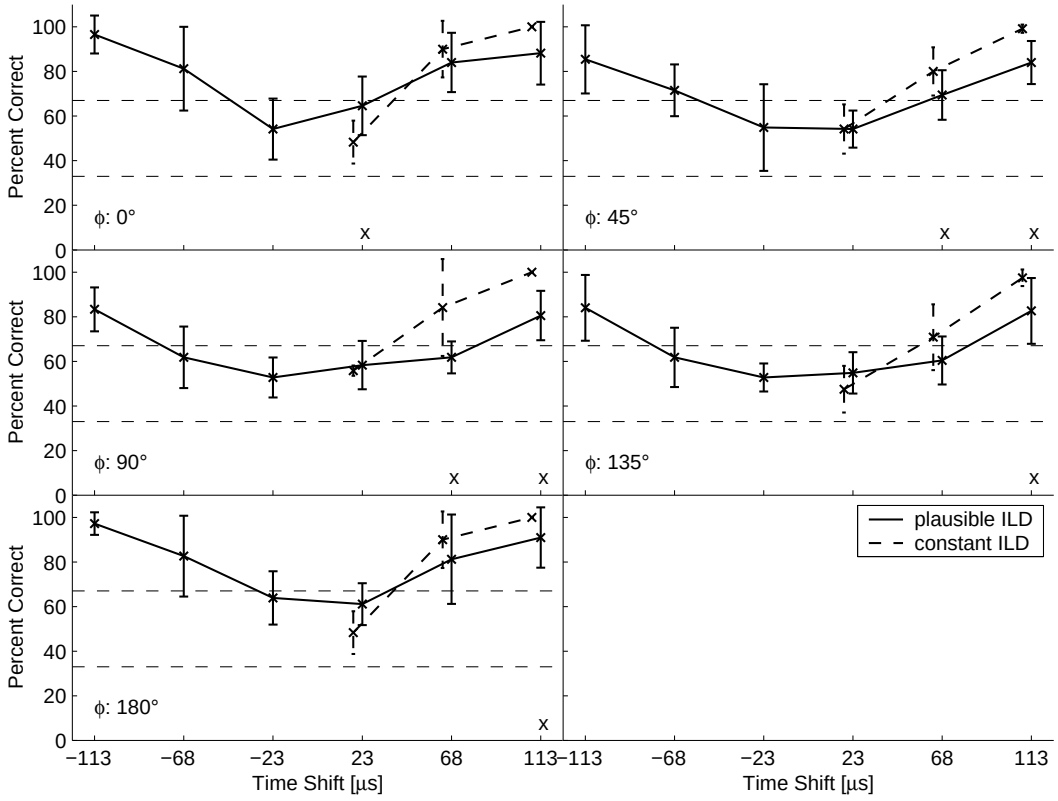


Figure 4.12: Percent correct responses averaged across subjects as a function of ITD variation for HRTF stimuli are depicted (solid lines, condition 'plausible ILD'). Dashed lines give percent correct responses for flat spectrum stimuli (condition 'constant ILD'). The dashed horizontal lines mark the thresholds for deviation from chance performance and the error bar indicate the inter-individual standard deviation.

Percentage correct identifications of manipulated (i.e. interaural time shifted) stimuli averaged across subjects are depicted in Figure 4.12. The percentage of correct responses is plotted as a function of the ITD variation for both measurement conditions. The solid lines connect data for the 'plausible ILD' condition and the dashed lines connects data for the 'constant ILD' condition. Significant differences ($p<0.05$) between the detection

rates in both conditions are marked by a small 'x' at the bottom of each sub-plot. The horizontal dashed line mark the thresholds for deviation from chance performance. Each subplot shows data from a different source positions in azimuth.
In the 'plausible ILD' condition the detection rate for $\Delta\tau = \pm 23\mu s$ is below the significance threshold for any angle of source incidence. However, for $\phi = 0°$ and $\phi = 180°$ percent correct responses are very near to the threshold. A delay of $\pm 68\mu s$ (3 samples) introduced to the lagging ear can be detected at $\phi = 0°, 45°, 180\circ$ but not significantly at $\phi = 90°, 135°$. If the delay is further increased (i.e. $\pm 113\mu s$ delay) it is significantly detected, independent from the angle of sound incidence. It should be noted, that the error bars for a delay of the lagging ear of $\pm 68\mu s$ show, that the sensitivity to the introduced delay varies considerably across subjects. Especially at $\phi = 0°$ and $\phi = 180°$ some subjects are below the significance threshold even for $\Delta\tau = 66\mu s$ and some are above the threshold for $\Delta\tau = 22\mu s$.
The detection rates for the flat spectrum stimuli (condition 'constant ILD') are higher for all ITD variations than the detection rates for the empirical stimuli if the detection rates deviate from chance performance. However, the detection rates are not significantly different for each condition, as can be seen in Figure 4.12 where only statistically significant differences are marked by the 'x' at the bottom of each plot.
The detection performance for $\Delta\tau = 23\mu s$ is always at chance level. In contrast, ITD variations of $68\mu s$ and $113\mu s$ were detectable for subjects independent from the source azimuth.

An estimate of the average ITD JNDs in both conditions was calculated from the

Azimuth [°]	0°	45°	90°	135°	180°
plausible ILD $\Delta\tau > 0$	23	63	78	81	30
plausible ILD $\Delta\tau < 0$	43	56	78	82	36
constant ILD $\Delta\tau > 0$	43	45	41	61	43

Table 4.3: ITD JNDs in microseconds obtained by calculating the intersection of the psychometric function with the detection threshold. The two different rows in the table indicate, if the target ITD was greater or smaller than the reference ITD.

curves given in Figure 4.12 by computing the intersection of the psychometric function with the detection threshold for deviation from chance performance. Since the psychometric functions were symmetrically measured around the reference ITD in the 'plausible ILD' condition ($\pm 1, 3, 5$ samples delay of the lagging ear), two thresholds for each angle of sound incidence are listed in Table 4.3.
Highest sensitivity to ITD changes can be observed for frontal sound in the 'plausible ILD' condition. The values for 0° azimuth are asymmetric, showing higher values for $\Delta\tau > 0$. An ANOVA on the individual data revealed that this effect is not significant (p=0.7).

The absolute size of the JND increases in the 'plausible ILD' condition by a factor of 3-4 as the angle of sound incidence is increased. An ANOVA was performed to assess the significance of the differences across angles. It shows, that the JNDs in the median plane are significantly different from the JNDs at lateral angles in the 'plausible ILD' condition ($p < 0.05$). No significant differences of the ITD JNDs were obtained for angles within the median plane (p=0.63). Furthermore, the differences in ITD JND at lateral angles are not significant. The tendency that the JND at 135° is higher than the JND at 45° is, therefore, also not significant (p=0.15).
In the 'constant ILD' condition the ITD JND at 135° deviates significantly from the ITD JND at the other azimuth positions (p<0.05).
In general, the detection rates in the 'plausible ILD' condition are smaller compared to the detection rates in the 'constant ILD' condition. Thus, two conclusions can be drawn from this finding. First, the ITD JND is influenced by the spectral shape of the stimuli and second, the ITD JND is decreased if additional spatial information is provided in the stimulus spectra. Hence, the results presented here support the hypothesis that the additional spatial information in the spectrum of the stimulus stabilizes the perceived location of the virtual object. In contrast, the alternative hypothesis that the additional spatial information generates a virtual object that is more focused in its spaciousness and that, therefore, spatial displacements are easier to detect, is not supported by the results of this experiment.

Comparison to the literature

To enable an comparison of the ITD JNDs obtained in the present study with data from the literature the ITD and ILD values of the reference stimuli are listed in Table 4.4. In the last row of the table mean values averaged across subjects are shown. For lateral source positions the ITD is increased to the range of approx. $500\mu s$ to $800\mu s$, whereas the reference ILD is increased to the range of approx 10 dB to 14 dB.
ITD JNDs were intensively investigated in the literature (s. (Durlach and Colburn, 1979) for a review). The JND of a 500 Hz tone was found to be around $10\mu s$ (Hershkowitz and Durlach, 1969) and for Gaussian noise between $12.3\mu s$ and $62.2\mu s$ depending on subjects (Mossop and Culling, 1995; Kinkel, 1990). To the knowledge of the authors, the lowest ITD JND was found for noise bursts to be approx. $6\mu s$ (Tobias and Zerlin, 1959).
In a study of Kinkel (1990) the ITD JND was measured for 1/3 octave bandpass noises with center frequencies of 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz as a function of the reference ITD and ILD. For zero ITD and ILD the obtained ITD JND is within the range of $20\mu s - 60\mu s$. An increase of the ITD and ILD reference values to $600\mu s$ and 15 dB, respectively, raised the ITD JND for most stimuli to the range of approx. $60\mu s$ to $160\mu s$. Only an increase of the reference ILD for the 500 Hz and 1 kHz stimuli did not result in a substantial increase of the ITD JND. Thus, on the average the ITD JND

is increased by a factor of approx. 3-4. This increase is consistent with the results of the present study. However, in this study a broadband noise was used. Therefore, a detailed comparison of the obtained ITD JNDs seems to be unappropriate.
The method used in this study limited the ITD JND to a minimum of $22\mu s$ due to the sampling rate of 44.1 kHz. The results from the literature show that lower JNDs have been found (e.g. (Domnitz, 1968; Hershkowitz and Durlach, 1969)). Therefore, the method was not appropriate for capturing the whole range of possible ITD JNDs. However, the results show that subjects were at or below the threshold in the most sensitive case. Hence, it can be assumed that for the stimuli used in this study, the obtained thresholds are representing the actual binaural temporal resolution.
In spite of differences in the methods, the ITD JNDs obtained in the current study are within expectations, both for the empirical and the flat spectrum stimuli.

Subjects \ Azimuth	ITD 0° ILD	ITD 45° ILD	ITD 90° ILD	ITD 135° ILD	ITD 180° ILD
RH	0 | 0.1	560| 11.2	800 |12.9	580| 7.8	0 | 0.7
IB	0 | 0.4	620| 13.0	840 |11.1	560| 6.6	0 | 0.6
HR	0 | 0.4	580| 12.8	820 |13.4	660| 10.5	0 | 0.5
HK	-60 | 3.3	460| 11.3	740 |14.0	640| 11.8	60| 0.0
JO	40 | 0.0	580| 11.1	780 |14.5	560| 5.5	-40| 3.1
MK	-40 | 0.3	500| 10.3	780 |16.0	620| 10.0	40| 0.5
Ø	23.3| 0.8	550| 11.6	793.3|13.7	603| 8.7	10| 0.9

Table 4.4: ITD and ILD of the individual reference stimuli. Dimensions of ITD and ILD are μs and dB, respectively.

4.7 Summary and general discussion

The general aim of the present study was to assess the sensitivity of subjects to deviations of the individual physical localization cues (described by HRTFs). Therefore, detection rates for manipulations of the individual HRTFs of 10 subjects were measured. Two kinds of spectral manipulations were applied to the HRTFs. The first manipulation (spectral detail reduction) is based on the work of Kulkarni and Colburn (1998). The findings of the present study are generally consistent with their results. A high amount of spectral detail can be removed from the HRTF spectra without affecting the perceived stimulus positions of a virtual stimulus (16 cestral coefficients were sufficient in the study of Kulkarni and Colburn and 16-32 were needed in the current study). For this amount of smoothing the frequency variation in the low frequency region is almost completely smoothed out. Hence, low frequency components seem to have a small contribution to the spatial perception because subjects were not able to detect a spatial displacement.

This finding is in contrast to the result from Asano et al. (1990) that the spatial information in the low frequencies (f<2 kHz) aid to resolve front-back confusions. It could be that due to the presentation of virtual stimuli already front-back confusion occurred and that subjects, therefore, were not able to detect changes in the stimuli. However, as pointed out before, the results of this study are consistent with the findings of Kulkarni and Colburn. In their study virtual stimuli were compared to a real sound source and, hence, front-back confusions introduced by smoothing would have been detected easily. Hence, it can be concluded that the lack of spectral information in the low frequency range does not affect the spatial perception at least for broadband stimuli.

The ILD deviations caused by cepstral smoothing at discrimination threshold are well correlated to the perceptual data, i.e. approximately the same ILD deviation in different situations (that roughly coincides with the ILD JND values from the literature) corresponds to the detected change in HRTF. Thus, the ILD deviation was used as a binaural distance measure for the differences of the physical localization cues. The results indicate that a mean ILD deviation (averaged across frequency channels of a Gammatone filter bank) of approx. 0.8 dB is detectable for sound incidence out of the median plane. This detection threshold is increased to 1.2 dB for sound incidence from lateral source positions.

The results given above are based on spatial displacements of the virtual stimuli because the source spectrum of the white noise was roved spectrally for each stimulus. The investigation was extended to non-spatial cues, like timbre, by presenting an unscrambled white noise and a click train stimulus to the subjects. Subjects were able to detect the manipulated HRTFs with higher detection rates compared to the scrambled white noise conditions for both, the white noise and the click train stimuli. Furthermore, the detection rates for click train stimuli were below the rates for white noise stimuli. This result is surprising because the spectral variation is the same for both stimuli. One hypothetical explanation is that in the click train condition subjects' attention was focused on the stimulus pitch of the click train which is introduced by the repetition rate of the clicks. This pitch is not changed by smoothing and, hence, subjects were less able to use the spectral variations as a detection cue. It can be concluded that for more complex stimuli than white noise (for instance, for music) the HRTF spectra can be smoothed by a higher degree without affecting the spectral timbre.

The detection thresholds for timbre variations, computed from the mean level differences of the smoothed and original HRTF spectra of the right ear, showed that for unscrambled white noise the monaural HRTF spectra may not deviate by more than 0.5 dB for sound incidence from the sides. An even lower threshold was computed for frontal sound incidence. For the click train condition a threshold could only be computed for 45° sound incidence (1 dB). In comparison to the threshold for the white noise stimulus it is increased by a factor of approx. 2.

In experiment II the sensitivity to a more complex spectral transformation was inves-

tigated that also shift the peaks and notches of the HRTF spectra. This was done by transforming the macroscopic spectral shape of the individual HRTF spectra to the shape of dummy head HRTF spectra ('spectral morphing'). As expected, subjects were very sensitive to the introduced manipulations. Again, the ILD deviation that is introduced by the 'spectral morphing' procedure served as a distance measure because a correlation analysis showed that the perceptual data is well correlated to this measure. The detection thresholds obtained from this distant measure are basically the same as for the spectral smoothing condition, if the scrambled white noise was presented to the subjects. For frontal sound incidence the thresholds are slightly lower than for the other directions. It can be assumed that the very obvious cue of front-back confusions aid to detect the manipulated HRTFs. This is likely because the 'spectral morphing' transformation distorts the spectral information (i.e. the center frequencies and the amplitudes of the peaks and notches) that is responsible for resolving the front-back confusions. In contrast, front-back confusions were not reported by the subjects if only the spectral detail of the HRTFs was reduced. However, although the front-back confusions introduce another detection cue, the thresholds described by the binaural distance measure (i.e. the deviation of the ILD that is introduced by 'spectral morphing') are almost the same for spectral detail reduction (for the scrambled white noise stimulus) and 'spectral morphing'. It can be concluded that the average ILD deviation across critical bands provides an appropriate measure for spatially relevant changes of the HRTF spectra.
In the last experiment presented in this study, the sensitivity to changes of the ITD was investigated. To investigate if the ITD JND is affected by the plausibility or consistency of the localization cues, two conditions were tested: First, detection rates for ITD variations of white noise stimuli convolved with individual HRIRs were measured. In a further condition, ITD JNDs were measured for white noise stimuli that exhibited the same ITD but had a constant ILD across frequency which is matched to the mean ILD (averaged across frequencies) of the individual HRTFs. Two hypotheses were tested to predict the differences of the detection rates. The first assumes that detection rates are higher for individual HRTFs because the virtual object is more focused in its spaciousness. The second predicts lower detection rates for the flat spectrum stimuli because the localization cues are less consistent and the virtual object is, therefore, less robust against distortions of one localization cue. The results showed that detection rates were *higher* for the less focused flat spectrum stimuli. Hence, more consistent localization cues seem to stabilize the virtual perception of a spatial acoustical object. This is a remarkable outcome since in traditional psychoacoustics it is assumed that interaural time discrimination is largely independent from object properties (such as, e.g. 'spatial diffusiveness'). For both conditions, the ITD JNDs calculated from the detection rates were within the expectations given by results found in the literature.

HRTFs for virtual auditory displays

The 'spectral morphing' procedure of the second experiment is of further interest. Using this method, perceptual relevant distances of individual HRTFs from different subjects can be described quantitatively. This can be done by calculating the morphing factor α, for which the ILD differences given by the distance measure D_{bin}, is above the appropriate detection threshold. For perceptually distant HRTFs low values of α are expected, whereas α is expected to be near to 1 for HRTFs that provide a similar spatial perception. Therefore, HRTFs can be grouped in perceptually similar HRTFs by using α as predictor for the perceptual distance.
However, the value α describes only perceptual distances of the ILD. The ITD is not taken into account by this measure. The results of the study show that thresholds for ITD deviations of non-individual HRTFs are well described by the threshold obtained in the literature. The ITD JND for empirical HRTFs is increased by additional localization cues and, therefore, the results for the ITD JND that can be found in the literature provide an lower limit for the ITD JND.

Chapter 5

Lead discrimination suppression in reverberant environments

Abstract

In reverberant environments the position of a sound source is dominated by the direct sound (the lead), whereas the spatial information of the reflections (the lag) is suppressed. Little attention has been paid in the literature to discrimination suppression of the direct sound in presences of a reflection and the results are not consistent. Thus, discrimination experiments were conducted to find out if in a natural listening environment the evaluation of the spatial information in the direct sound is processed in the same way as in a non-reverberant environment. The task of the subjects was to detect manipulations in the spatial information of the direct sound under reverberant and non-reverberant conditions. A 500 ms white noise stimulus was convolved with individual head related impulse responses (HRIRs) under the non-reverberant condition. In the reverberant condition binaural impulse responses of a seminar room (excluding the direct sound) were added to the HRIRs. Three manipulations were applied to the HRIRs: I) spectral smoothing, II) transformation of the macroscopic spectral shape ('spectral morphing') and III) ITD variations. The results show that for all three experiments the detection rate of the manipulations of the direct sound are significantly reduced under the reverberant condition. Thus, in a reverberant environment the contribution of the direct sound to the spatial perception is reduced. It is hypothesized that the lead discrimination suppression is due to further localization cues in the reflections that stabilize the perceived localization of the stimulus and make it more robust to changes in the direct sound. Due to the discrimination suppression in the spatial information of the lead less individual information is, therefore, needed in the direct sound in reverberant environments.

5.1 Introduction

One of the most important phenomenons of auditory localization for our daily life is the ability to localize the position of a sound source in a reverberant environment. The acoustical reflections produced by the environment are delayed, transformed (e.g. by absorption) copies of the original signal that are added to the direct sound in the ear canal. The sound originating from the source position is always leading the sequence of signals reaching the ear. This fact is used by the auditory system to localize the sound source position by giving precedence to the first wave front (see e.g. (Wallach *et al.*, 1949; Blauert, 1971)) and suppressing the spatial information of the lagging sounds. This effect is, therefore, called precedence effect and most often investigated by two stimuli measurement paradigms in which the reduction of spatial information in the second stimulus (the lag) in presence of the first stimulus (the lead) is investigated (e. g. (Wallach *et al.*, 1949; Perrott *et al.*, 1989; Haas, 1949; Zurek, 1980; Shinn-Cunningham *et al.*, 1993)). A comprehensive review of the precedence effect is given by Litovsky et al. (1999). From these investigations it is known, that the lag contributes only little to the perceived azimuth position of the sound but affects non-spatial cues like loudness and stimulus timbre (Blauert, 1974; Freyman *et al.*, 1998).

Nevertheless, the spatial perception is enhanced in reverberant environments. The energy ratio between the direct sound and the reflections is used by the auditory system as a cue for distance perception (e.g. (Bronkhorst and Houtgast, 1999)). Thus, the localization cue provided by reverberation differs considerably from those provided by the direct sound only. In a non-reverberant environment the direct sound recorded at the eardrum is equivalent to the head related impulse responses (HRIRs) convolved with the sound emanated from the source position. The HRIRs (or their frequency domain representations, the head related transfer functions (HRTFs)) contain all spatial information that can be used by the auditory system to estimate the source position. HRTFs describe binaural (interaural time difference, ITD and interaural level difference, ILD) and monaural (spectral filtering due to interference effects and pinna filtering) localization cues that can be exploited to calculate an estimate of the source direction. However, distance cues are only rudimentarily inherent in the head related transfer transfer functions (HRTFs) for source positions at a distance below 1 m especially in the median plane (Brungart and Rabinowitz, 1995). It is likely that the position of a sound source is determined by integrating over all available localization cues. This implies that redundant or additional localization cues increase the robustness of the spatial perception against distortions in one localization cue. Therefore, the sensitivity to variations of the localization cues is expected to be decreased in reverberant environments.

Furthermore, if the sensitivity to differences of individual HRTFs is reduced by adding reverberation, the amount of individual information needed in the direct sound is expected to be decreased. Thus, by incorporating reverberation to virtual auditory displays

not only the distance perception is enhanced but also the need for individual HRTFs to generate the direct sound is expected to be decreased. This would save costs and effort for the development of individual virtual environment generators.
The localization cues which can be extracted from HRTFs are ILD, ITD and monaural spectral filtering. Thus, to test the hypotheses given above three different discrimination experiments are conducted in which the detection performance of subjects to changes of the localization cues in the direct sound is compared under reverberant and non-reverberant conditions. The assumption is that if reverberation stabilizes the perceived position of an acoustical object higher variations of the localization cues can be introduced in the reverberant condition compared to the non-reverberant condition, without affecting the spatial perception.
To incorporate a test of the second assumption that less individual information is needed in the direct sound in a reverberant environment, two spectral manipulations of the HRTF spectra of the direct sound were chosen that reduce the amount of individual spectral information. In the first experiment cepstral smoothing was applied to the HRTF spectra to reduce the spectral detail. The investigation on the inter-individual standard deviation of the HRTF spectra across subjects given in Chapter 3 shows that the individual information in the HRTF spectra is reduced by cepstral smoothing. Hence, if a higher amount of spectral detail can be reduced in the reverberant condition without affecting the spatial perception, less individual spectral information is needed.
The second manipulation transforms the individual spectral shape of the HRTF spectra to the shape of dummy head HRTF spectra ('spectral morphing') which deviate strongly from individual ones (see Chapter 3). Again, if less individual information is needed in the direct sound in a reverberant conditions it is expected that more non-individual spatial information can be introduced to the stimuli without causing a spatial displacement.
In a further experiment the sensitivity to variations of the ITD is investigated under reverberant and non-reverberant conditions. In Chapter 3 it is shown that the inter-individual differences of ITDs obtained from different subjects averaged across source locations in the horizontal plane is approx. $40\mu s$. This values is within the range of the ITD JND (e.g. (Koehnke *et al.*, 1995)). If the ITD JND is further increased in reverberant environments it can be assumed that in this case individual ITD information is not needed for creating perceptually accurate virtual acoustic stimuli.

Related studies

Investigations that are related to the current study compare the absolute localization performance in reverberant and non-reverberant conditions or investigate discrimination suppression of the lead in presence of a lag.
In a study of Hartmann (1983) it was found that the absolute localization accuracy of a 500 Hz tone is *not* affected by changing the amount of reflections of a concert hall from

an absorbing condition to a reflecting condition. On the other hand, Begault (1992) observed that the localization acuity to speech stimuli created with non-individualized HRTFs *is* reduced if synthetic reverberation is added to the stimuli but the distance perception was enhanced. Thus, from absolute localization experiments conducted in the literature a clear picture concerning the differences in the localization accuracy under reverberant and non-reverberant conditions can not be extracted.
Although discrimination tasks are more sensitive to changes of the stimuli than absolute localization tasks, studies conducting discrimination experiments also do not show consistent results. In a study by Litovsky and Macmillan (1994) the change of the minimum audible angle (MAA) of the lead with and without the presence of the lag was investigated. No significant influence of the lag on the MAA of the lead was found. In a later study (Litovsky, 1997) a slight reduction of the MAA of the lead in the presence of the lag was observed. In this study longer stimuli were used and different groups of subjects with respect to their age.
In a study of Tollin et al. the ITD JND was measured for click stimuli with and without the presence of a lag. It was shown, that the ITD JND of the lead is increased by a factor of two if a lag was present. However, in a reverberant environment multiple reflections are following the direct sound. For distance perception it is likely that the auditory system averages across the first 6 ms (Bronkhorst and Houtgast, 1999). Therefore, it can be assumed that the decrease of the ITD JND in presence of a lag is higher for multiple reflections.

5.2 Methods

The general task of the subjects was to identify spatial displacements of virtual stimuli that were created by manipulated HRIRs convolved with a white noise stimulus. A two interval-two alternative forced choice (2I-2AFC) measurement paradigm was used for the experiments I-III. A stimulus sequence of four stimuli grouped in two intervals was presented to the subjects. The task of the subjects was to identify the interval in which one of the two stimuli deviated with respect to its spatial position.
In each experiment both reverberant and non-reverberant stimuli were presented in separate measurement sessions. The non-reverberant stimuli were created by applying individual HRTFs measured in an anechoic room to a reproducible scrambled white noise stimulus (see Chapter 3 for a description of the HRTF measurements).
For the generation of the reverberant stimuli non-individual binaural room impulse responses of an asymmetric seminar room were added to HRIRs by exchanging the direct sound of the room impulse responses with the HRIRs. The non-individual room impulse responses were measured for one selected listener (subject 'JO'). It can be assumed that the non-individual room impulse responses do not restrict the generality of the results

for the following reason: The reflections are spectrally filtered copies of the direct sound radiated from directions that primarily depend on the room and the orientation of the source relative to the listener. If non-individual room impulse responses are added to the direct sound, the reflections are filtered with non-individual HRIRs, which deviate in ILD and ITD in comparison to the corresponding values obtained from individual HRTFs. However, it is known from investigations on the precedence effect that ILD and ITD JNDs are increased for the lagging sound indicating that the sensitivity to individual cues in the reverberation process is greatly reduced. For instance, the ITD JND is decreased by a factor of 4-5 for reflections within the first five milliseconds (Tollin and Henning, 1998) and the MAA of the lag is increased by a factor of 2-6 depending on the time delay between lead and lag (Perrott *et al.*, 1989). Hence, it can be concluded that the differences in the localization cues introduced by using non-individual room impulse responses should not affect the results.

5.2.1 Subjects

A total number of six subjects (1 female and five male) participated in the experiments. The subjects were aged from 27 to 34 years and had normal hearing. The number of subjects participating in each of the three experiments is listed in the second row of Table 5.1. All subjects were members of the Physics and Psychology department of the University of Oldenburg and had extensive experience in psychoacoustic tasks. Each subject participated in both reverberant and non-reverberant experimental conditions. The author participated in all measurements.

	Exp I	Exp II	Exp III
Trials p. Cond.	40	30	24
Subjects	6	5	5
Sessions	4	6	6

Table 5.1: Number of trials per stimulus condition and measurement situation (row I), number of subjects per measurement condition (row II) and number of sessions (row III).

5.2.2 Stimuli

The same frozen white noise stimulus was used for all experiments. The noise sample had a duration of 500 ms. The on- and offsets were ramped by 5 ms squared cosine ramps. In the experiments I and II the spectrum of the white noise was scrambled randomly before it was convolved with the target or reference HRTFs. Scrambling was performed

in 1/6 octave bands by up to ± 5 dB. In experiment III the noise spectrum was left unchanged.
For each of the experiments I-III, anechoic and reverberant virtual stimuli were prepared. The first group consisted of the white noise sample convolved with manipulated HRTFs without reverberation. Under the reverberant condition reverberation was added to the manipulated HRTFs of the first group. After preparation of the target and reference HRTFs, they were convolved with a white noise sample.

5.2.2.1 Non-reverberant stimuli

Individual HRTFs were measured for each subject (see Chapter 3). Three different kinds of manipulation were applied to the HRTFs.

Experiment I: Reduction of the spectral HRTF details. The spectral detail of the HRTF spectra is reduced by cepstral smoothing. To smooth out the HRTF spectra the logarithm of the absolute HRTF spectra is reconstructed by a Fourier Series

$$log(|\hat{H}(k)|) = \sum_{n=0}^{M} \tilde{C}(n) \cos \frac{2\pi nk}{N} \tag{5.1}$$

where $\tilde{C}(n)$ can be obtained from the cepstrum $C(n)$ of the HRTF spectrum $H(k)$

$$C(n) = \sum_{k=0}^{N-1} \log |H(k)| e^{\frac{i2\pi kn}{N}} \tag{5.2}$$

$$\tilde{C}(n) = \begin{cases} \frac{(C(1)+C^*(1))}{2} & : \quad n = 0 \\ (C(n) + C^*(n)) & : \quad 1 <= n <= N/2 \end{cases}$$

The upper limit M of the series defines how many cosine terms are used for a reconstruction of the spectrum. If M equals $N/2$ (N is the length of the corresponding impulse response) no smoothing occurs. For $M < N/2$ cosine terms representing amplitude fluctuations of higher orders are neglected. Therefore, the spectrum is smoothed out by decreasing M.
The reference stimulus was created by using $M = 128$ coefficients to reconstruct the HRTF spectrum. Target stimuli had HRTF spectra with $M = 8, 16, 32, 64$ terms of the Fourier Series. The phase of each HRTF was calculated from $\hat{H}(k)$ as minimum phase plus a frequency independent group delay to incorporate the ITD.

Experiment II: Transformation of the macroscopic spectral shape ('spectral morphing'). The macroscopic shape of the target HRTF spectra was manipulated by transforming the individual HRTF spectra to the corresponding HRTF spectra of dummy head HRTFs. A description of the dummy head is given by (Trampe, 1988)).

This process is called *'spectral morphing'* throughout the study. It replaces the individual macroscopic spectral HRTF shape by the structure obtained from the HRTF of a dummy head. By Equation 5.3 the absolute spectrum of the individual HRTF $|H|$ is transformed into $|\hat{H}_\alpha|$. The parameter α describes the degree of morphing. $|H_{MS}|$ and $|D_{MS}|$ are representing the macroscopic spectral shape of the individual and the dummy head HRTF, obtained by 6^{th} octave smoothing. By increasing α from zero to one the proportion of the macroscopic dummy head spectra is increased. For $\alpha = 0$ $|H|$ equals $|\hat{H}_\alpha|$ and for $\alpha = 1$ the individual macroscopic shape is completely replaced by the dummy head shape.

$$|\hat{H}_\alpha| = (1-\alpha)|H| + \alpha|H|\frac{|D_{MS}|}{|H_{MS}|} \tag{5.3}$$

The reference HRTF was created by $\alpha = 0$ and the targets were calculated by setting α to 0.1-0.9 with $\Delta\alpha = 0.2$. The phase of the HRTFs is calculated from $|\hat{H}_\alpha|$ as minimum phase plus a frequency independent group delay.

Experiement III: ITD variation. In this experiment the interaural time delay between the left and right ear HRTFs was manipulated. The ITD of the reference stimuli were given by the ITDs of the empirically measured HRTFs. Targets were created by shifting the impulses responses of the lagging ear (left) by $\pm 1, 3, 5$ samples. Due to the sampling frequency of 44.1 kHz ITD variations of approx. $\pm 22\mu s$, $67\mu s$ and $110\mu s$ were introduced.

5.2.2.2 Reverberant stimuli

In each of the experiments I-III non-reverberant stimuli and reverberant stimuli were presented in separate sessions. The non-reverberant stimuli were noise samples convolved with the target or reference HRIRs as described before. Under the reverberant condition reflections were added to the HRTFs and then convolved with the noise sample. To illustrate the time pattern of the room reflections the envelope of the room impulse responses measured by microphones in the ear canals of subject 'JO' is shown in Figure 5.2. Each panel shows the first 40 ms of the impulse response measured in the left (thin lines, shifted in amplitude for visibility) and right ear canals (thick lines) for the source azimuth given in the panel (see Figure 5.1 for a sketch of azimuth positions in the room). It can be seen from this figure that the direct sound is clearly separated from the early reflections. The direct sound is located at approx. 6 ms at the right ear and shifted by the ITD at the left ear. At approx. 11 ms two first reflections separated by approx 1 ms, can be identified. Because the time delay between direct sound and the first two reflections is independent of azimuth, it is likely that these are reflections from

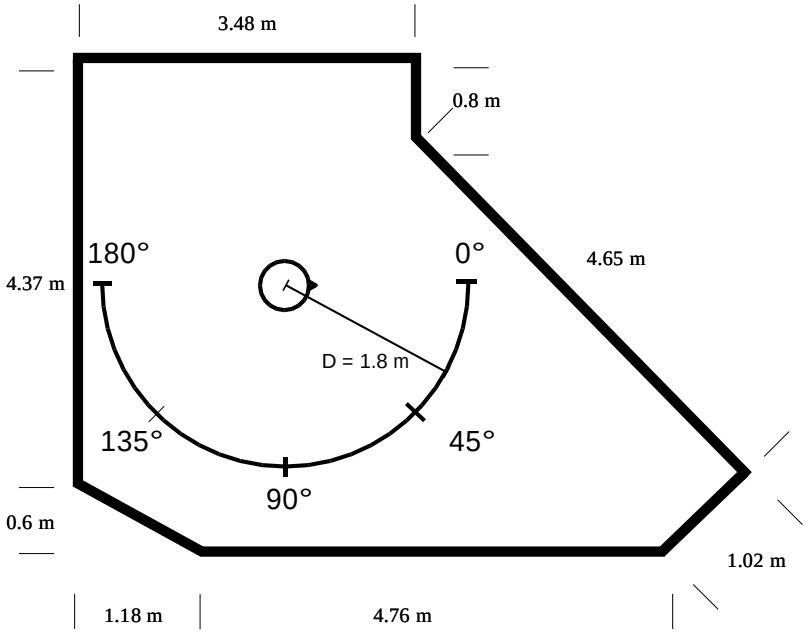


Figure 5.1: Floor plan of the room in which impulse responses were measured. The position of the center of the head was chosen by the restriction that a half circle with a radius of 1.8 m can be installed in the right hemisphere. Impulse responses were measured at the positions marked on the half circle.

the floor and the ceiling. Various reflections from different azimuths are succeeding the first reflections in intervals of 3 ms to 10 ms. For lateral angles, a prominent reflection at the left ear at approx. 12 ms can be identified. From Figure 5.1 it can be seen that this reflection is originated from the wall on the left side of the dummy head. After 40 ms late reflections evolve into a 'noisy' part of the impulse response (not shown here).
Target and reference stimuli in the reverberant condition were created by replacing the direct sound of the room impulse responses with the target and reference HRIRs, respectively. To give an example, the complete process of creating the reverberant stimuli in experiment I is described. First, HRTFs from all relevant azimuthal positions were measured for each subject individually in the anechoic room (see Chapter 3). Smoothed versions of the HRIRs were calculated from the HRIRs by applying cepstral smoothing to the HRTF spectra (see Equation 4.2 in Section 4). Then, room impulse responses were measured from a selected subject ('JO'). The speaker for obtaining the room impulse responses was located at the same azimuths as it was for the HRTF measurements. Subsequently, the direct sound of the room impulse responses was replaced by the previously obtained HRIRs. The reflections were scaled in amplitude in a way that the direct sound and the HRIRs of the right ear have the same RMS values. Preparing the stimuli in this way ensured that the HRIRs in the non-reverberant measurement condition and the direct sound in the reverberant condition were the same.

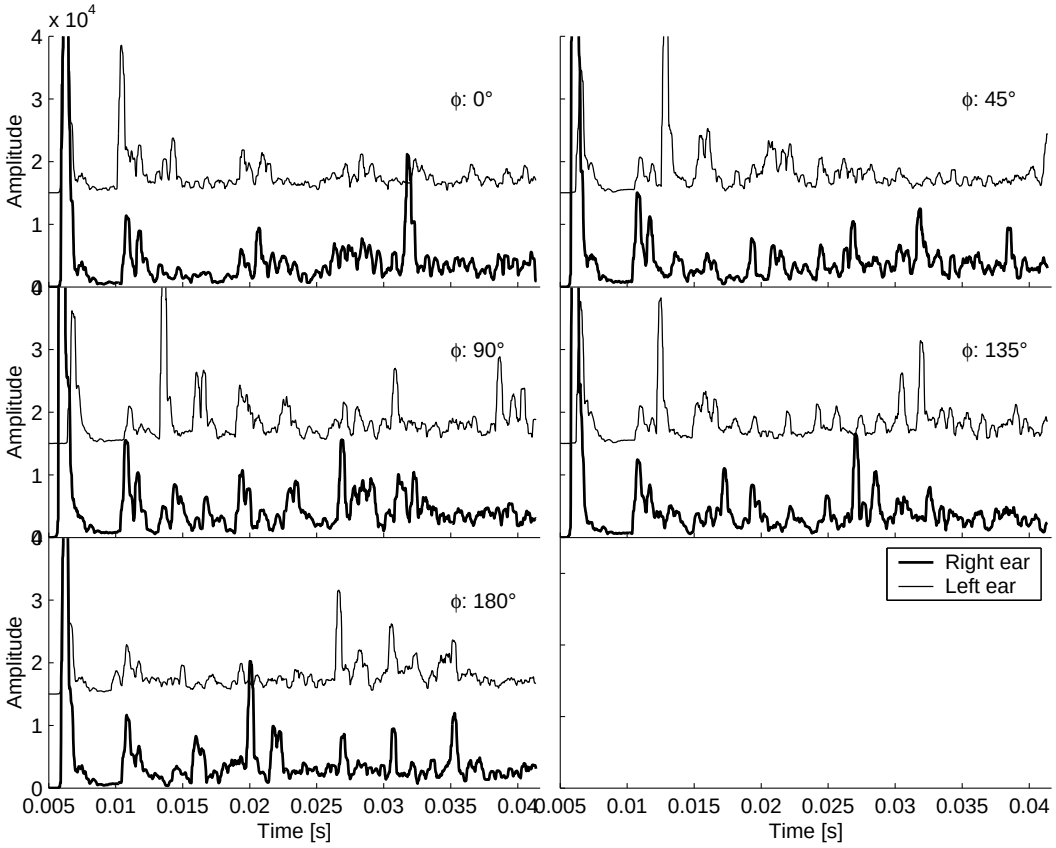


Figure 5.2: RMS values (averaged across 313μs time frames) of room impulse responses measured in the right (thick lines) and left (thin lines, shifted in amplitude for better visibility) ear canal. The sound source was positioned at the azimuth positions in the environment shown in Figure 5.1

5.2.3 Procedure

Subjects were seated in a sound isolated booth (IAC, Model No. 405A) in front of a window. The monitor of the computer controlling the experiments by running a MatLab script was located behind the window. Stimuli were presented to the subjects over a headphone (AKG 501) which was plugged into the output of a sound card (Soundblaster 128). The presentation level was set to a comfortable level for the subjects (approx. 70 dB A measured at the right ear of a dummy head for frontal sound incidence.).
For each trial a stimulus sequence consisting of four stimuli within two intervals was presented. One of the two intervals consisted of one reference and one target stimulus and the other interval consisted of two reference stimuli. The task of the subject was to identify the interval containing the target stimulus. The keyboard of the computer was used as input device. The position of the target within the stimulus sequence was chosen at random. Intervals were separated by 300 ms pauses and stimuli within intervals by 100 ms delays.

For each trial the horizontal location of the stimulus sequence was randomly chosen out of five different azimuth positions ($\phi = 0°, 45°, 90°, 135°, 180°$). The stimulus positions were the same for all three experiments. The number of stimulus repetitions per stimulus condition is listed in the first row of Table 5.1. In the second row the number of subjects participating in each experiment is shown and in the last row the number of sessions that each subject had to attend. Two experimental conditions were conducted in each experiment. In the first condition non-reverberant stimuli were used and in the second condition reverberant stimuli were presented to the subjects.

5.3 Results

In Figures 5.3-5.5 the results of the three discrimination experiments are shown for both stimulus conditions. The organization of the plots is the same for each of the three experiments. In each subplot the percentage of correct responses is shown as a function of the manipulation parameter for a different azimuth angle. Data for non-reverberant stimuli are represented by crosses and for the reverberant stimuli by open rhombi. In all conditions mean values across subjects are shown. The horizontal dashed lines indicate the 95% significance threshold for deviation from chance performance.
No standard deviations or error bars are shown in order to simplify the plots. To analyze the significance of the differences between the reverberant and non-reverberant conditions, a non-parametric ANOVA (Kruskal-Wallis) was computed. If the differences are significant ($p < 0.05$) a box plotted by dashed lines is enclosing the corresponding data points. For high significance ($p < 0.01$) the box is plotted by solid lines.

5.3.1 Experiment I: HRTF smoothing

In Figure 5.3 the results for detecting the target stimuli with smoothed HRTF spectra are shown for the reverberant and non-reverberant conditions. Percentage of correct responses are plotted as a function of the number of smoothing coefficients.
The figure illustrates that 16 ($\phi = 0°, 90°, 180°$) to 32 ($\phi = 45°, 135°$) cepstral coefficients are sufficient for providing all spatial information in the non-reverberant condition. Significant reductions of the detection rates occur for all angles of sound incidence in the reverberant condition. The detection rates are not significantly different from chance performance for all angles of azimuth, except for 135°. For this azimuth the detection rates are above the threshold for 8 cepstral coefficients.
The differences in the detection rates between the non-reverberant and the reverberant condition are significant for 8 cepstral coefficients for all angles of sound incidence. The largest differences occur for 135° azimuth were they are highly significant for 8 to 32 smoothing coefficients. To quantify the detection differences in the reverberant and non-

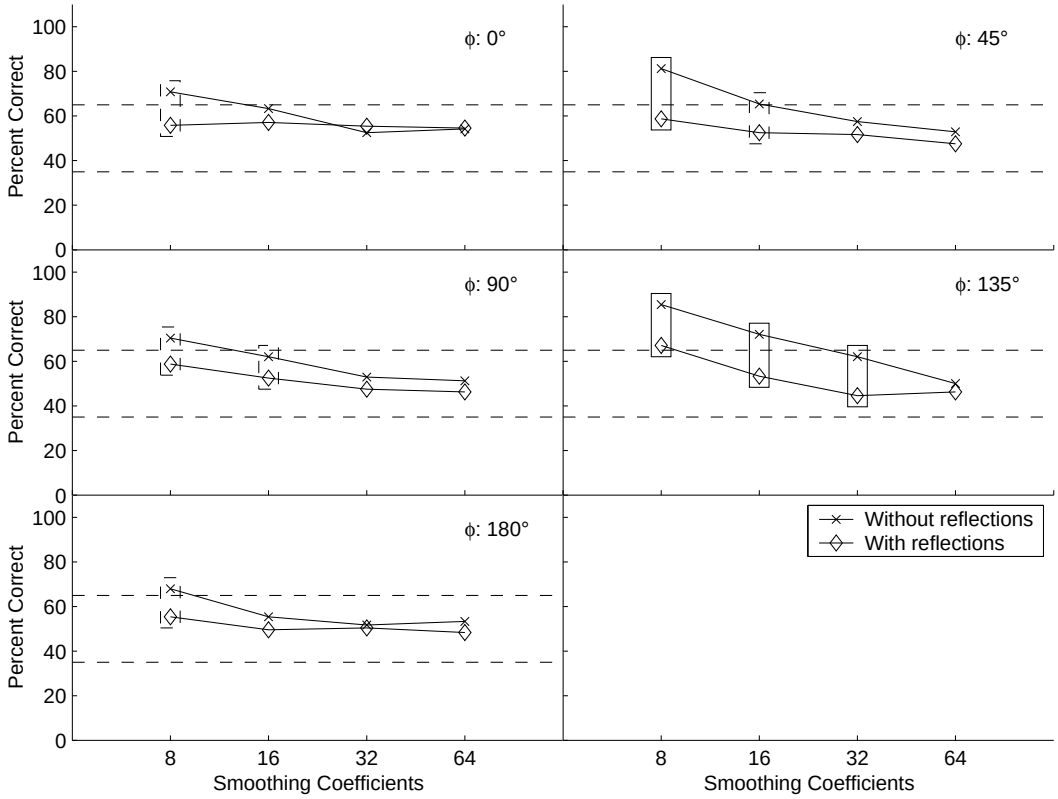


Figure 5.3: Detection rates for stimuli with smoothed spectra in reverberant (open rhombi) and non-reverberant (crosses) conditions. If the symbols are enclosed by a box the differences in the detection rates are significant (i.e. $p<0.05$) as indicated by dashed lines and highly significant (i.e. $p<0.01$) as indicated by solid lines.

reverberant condition, detection thresholds were computed and are listed in Table 5.2 for the reverberant (R) and non-reverberant (NR) condition. The thresholds are given in terms of the ILD deviation between the ILDs of the reference and target HRTFs. To calculate the thresholds, the psychometric functions were plotted as a function of the ILD deviation (averaged across frequency). The threshold was defined to be the ILD deviation for which the linear interpolation of the detection rates as a function of the corresponding ILD deviation intersects the 95% significance threshold for deviation from chance performance.

The ILD deviations were obtained by computing the ILDs of the target and reference

Condition \ Azimuth	0°	45°	90°	135°	180°
ILD deviation, R[dB]	>1.1	>2.1	>1.6	3.2	> 0.87
ILD deviation, NR[dB]	0.9	1.5	1.44	1.4	0.85

Table 5.2: ILD deviation thresholds for the detection of smoothed HRTFs in reverberant and non-reverberant conditions. If the detection rate was below threshold even for the strongest cue the thresholds are marked by a $'>'$ sign.

HRTFs in each filter bank channel of a Gammatone filter bank. ILD differences between target and reference stimuli were computed in each filter bank channel and averaged across frequency. This threshold was computed because the outcome of a correlation analysis was that the ILD deviation calculated in this way shows the highest correlation to the perceptual data in the non-reverberant condition (see Section 4.4).
If the detection rate is below the threshold for eight cepstral coefficients, the ILD deviation for this degree of smoothing is listed and marked by a $'>'$ sign to indicate that the threshold is above the listed value. Only for 135° of azimuth the detection rate is above threshold in the reverberant condition for eight cepstral smoothing coefficients. The threshold for this source direction is raised by a factor of two in this case. For the other angles of sound incidence it can be speculated that similar thresholds reductions occur.

5.3.2 Experiment II: Spectral morphing

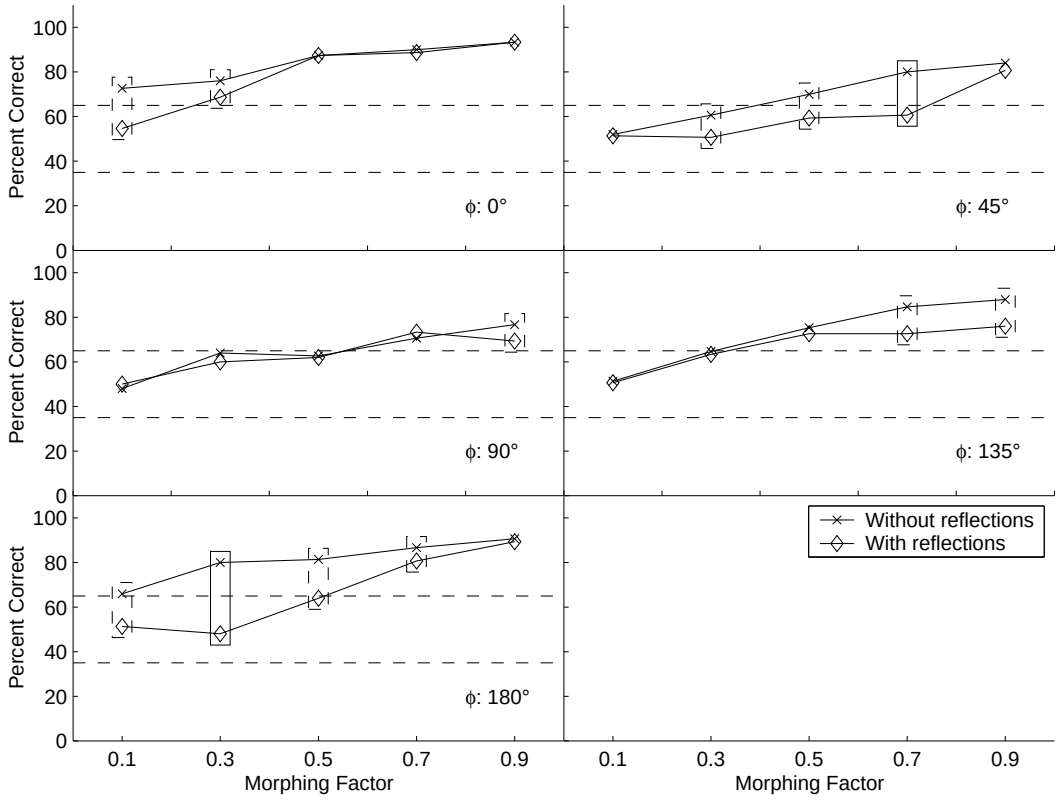


Figure 5.4: Detection rates for stimuli created with spectrally morphed HRTFs in reverberant (open rhombi) and non-reverberant (crosses) conditions.

The results of the 'spectral morphing' experiment are presented in Figure 5.4. The percentage of correct responses in the reverberant (rhombus symbol) and non-reverberant (crosses) condition are plotted as a function of the morphing factor α.

Condition \ Azimuth	0°	45°	90°	135°	180°
ILD deviation, R[dB]	0.75	2.2	1.51	1.63	1
ILD deviation, NR[dB]	<0.41	1.3	1.46	1.34	<0.32

Table 5.3: ILD deviation thresholds for the detection of spectrally morphed HRTFs in reverberant and non-reverberant conditions. If the detection rate was above the threshold even for the smallest cue the thresholds are marked by a $'<'$ sign.

It can be seen that subjects are highly sensitive to the 'spectral morphing' manipulation in the non-reverberant condition for sound incidence out the median plane (i.e., $\phi = 0°$ and $\phi = 180°$). The detection rates deviate from chance performance even for $\alpha = 0.1$. For lateral angles the sensitivity to the manipulation is reduced being lowest at 90° azimuth.
In the reverberant condition the pattern of the sensitivity as a function of source direction is changed. The lowest sensitivity can be observed at 45° azimuth and the highest for 0° and 135°.
Significant reduction of the detection rates (in comparison to the non-reverberant condition) can be seen for all angles of azimuth in the reverberant condition. The highest differences occur for $0°, 45°, 180°$ azimuth. However, the detection rates for 90° and 135° azimuth are nearly identical (i.e., at chance level for low values of α). Only at higher values of α significant differences can be seen. The thresholds listed in Table 5.3 were computed in the same way as in the HRTF smoothing experiment (s. Section 5.3.1). If the detection rate is not below the significance threshold (for instance at zero degree azimuth, non-reverberant condition) the ILD deviation for the lowest value of α is presented and marked by a $'<'$ sign. It can be seen that the detection thresholds are decreased by a factor of approx 1.7 for $\phi = 0°$ and 45°. For 180° of azimuth even stronger reduction of the sensitivity to the manipulation can be observed (> 3). As pointed out before no significant threshold differences between the reverberant and non-reverberant condition can be seen for $\phi = 90°$ and $\phi = 135°$.

5.3.3 Experiment III: ITD variation

In Figure 5.5 the results for the ITD variation experiment are shown. The number of correct responses in percent is plotted as a function of the ITD variation $\Delta\tau$. The crosses represent the non-reverberant condition and the rhombi represent the reverberant condition.
In general, the sensitivity to the ITD variation is reduced in the reverberant condition. For sound incidence out of the median plane the shape of the psychometric function is maintained, but the percent correct score is decreased by a nearly constant factor for all $\Delta\tau$. The differences between the percentage of correct responses in the reverberant and

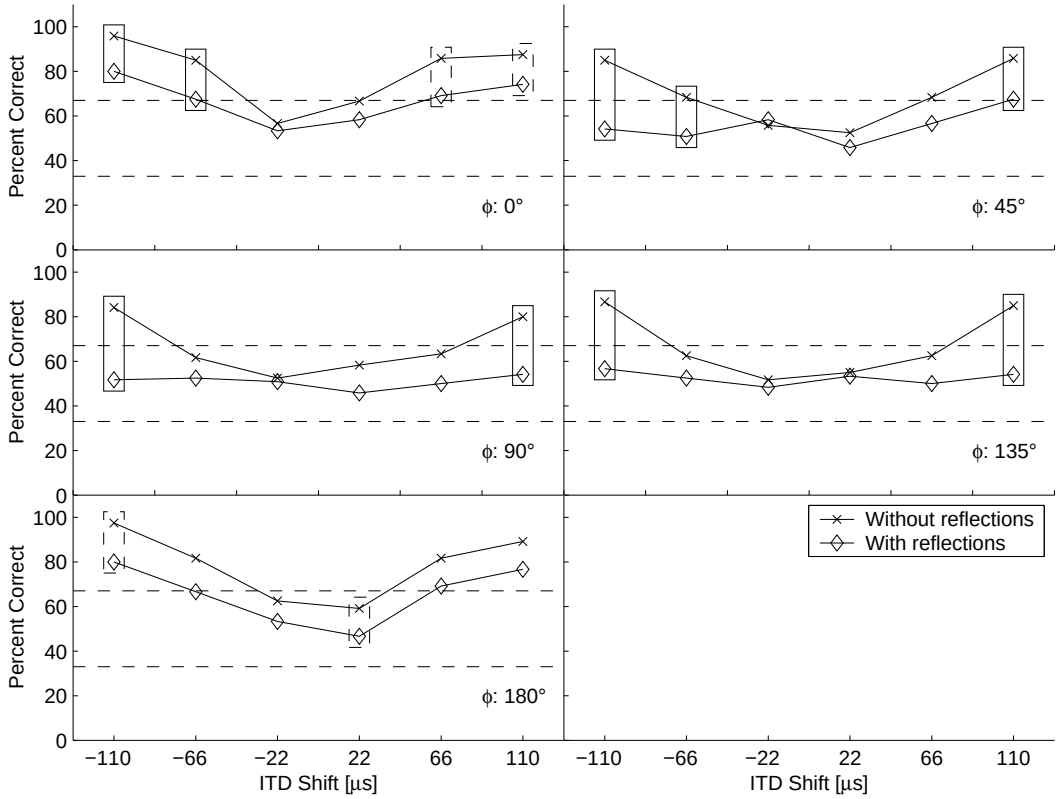


Figure 5.5: Detection rates for stimuli with shifted ITDs in reverberant (open rhombi) and non-reverberant (crosses) conditions.

non-reverberant condition at 0° of azimuth are significant for $\Delta\tau >= 66\ \mu s$. The reduction in sensitivity is similar at 180° compared to the frontal hemisphere and is significant for two introduced ITD variations. For lateral sound incidence ($\phi = 45°, 90°, 135°$) the sensitivity to the ITD variation in the reverberant condition is decreased below the significance threshold for all $\Delta\tau$. Only at 45° the average detection rate is slightly above the threshold for $\Delta\tau = 110\ \mu s$. The differences in percent correct responses in the non-reverberant and reverberant condition are highly significant for $\Delta\tau = 110\mu s$ at all lateral source positions. Detection thresholds were computed by calculating the intersection of the psychometric functions with the detection thresholds marked by the horizontal dashed lines. For the reverberant condition (R), this procedure was only applicable for source locations in the median plane. At lateral positions the detection rate is below the detection threshold. Threshold were calculated where possible and summarized in Table 5.4. Thresholds for source locations in the median plane are increased by a factor of approx. two in the reverberant condition.

Condition/Azimuth	0°	45°	90°	135°	180°
NR: $\Delta\tau < 0$	44	55	79	79	30
NR: $\Delta\tau > 0$	28	60	81	82	36
R: $\Delta\tau < 0$	66	> 110	79	> 110	69
R: $\Delta\tau > 0$	58	111	30	> 110	64

Table 5.4: Average detection thresholds of ITD variation manipulation are computed from the intersection of the psychometric function in Figure 5.5 with the 95% confidence level *NR* indicates the non-reverberant condition and *R* the reverberant case. Thresholds are given in μs.

.

5.4 Discussion

The general aim of this study was to investigate if the localization cues contained in the HRTFs (of the direct sound) are evaluated differently in a reverberant environment in comparison to a non-reverberant condition. Therefore, the sensitivity of three different types of manipulations were measured in both conditions. In the first experiment the spectral details of the HRTFs were reduced. The results show that the detection performance was reduced in the reverberant condition in comparison to the non-reverberant condition for all angles of sound incidence. Only at 135° azimuth the detection performance was above chance level in the reverberant condition for eight cepstral coefficients. For the other source positions no significant detection could be observed in the reverberant condition. For these source positions 4 or even 2 cepstral coefficients could be sufficient for providing all spatial information in the reverberant condition. The thresholds computed for 135° of azimuth showed, that the ILD deviation of the target to the reference ILD can be two times higher in the reverberant condition than in the non-reverberant condition.

The 'spectral morphing' experiment was intended to disturb the individual information in the macroscopic structure of the HRTFs. This was done by a stepwise transformation of the individual HRTF spectrum to the spectrum of a dummy head. The manipulation distorts the individual spatial information in the center frequencies of the peaks and notches of the HRTFs. The results of this transformation show, that for source positions in the median plane and for 45° azimuth the sensitivity to the manipulation is reduced in the reverberant condition. For 90° and 135° azimuth no reduction in threshold can be observed in the reverberant condition. Hence, the sensitivity of the auditory system to ILD changes (or changes to the spectral composition of the ILD) is not reduced due to reflections by the same amount for each angle of sound incidence. This result differs from the smoothing experiment, where the sensitivity was reduced for all angles of sound incidence.

In the third experiment the sensitivity to ITD variations was compared in reverberant and non-reverberant conditions. The sensitivity to the manipulation was found to be reduced significantly for all angles of azimuth. For source positions in the median plane the ITD JND is increased by a factor of two. If the sound is emanated from lateral positions, the detection rates for the target stimulus does not deviate from chance performance. However, the ITD JNDs are at least increased to $110\mu s$. This corresponds to an increase of the ITD JND by a factor of 1.4 for lateral positions. From these results it can be concluded that the detection thresholds are elevated by a factor of approximately two, both for the spectral variations and the variation of the ITD.

Relations to the precedence effect

To conclude, the sensitivity to changes in the binaural localization cues contained in the HRTF (the direct sound) is reduced in a reverberant environment. This can be explained by the following simple assumption. If the precedence effect would operate 'correctly' than the spatial information contained in the reflections would not disturb the localization perception. The spatial information would only be taken from the direct sound of the stimulus and, hence, the detection performance for reverberant stimuli would be the same as for non-reverberant stimuli. Because the sensitivity to the manipulations in the reverberant condition *is* reduced, it can be concluded that the precedence effect failes to operate 'correctly'. This means, that the early reflections influence the spatial perception of the direct sound to a certain degree.
In a study of Litovsky et al. (1994) no significant reduction of the minimum audible angle (MAA) of the direct sound (the lead) in presence of a reflection (the lag) has been found. MAA were measured in a single burst condition (without the presentation of the lag) and compared to the MAA of the lead in a lead-lag condition. Although slight differences occurred in the MAA, they were not significant. In this study very short noise stimuli (6 ms) were used. In a later study (Litovsky, 1997) the same stimulus conditions were compared to each other using longer stimuli and different groups of subjects (children and adults). Significant differences between the single burst and lead discrimination task were found for 25 ms noise bursts. In this case the MAA increased from 0.78° to 1.15°. This increase was even higher for children (1.55° to 4.4° for five year old children and 5.65° to 23.05° for 18 month old children). Hence, this investigation is consistent with the results of our study, showing that in a reverberant environment the sensitivity to changes in the direct sound is reduced.
In a MAA experiment both the ILD and the ITD of the target signal are varied. Hence, the reduction in sensitivity could be caused by a reduction of the ILD or ITD JND. To the knowledge of the author, the effect of the ILD JND increase in the direct sound in the presence of reflections has not been investigated yet by other studies. However, in a study of Tollin et al. (1998) the ITD JND of the lead was measured for precedence

effect stimuli (two clicks) and for single clicks. Threshold elevation factors (TEFs) were computed by calculating ratios of the ITD JND of the lead to the single click ITD JND. It was found that the TEFs increase as a function of the inter-click interval (ICI) separating the lead from the lag. For ICIs of 0.8 ms to 10 ms the TEF is approx. two. This indicates, that the ITD JND of the lead is two times higher in presence of the lag than without. This finding is confirmed by the results of experiment III, where the ITD is also increased by a factor of two in the reverberant condition (for sound incidence out of the median plane). It is remarkable that although realistic reverberation was used here instead of a single click in the study by Tollin et al. (1998) the reduction in sensitivity is the same in both studies. It can be concluded, that the effect is caused by early reflections of the room impulse responses and that the later reflections play a minor role.

Hypothesis: Perceptual stabilization by reflections

In a non-reverberant environment the localization acuity is determined by the spatial information contained in the HRTFs and the ability of the auditory system to extract the spatial cues from the HRTFs. In a reverberant environment the direct sound, (which is identical to the HRIR for ideal click stimuli) also determines the spatial perception of the source location. However, the results of this study show, that due to reflections of the sound by the environment, the sensitivity to HRTFs manipulations is reduced. This does not mean a priori that the localization performance in an absolute localization task is reduced by reflections.

Two different hypothesis can be supposed: First, it can be assumed that subjects were less sensitive because the spatial perception of the stimulus is in a way stabilized by the reflections. This means, that the reflections add information that can be used by the localization process to build the spatial object. The relative contribution of the HRTF information in the direct sound would be decreased and, therefore, the JNDs to HRTF manipulations would be increased.

The second hypothesis assumes that reflections with different azimuth position than the azimuth position of the direct sound confuse the auditory system and lead to a more fuzzy perception of the acoustic object. Manipulations to the HRTFs cues of the direct sound are then less detectable to the subjects in the reverberant condition, because the stimulus is less concentrated in its spaciousness. However, the results of the current study show, that the sensitivity to HRTF manipulations is approx. reduced by a factor of two. Translated to absolute localization experiments this would result in a localization blur being two times higher in the reverberant condition. To the knowledge of the author this has not yet been found by localization experiments. Hence, this hypothesis seems less plausible than the first hypothesis.

The first hypothesis is further supported by the anecdotal report of the subjects that in the reverberant condition front/back confusion in the 'spectral morphing' experiment

III were less often observed than in the non-reverberant condition. In order to further analyze the reason for the stabilization of the spatial percept due to reflections, the physical cues provided by the reflections have to be considered further. It is known from the literature that reflections add distance information to the stimulus (Békésy, 1938; Mershon and King, 1975; Sheeline, 1983) which is only basically inherent in the HRTFs in non-reverberant environments (Brungart and Rabinowitz, 1995). Furthermore, the first reflections can have the same source azimuth as the direct sound. This holds especially for reflections from the floor and a low ceiling as was pointed out before by Hartmann (1983). In this case, the reflections reinforce the information of the direct sound. Thus, manipulations in the localization cues of the direct sound are in a way corrected by the reflections. Hartman (1983) found that localization performance of a rectangular gated tone (500 Hz) was decreased for a higher ceiling compared to a lower ceiling. It was concluded that the reflections from the ceiling have the same source azimuth as the source and therefore facilitate the localization of the sound.
The azimuth direction of the first reflections reaching the ear of the listener, however, is depending on the geometry of the room and the orientation of the source and listener position within this room.
More information about the source position in the stimulus (provided by reflections) increase the robustness against distortions in the binaural cues of the direct sound. However, this argument would only hold true if the localization acuity is not dramatically reduced in a reverberant environment. In the study of Hartmann (1983) it was found that the localization accuracy of a 500 Hz tone was independent of the reverberation time of a room with variable acoustics. It was concluded, that for the different degrees of reverberation and absorption (7 dB difference in the level of the reflections) the localizations performance did not change. On the other hand, it was shown in a study of Begault (1992) that for speech stimuli with synthetic reverberation, the localization acuity was reduced compared to the non-reverberant condition, while distance perception was enhanced. Hence, although the stability argument seems to be plausible, no compelling data can be found in the literature that support this view.

Consequences for the use of HRTFs in reverberant environments

Independent from the hypothesis of how the reduction in sensitivity is caused by the reflections it can be concluded from the results of this study that less spatial information in the direct sound is needed in reverberant environments. Even for only eight cepstral smoothing coefficients the detection rate is below the threshold in the reverberant condition. From Chapter 3 of this thesis it can be seen that for eight cepstral coefficients the inter-individual standard deviation of the HRTF spectra across subjects is reduced. It can be concluded, that less individual information in the HRTFs is needed, if reverberation is added to the HRTFs of the direct sound. This is supported by the results of

experiment II. For three of five source positions the sensitivity to the 'spectral morphing' procedure is reduced in the reverberant condition. However, for higher values of the morphing factor α, the detection rate is above the threshold. Therefore, although the sensitivity of subjects to individual information is reduced, the reduction is not sufficient for using dummy head HRTFs without a change in the spatial perception.
The investigations of the HRTFs presented in Chapter 3 show that the spectral difference between subjects are smaller than the differences between a subject and the dummy head. Therefore, it can be assumed that the reduction in sensitivity to the spectral cues is sufficient for using non-individualized HRTFs. This can be investigated by using non-individualized HRTFs for the spectral morphing' procedure rather than dummy head HRTFs. If the detection performance is below the threshold even for higher values of α this would indicate that in reverberant conditions individual spectral information is not needed. However, this has to be investigated.
The sensitivity to ITD variations is also reduced in reverberant environments. From the investigation in Chapter 3 on the ITD standard deviation across subjects ($\bar{\sigma} = 40.1\ \mu s$) it can be seen that the differences between subjects are within the dimension of the ITD JND estimated in experiment III. Hence, in reverberant conditions the need for individual ITD information in the direct sound is reduced.

5.5 General conclusion

The results of this study can be summarized as follows.

- Sensitivity to manipulations of the ILD and ITD is reduced by a factor of approx. two in reverberant conditions
- Therefore, less individual spatial information is needed in the direct sound of a stimulus compared to non-reverberant environments
- The reduction in sensitivity could be caused by additional localization cues provided by the reflections that enhance the robustness against distortions of the spatial information in the direct sound.

Chapter 6

Spatial elevation perception of a spectral source cue

Abstract

Spectral scrambling is applied to the spectrum of the stimulus in localization experiments to prevent the subject from using spectral timbre variations as a cue. It has not been investigated yet, if the spectral scrambling introduces a localization cue that affects the apparent stimulus position. The spectral scrambling of the source spectrum could introduce a monaural cue that influences the elevation perception. Therefore, in the experiment presented here, the influence of a spectral cue in the source spectrum on the perceived elevation was studied for a noise stimulus (500 ms length) that is projected to the horizontal plane by using virtual acoustics. The spectrum of the source sound contains a monaural spectral cue that points to an elevation in the range of $-40°$ *to* $60°$*. The task of the subject was to judge the perceived elevation in an absolute localization paradigm as a function of the spectral cue in the source spectrum. The results show that the spectral cue in the source spectrum significantly influences the perceived elevation with a maximum effect of* $20°$*. Hence, there is a need for developing scrambling methods that only change the perceived timbre but not the perceived localization of a given sound.*

6.1 Introduction

The localization performance of the auditory system of human subjects is normally measured by presenting a sound source at a certain stimulus position and asking the subject to report the perceived source location (see Chapter 2.3). To estimate the source position of the stimulus the subjects can used binaural cues (interaural time differences, ITD and interaural level differences, ILD) as well as monaural spectral cues that are introduced by interference effects and pinna filtering. However, the monaural cues can

serve as a spectral timbre cue that could be used by subjects to learn the timbre that corresponds to a certain stimulus position. To prevent the subjects from using timbre cues for the identification of the stimulus position, the spectrum of the sound source is often randomly scrambled in a certain level range before the stimulus is presented to the subjects.

Spectral scrambling has been used in a variety of localization studies. For instance, Wightman and Kistler varied the spectral amplitude of the source stimulus in critical bands by up to 20 dB in absolute localization experiments (Wightman and Kistler, 1989b; Wightman and Kistler, 1992; Wightman and Kistler, 1997). The same manipulation of the source spectrum was used by Wenzel et al. (1993). Kulkarni et al. scrambled the source spectrum in 1/3 octave bands by up to $\pm 5\ dB$ to prevent the subject from using non-spatial cues in a real/virtual source discrimination task. Langendijk and Bronkhorst varied the stimulus spectrum in 1/3 octave bands in order to investigate if subjects are able to virtual stimuli generated with interpolated HRTFs (Langendijk and Bronkhorst, 2000). In the experiments described in Sections 4 and 5 the stimulus spectrum was scrambled in 1/6 octave bands by up to $\pm 5\ dB$.

However, the spectral shape, that is introduced to the source spectrum by scrambling, could contain spatial information that could be processed by the auditory system. Therefore, the perceived stimulus position could change depending on the scrambled source spectrum. Hence, to quantify the bias that could be introduced by spectral scrambling, the affect of a monaural spectral cue in the spectrum of a broadband stimulus is investigated in the current study by performing an absolute localization experiment.

It is well known that narrow band stimuli can cause confusions to the auditory system. The apparent source position can be determined by the center frequency of the narrow band sound independent of the actual source position. For instance, Blauert (1969) found that the perceived source position of 1/3-octave band noise signals, emanated from locations in the median plane, is only determined by the center frequency of the stimuli. Butler and Helwig (1983) varied the center frequency of 1 kHz wide noises and showed that the perceived spatial position in the median plane goes from front to back, as the center frequency is increased from 4 kHz to 12 kHz. Musicant and Butler (1985) showed that the center frequencies of 1 kHz wide noise also determine the perceived localization in the horizontal plane, where binaural cues are usable for the subjects.

The physical properties that relate to the judged locations of the narrow band stimuli can be found in the head related transfer functions (HRTFs). They describe the directional dependent transformation of a sound from its source location to a point in the ear canal. The HRTFs for the judged locations have peaks at frequencies that correspond to the center frequencies of the narrow band signals. Depending on studies these peaks are called 'boosted bands' (Blauert, 1969), 'covert peaks' (Flannery and Butler, 1981; Musicant and Butler, 1984) or 'proximal stimulus spectra' (Middlebrooks, 1992). Thus, salient peaks in the spectra of a stimulus bias the apparent stimulus location to the po-

sition for which the HRTF spectra have a peak in the corresponding frequency range. It is likely that the spectrum of scrambled broadband stimuli has prominent peaks in some frequency bands. Therefore, the auditory system could relate the peaks introduced in the source spectrum by spectral scrambling to spectral filtering by the HRTFs. As a result the apparent position of the stimuli would vary randomly for each scrambled stimulus spectrum. Although this explanation seems to be plausible, it has not been investigated in a systematic way if monaural source cues in a broadband stimulus spectrum affect the spatial perception.

Hence, in the study presented here it is investigated if the elevation perception of a broad band stimulus is affected by a monaural cue in the source spectrum. A noise stimulus is randomly presented from one of five different azimuth positions in the horizontal plane by using the methods of virtual acoustics (e.g. (Wightman and Kistler, 1989a; Hammershoi, 1995)). Before the virtual stimulus is generated the spectrum of the noise is multiplied with the spectrum of a HRTF of one ear measured at the same azimuth position. However, the elevation of the HRTF spectrum was chosen from $\theta = -40°, -20°, 0°, 20°, 40°, 60°$. Hence, a broad band stimulus is projected to the horizontal plane that contains a monaural source cue that points to a different elevation at the same azimuth. The task of the subjects was to judge the source location as a function of the monaural source cue in an absolute localization task. If the perception of the stimulus is independent from its spectrum then the subjects should localize each stimulus in the horizontal plane. If, however, the monaural source cue influences the perception of the stimulus, the judged elevation should increase as the elevation of the monaural source cue is increased. Two experiments were conducted. In the first experiment the monaural spectrum of the left ear HRTF was applied to the white noise source and in the second experiment the monaural spectrum of the right ear HRTF was applied to the stimulus spectrum.

6.2 Method

Subjects

Six subjects (four male and two female) aged from 28-35 participated voluntarily in the localization experiment. All had normal hearing and extensive experience in psychoacoustic tasks. They were members of the physics and psychology departments of the University of Oldenburg.

Stimuli

A catalogue of individual HRTFs measured at a high number of source positions was recorded in a separate session (see Chapter 3 for a description of the measurements). All

HRTFs needed to generate the source stimuli were taken from this database.
A frozen white noise sample (500 ms length, ramped with 5 ms squared cosines) was used as a source stimulus. The spectrum of the noise was multiplied with the spectrum of the individual HRTF of one ear at azimuth ϕ and elevation θ. The spectrally transformed noise sample was then convolved with the left and right ear HRTFs measured at the same azimuth ϕ but the elevation was set to the horizontal plane. Hence, the monaural information in the source spectrum points to a different elevation than the HRTFs used to project the noise sample to the horizontal plane. The azimuth positions were chosen from $\phi = 0°, 45°, 90°, 135°, 180°$. The HRTF spectra for the monaural source cue were obtained from the elevations $\theta = -40° - +60°, \Delta\theta = 20°$ at each of the listed azimuths. Two sets of stimuli were created. In the first set, the monaural elevation cue of the left ear was used to transform the source spectrum (condition ML) and in the second set the monaural cue of the right ear (condition MR) was applied. No headphone correction was performed. To illustrate the spectra multiplied to the spectrum of the sound source

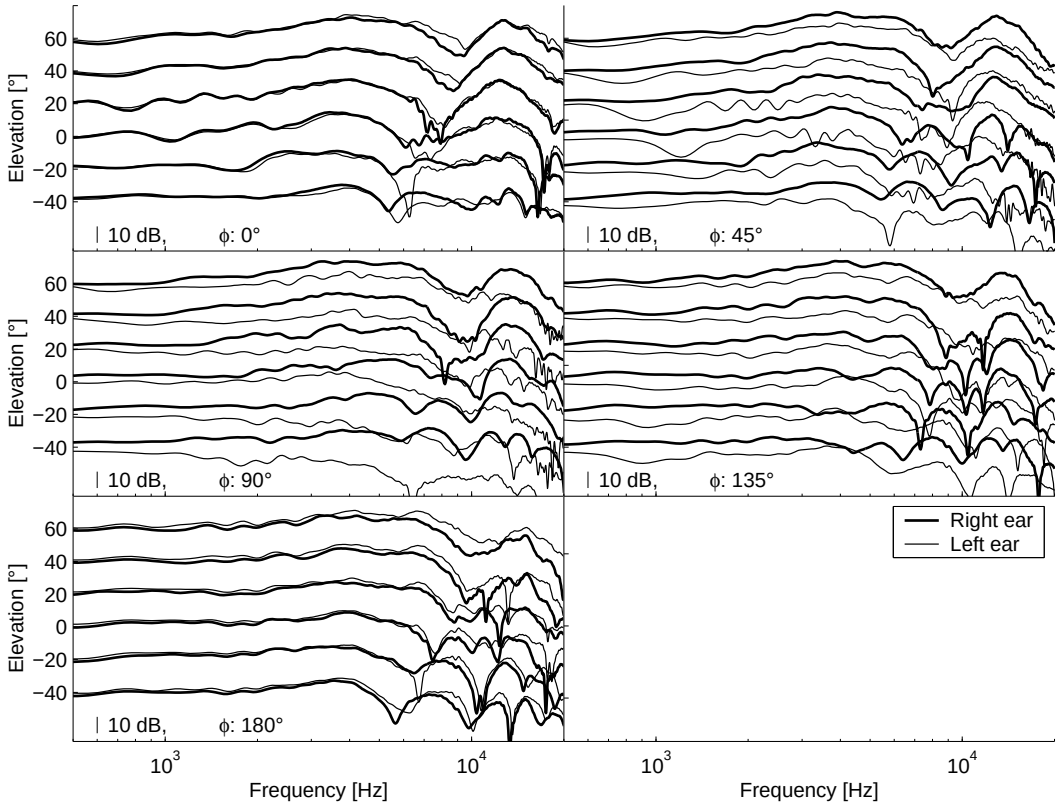


Figure 6.1: HRTF spectra of one subject used for the transformation of the source spectrum for this subject. The thick solid lines show spectra for the right ear and the thin solid lines are representing the left ear.

(in order to generate the monaural source cue), in Figure 6.1 the HRTF spectra of one subject are shown. In each subplot HRTFs at azimuth ϕ for the elevations $\theta = -40° - +60°, \Delta\theta = 20°$ are presented. Right ear HRTFs are plotted by thick solid lines

and left ear HRTFs by thin solid lines. The spectra show the typical shape of HRTFs as a function of azimuth and elevation.

Procedure

The localization experiments were conducted in a sound isolated booth (IAC 405A). The subjects were seated on a small chair in front of a window. The IBM compatible computer that controlled the experiments was located outside the room behind the window. A WinShell batch program controlled the stimulus presentation and the recording of the subjective data.
The stimuli were computed off-line and stored on the harddisc of the computer. An AKG 501 headphone served to present the stimuli to the subject. It was directly plugged to the output of a SoundBlaster 128 sound card. The presentation level was approx. 70 dB(A) (measured at the right ear of a dummy head for frontal sound incidence). After stimulus presentation the subjects recorded the perceived stimulus location by using the GELP technique (see Chapter 2 for a comprehensive description of the GELP technique). The recording device consisted of a sphere with a diameter of 30 cm located in front of the subject. The subject had to point to a location on the sphere that corresponds to the perceived stimulus position, as if the subject would be sitting in the center of the sphere. A Polhemus Inside Track was used to capture the position of the input device (a receiver of the Inside Track) on the spherical surface. The computer recorded the position of the receiver on the surface when it was placed there for at least one second. To acknowledge the recording, a brief signal was presented to the subject by headphone.
The azimuth position as well as the elevation of the monaural source cue was chosen at random for each trial. Each stimulus condition was repeated 10 times.
Subjects conducted two separate sessions. In the first session the source spectrum was transformed by the left ear HRTF spectrum (condition ML) and in the second session the source spectrum was transformed by the right ear spectrum (condition MR).

6.3 Results

An inspection of the individual responses revealed that the subjective judgments for each azimuth show an individual bias in elevation, either to lower or higher elevations, depending on the subject. This bias might be due to the monaural elevation cue in the source spectrum that confuses the auditory system and leads to an inaccurate elevation perception. In order to eliminate the bias, it was individually subtracted from the localization data before averaging across subjects. Note, that the elimination of the bias does not affect the shape of the curves presented in Figure 6.2 but the absolute position of the curves is shifted vertically. Furthermore, the inter-individual standard deviation is

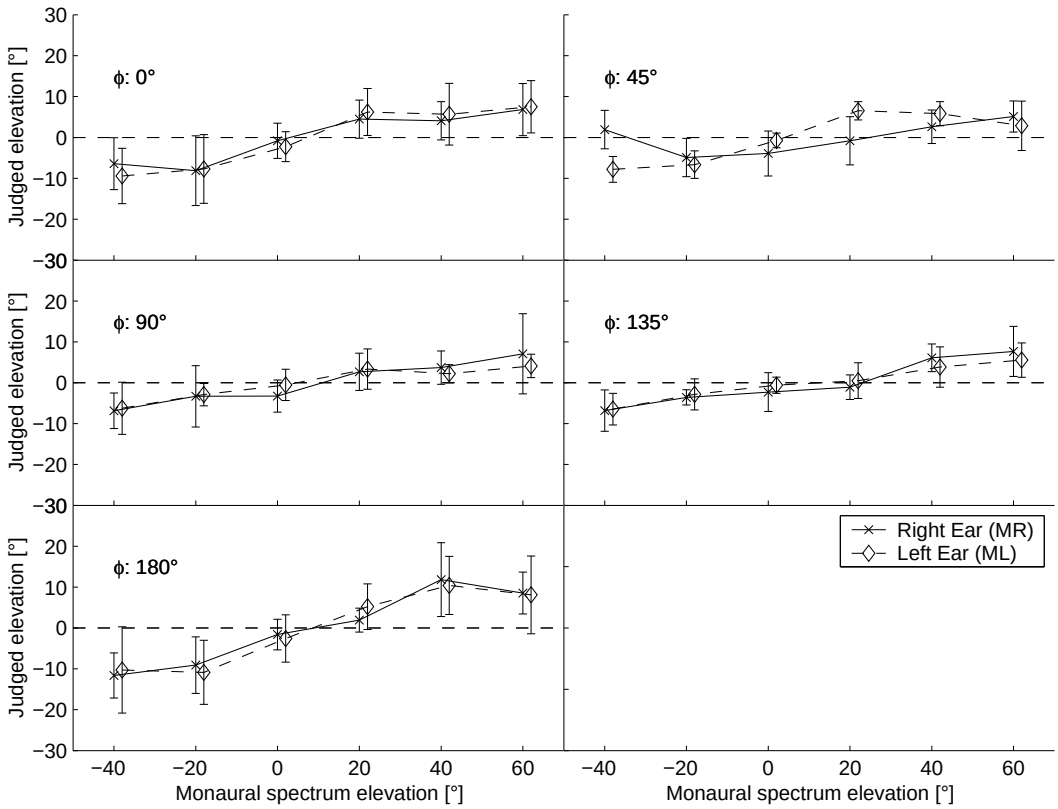


Figure 6.2: Perceived elevation as a function of the monaural source cue for different angles of azimuth averaged across subjects. Elevation judgements for monaural source cues taken from the right ear are connected by solid lines and by dashed lines for monaural cues from the left ear HRTF. The error bars indicate the inter-individual standard deviation.

reduced. The amount of this bias is listed in Table 6.1 for each subject individually. The bias was eliminated because the absolute localization error in elevation was less relevant for the present study than the change in elevation perception for stimuli with different monaural source cues.

The results of the localization experiment are summarized in Figure 6.2. Each subplot shows the mean perceived elevation averaged across subjects after removing the bias as a function of the source cue elevation. The solid line shows data for the MR condition (right ear HRTF spectrum) and the dashed line shows data for the ML condition (left ear HRTF spectrum).

The shape of the curve is nearly identical for all azimuth positions. The lowest perceived elevation angle is approx. $-10°$, increasing as the source cue elevation is increased from $-20°$ to $60°$. At $180°$ azimuth no further increase of the perceived source elevation can be observed for a source cue elevation increasing from $40°$ to $60°$. The maximum range of the perceived elevation is approx. $20°$, with a trend to a smaller range at lateral azimuth positions compared to positions in the median plane.

Subjects\Condition	ML					MR				
Azimuth	0°	45°	90°	135°	180°	0°	45°	90°	135°	180°
HK	23	43	16	-13	-23	24	27	4	-4	-13
JD	-3	2	-1	-1	3	4	-2	-3	2	3
HR	4	30	3	1	19	0	15	-7	-8	5
RH	3	4	-1	7	9	10	-2	-8	6	27
MK	10	19	8	-4	6	43	28	-5	5	50
IB	-1	9	3	-1	-7	5	3	-6	-10	-1

Table 6.1: The bias in degree (i.e. the mean localization error in elevation) is listed for each subject and each angle of azimuth for both measurement conditions.

The shape of the curve for the MR and ML condition is nearly identical. Only at 45° of azimuth the left ear source cue and the right ear source cue provide a different perception of the stimulus elevation.
The significance of the perceived differences in stimulus elevation as a function of the source cue elevation was analyzed by a non-parametric ANOVA (Kruskall-Wallis). The null hypothesis was that the judged elevation for a monaural source cue from $-40°$ is equal to the judged elevation for a monaural source cue, obtained from higher elevations. The analysis reveals that the elevation judgements for a monaural source cue from 20° elevation are significantly different from the $-40°$ elevation for all angles of azimuth ($p < 0.05$), except for 135° azimuth. However, at this azimuth the mean judgements between $-40°$ and 40° are significantly different ($p < 0.05$).
The differences in the judged elevations can be related to differences in the monaural source cue. This is illustrated in Figure 6.3. Here, the absolute level differences between the HRTFs spectra at $-40°$ and higher elevations are shown separately for the left (thin solid lines) and right (thick solid lines) ear. Each subplot shows level differences in dB averaged across subjects as a function of frequency for one azimuth.
The variation of the monaural spectra as a function of elevation is concentrated on the frequency range around 6-8 kHz for azimuth locations in the median plane ($\phi = 0°, 180°$). The level differences in this frequency band increase steadily for increasing elevations. Only minor effects can be seen in other frequency areas. Due to the symmetry of the head with respect to the median plane, similar differences for both ears can be observed. At 45° of azimuth an increase of the source elevation raises the level in the 6-8 kHz area of the left ear. Smaller changes in the same frequency region can be observed for the right ear. Compared to the other azimuth angles the variation of the right ear HRTF as a function of source elevation is rather small and can, thus, explain the deviating elevation perception for this direction (Figure 6.2, upper right panel, solid line). For azimuth positions at 90° and 135° the main effects are a general increase in the level of the left ear. Furthermore, an increase of the level in the frequency range around 9 kHz

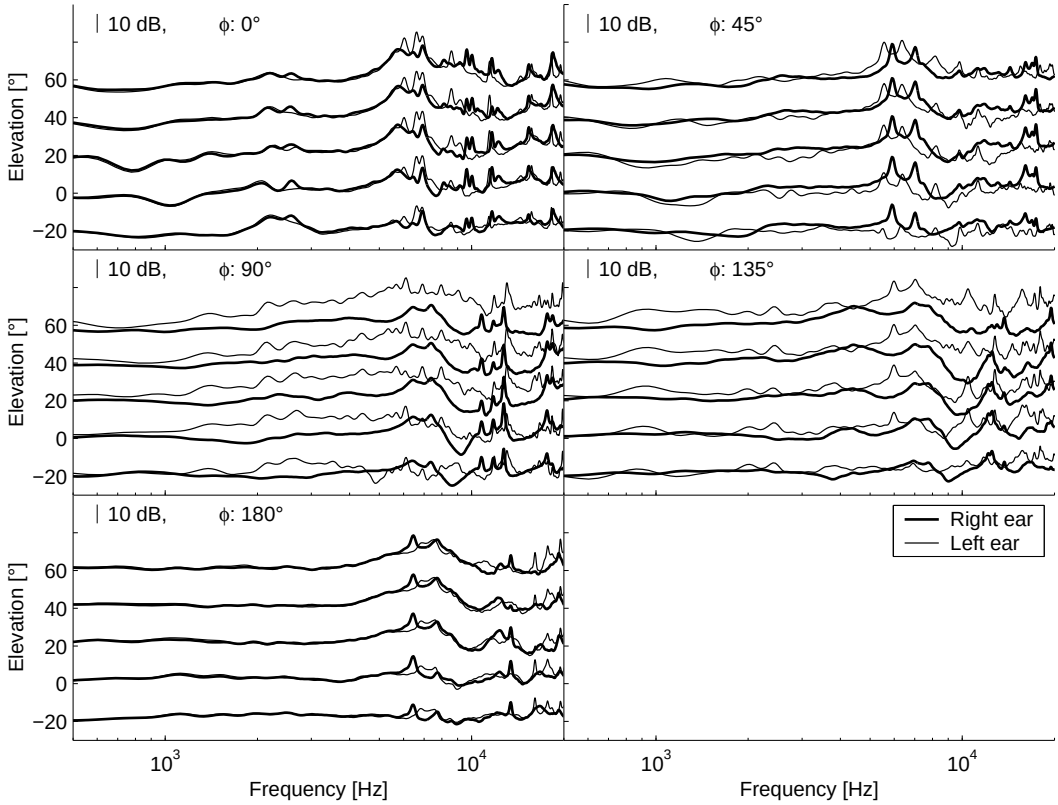


Figure 6.3: Level differences between the HRTF spectra at −40° elevation and higher elevations for different angles of azimuth are shown. Level differences for the right ear are plotted by thick solid lines, and for the left ear by thin solid lines.

at the right ear is introduced by higher source elevations.

6.4 Discussion

The results of the localization experiment show that a monaural cue in the spectrum of the source can significantly influence the elevation perception of a broad band stimulus. The monaural cues taken from the left and right ear HRTFs of sources at −40° to 60° elevation, modify the perceived elevation in a range of approximately 20° with a tendency of a smaller range for lateral source positions.

The analysis of the physical cues that caused the different elevation judgements shows that for most azimuth angles of sound incidence level changes occurred primarely in the frequency band from 6-8 kHz. The range of the level differences in theses frequency bands is approx. 10 dB. Level differences in a broader frequency area can only be observed in the HRTF spectra of the contralateral ear at 90° and 135°. The frequency band around 8 kHz was specified by Blauert as a boosted band for the above direction (Blauert, 1969). The results of the analysis of the physical cues confirms the finding that an increase in the level in this band increases the perceived stimulus elevation. However,

a general level increase in a broader frequency range, as observed for the left ear at 90° and 135°, also seems to cause an increased elevation perception.

The general intention of this study was to asses the effect of spectral scrambling on the elevation perception of a stimulus. Most methods for scrambling the source spectrum rove the spectrum in a range ($> 10\ dB$) that seems to be sufficient for producing peaks in the source spectrum yielding a variation of the perceived elevation. Furthermore, the spectral source cue in the current study only introduced level differences in the 'above' band. It is likely that the random level variation in the source spectrum introduced by spectral scrambling also introduces peaks and notches in other directional bands. Therefore, it can be assumed that spectral scrambling causes further variations of the perceived stimulus location, for instance, in the front-back dimension.

The use of virtual acoustics for the presentation of stimuli does not necessarily limit the generality of the results. It can be assumed that an equivalent experiment in the free-field would show similar results. However, it is possible that even small rotations of the head during stimulus presentation could lead to a breakdown of the effect. If the head position is changed during the stimulus presentation, the auditory system could estimate the source spectrum by computing the amplitude ratio of the stimulus for two different head positions. This would enable the auditory system to distinguish between the source spectrum and the spectral transformation of the HRTFs.

Therefore, it would be of interest to investigate if the change in the perceived elevation as a function of the monaural source cue can also be observed for stimulus presentations in the free-field. However, the head should be fixed by using a bite bar to provide an experimental condition that is equivalent to the virtual stimulus presentation.

The results of this study have consequences for absolute localization experiments as well as for localization discrimination tasks. If the source spectrum is scrambled before each stimulus presentation in an absolute localization experiment, the perceived source location may subjectively change as a function of source scrambling. An increased variabiliy of the results of localization experiments are in this case not only introduced by the localization uncertainty but also by the spectral scrambling of the source sound. Hence, the localization uncertainty measured by scrambled sound stimuli will never be smaller than the spatial variation of the sound source as a function of scrambling.

In spatial discrimination tasks subjects are asked to identify a target stimulus with a spatial displacement relative to a reference stimulus. To avoid the subjects from using non-spatial cues, both the target and reference stimulus are scrambled in their spectral content. Again, scrambling could affect the spatial position of both stimuli. Thus, only spatial displacements greater than the spatial variation as a function of source scrambling can be significantly detected, because both stimuli are changing their position in each trial.

The investigation presented in this paper shows that the source spectrum affects the perceived source position even if all binaural and monaural information is provided by

the HRTFs. Therefore, further research is needed to develop scrambling methods that introduce no spatial cues but only change the timbre of the stimulus. Without such a procedure, the sensitivity of localization measurements to the spatial resolution is always restricted to the spatial variation caused by spectral scrambling. The effect of random scrambling on the perceived source position is hardly predictable, because variations in a high dimensional parameter space are performed. Therefore, an empirical investigation of the effect of spectral scrambling on the apparent source position might be appropriate. This could be done by an empirical procedure outlined as follows: A pair of noise signals could be presented via virtual acoustics to an arbitrary angle of azimuth. In each trial the spectrum of one of the stimuli is randomly scrambled in a way similar as described in the present study. The task of the subject would be to indicate if both stimuli are perceived at the same position. If the amplitudes of the spectral variation in the frequency bands of the scrambled noise stimulus are recorded for each trail, the spectrally roved stimuli for which the source position did not change could be extracted. This information could be used to generate a scrambling procedure that does not affect the spatial stimulus position.

6.5 Conclusions

From the investigation presented here the following conclusions can be drawn:

- Appropriate spectral cues in the source spectrum can bias the perceived localization of a broadband sound by up to 20° in elevation in virtual acoustic localization tasks.
- The physical cues in the source spectrum that produce these elevation are consistent with the directional bands for the 'upward' direction found by Blauert (1969).
- New spectral scrambling methods should be developed in the future where only a change in the perceived timbre, but no change in perceived elevation or localization, respectively, is produced.

Chapter 7

Summary and Conclusion

The general aim of this thesis was to assess the perceptual robustness of the auditory system to variations of the respective physical localization cues and (closely related) to characterize the amount of individual information that is needed in head related transfer functions (HRTFs) to achieve an accurate perception of virtual acoustic stimuli.
To measure individual HRTFs a mechanical setup (the TASP system) was constructed that allows for a rapid and accurate positioning of a physical sound source on a sphere of possible source locations (see Chapter 2). The usability and reliability of this measurement system was investigated by conducting two free-field localization experiments. The results were similar to comparable experiments from the literature even though the measurement setups differed in several aspects. Therefore, it can be concluded that the experimental setup used here does not substantially influence the measured localization accuracy. However, it seems necessary to re-establish the subjects head position at the center of the TASP system after each stimulus presentation. Since this was not the case in the present experiments, the data data did not show the increased localization accuracy for frontal sound incidence usually obtained with a fixed head position.
The individual subjects localization decision was recorded using the Gods eye view localization pointing (GELP) technique which was evaluated in different experimental series with respect to its accuracy. Even though the subjects were less accurate in handling the GELP technique in a darkened room than in a lighted room, the recorded localization accuracy of acoustical targets is obviously not restricted by using the tactile sense to handle the GELP technique.
In order to analyze physical differences between HRTFs of different subjects, the TASP system was used to measure individual HRTFs from 11 subjects and one dummy head (see Chapter 3). The results show high inter-individual differences in the monaural and binaural cues. Hence, dummy head HRTFs can in principle not replace individual HRTFs since the former would lead to an inaccurate spatial perception for the majority of subjects. Furthermore, the inter-individual differences between individual HRTFs are

smaller than the differences between individual and dummy head HRTFs. Hence, if individual HRTFs are not available for building virtual acoustic displays, non-individual HRTFs of a selected listener (showing a high localization performance in free-field localization experiments) instead of dummy head HRTFs are recommended for the best expected localization performance under theses circumstances.

In the process of realizing HRTFs as digital filters, the HRTF spectra are often smoothed to reduce the computational effort to process these filters. In the second section of Chapter 3 the effect of two different smoothing procedures (cepstral smoothing and $1/N$ octave smoothing) on the localization cues was investigated. It is shown that both the monaural and binaural localization cues are affected by smoothing. A comparison of the effect of cepstral smoothing and $1/N$ octave smoothing on the ILD revealed that the logarithmic $1/N$ octave smoothing is less appropriate for reducing the spectral detail of HRTF spectra because comparably high ILD deviations are introduced. Furthermore, the reduction of the filter length is less effective with respect to $1/N$ octave smoothing in comparison to cepstral smoothing.

In Chapter 4 the investigation is expanded to perceptual consequences of spectral and temporal HRTF manipulations. Discrimination experiments were conducted to obtain thresholds for perceptually just noticeable deviations of the respective physical localization cues. The results indicate that a high amount of spectral detail can be eliminated from the HRTF spectra without affecting the *spatial* perception. However, if the subjects were able to use spectral timbre cues, the detection thresholds are substantially decreased. Furthermore, subjects were very sensitive to variations of the macroscopic structure of the HRTF spectra which were varied by introducing non-individual spectral information of dummy head HRTFs, especially for source positions in the median plane. It can be concluded that subjects are more sensitive to distortions of the individual localization cues if the center frequencies of the perceptually relevant peaks and notches of the HRTF spectra are affected.

A correlation analysis showed that the average ILD deviation that is introduced by manipulating the HRTF spectra correlates well to the perceived spatial displacement. Hence, the ILD deviation was used as a distance measure for obtaining thresholds for perceptually just noticeable variations of the HRTF spectra. Thus, an important conclusion is that the ILD deviation can be used for predicting the perceptual effect of spectral manipulations.

The introduced 'spectral morphing' method could be used to quantify perceptually relevant distances of individual HRTFs by calculating the 'morphing' factor for which the corresponding ILD deviation exceeds the appropriate thresholds which were obtained in the psychoacoustic experiments. In this way, perceptually relevant distances of HRTF spectra from different subjects could be quantified.

It was furthermore shown, that the ITD JND is increased if the stimuli have a plausible ILD that is determined from individually measured HRTFs. It can be assumed that the

additional spatial information provided by the plausible frequency composition of the ILD enhances the perceptual robustness to ITD variations of the virtual acoustic object.
In Chapter 5 it is investigated if reflections and reverberation affect the evaluation of the localization cues in the direct sound (which is assumed to be the primary factor in localization). To do so, the sensitivity of subjects to manipulations of the localization cues in the direct sound is compared under reverberant and non-reverberant conditions. The results show that the detection thresholds increase by a factor of approx. 2 in the reverberant condition for spectral and temporal manipulations. Hence, reflections do influence the evaluation of the localization cues in the direct sound. It can be assumed that the increased thresholds for manipulations of the localization cues in the direct sound are not caused by an increased localization blur but by an increased perceptual robustness of the acoustical object due to additional spatial information provided by the reflections.
Hence, both the results of the investigation on the ITD JND and the results obtained in Chapter 5 suggest that additional spatial information provided to the auditory system (which is available in real acoustics, but not necessarily in virtual acoustics) stabilizes the perception of virtual acoustic stimuli and enhance the robustness to distortions or deviations of the localization cues.
The physical differences between individual and non-individual HRTFs can be interpreted as distortions of the individual localization cues. Therefore, the results of this thesis show that the need for individual HRTFs is substantially reduced if additional localization cues (e.g. distance information provided by reflections or dynamic cues due to head rotation) can be used by the auditory system.
In some of the experiments presented above, the perceptual dimension was restricted to spatial cues by scrambling the source spectrum of the stimulus. Subjects anecdotally reported that not only the target stimulus was spatially displaced in the discrimination experiments but also the reference stimuli were moving depending on the spectral timbre variation that was introduced by spectral scrambling. Hence, in Chapter 6 the effect of a source spectrum variation on the perceived stimulus location was investigated. The results of an absolute localization experiment show that a spectral cue in the source spectrum can vary the perceived stimulus elevation in a range of 20°. This effect is mainly caused by an increase in level in a comparatively small frequency band that corresponds to the directional 'above' band specified by Blauert (1969). It can be concluded that spectral scrambling can vary the perceived source location in the estimated range.
This outcome has consequences for the interpretation of the results obtained in Chapters 4 and 5. If spectral scrambling was applied to the stimuli, random variations of the perceived stimulus position could have been perceived by the subjects. It is likely that if an ideal scrambling procedure which only varies the spectral timbre of the stimuli would have been used, subjects would have been able to detect smaller spatial changes of the stimulus. Hence, it can be concluded that in this case the ILD deviation thresholds for

perceptually just noticeable distortions of the HRTF spectra would have been slightly decreased.

Taken together, the current thesis has provided new methods for characterizing the human ability to localize sounds both in virtual and real acoustical environments. It has also shown the limits of the physical representation of sound stimuli that have to be obeyed when simulating an acoustical scene. A particular noteworthy aspect is the stabilization of spatial perception produced by the "first wavefront" by adding natural acoustical features, such as a natural ILD pattern across frequency and reflections/reverberations. It can be expected that the current results will be useful not only for gaining more insights about the nature of human sound processing, but also for the design and verification of systems employing virtual acoustics, for example in telecommunication, remote control and home entertainment.

Appendix A

A.1 Free field localization experiments in the literature

The studies of Wightman and Kistler (1989a), Gilkey et al. (1995) and Makous & Middlebrooks (1990) are summarized briefly. The results of these studies are used for a comparison of the results of the current study in Section 4.

Wightman and Kistler compared the localization performance under free-field and virtual conditions. In the free-field condition, noise stimuli were presented over six speakers, mounted on a semicircular steel arc. The ends of the arc were mounted above and below the subject. Source positions in elevation were chosen by using one of six speakers. The stimulus spectrum (a pulsed train of gaussian noise, eight 250 ms pulses separated by 300 ms) was roved in level (20 dB range in critical bands) to prevent the subject from learning the stimulus timbre. The stimulus locations were spaced by 10° in azimuth and by 18° in elevation ranging from −36° to 54°. Before data collection, the subjects were intensively trained to learn the verbal report technique of azimuth and elevation angles.

In order to introduce and validate the GELP technique, Gilkey et al. conducted two experiments. In the first one, click train stimuli (300 ms length) were presented by one of a pool of 239 loudspeakers, distributed on the surface of a sphere with a diameter of 4.3 meters. The speaker resolution on the surface was between 8° and 15°c surrounding the subject in azimuth and covering the elevation range from −45° to 90°. The GELP technique was used to record the subjectively perceived source location. In a second experiment, the source locations were verbally conveyed by the experimenter. In both experiments the head of the subjects was centered in the localization dome by a bite bar. A comparison of experiment I and II was used to validate the GELP technique. In order to compare their work with the literature, Gilkey et al. re-analyzed data from Wightman & Kistler.

In the study from Makous and Middlebrooks the stimuli were presented via one of 36 loudspeakers (10° spacing) mounted on a circular hoop with a diameter of 2.4 meters.

The rotation axis lies within the interaural axis. The source locations covered all azimuths (10° resolution) and elevation within −45° to 55° with respect to a double pole system of coordinates. The stimuli had a flat spectrum within 1.6 kHz to 16 kHz and pseudo random phase. The subjectively perceived source locations were recorded by monitoring the subjects head position. The subjects were asked to point to the source location with their noses. Between measurement trials an acoustical reference stimulus was presented in front of the subject to allow for a realignment of the subjects head. Two measurement conditions were used to measure the individual localization performance (Open Loop condition) and the inherent motor response (Closed loop condition). In the former condition the stimulus duration was 150ms and in the latter the stimulus lasted until the subjective data was recorded. However, it should be noted that in all previous mentioned studies the stimuli were not only distributed in the horizontal and median plane as in the current study.

A.2 Correlations between distance measures and percent correct responses for Chapter 4

Correlation coefficients between the perceptual data (obtained in Chapter 4) and different distance measures are given in Tables A.2-A.2. The distance measures are described below. Note that the distance measures $D10$, $D11$ are D_{mon} and D_{bin}, respectively.

Description of the distance measures

1. D1: Absolute difference between ILDs of reference and target HRTFs averaged across frequency bins on a linear scale.

2. D2: Absolute difference between right ear HRTF spectra of target and references HRTFs averaged across frequency bins on a linear scale.

3. D3: Absolute difference between left ear HRTF spectra of target and references HRTFs averaged across frequency bins on a linear scale.

4. D4: Maximum absolute difference between ILDs of reference and target HRTFs across frequency bins on a linear scale.

5. D5: Maximum absolute difference between right ear HRTF spectra of target and references HRTFs across frequency bins on a linear scale.

6. D6: Maximum absolute difference between right ear HRTF spectra of target and references HRTFs across frequency bins on a linear scale.

7. D7: Maximum absolute difference between ILDs of reference and target HRTFs across frequency channels of a Gammatone filter bank.

8. D8: Maximum absolute difference between right ear HRTF spectra of reference and target HRTFs across frequency channels of a Gammatone filter bank.

9. D9: Maximum absolute difference between left ear HRTF spectra of reference and target HRTFs across frequency channels of a Gammatone filter bank.

10. D10: Mean absolute difference between ILDs of reference and target HRTFs averaged across frequency channels of a Gammatone filter bank. This distance measure is called D_{bin} throughout the study.

11. D11: Mean absolute difference between right ear HRTF spectra of reference and target HRTFs averaged across frequency channels of a Gammatone filter bank. This distance measure is called D_{mon} throughout the study.

12. D12: Mean absolute difference between left ear HRTF spectra of reference and target HRTFs averaged across frequency channels of a Gammatone filter bank.

SS I

Distance measure / Azimuth	0°	45°	90°	135°	180°	Ø
D1	0,2	0,47	0,74	0,8	0,57	0,7
D2	0,21	0,55	0,78	0,81	0,75	0,78
D3	0,26	0,41	0,65	0,77	0,64	0,69
D4	0,18	0,41	0,66	0,75	0,42	0,61
D5	0,04	0,57	0,68	0,66	0,64	0,66
D6	0,26	0,41	0,65	0,77	0,64	0,69
D7	0,12	0,42	0,59	0,85	0,38	0,61
D8	0,02	0,37	0,78	0,86	0,76	0,8
D9	0,25	0,45	0,46	0,84	0,35	0,55
D10	0,12	0,43	0,75	0,79	0,65	0,73
D11	0,14	0,42	0,81	0,85	0,8	0,82
D12	0,33	0,48	0,69	0,83	0,76	0,76

Table A.1: Correlation coefficients between percentage of correct responses in condition 'SS I' and twelve different distance measures are shown. Mean values averaged across source azimuths from $90° - 180°$ are given in the last column.

SS II

Distance measure / Azimuth	0°	45°	90°	135°	180°	Ø
D1	0,22	0,58	0,37	0,44	0,13	0,31
D2	0,31	0,72	0,14	0,23	0,26	0,21
D3	0,19	0,42	0,48	0,53	0,22	0,41
D4	0,19	0,65	0,41	0,47	-0,01	0,29
D5	0,15	0,72	0,08	0,14	0,16	0,13
D6	0,19	0,42	0,48	0,53	0,22	0,41
D7	-0,05	0,58	0,35	0,45	0,37	0,39
D8	0,06	0,58	0,15	0,27	0,29	0,24
D9	0,32	0,48	0,27	0,43	0,39	0,36
D10	0,11	0,54	0,23	0,46	0,49	0,39
D11	0,18	0,63	0,19	0,29	0,41	0,3
D12	0,31	0,52	0,38	0,49	0,45	0,44

Table A.2: Correlation coefficients between percentage of correct responses in condition 'SS II' and twelve different distance measures are shown. Mean values averaged across source azimuths from $0° - 180°$ are given in the last column.

SS III

Distance measure / Azimuth	0°	45°	90°	135°	180°	Ø
D1	0,7	0,87	0,68	0,66	0,45	0,67
D2	0,72	0,81	0,68	0,63	0,56	0,68
D3	0,75	0,8	0,72	0,64	0,49	0,68
D4	0,73	0,77	0,67	0,53	0,36	0,61
D5	0,71	0,78	0,67	0,74	0,4	0,66
D6	0,75	0,8	0,72	0,64	0,49	0,68
D7	0,71	0,83	0,69	0,69	0,59	0,7
D8	0,57	0,67	0,67	0,66	0,56	0,63
D9	0,51	0,77	0,66	0,57	0,45	0,59
D10	0,66	0,88	0,71	0,75	0,59	0,72
D11	0,72	0,74	0,67	0,65	0,57	0,67
D12	0,63	0,86	0,71	0,75	0,48	0,69

Table A.3: Correlation coefficients between percentage of correct responses in condition 'SS III' and twelve different distance measures are shown. Mean values averaged across source azimuths from $0° - 180°$ are given in the last column.

Spectral morphing

Distance measure / Azimuth	0°	45°	90°	135°	180°	Ø
D1	0,62	0,85	0,7	0,74	0,62	0,71
D2	0,7	0,81	0,79	0,63	0,48	0,68
D3	0,55	0,77	0,68	0,73	0,41	0,63
D4	0,66	0,8	0,67	0,76	0,57	0,69
D5	0,78	0,7	0,68	0,53	0,55	0,65
D6	0,55	0,77	0,68	0,73	0,41	0,63
D7	0,64	0,81	0,64	0,81	0,59	0,7
D8	0,74	0,72	0,77	0,58	0,5	0,66
D9	0,55	0,75	0,72	0,74	0,24	0,6
D10	0,68	0,79	0,74	0,75	0,67	0,73
D11	0,69	0,82	0,77	0,59	0,43	0,66
D12	0,54	0,78	0,76	0,75	0,23	0,61

Table A.4: Correlation coefficients between percentage of correct responses in condition 'spectral morphing' and twelve different distance measures are shown. Mean values averaged across source azimuths from $0° - 180°$ are given in the last column.

References

Alrutz, Herbert (**1983**). *Über die Anwendung von Pseudorauschfolgen zur Messung an linearen Übertragungssystemen*, (Phd-thesis). Ernst-August-Universität Göttingen.

Asano, F. (**1990**). Role of spectral cues in median plane localization. *J. Acoust. Soc. Am.*, **88**(1):159–168.

Begault, D. R. (**1992**). Perceptual effects of synthetic reverberation on three-dimensional audio systems. *J. Aud. Eng. Soc.*, **40**(11):895–904.

Békésy, G. v. (**1938**). Über die Entstehung der Entfernungsempfindung beim Hören. *Akustische Zeitschrift*, **3**:21–31.

Blauert, J. (**1969**). Sound localization in the median plane. *Acustica*, **22**:205–213.

Blauert, J. (**1971**). Localization and the law of the first wav front in the median plane. *J. Acoust. Soc. Am.*, **50**:466–470.

Blauert, J. (**1974**). *Räumliches Hören.* S. Hirzel Verlag.

Blauert, J. (**1998**). *verbal communication.*

Borish, J. and J. B. Angell (**1983**). An efficient algorithm for measuring the impulse response using pseudorandom noise. *J. Audio Eng. Soc.*, **31**(7):478.

Bronkhorst, A. W. (**1995**). Localization of real and virtual sources. *J. Acoust. Soc. Am.*, **98**:2542–2553.

Bronkhorst, A. W. and T. Houtgast (**1999**). Auditory distance perception in rooms. *Nature*, **397**(6719):517.

Brungart, D. S. and W. M. Rabinowitz (**1995**). Auditory localization of nearby sources. head related transfer functions. *J. Acoust. Soc. Am.*, **16**:331–353.

Butler, R. A. and C. C. Helwig (**1983**). The spatial attributes of stimulus frequency in the median sattigal plane and their role in sound localization. *Am. J. Otolaryngol.*, **4**:165–173.

Butler, R.A. and K. Belendiuk (**1977**). Spectral cues utilized in the localization of sound in the median sattigal plane. *J. Acoust. Soc. Am.*, **61**:1264–1267.

D., Musicant A. and R.A. Butler (**1985**). Influence of monaural spectral cues on binaural localization. *J. Acoust. Soc. Am.*, **77**:202–208.

Domnitz, R. (**1968**). The interaural time JND as a simultaneous function of interaural time and interaural amplitude. *J. Acoust. Soc. Am.*, **50**:1549–1552.

Durlach, N. I. and H.S. Colburn (**1979**). *Binaural phenomena.*, Chp. 10, pp. 365–466. Academic Press.

Fisher, I.N., T. Lewis and B.J. Embleton (**1987**). *Statistical analysis of spherical data.* Cambrigde: Cambridge University Press.

Flannery, R. and R. A. Butler (**1981**). Spectral cues provided by the pinna for monaural localization. *Percept. Psychophys.*, **29**:438–444.

Freyman, R. L., D. D. McCall and R. K. Clifton (**1998**). Intensity discrimination for precedence effect stimuli. *J. Acoust. Soc. Am.*, **103**(4):2031–2041.

Gilkey, R.H., M.D. Good, M. A. Ericson, J. Brinkman and J.M. Steward (**1995**). A pointing technique for rapidly collecting responses in auditory research. *Behav. Res. Meth. Instr. and Comp.*, **27**:1–11.

Good, M. and R. H. Gilkey (**1996**). Sound localization in noise: The effect of signal to noise ratio. *J. Acoust. Soc. Am.*, **99**:1108–1117.

Haas, H. (**1949**). The influence of a single echo on the audibility of speech. *J. Audiol. Eng. Soc.*, **20**:145–159.

Hammershoi, D. (**1995**). *Binaural technique - a method of true 3D sound reproduction.*, (Phd-Thesis). Aalborg University Press: Aalborg University.

Hammershoi, D. and H. Møller (**1996**). Sound transmission to and along the ear canal. *J. Acoust. Soc. Am.*, **100**:408–427.

Hartmann, W. M. (**1983**). Localization of sound in rooms. *J. Acoust. Soc. Am.*, **74**:1380–1391.

Hebrank, J. and D. Wight (**1974**). Spectral cues used in the localization of sound sources in the median plane. *J. Acoust. Soc. Am.*, **56**:1829–1834.

Hershkowitz, R. M. and N. I. Durlach (**1969**). Interaural time and amplitude JND's for a 500-hz tone. *J. Acoust. Soc. Am.*, **46**:1464–1467.

Kinkel, M. (**1990**). *Zusammenhang verschiedener Parameter des binauralen Hörens bei Normal und-Schwerhöringen.*, (Phd-Thesis). Georg-August-Universität zu Göttingen.

Koehnke, J., C. P. Culotta, Hawley M. L. and H. S. Colburn (**1995**). Effects of reference interaural time and intensity differences on binaural performance in listeners with normal and impaired hearing. *Ear and Hearing*, **16**:331–353.

Kuhn, G.F. (**1977**). Model for the interaural time differences in the azimuthal plane. *J. Acoust. Soc. Am.*, **62**(1):157–167.

Kulkarni, A. and H.S. Colburn (**1998**). Role of spectral detail in sound localization. *Nature*, **396**:747–749.

Kulkarni, A., Isabelle S. K. and H. S. Colburn (**1999**). Sensitivity to head-related transfer function phase spectra. *J. Acoust. Soc. Am.*, **105**(5):2821–2840.

Langendijk, E. H. A. and A. W. Bronkhorst (**1997**). Collecting localization responses with a virtual acoustic pointer. *J. Acoust. Soc. Am.*, **101**:3106.

Langendijk, E. H. A. and A. W. Bronkhorst (**2001**). *Contribution of spectral cues to human sound localization*. Submitted to JASA.

Langendijk, E. H. A. and A.W. Bronkhorst (**2000**). Fidelity of three-dimensional sound reproduction using a virtual auditory display. *J. Acoust. Soc. Am.*, **107**:528–537.

Litovsky, R. Y. (**1997**). Developmental changes in the precedence effect: Estimates of minimum audible angle. *J. Acoust. Soc. Am.*, **102**:1739–1745.

Litovsky, R. Y., H. S. Colburn, W. A. Yost and S. J. Guzman (**1999**). The precedence effect. *J. Acoust. Soc. Am.*, **106**(4):1633–1654.

Litovsky, R. Y. and N. A. Macmillan (**1994**). Sound localization under conditions of the precedence effect: Effects of azimuth and standard stimuli. *J. Acoust. Soc. Am.*, **96**:752–758.

Lorenzi, C., S. Gatehouse and C. Lever (**1999**). Sound localization in noise in normal hearing listeners. *J. Acoust. Soc. Am.*, **105**(3):1810–1820.

Makous, J. C. and J. C. Middlebrooks (**1990**). Two-dimensional sound localization by human listeners. *J. Acoust. Soc. Am.*, **87**:2188–2200.

Mehrgardt, S. and V. Mellert (**1977**). Richtungshören in der Medianebene und Schallbeugung am Kopf. *J. Acoust. Soc. Am.*, **61**:1567–1576.

Mershon, D. H. and E. King (**1975**). Intensity and reverberation as factors in the auditory perception of egocentric distance. *Perception and Psychophysics*, **18**:409–415.

Middlebrooks, J.C. (**1992**). Narrow-band sound localization related to external ear acoustics. *J. Acoust. Soc. Am.*, **92**:2607–2624.

Middlebrooks, J.C. (**1999a**). Individual differences in external-ear transfer functions reduced by scaling in frequency. *J. Acoust. Soc. Am.*, **106**:1480–1492.

Middlebrooks, J.C. (**1999b**). Virtual localization improved by scaling non-individualized external-ear transfer functions in frequency. *J. Acoust. Soc. Am.*, **106**:1493–1510.

Middlebrooks, J.C. and D.M. Green (**1991**). Sound localization by human listeners. *Annu. Rev. Psych.*, **42**:135–159.

Middlebrooks, J.C., E. A. Macpherson and Z. A. Onsan (**2000**). Psychophysical customization of directional transfer functions for virtual sound localization. *J. Acoust. Soc. Am.*, **108**(6):3088–3091.

Møller, H., M. F. Sørensen, D. Hammershøi and C. B. Jensen (**1995**). Head-related transfer functions of human subjects. *J. Audio. Eng. Soc.*, **43**(5):300–321.

Moore, B. C. J., R. W. Peters and Glasberg B. R. (**1990**). Auditory filter shapes at low center frequencies. *J. Acoust. Soc. Am.*, **88**:132–140.

Morimoto, M. and H. Aokata (**1984**). Localization cues of sound sources in the upper hemisphere. *J. Acoust. Soc. Jpn*, **5**(3):165–173.

Mossop, J. E. and J. F. Culling (**1995**). Lateralization of large interaural delays. *J. Acoust. Soc. Am.*, **16**:331–353.

Musicant, A. D. and R. A. Butler (**1984**). The psychophysical basis of monaural localization. *Hear. Res.*, **14**:185–190.

Oldfield, S.R. and S. A. Parker (**1984a**). Acuity of sound localization: a topography of auditory space I. Normal hearing condition. *Perception*, **13**:581–600.

Oldfield, S.R. and S. A. Parker (**1984b**). Acuity of sound localization: a topography of auditory space II. Pinna cues absent. *Perception*, **13**:601–617.

Oppenheim, A.V. and R.W. Schafer (**1975**). *Digital Signal Processing*. Prentice-Hall.

Otten, J. (**1997**). *Psychoakustische Messungen zur Lokalisationsfähigkeit beim Menschen.*, (Diplom-Thesis). University of Oldenburg.

Perrott, D. R., K. Marlborough and P. Merrill (**1989**). Minimum audible angle thresholds obtained under conditions in which the precedence effect is assumed to operate. *J. Acoust. Soc. Am.*, **85**(1):282–288.

Pösselt, C., J. Schröter, M. Opitz, P. L. Diverny and J. Blauert (**1986**). Gerneration of binaural signals for research and home entertainment. *Proc.* 12^{th} *Int. Cong. on Acoustics (Toronto).*

Rayleigh, Lord (**1907**). On our perception of sound direction. *Philos. Ma.*, **13**:214–232.

Rife, Douglas D. and J. Vanderkooy (**1993**). Transfer-function measurement with maximum-length sequences. *J. Audio Eng. Soc.*, **37**(6):419.

Roffler, S. K. and R. A. Butler (**1968**). Factors that influence the localization of sound in the vertical plane. *J. Acoust. Soc. Am.*, **43**:1255–1259.

Shaw, E. A. G. (**1974**). Transformation of the sound pressure level from the free-field to the eardrum in the horizontal plane. *J. Acoust. Soc. Am.*, **56**:1848–1861.

Shaw, E. A. G. (**1997**). *Binaural and Spatial Hearing in Real and Virtual Environmentes*, Chp. 2, pp. 25–47. Lawrence Erlbaum Associates.

Shaw, E.A.G. and R. Teranishi (**1968**). Sound pressure generated in an external-ear replica and real human ears by a nearby point source. *J. Acoust. Soc. Am.*, **44**(1):240–249.

Sheeline, C. W. (**1983**). *An investigation of the effects of direct and reverberant signal interaction on auditory distance perception*, (Phd Thesis). Stanford University.

Shinn-Cunningham, B. G., P. M. Zurek and N. I. Durlach (**1993**). Adjustment and discrimination measurements of the precedence effect. *J. Acoust. Soc. Am.*, **93**(5):2923–2932.

Teranishi, R. and E.A.G. Shaw (**1968**). External ear acoustics models with simple geometry. *J. Acoust. Soc. Am.*, **44**:257–263.

Tobias, J. V. and S. Zerlin (**1959**). Lateralization threshold as a function of stimulus duration. *J. Acoust. Soc. Am.*, **31**:1591–1594.

Tollin, D. J. and G. B. Henning (**1998**). Some aspects of the lateralization of echoed sound in man. I. the classical interaural-delay based precedence effect. *J. Acoust. Soc. Am.*, **104**:3030–3038.

Trampe, Ulrich (**1988**). *Akustische Übertragung des durchschnittlichen antropomorphen Außenohrs*, (Diplom Arbeit). Unversität Oldenburg.

Wallach, H., E. B. Newman and M. R. Rosenzweig (**1949**). The precedence effect in sound localization. *Am. J. Psychol.*, **LXII**:315–336.

Wenzel, E.M., M. Arruda, D.J. Kistler and F.L. Wightman (**1993**). Localization using non-individualized transfer functions. *J. Acoust. Soc. Am.*, **94**(1):111–123.

Wiener, F.M. and D.A. Ross (**1946**). The pressure distribution in the auditory canal in a progressive sound field. *J. Acoust. Soc. Am.*, **18**:401–498.

Wightman, F. I. and D. Kistler (**1989a**). Headphone simulation of free-field listening: I Stimulus synthesis. *J. Acoust. Soc. Am.*, **85**:858–867.

Wightman, F. I. and D. Kistler (**1989b**). Headphone simulation of free-field listening: II Psychoacoustic validation. *J. Acoust. Soc. Am.*, **85**:858–867.

Wightman, F. I. and D. Kistler (**1997**). Monaural sound localization revisited. *J. Acoust. Soc. Am.*, **101**:1050–1063.

Wightman, F.L. and D.J. Kistler (**1992**). The dominant role of low-frewuency interaural time differences in sound localization. *J. Acoust. Soc. Am.*, **91**:1648–1661.

Woodworth, R.S. (**1954**). *Experimental psychology.* Rinehart & Winston.

Zurek, P. M. (**1980**). The precedence effect and its possible role in the avoidance of interaural ambiguities. *J. Acoust. Soc. Am.*, **67**(3):952–964.

Zwicker, E. and H. Fastl (**1990**). *Psychoacoustics: Facts and Models.* Heidelberg, Germany: Springer-Verlag.

Erklärung

Hiermit erkläre ich, daß ich die vorliegende Arbeit selbständig verfaßt und nur die angegebenen Quellen und Hilfsmittel verwendet habe.

Oldenburg, den 13. Juli 2001 Jörn Otten

Danksagung

Mein herzlichster Dank gilt all den Menschen, die mir geholfen haben, diese Arbeit zu verrichten und erfolgreich zu beenden. Insbesondere bedanke ich mich bei

Prof. Dr. Dr. Birger Kollmeier für die Ermöglichung dieser Arbeit und für die Schaffung der hervorragenden (menschlichen wie materiellen) Arbeitsbedingungen in der AG Medizinische Physik. Seine hilfreichen und führenden Kommentare insbesondere zum Schluß dieser Arbeit haben wesentlich zum Gesamtbild beigetragen.

Prof. Dr. Volker Mellert für die freundliche Übernahme des Koreferats.

Prof. Dr. Steven Colburn für anregende Diskussionen über das Richtungshören und den Beitrages der HRTFs zum selbigen.

Dr. Adelbert Bronkhorst für seine konstruktiven Kritiken und sein Interesse an dieser Arbeit.

Michael Kleinschmidt und Rainer Huber dafür, dass sie es mit mir jahrelang in einer Sardinenbüchse ausgehalten haben.

Dr. Thomas Brand, für kleine und große Hilfen in jeder Lebenslage, der bereitwilligen Preisgabe seiner unglaublichen 'Datenbank' an Informationen und der Teilung eines Hobbies, welches langsam 'High-End' wird...

Holle Kirchner, die mit ihrem klarem Blick und kritischen Bemerkungen in den Anfängen dieser Arbeit eine wesentliche Unterstützung war.

den Mitgliedern der AG Medizinische Physik, die durch das freundliche Miteinander eine entspannte und weiträumige Arbeits-Atmosphäre geschaffen haben.

den Versuchspersonen Helmut Riedel, Michael Kleinschmidt, Rainer Huber, Ingo Baumann, Holle Kirchner, Karin Troidl, Dirk Junius, Dr. Carsten Reckhardt und Jörg Damaschke.

Jonas und Julian, die alle Probleme vergessen lassen.

meiner Frau Sandra. Ohne ihre selbstlose Unterstützung wäre diese Arbeit nicht möglich gewesen.

Lebenslauf

Am 22.06.1970 wurde ich, Jörn Otten, in Leer/Ostfriesland als drittes Kind von Imbke Otten, geb. Hörmann, und Otto Otten geboren. In dem Zeitraum von 1976 - 1980 besuchte ich die Ludgeri Grundschule in Leer und wechselte nach der Orientierungstufe (1980 - 1982) auf das Ubbo-Emnius Gymnasium, welches ich ab 1982 besuchte. Die schulische Ausbildung konnte ich 1990 mit dem Abschluß der Allgemeinen Hochschulreife beenden. Den Zivildienst beim Paritätischen Wohlfahrtsverband absolvierte ich in dem Zeitraum von Mai 1990 bis Juli 1991. Das Studium der Physik wurde an der Universität Oldenburg im Oktober 1991 begonnen und mit dem Abschluß als Diplom Physiker in Juni 1997 beendet. Die vorliegende Dissertation wurde seit Juli 1997 als Stipendiat des Graduiertenkollegs 'Psychoakoustik' unter der Leitung von Prof. Dr. Dr. Birger Kollmeier angefertigt. Seit dem 01.05.2001 bin ich als wissenschaftlicher Mitarbeiter am Hörzentrum Oldenburg tätig.