Audibility of differences in adjacent head-related transfer functions

Hoffmann, Pablo F.; Møller, Henrik

Published in:
Acustica United with Acta Acustica

DOI (link to publication from Publisher):
10.3813/AAA.918111

Publication date:
2008

Acta Acustica united with Acustica
Vol. 94 (2008) 945 – 954

DOI 10.3813/AAA.918111

Audibility of Differences in Adjacent Head-Related Transfer Functions

Pablo F. Hoffmann, Henrik Møller
Acoustics, Department of Electronic Systems, Aalborg University, Fredrik Bajers Vej 7 B5, 9220 Aalborg Ø, Denmark. pfh@es.aau.dk

Summary
The smallest directional change that can reliably be perceived provides a useful measure to assess the required spatial resolution for virtual spatial sound. Here, the ability of naive listeners to discriminate changes in the characteristics of HRTFs was measured. In one experiment the smallest angular separation needed to discriminate between the magnitude spectrum of HRTFs was determined. In a second experiment the smallest change in interaural time difference (ITD) that could just be audible was determined. Generic HRTFs were used for both experiments. Results showed a large inter-subject variability, which was particularly pronounced for discrimination of changes in ITD. Mean thresholds for changes in ITD ranged from 87.8 to 163 μs. Mean thresholds for discrimination of spectral differences ranged from 2.4 to 11°, and significant differences were found depending on the direction of change. Results suggest that ITDs do not seem to require very high resolutions, and that spatial resolution for spectral characteristics is not uniform meaning that different resolutions are needed depending on sound direction.

PACS no. 43.60.Dh, 43.66.Lj, 43.66.Pn

1. Introduction

It is well known that the directional characteristics of virtual spatial sound can be effectively synthesized using the head-related transfer function (HRTF) [1, 2, 3]. The procedure, most commonly referred to as binaural synthesis, is accomplished by convolving the HRTF in the time domain with an anechoic recording and delivering the result typically over headphones. An important aspect in spatial sound rendering is the resolution at which HRTFs represent auditory space.

If the spatial resolution is higher than our perception so that differences between adjacent HRTFs are much below audibility, then the effort of producing such a high resolution is wasted. In contrast, a resolution that is too low will degrade our auditory spatial perception. Therefore, it seems that from a perceptual viewpoint an appropriate resolution would be a resolution that is equal to, or just higher than, the minimum audible difference between HRTFs.

The minimum audible angle (MAA) [4] is probably the most typical measure of auditory spatial resolution. MAA is defined as the smallest displacement in the position of a sound that can consistently be detected from no displacement. Typically, two sounds are presented sequentially and the listener has to judge the location of the second sound relative to the first. For example, in case of changes in azimuth (horizontal MAA) the task is to detect whether the second sound was to the left or to the right of the first sound. Horizontal MAA is about 1° for a 500-Hz tone presented from a loudspeaker in front of the listener [4]. This spatial acuity is maintained for broadband stimuli reproduced both over loudspeakers [5] and over headphones using artificial-head binaural recordings [6]. Vertical MAA is approximately 4–6° for the forward direction, and, in general, is larger than the horizontal MAA [6, 7].

In MAA experiments all spatial cues are available to the listener. To estimate listeners' ability to discriminate changes in individual cues, experiments have typically measured what is called the just-noticeable differences (JNDs) in interaural time difference (ITD) and interaural level difference (ILD). In optimal testing conditions JNDs are about 10-20 μs for changes in ITD and 1 dB for changes in ILD [8]. The purpose of the present study is to measure the ability of listeners to discriminate differences in the characteristics of HRTFs. We attempt to estimate the largest possible angle for which listeners cannot distinguish between adjacent HRTFs. And this is done for the time and spectral characteristics of the HRTF separately. All experiments were conducted using HRTFs measured on an artificial head with a directional resolution of 2°[9].

1.1. Characteristics of the HRTF

Characteristics of the HRTF can be classified such that time characteristics are associated to the interaural time difference (ITD), and spectral characteristics to the magnitude spectrum. Based on this classification, a common

Received 28 February 2008,
accepted 1 October 2008.

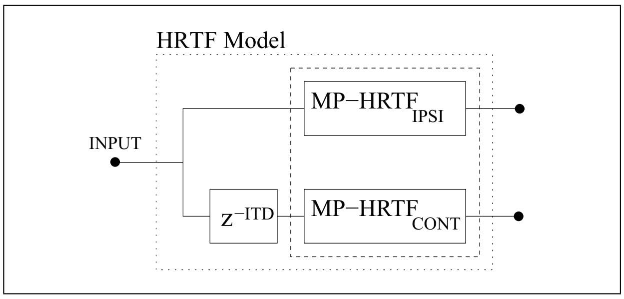

Figure 1. Minimum-phase and frequency-independent ITD model of the HRTF. Minimum-phase filters are enclosed in the dashed box. The IPSI and CONT sub-indices indicate the ipsilateral and contralateral components respectively. The ITD is implemented by cascading the delay to the contralateral component of the HRTF.

model of the HRTF is built as a pair of minimum-phase filters — one filter for each ear — with a pure delay cascaded to the filter representing the contralateral component of the HRTF [10, 11]. Here, the contralateral component refers to the ear farther from the sound source for directions off the median plane. A diagram of this model is shown in Figure 1. The function of the delay is to control the ITD, and it reflects differences in the linear-phase and all-pass components of the HRTFs. Although the phase of all-pass components is not linear, it has been shown that the approximation by a pure delay equal to the interaural difference in the low-frequency group delay, does not have audible consequences [12, 13]. The minimum-phase filters produce the same magnitude spectrum of the measured HRTF. That is, they control monaural spectral cues to both ears, and thereby they also control interaural spectral difference cues (ISD). For practical purposes the minimum-phase filters are generally implemented as finite-impulse-response (FIR) filters. This HRTF model has proven to be perceptually valid from experiments comparing stimuli filtered with empirical HRTFs and stimuli filtered with modeled HRTFs. Results from experiments involving discrimination tasks [14], and sound localization tasks [15], have shown that empirical and modeled HRTFs are indistinguishable and that they generate the same spatial percept.

1.2. Goal of the study

The HRTF model based on minimum-phase filters and pure delay provides the means to measure audibility of differences in HRTFs for spectral and time characteristics independently. In this context, the present study is divided into two experiments. Experiment I measures audibility thresholds for spectral changes in HRTFs, i.e., only the magnitude spectrum is varied while ITD remains constant. Experiment II measures audibility thresholds for changes in ITD while the magnitude spectrum remains unchanged. Unlike typical experiments on auditory spatial resolution, here, sensitivity to differences in the characteristics of HRTFs are based on any possible cue, and not exclusively on a perceived directional shift of the stimuli. Audibility thresholds are measured for a number of directions.

2. Experiment I: Audibility of spectral differences

2.1. Method

2.1.1. Subjects

Ten subjects, five males and five females, participated in the listening test. Subjects were paid for their participation and their age ranged from 21 to 32. Subjects had normal hearing and they were selected by means of an audiometry screening at less than 10 dB HL for frequencies ranging from 250 Hz to 4 kHz in octave steps, and less than 15 dB HL for 8 kHz. All subjects had little or no experience in listening experiments.

2.1.2. Apparatus

Stimuli were processed and played back using a PC equipped with a professional audio card RME DIGI96/8 PST. The digital output of the audio card was connected to a 20-bit D/A converter (Big DAADi) set at a 48 kHz sampling rate. From the D/A converter the signal was sent to a stereo amplifier (Pioneer A-616) modified to have a calibrated gain of 0 dB. A 20-dB passive attenuator was connected to the output of the amplifier in order to reduce the noise floor. Finally, the stereo signal from the output of the attenuator was delivered to the listener through a pair of equalized Beyerdynamic DT-990 circumaural headphones.

2.1.3. Stimuli and spatial synthesis

Five minutes of broadband pink noise, with a bandwidth of 20-16000 Hz, was used as the source signal. This signal was convolved with the headphone equalization filters and stored as a two-channel audio file. The equalization filters were derived from headphones transfer functions measured on 23 subjects, and they were implemented as 256-tap minimum-phase FIR filters. The overall gain of the system was set so that the source signal simulated a level equivalent to that of a free-field source at a sound pressure level of approximately 68 dB.

To simulate directional sound, HRTFs measured with a resolution of 2° on an artificial head were used [9]. Nine positions were selected in the left half of the upper hemisphere. Directions are given as (lateral angle, polar angle) in a polar coordinate system with interaural axis and left-right poles. In this system, referred to as the interaural-polar coordinate system, positions with the same ITD have approximately the same lateral angle, and the polar angle is used to specify source position around the cone determined by the ITD. The convention used here is that 90° and -90° lateral angle correspond to left and right sides, 0° polar angle to the anterior portion of the horizontal plane, 180° polar angle to the posterior portion of the horizontal plane, and 90° polar angle to the upper portion of the frontal plane. In this study, five positions were selected in the median plane (0° lateral angle) at 0°, 44°, 90°, 136° and 180° polar angle. Three positions were selected in an iso-ITD contour to the left ((58°, 0°), (46°, 90°) and (54°, 180°)). The positions with polar angles of 0° and

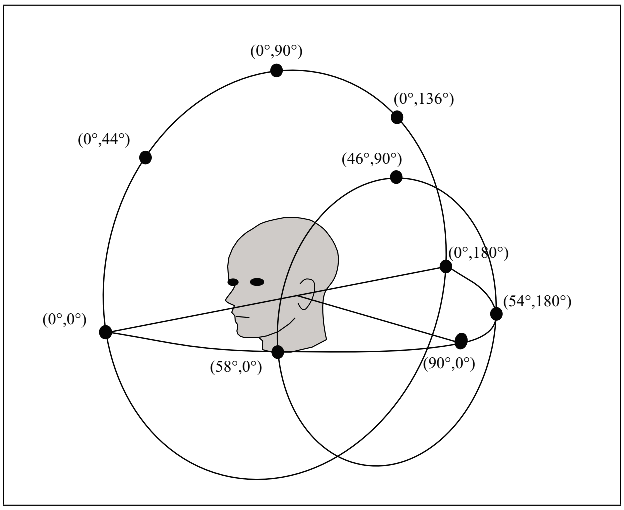

Figure 2. Nominal positions employed in the listening experiment. These positions serve as reference in the experiment. Spatial coordinates are indicated in an interaural-polar coordinate system.

180° were chosen to match as closely as possible the ITD for (46°,90°). Because iso-ITD contours are not geometrically perfect, lateral angle varied slightly with polar angle. The position at 90° lateral angle was also included. In the remainder of this article, these positions will be referred to as *nominal positions* and they are shown in Figure 2.

The measured HRTFs were represented as minimum-phase FIR filters with the ITD calculated separately and inserted to the contralateral impulse response. Filters' length was 1.5 ms (72 coefficients at 48 kHz), and, to control the low-frequency part of the HRTFs, the DC value of each HRTF was set to unity gain as described in [16, section 5.2]. Minimum-phase representations were calculated using homomorphic filtering [17, ch. 12]. Figure 3 shows the magnitude spectra of HRTFs corresponding to the selected nominal positions. ITDs were derived from the interaural differences in group delay of the excess-phase components of the HRTFs evaluated at 0 Hz [18], and they were $-625\ \mu$s for (90°,0°), $-437.5\ \mu$s for the directions in the iso-ITD contour and $0\ \mu$s for the directions in the median plane.

2.1.4. Psychophysical method

Audibility of spectral differences in HRTFs was determined in a three-interval, three-alternative forced-choice task using the method of constant stimuli. The duration of both the stimulus and the inter-stimulus interval was 300 ms. On a single trial, a segment of the pink-noise, already equalized for the headphones, was randomly selected and 10-ms raised-cosine ramps were applied to the onset and offset. The same noise segment was used for the three stimulus intervals (frozen noise). In two of the intervals the noise burst was filtered with an HRTF corresponding to a nominal position. In the remainder interval, selected at random with equal *a priori* probability, the noise burst was filtered with an HRTF that produced a directional shift from the nominal position at possible angular distances of 0.5°, 1°, 2°, 4°, 8° and 16°. The subjects' task was to identify the interval that contained the deviating stimulus. They had to push one of three buttons in a response-box to indicate their choice. Intervals were signaled by lights that were also used as feedback. After a 1-s silence interval a new trial was presented.

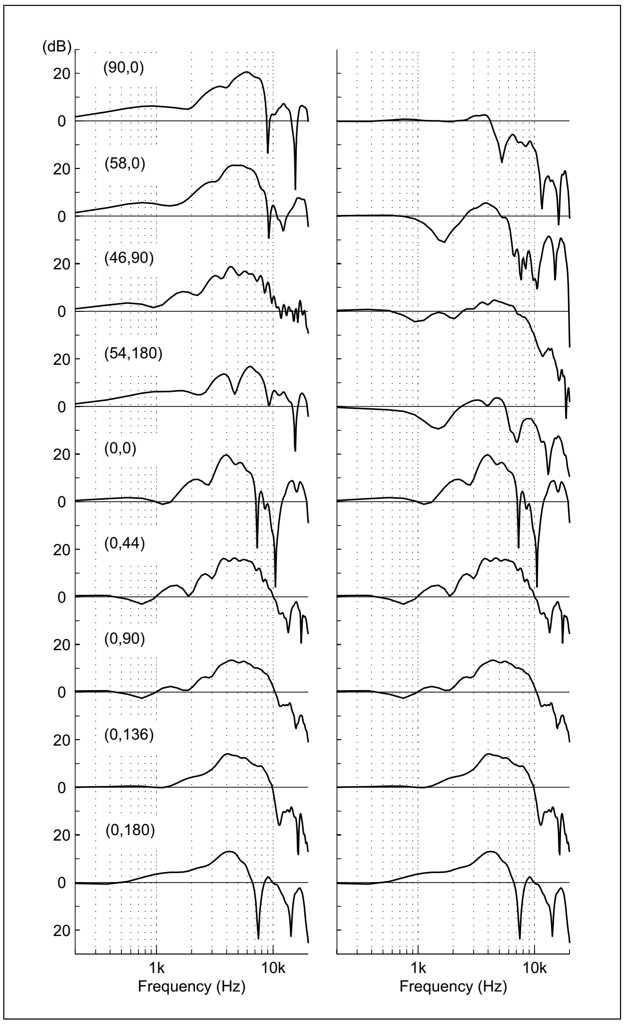

Figure 3. HRTFs used for the nominal positions. Left and right columns represent HRTFs' components for the left- and right-ear respectively.

Different modes of directional change were used to shift between HRTFs, and the selection of these modes depended to some extent on nominal position. For most positions the modes were changes in lateral angle, denoted by left/right, and changes in polar angle, denoted by up/down. For three positions, other modes were required so that the directional changes were physically meaningful. Specifically, for the position (90°,0°) and changes in lateral angle the left/right modes were exchanged by back/forth modes. Also note that for this position up/down modes do not reflect changes in polar angle but actual up and down changes along the frontal plane. For the nominal positions (46°,90°) and (0°,90°) and changes in polar angle, up/down modes were replaced by back/forth modes.

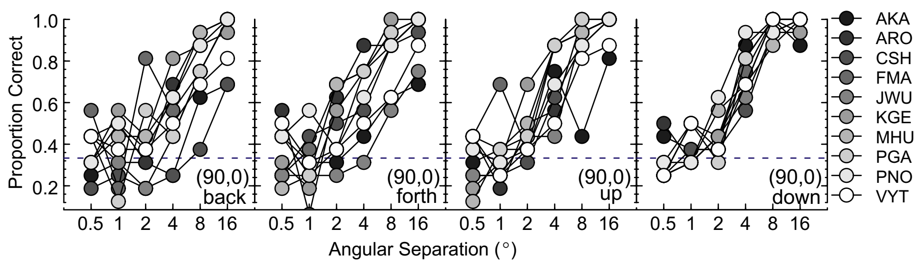

Figure 4. Proportion of correct responses for spectral differences in HRTFs for the position at 90° lateral angle. Results for each mode of directional change are plotted on individual panels. The dashed line indicates chance level.

Recall that here HRTFs refer to minimum-phase filters and thus the deviating stimulus did not include a change in ITD but this remained equal to the ITD of the nominal position. HRTFs for angular distances of 0.5° and 1° were not available from measurements, and therefore, they were obtained from linear interpolation between the nominal position and the position separated by 2°. The interpolation was done in the time domain since the minimum-phase impulse responses are optimally aligned. For the HRTFs used in this study, linear interpolation between minimum-phase impulse responses separated by 2° is considered perceptually correct [19].

2.1.5. Experimental design

Subjects were tested individually in a sound-insulated cabin with absorbing walls specially designed for psychoacoustic experiments. Once in the cabin subjects were provided with written instructions about the task to perform. Subjects were then presented with a few trials in order to acquaint them with the task and the procedure. To further familiarize the subjects a block of sixteen trials was employed as practice. The HRTF of the nominal position (0°,0°) was used for the reference stimulus and only the angular distance of 16° and a downward directional change were employed. Practice blocks were repeated until subjects could respond correctly at least fifteen out of the sixteen trials. In general, practice took about 30 to 45 minutes to complete and since the purpose of the experiment was to use naive subjects no further practice was given.

In the main experiment, nominal position and mode of directional change were held constant within a block of trials. Sixteen repetitions were presented at each angular distance. The order in which they were presented was fully randomized. At the beginning of each block four trials using 20° of angular distance were used as warm-up trials. Each block consisted of 100 trials, and one block took between 7 to 8 minutes to complete. At the end of each block subjects were instructed to remove the headphones. A pause of 1–2 minutes was normally used between blocks but subjects were free to have longer pauses if necessary. After completion of three blocks subjects were instructed to hold a break. The entire experiment was completed in 3 to 4 two-hour sessions with each session held on a different day.

2.1.6. Data analysis

Audibility thresholds were defined as the angular distance for which subjects' performance was equal to half way between chance performance and perfect performance. Since the experiment used a three-alternative forced-choice method the theoretical performance range from 0.33 to 1.0, and therefore the threshold was defined as 0.66 performance. The proportion of correct responses for each angular distance follows a binomial distribution. By repeating each condition 16 times we make sure that for a performance equal to 0.66 or greater, the null hypothesis of the proportion being equal to chance performance is rejected at a significant level of $p < 0.01$. This is done in order to statistically support the threshold definition.

Thresholds were estimated by fitting a logistic function to the proportion of correct responses using a least-square criterion [20]. The logistic function is given by

$$p(x) = \lambda + (1-\lambda)\left(1 + \mathrm{e}^{-(x-\alpha)/\beta}\right)^{-1}, \qquad (1)$$

where $p(x)$ is the proportion of correct responses, x is the independent variable (angular distance), α is the threshold and β is the slope parameter. During the fitting procedure both parameters (α and β) are actually estimated but only α will be reported. The parameter λ represents chance performance and it was not estimated but fixed to 0.33. This performance is expected when listeners cannot detect the deviating stimulus. Psychometric functions were fitted for each subject and each condition, and all thresholds were estimated on the logarithm of the angular distance.

2.2. Results

Figures 4, 5, and 6 show proportions of correct responses for each listener and each condition. Each figure shows results for nominal positions with the same ITD. The abscissa represents angular separation in degrees, and is presented in a logarithmic scale. The ordinate represents subject's performance (given at the different angular separations). For directions in the iso-ITD contour nominal positions are arranged in rows, and modes of directional changes are separated in columns. In general, performance tended to increase monotonically with increasing angular separation. However, for several conditions and subjects, performance did not reach 100% at the largest angular separation employed (16°). Also note that for directions in the median plane overall performance was poorer with

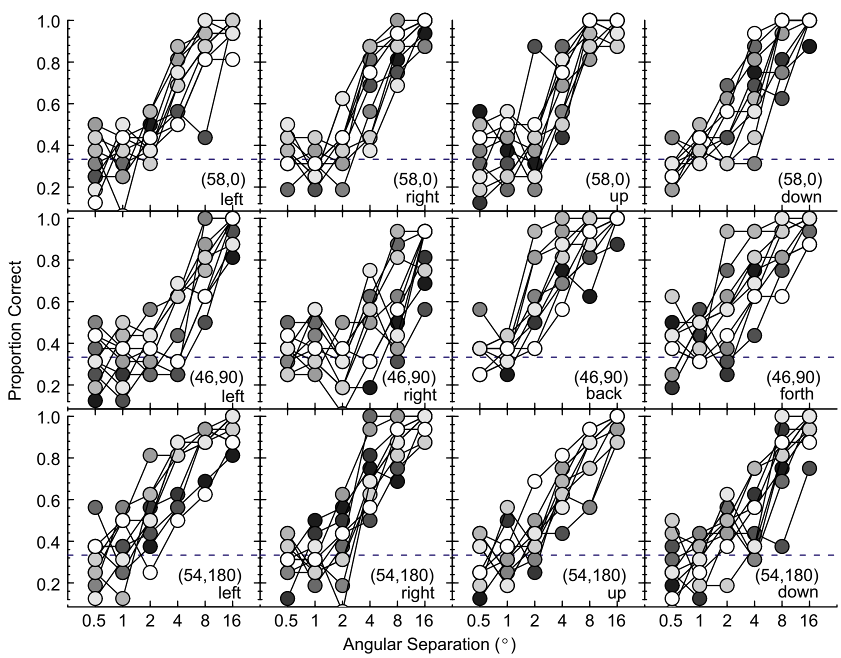

Figure 5. Proportion of correct responses for spectral differences in HRTFs for positions in the iso-ITD contour. Corresponding nominal position and mode of directional change are indicated on each panel.

Table I. Mean thresholds across subjects for the discrimination of spectral differences in HRTFs. Thresholds based on less than ten subjects are shown with a subscript that indicates the number of subjects used to compute the average.

ITD (μs)	Nom. Dir.	Threshold (°)					
		left	right	up	down	back	forth
-625	(90°,0°)	-	-	4.0	3.4	6.0	4.6
-437.5	(58°,0°)	4.3	4.2	3.6	4.0	-	-
	(46°,90°)	5.9	8.1_9	-	-	2.8	3.2
	(54°,180°)	3.5	3.9	4.7	5.8	-	-
0	(0°,0°)	6.6_8	-	2.7	2.4	-	-
	(0°,44°)	7.4_8	-	11.0_6	8.8	-	-
	(0°,90°)	7.2_7	-	-	-	-	-
	(0°,136°)	8.5_6	-	9.4_9	6.5_9	-	-
	(0°,180°)	7.1	-	4.9	5.8	-	-

higher elevations and this was more evident for discrimination along the polar angle. Poorest performance was observed for (0°,90°) with back/forth modes of directional change. In these conditions, proportion of correct response did not depart from chance for almost all subjects and angular separations. Only one subject (JWU) had a percent correct slightly above threshold for the largest angular separation and downward change. This subject is not the same subject (MHU) who was clearly the most sensitive to leftwards changes for the same nominal position. For angular separations of 0.5° and 1°, performance was at chance for the majority of conditions and for all subjects.

Psychometric functions were fitted only to proportion data for which performance exceeded 0.66 within the range of angular separations employed. Based on this criterion, 12.3% of the total pool of individual thresholds could not be estimated. Individual thresholds were averaged across subjects, and the obtained mean values are summarized in Table I. Thresholds based on less than the total number of subjects are shown with a subscript that indicates the number of subjects used to compute that mean. The smallest mean threshold was 2.4° for (0°,0°) and downward change, and the largest could not be estimated for (0°,90°) and back/forth modes.

2.3. Discussion

Audibility of spectral differences in HRTFs was estimated by measuring how well subjects could discriminate between minimum-phase HRTFs from adjacent positions. Thresholds for changes in polar angle increase as the polar angle approaches 90° for positions in the median plane, but they decrease for positions in the iso-ITD contour. For changes in lateral angle, thresholds also increase with elevation and this is seen in both the median plane and iso-ITD contour. The direction dependency and range of

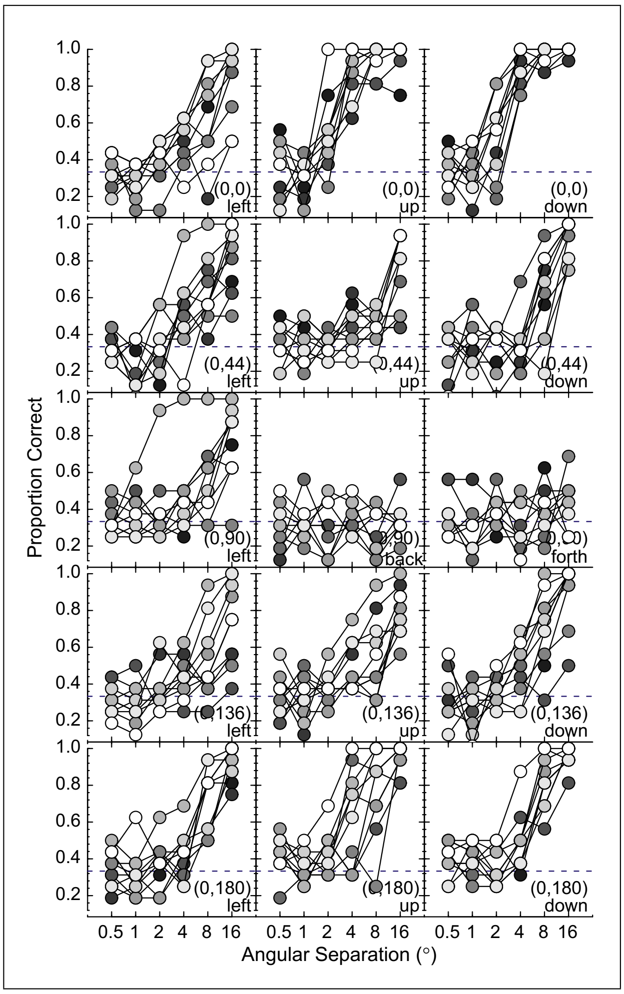

Figure 6. Proportion of correct responses for spectral differences in HRTFs for positions in the median plane. Corresponding nominal position and mode of directional change are indicated on each panel.

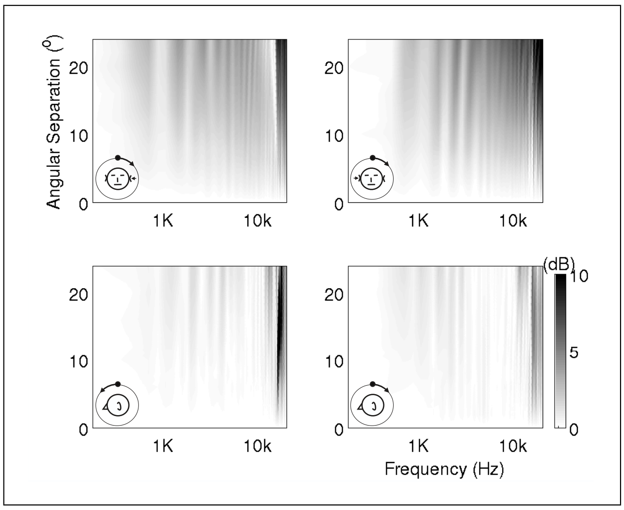

Figure 7. Spectral magnitude differences as a function of angular separation for nominal direction (0°,90°). Differences are given in absolute dB values. Top panels show differences for changes in lateral angle and bottom panels show differences for changes in polar angle. For clarity, the direction of change is illustrated on each panel. Top-left: changes to the left and magnitude differences for the left HRTF (indicated by the small arrow); top-right: changes to the left and magnitude differences for the right HRTF; bottom-left: changes from above to the front and magnitude differences for the left HRTF; bottom-right: changes from above to the rear and magnitude differences for the left HRTF.

thresholds observed in this study are comparable to those from a study conducted by Hoffmann and Møller [21], who examined sensitivity to HRTF magnitude using a similar procedure.

In the median plane, thresholds increased more rapidly as a function of nominal position for changes in polar angle than in lateral angle. In fact, at (0°,90°) (above the head) subjects were unable to perform above chance level for any of the two modes of directional change along the polar angle. These thresholds are in agreement with the observed increase in localization blur with elevation [22]. The decrease in sensitivity to changes in magnitude as elevation moves towards 90° for differences in polar angle can be explained in pure physical terms by comparing the extent to which the magnitude of the HRTFs changes as a function of angular separation for the different modes of directional change. Figure 7 shows differences in dB (expressed in absolute values) for the nominal direction (0°,90°) and for changes in lateral and polar angles. It is clear that when HRTFs are changed along the lateral angle (i.e. changes to the left) a small angular separation produces larger spectral differences than when the change is in polar angle, being either a backward or a forward change. Note that for downward changes, there are almost no differences in the frequency range 5–12 kHz.

For positions in the median plane the significance of the effects was evaluated in a two-way analysis of variance (nominal position x mode of directional change). Because thresholds for (0°,90°) and changes in polar angle could not be estimated, separate ANOVAs were done for lateral and polar changes. For lateral changes, main effect of nominal position was not significant. For polar changes, main effect of nominal position was significant ($F(3,25) = 28.9$, $p < 0.001$), and main effect of directional change was not significant. There was a slightly significant interaction ($F(3,20) = 4.3$, $p = 0.016$). This may be attributed to the fact that thresholds for the downward condition were lower than the upward condition for all directions but (0°,180°).

A two-way within-subject analysis of variance on thresholds for positions in the iso-ITD contour showed that main effect of nominal position was not significant and main effect of mode of directional change was slightly significant ($F(3,26) = 4.5$, $p = 0.012$). The interaction between nominal position and directional change was highly significant ($F(6,53) = 10.4$, $p < 0.001$). Thresholds for left/ right changes increased towards 90° polar angle whereas up/down thresholds decreased. Note that the effect of elevation on changes in polar angle was opposite to the effect observed in the median plane. This suggests that sensitivity to changes in polar angle seems to increase as the sagittal plane moves to lateral positions.

Thresholds for (90°,0°) were significantly lower for back/forth than left/right changes ($p < 0.01$). This result is consistent with vertical MAAs being generally smaller than horizontal MAAs for the most lateral positions in the horizontal plane [7, 23, 24]. One difficult aspect to evaluate is whether the prominent cues were provided by changes in the ipsilateral or contralateral component of the HRTF. On the one hand, the contralateral component is much more sensitive to directional shifts than the ipsilateral one. On the other hand, the overall interaural level difference of roughly 15–20 dB makes unlikely that naive listeners could have made an effective use of spectral differences in the contralateral component.

3. Experiment II: Audibility of time differences

3.1. Method

Twelve subjects participated in this experiment. Five subjects had previously participated in experiment I and the other seven had no previous experience in listening experiments. The experimental method was essentially the same as described in experiment I. For the discrimination of changes in ITD, the three intervals were filtered with the same HRTF corresponding to a given nominal direction. The target stimulus was generated by either adding or subtracting an extra delay to the ITD of the nominal position. The amount of delay could be selected from a set of five pre-specified values that corresponded to 20.8, 41.6, 83.3, 166.6, 333.3 μs; or 1, 2, 4, 8 and 16 samples at a 48-kHz sampling frequency respectively. These delays are referred to as ΔITDs. For the nominal direction (90°,0°) ΔITDs were only subtracted from the nominal ITD. For the positions located in the iso-ITD contour ΔITDs were both added and subtracted, and for positions in the median plane the ΔITDs were only added. Combining nominal positions with corresponding addition and subtraction of ΔITD, a total of twelve conditions were tested (90°lateral angle x 1 ITD shift + 3 iso-ITD positions x 2 ITD shifts + 5 median-plane positions x 1 ITD shift). For the 16-trials practice blocks the position (0°,0°) with a ΔITD of 416 μs (20 samples) was presented.

3.2. Results

Proportion of correct responses for the tested conditions are shown in Figures 8, 9, and 10. The abscissa specifies the ΔITD in μs, and is given in a logarithmic scale. Results for 90° lateral angle (Figure 8) refer to decrements from the -625-μs nominal ITD. For positions in the iso-ITD contour (Figure 9) the left column represents increments in ITD and the right column represents decrements in ITD.

For positions in the median plane (Figure 10) results refer to increments in ITD. Generally, performance tended to improve with increasing ΔITD but substantial differences were observed across subjects. In addition, a large portion of the percent-correct responses for several conditions did not reach perfect performance for the largest ΔITD.

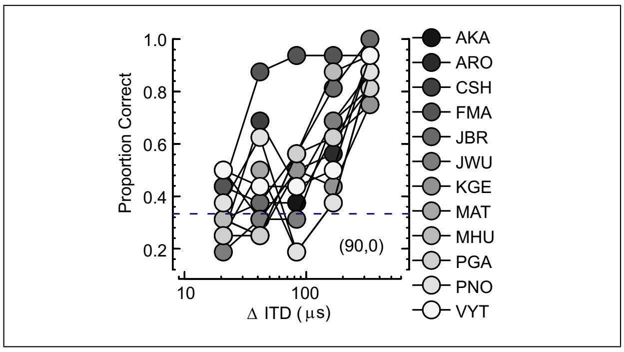

Figure 8. Proportion of correct responses for discrimination of ITDs for the position at 90° lateral angle. Results correspond to a decrement in ITD condition.

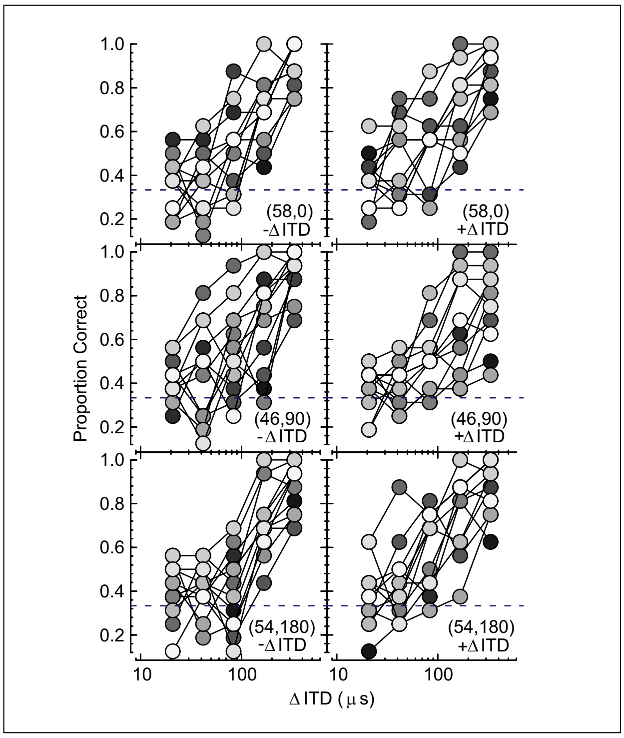

Figure 9. Proportion of correct responses for discrimination of ITDs for positions in the iso-ITD contour. Nominal positions are indicated on each panel. Negative ΔITD represents an increment from the nominal ITD, and positive ΔITD represents a decrement.

Thresholds for each subject and condition were estimated using a logistic regression in the same manner as for thresholds on spectral differences. Subjects' sensitivities were significantly different as shown by an analysis of variance with subjects as factor ($p < 0.001$). A post hoc analysis (Tukey HSD) revealed that there were primarily two subjects (JBR, PGA) who had significantly lower thresholds compared to nine and eight other subjects respectively.

Mean thresholds were calculated across subjects and are summarized in Table II and plotted on Figure 11 along with individual thresholds. Data are grouped by ITD and the abscissa represents the polar angle of the nominal po-

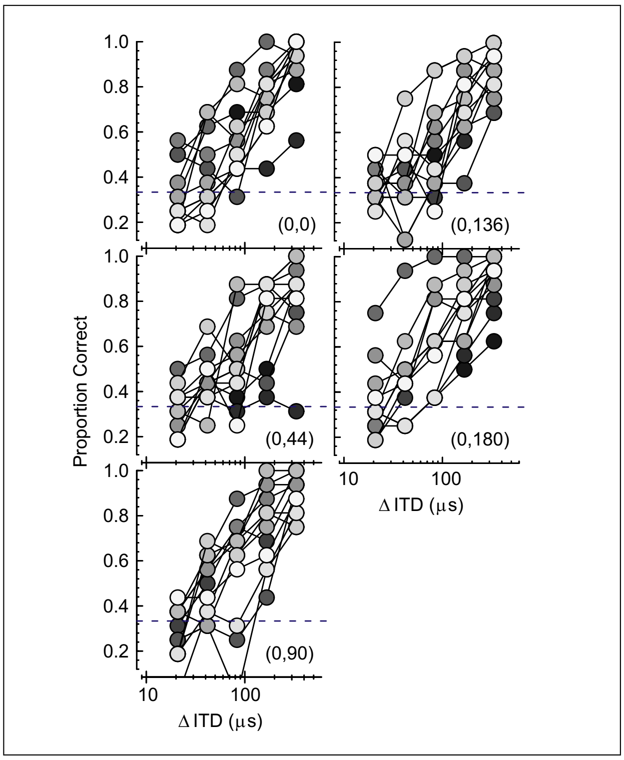

Figure 10. Proportion of correct responses for discrimination of ITDs for the positions in the median plane. Nominal positions are indicated on each panel.

Table II. Average thresholds for discrimination of time differences in HRTFs. Thresholds are given in μs. Average thresholds obtained from less than twelve subjects are shown with a subscript that indicates the number of subjects used to compute the average.

ITD (μs)	Nom. Dir.	Threshold (μs)	
		Addition	Subtraction
-625	(90°,0°)	–	160.8
-437.5	(58°,0°)	137.3	127.2
	(46°,90°)	118.2	163.8_{10}
	(54°,180°)	154.4_{11}	109.2
0	(0°,0°)	104.0_{11}	–
	(0°,44°)	128.1_{11}	–
	(0°,90°)	109.2	–
	(0°,136°)	134.4	–
	(0°,180°)	87.8_{11}	–

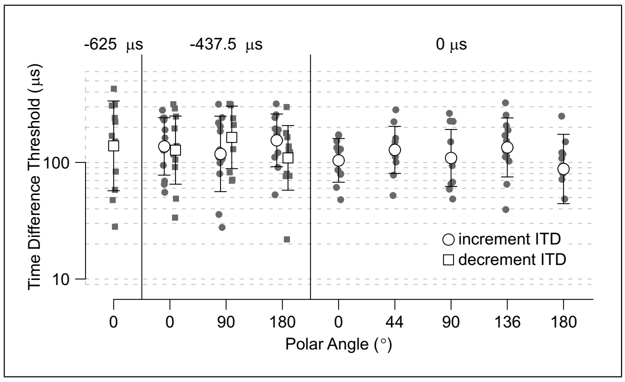

Figure 11. Individual (grey color symbols) and mean thresholds on time differences in HRTFs. Data is grouped by nominal ITD and the abscissa describes the polar angle for a given ITD. Thresholds for increments of ITD (circles) are observed for directions with ITDs 0 μs and -437.5 μs. Thresholds for decrements in ITD (squares) are observed for directions with ITD $-437.5\,\mu$s and $-625\,\mu$s. Error bars indicate 1 ± standard deviation.

sition. For positions in the median plane mean thresholds ranged from 87.8 to 134.4 μs. There was not significant effect of nominal position. For directions in the iso-ITD contour a two-way analysis of variance with sign of ΔITD and nominal position as factors, showed that there was no significant difference between increments and decrements of ITDs nor was the difference between nominal position significant. Mean thresholds ranged from 109.3 to 163.8 μs. For (90°,0°), in which ΔITDs were subtracted from the nominal ITD, the mean threshold was 160.8 μs.

3.3. Discussion

Early experiments on just-noticeable differences in ITDs show that listeners' sensitivity is quite remarkable for stimuli presented in optimal conditions. These experiments found thresholds around 10-20 μs for pure tone signals between 500 Hz and 1 kHz with a reference ITD of 0 μs [25, 26]. For click-like stimuli, thresholds have been found to be in the range of 20-40 μs as the nominal ITD increases from 0 μs to around 500 μs [27]. These values may roughly apply to broadband stimuli.

Our results show mean thresholds in a range of about 87.8–163.8 μs. Differences between our data and the literature may stem from factors such as different types of stimuli and the level of training of the subjects. Regarding differences in stimuli there is the possibility that the filtering imposed by the HRTFs may have had an effect on the thresholds. An unfiltered noise stimuli as a control condition could have helped in revealing any possible influence of the HRTFs. Even though this factor is a perfectly valid possibility, in the authors' view, it seems unlikely that HRTF filtering have had a significant effect.

In terms of subject's experience the difference between our results and previous ones could be because sensitivity to ITDs has often been measured on highly trained, and selected, subjects. This factor is considered as part of the optimal conditions previously mentioned. In the present study subjects did not go through an extensive practice phase but a relatively short practice. Other studies employing subjects with little or no experience have reported thresholds in the range of 70–80 μs [28, 29, 30]. Large differences between subjects have also been observed. In a study by [31] performance on several tasks involving binaural processing was measured. Results on just-noticeable differences in ITD showed that for subjects with extensive experience the range was 9.8–10.2 μs and for less experienced subjects the range was 49.7–102.5 μs. Thresholds obtained here are comparable to those from the less experienced subjects.

4. General discussion

4.1. Comparison between spectral and time thresholds

In this study we attempted to measure the lowest directional resolution — or largest directional change — for which listeners could not distinguish between adjacent directions by using any criterion whatsoever. Performance in the task involving discrimination of changes in ITD was particularly poor, and this may be partially attributed to the naiveness of the listeners regarding tasks involving binaural processing [31]. Approximating ITD thresholds to their corresponding change in degrees, and comparing them to those for spectral differences, indicates that thresholds for spectral differences are substantially lower than those for time differences. This would imply that in terms of pure discrimination listeners give priority to spectral differences over time differences.

A possible explanation for the differences in ITD and spectral thresholds may be the actual number and type of cues available for each one. Recall that subjects could use any potential difference in the stimuli as criterion for discrimination. It is reasonable to think that the most direct cue for the discrimination of ITDs is a lateral displacement in the apparent source position. Also, the fact that changes in ITD were not accompanied by the corresponding changes in ILD may have provided another cue. Presenting ITDs and ILDs in conflict can generate perceptions associated to broadened sound images or even multiple sound images [32]. Experiments about the relative potency of the individual interaural cues, referred to as time-intensity trading and measure in μs/dB, have shown large variability in the reported values (1 to 300 μs/dB) as well as large intersubject differences [33]. This observation may also partially account for the large thresholds and large intersubject variability observed in our ITD thresholds. However, because these time-intensity trading cues may equally apply to the discrimination of spectral differences, at least in terms of changes in ILD, other cues may have mediated discrimination of spectral differences. Additional cues for spectral differences relative to ITD are thought to be loudness and interaural as well as monaural spectral cues.

Spectral differences, and particularly small differences, may first result in a perceived change in timbre, and as the differences increase, a perceived shift in the apparent location of the sound may also occur. This is consistent with a study by Langendijk and Bronkhorst [1] who examined the required spatial resolution for measured HRTFs so that interpolated HRTFs generate the same spatial percept. They found that a resolution of 6° was required in a condition where stimuli level was fixed. In a second condition where the stimuli's spectrum was scrambled, that is, levels at different third-octave bands were randomized so that the use of timbral cues was minimized, the required spatial resolution increased to 10–15°. Our finding for the forward direction that sensitivity to polar changes was higher than to lateral changes is in agreement with this observation.

Here, the audibility of spectral and time differences has been tested separately. For changes in lateral angle, a natural progression of this study would be to examine listener's sensitivity to the combination of both spectral and time differences. Could we be more sensitive to HRTF differences if ITD and spectrum work together at the same time? This paradigm corresponds to a more realistic situation, and thereby it makes possible a more direct comparison with measurements of human spatial resolution such as the MAA.

4.2. Implications in spatial resolution of HRTFs

For the simulation of stationary sound sources in three-dimensional space, our findings suggest that the required spatial resolution for ITD is different than the required spatial resolution for spectral information of HRTFs. Minimum-phase HRTFs require a higher spatial resolution than ITDs implemented as pure delays.

In addition to stationary sound sources, three-dimensional sound systems also incorporate simulation of moving sound. In this context, HRTFs must be constantly updated to compensate for the positional changes of the moving sound. In terms of ITDs, update rates for dynamic ITD changes are typically set equal to the sampling frequency. That is, delays are updated at every new sample, e.g., for a 48-kHz sampling frequency delays would be updated approximately at every 21 μs. Here, the results from discrimination of changes in ITD range between values that are 4–6 times larger than 21 μs, and this would imply that delays can be updated at slower rates. However, the high update rate is mainly because audible artifacts may be produced when switching between ITDs, and, in fact, the audibility of these artifacts becomes a more critical aspect of the required resolution for time-varying delays than the audibility of changes in ITD [34].

In terms of audibility of spectral differences our findings indicate that for high elevations the number of HRTF filters may be reduced as compared to lower elevations. This is in line with the results from Minnaar *et al.* [19] who studied the required directional resolution such that the error introduced by linear interpolation between minimum-phase representation of HRTFs was inaudible. Though our results show that a low resolution is required for elevated positions, it is important to emphasize that for high elevations, sensitivity to changes in HRTF magnitude is more dependent on the direction of change than for lower elevations.

5. Conclusions

For the positions used in this study and for naive listeners, differences between magnitude spectra of adjacent HRTFs become audible at smaller angular separations than those corresponding to changes in ITD. This result can be attributed in part to the fact that changes in ITD constitute an auditory spatial cue only, whereas other non-spatial cues such as changes in timbre are available for the discrimination of, particularly small, spectral changes. Opposite to thresholds for ITD, thresholds for spectral differences change significantly as a function of direction. In summary, some of the implications of these results on synthesis of virtual spatial sound are that, spatial resolution of spectral characteristics depends upon the position and trajectory of the sound source, and that ITDs do not seem to require very high spatial resolutions. In this study, experiments were based on non-individualized HRTFs, and thus, if individualized HRTFs were used, results may differ.

Acknowledgements

Economic support from the Danish Technical Research Council and the Research Council for Technology and Production Science is greatly acknowledged.

References

[1] E. H. A. Langendijk, A. W. Bronkhorst: Fidelity of three-dimensional-sound reproduction using a virtual auditory display. J. Acoust. Soc. Am. **107** (2000) 528–537.
[2] C. I. Cheng, G. H. Wakefield: Introduction to Head Related Transfer Functions (HRTFs): Representations of HRTFs in time, frequency and space. J. Audio Eng. Soc. **49** (2001) 231–249.
[3] F. L. Wightman, D. J. Kistler: Measurement and Validation of Human HRTFs for Use in Hearing Research. Acta Acustica united with Acustica **91** (2005) 429–439.
[4] A. W. Mills: On the minimum audible angle. J. Acoust. Soc. Am. **30** (1958) 237–246.
[5] D. R. Perrott, S. Pacheco: Minimum audible angle thresholds for broadband noise as a function of the delay between the onset of the lead and lag signals. J. Acoust. Soc. Am. **85** (1989) 2669–2672.
[6] D. W. Grantham, B. W. Y. Hornsby, E. A. Erpenbeck: Auditory spatial resolution in horizontal, vertical and diagonal planes. J. Acoust. Soc. Am. **114** (2003) 1009–1022.
[7] D. R. Perrott, K. Saberi: Minimum audible angle thresholds for sources varying in elevation and azimuth. J. Acoust. Soc. Am. **87** (1990) 1728–1731.
[8] R. M. Hershkowitz, N. I. Durlach: Interaural time and amplitude jnds for 500-Hz tone. J. Acoust. Soc. Am. **46** (1969) 1464–1467.
[9] B. P. Bovbjerg, F. Christensen, P. Minnaar, X. Chen: Measuring the head-related transfer functions of an artificial head with a high directional resolution. 109th Convention of the Audio Engineering Society, Los Angeles, California, USA, 2000, convention paper 5264.
[10] J.-M. Jot, V. Larcher, O. Warusfel: Digital Signal Processing in the Context of Binaural and Transaural Stereophony. 98th Convention of the Audio Engineering Society, Paris, France, 1995, convention paper 3980.
[11] A. Kulkarni, S. K. Isabelle, H. S. Colburn: On the minimum-phase approximation of head-related transfer functions. Proc. of the ASSP (IEEE) Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, USA, 1995, 84–87.
[12] P. Minnaar, H. Møller, J. Plogsties, S. K. Olesen, F. Christensen: Audibility of all-pass components in binaural synthesis. 106th Convention of the Audio Engineering Society, Munich, Germany, 1999, convention paper 4911.
[13] J. Plogsties, P. Minnaar, S. K. Olesen, F. Christensen, H. Møller: Audibility of all-pass components in head-related transfer functions. 108th Convention of the Audio Engineering Society, Paris, France, 2000, convention paper 5132.
[14] A. Kulkarni, S. K. Isabelle, H. S. Colburn: Sensitivity of human subjects to head-related transfer function phase spectra. J. Acoust. Soc. Am. **105** (1999) 2821–2840.
[15] D. J. Kistler, F. L. Wightman: A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. J. Acoust. Soc. Am. **91** (1992) 1637–1647.
[16] D. Hammershøi, H. Møller: Binaural technique, basic methods for recording, synthesis and reproduction. – In: Communication Acoustics. J. Blauert (ed.). Springer Verlag, Berlin, Germany, 2005, 223–254.
[17] A. V. Oppenheim, R. W. Schafer: Discrete-time signal processing. Prentice Hall, New Jersey, NJ, USA, 1989.
[18] P. Minnaar, J. Plogsties, S. K. Olesen, F. Christensen, H. Møller: The interaural time difference in binaural synthesis. 108th Convention of the Audio Engineering Society, Paris, France, 2000, convention paper 5133.
[19] P. Minnaar, J. Plogsties, F. Christensen: Directional Resolution of Head-Related Transfer Functions Required in Binaural Synthesis. J. Audio Eng. Soc. **53** (2005) 919–929.
[20] D. M. Bates, D. G. Watts: Nonlinear regression analysis and its aplications. John Wiley & Sons, Inc., New York, NY, USA, 1988.
[21] P. F. Hoffmann, H. Møller: Some observations on sensitivity to HRTF magnitude. J. Aud. Eng. Soc. (2008) in print.
[22] J. Blauert: Spatial hearing: The psychophysics of human sound localization. MIT Press, Cambridge, Massachusetts, USA, 1997.
[23] K. Saberi, L. Dostal, T. Sadralodabai, D. R. Perrott: Minimum audible angles for horizontal, vertical, and oblique orientations: Lateral and dorsal planes. Acustica **75** (1991) 57–61.
[24] A. W. Bronkhorst: Horizontal and vertical MAAs for a wide range of sound source locations (A). J. Acoust. Soc. Am. **93** (1993) 2351.
[25] R. G. Klumpp, H. R. Eady: Some measurements on interaural time difference thresholds. J. Acoust. Soc. Am. **28** (1956) 859–860.
[26] J. Zwislocki, R. S. Feldman: Just noticeable differences in dichotic phase. J. Acoust. Soc. Am. **28** (1956) 860–864.
[27] E. R. Hafter, J. D. Maio: Difference thresholds for interaural delay. J. Acoust. Soc. Am. **57** (1975) 181–187.
[28] J. Koehnke, C. P. Culotta, M. L. Hawley, H. S. Colburn: Effects of reference interaural time and intensity differences on binaural performance in listeners with normal and impaired hearing. Ear Hear. **6** (1995) 331–353.
[29] B. A. Wright, M. B. Fitzgerald: Different patterns of human discrimination learning for two interaural cues to sound-source location. Proc. Natl. Acad. Sci. **98** (2001) 12307–12312.
[30] L. R. Bernstein, C. Trahiotis, E. L. Hyde: Inter-individual differences in binaural detection of low-frequency or high-frequency tonal signals masked by narrow-band or broad-band noise. J. Acoust. Soc. Am. **103** (1998) 2069–2078.
[31] J. Koehnke, H. S. Colburn, N. I. Durlach: Performance in several binaural-interaction experiments. J. Acoust. Soc. Am. **79** (1986) 1558–1562.
[32] E. R. Hafter, L. A. Jeffress: Two-images lateralization of tones and clicks. J. Acoust. Soc. Am. **44** (1968) 563–569.
[33] N. I. Durlach, H. S. Colburn: Binaural phenomena. – In: Handbook of Perception, Vol. IV. E. C. Carterette, M. P. Friedman (eds.). Academic Press, New York, NY, USA, 1978, 366–466.
[34] P. F. Hoffmann, H. Møller: Audibility of direct switching between head-related transfer functions. Acta Acustica united with Acustica **94** (2008) 955–964.