Audibility of time differences in adjacent head-related transfer functions (HRTFs)

Hoffmann, Pablo F.F.; Møller, Henrik

Published in:
Audio Engineering Society Convention Papers

Publication date:
2006

Audio Engineering Society

Convention Paper 6914

Presented at the 121st Convention
2006 October 5–8 San Francisco, CA, USA

Audibility of time differences in adjacent head-related transfer functions

Pablo F. Hoffmann[1] and Henrik Møller[1]

[1] *Department of Acoustics, Aalborg University*

Correspondence should be addressed to Pablo F. Hoffmann (pfh@acoustics.aau.dk)

ABSTRACT

Changes in the temporal and spectral characteristics of the sound reaching the two ears are known to be of great importance for the perception of spatial sound. The smallest change that can be reliably perceived provides a measure of how accurate directional hearing is. The present study investigates audibility of changes in the temporal characteristics of HRTFs. A listening test is conducted to measure the smallest change in the interaural time difference (ITD) that produces an audible difference of any nature. Results show a large inter-individual variation with a range of audibility thresholds from about 20 μs to more than 300 μs.

1. INTRODUCTION

A head-related transfer function (HRTF) can be represented as a pair of minimum-phase filters - one for each ear - with a pure delay cascaded to the contralateral component of the HRTF. This type of delay corresponds to a linear phase function in the frequency domain (frequency-independent delay). The minimum-phase filters produce the same magnitude spectrum of the measured HRTF, thus, they control spectral cues and interaural level difference (ILD) cues. The function of the delay is to control the interaural time difference (ITD). A diagram of this model is shown in Fig. 1. The 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 [1], and sound localization tasks [2], have shown that empirical and modeled HRTFs are indistinguishable. This, provided that the ITD has been calculated correctly [3].

In a three-dimensional sound system, the implementation of ITD as a pure delay is straightforward as long as its required value can be represented as a multiple integer of the sampling time of the system. This means that the ITD is controlled by a delay line whose units are given in samples. For example,

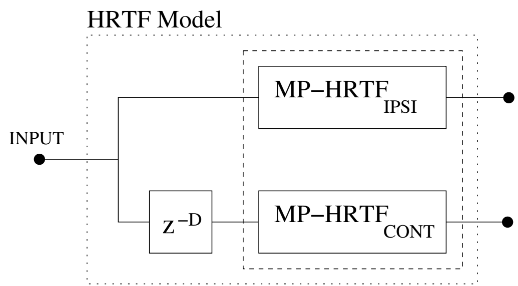

Fig. 1: Minimum-phase and frequency-independent ITD model of the HRTF. Minimum-phase filters are enclosed in the dashed box. The IPSI and CONT sub-indices indicate the ipsilateral and contralateral components respectively. The ITD is implemented by cascading the delay to the contralateral component of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

if a system is running at a sampling frequency of 48 kHz, the smallest ITD that can be achieved is about 20.8 μs (which corresponds to 1/48000 s). This resolution might be sufficient for the simulation of static sound images. For the implementation of scenarios involving moving sound, where the directional information is constantly updated, ITDs may need a higher resolution. This could also include time-varying propagation delays for the simulation of Doppler effects [4]. The objective of increasing the resolution is to avoid audible artifacts during the update [5, 6], and therefore, to allow a smooth transition along the path of the moving sound.

An intuitive approach to increase the time resolution is to increase the sampling frequency of the system, thus, making the sampling time smaller. However, if the required resolution needs to be increased even more, it can be observed that this approach quickly becomes impractical. An alternative approach, which is most commonly used, interpolates between ITDs by using fractional delay filters [7]. These filters can provide delays which are smaller than the sampling time. However, extra care must be taken, since, depending on the implementation, fractional delay filters might also affect the magnitude spectrum; especially for broadband sounds. In addition, the use of fractional delay filters unavoidably increases the complexity of the system when compared to a sample-based implementation. So it is relevant to ask how high this resolution should be set. It seems that perceptual aspects of the auditory system must be revised, so as to find the middle ground between perceptual adequacy of the system and its computational complexity counterpart.

In relation to the use of HRTFs for implementing dynamic three-dimensional sound, a measure of the ability of human listeners to distinguish between adjacent sound locations can provide the perceptual basis for the required spatial resolution of the HRTF filters. Based on the characterization of HRTFs explained in the first paragraph, it is hypothesized that ITD and minimum-phase transfer functions may require different spatial resolutions. Audibility of differences between minimum-phase components of adjacent HRTFs has been investigated in a previous work [8]. Thresholds were found to range from about $\pm 2°$ for directions close to the forward direction, to $\pm 8°$ for directions in the rear and elevated part of the median plane.

This article will report on a psychoacoustic experiment designed to measure audibility of ITD changes when using broadband noise spectrally shaped by different HRTFs. The present study focuses on finding the smallest change in ITD that produces a perceivable difference of any nature.

2. METHOD

2.1. Subjects

Twelve subjects, five males and seven females, participated in the listening experiment. Subjects were paid for their participation. Their ages ranged from 21 to 32. Subjects' normal hearing was assessed by 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.

2.2. 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 D/A converter with 16 bit resolution at a sampling rate of 48 kHz. From the D/A converter the signal went to a stereo amplifier (Pioneer A-616) modified to have a calibrated gain of 0 dB. A passive attenuator of -20 dB was then connected to the output of the amplifier in order to reduce its noise floor. Finally, the stereo output signal from the attenuator was delivered to the listener through a pair of Beyerdynamic DT-990 circumaural headphones. The overall gain of the system was calibrated so as the unprocessed sound (see 2.3) generates an equivalent free-field sound pressure level of 68 dB approximately.

2.2.1. Headphone Equalization

Two 256-coefficients minimum-phase filters were employed in order to compensate for the left and right headphone transfer functions respectively. The design of the equalization filters was based on headphone transfer functions (PTFs) measured on blocked ear canal from 23 subjects (none of them participated in this listening test). Five PTFs were obtained from each ear and subject, and subjects were asked to reposition the headphones between measurements. PTFs were measured by using a maximum-length sequence technique (MLS) [9]. PTFs were averaged on a sound power basis, and a minimum-phase representation of the inverse of the average PTF was computed for each ear. More details on the measurement and equalization techniques used can be found in [10].

2.3. Stimuli

Five minutes of broadband pink noise, with a bandwidth of 20-16000 Hz, was digitally generated at 48 kHz (unprocessed sound), convolved with the headphone equalization filters and stored as an audio file.

2.3.1. Directional sound synthesis

To simulate directional sound, HRTFs measured with a high resolution on an artificial head were used [11]. Nine directions were selected in the left half of the upper hemisphere. Directions are given as (azimuth ϕ, elevation θ) in a polar coordinate system with interaural axis and left-right poles. This system is also referred to as the interaural-polar coordinate system. Its use is advantageous under the assumption that the azimuth indicates the lateral position of a contour with constant ITD, i.e. an iso-ITD contour, and the elevation angle indicates

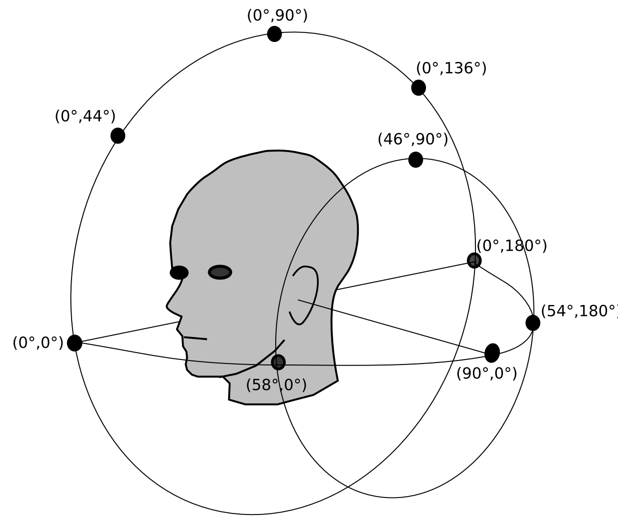

Fig. 2: Reference directions employed in the listening experiment. Azimuth and elevation are indicated in an interaural-polar coordinate system.

the precise location in the contour [12]. In this system 90° and -90° azimuth correspond to left and right sides, 0° elevation to the frontal part of the horizontal plane, 180° elevation to the rear part of the horizontal plane, and 90° elevation to the upper part of the frontal plane. Five directions were selected in the median plane (0° azimuth; 0°, 44°, 90°, 136° and 180° elevation). Three directions were selected on a cone of confusion ((58°, 0°), (46°, 90°) and (54°, 180°)). These three directions were chosen to be in an iso-ITD contour rather than being on the same geometrical cone, thus, their azimuth varies with elevation. The direction at 90° azimuth was also selected. In the following of this article, these directions are referred to as the *reference directions*. Fig. 2 shows these reference directions.

2.3.2. HRTFs processing

The measured HRTFs were available as impulse responses of 256 coefficients. These impulses responses included an initial delay due to the distance from the loudspeaker to the in-ear microphones in the artificial head when the measurements were carried out. The initial delay was removed and the remaining part of the impulse response was truncated using a rectangular window of 72 coefficients. This corre-

Table 1: Reference directions and their associated ITDs. These ITD values correspond to the reference ITDs used in the listening experiment.

Reference direction (ϕ, θ)	ITD (μs)
(0°,0°)	0
(0°,44°)	
(0°,90°)	
(0°,136°)	
(0°,180°)	
(58°,0°)	-437.5
(46°,90°)	
(54°,180°)	
(90°,0°)	-625

sponds to a length of 1.5 ms at 48 kHz. It has been shown that for noise-like stimuli this length is sufficient to avoid audible effects of the truncation [13]. The DC value of each new impulse response was set to unity gain (section 5.2 in [14]). Minimum-phase representations were calculated using homomorphic filtering as described in [15]. ITD values were derived from the interaural differences in group delay of the excess-phase components of the original HRTFs evaluated at 0 Hz [3]. The obtained ITDs were rounded to the nearest sample and inserted to the contralateral component of the minimum-phase impulse responses. These ITDs are also referred to as *reference ITDs*. Table 1 shows the calculated ITDs for the selected directions.

2.4. Psychometric Method

Audibility of time differences were determined in a three-interval, three-alternative forced-choice task using the method of constant stimulus. Both the length of the stimulus and the inter-stimulus interval were 300 ms. On a single trial, a 300-ms segment from the headphone-equalized pink-noise sequence was randomly selected, and filtered with a reference direction. Raised-cosine fade-in and -out ramps of 10 ms were applied to the segment. In one of the intervals, which was selected at random with equal *a priori* probability, the right channel of the stimulus was shifted by a pure delay relative to the left channel. 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. Feedback lights were used to immediately show the correct response to the subjects. After a silence interval of 1 s a new trial was presented.

The pure delay used to generate the target stimulus was selected from a set of five values. These values are referred to as ΔITDs in the remaining of this article. ΔITDs, in absolute value, were 20.8, 41.6, 83.3, 166.6, 333.3 μs, which correspond to 1, 2, 4, 8 and 16 samples at 48 kHz respectively. For the directions located in the iso-ITD contour ΔITDs could be applied in two modes. One mode corresponded to a negative ΔITD in which the reference ITD was incremented by an amount equal to the selected ΔITD. The other mode corresponded to a positive ΔITD, meaning that the delay difference between the channels was decremented by an amount equal to the selected ΔITD. Since for directions in the median plane the reference ITD is 0 μs, ΔITDs were applied in order to increase the ITD (left channel was always leading in the target stimulus). For the direction at 90° azimuth ΔITDs were applied so the reference ITD was decremented. Including reference directions and the modes in which ΔITDs were applied, a total of twelve conditions were tested.

2.5. Experimental Design

Subjects were 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 they were to perform. Subjects were then presented with a few trials in order to acquaint them with the task and the procedure. Posteriorly, a run consisting of sixteen trials was employed as practice. The forward direction was used as reference, and only a ΔITD of 333.3 μs was applied to generate the stimulus that deviates. It was assumed that this ΔITD was large enough so the deviating stimulus was easy to discriminate. Practice runs were repeated until subjects could respond correctly at least fifteen out of the sixteen trials. In general, practice sessions lasted from about 30 to 45 minutes. Since the purpose of the experiment was to use naive subjects no further practice (or any kind of training) was performed.

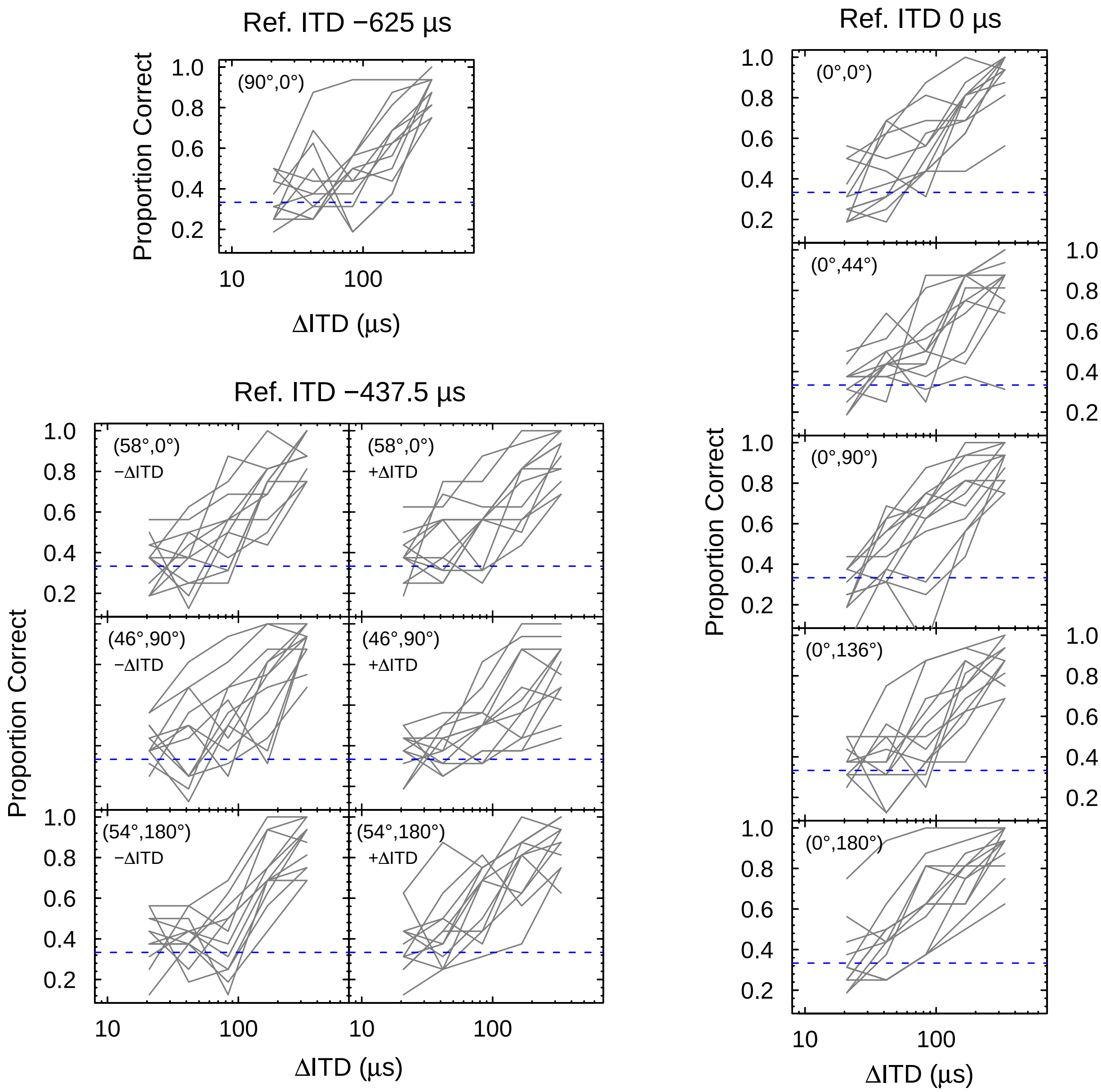

Fig. 3: Psychometric functions of all subjects and conditions. Psychometric functions from one condition are plotted on a single panel. The panels are grouped by reference ITD. The upper-left panel shows psychometric functions for the direction at 90° azimuth. The group of six panels in the lower-left shows psychometric functions for directions in the iso-ITD contour. Negative and positive ΔITDs indicate increment and decrement from the reference ITD respectively. The five panels to the right show psychometric functions for the directions in the median plane. The dashed line indicates chance level.

In the main experiment, runs of 100 trials were employed to test each condition. Reference direction was held constant within a run, and sixteen repetitions were obtained from each ΔITD. The order in which they were presented was randomized. Sixteen "catch trials", in which the three intervals were the same, were also included. At the beginning of each run four trials using a ΔITD of 416 μs (20 samples) were used as warm up trials. A run took between 7 to 8 minutes to complete. After a run was finished subjects were instructed to remove the headphones. A pause of between 1 to 2 minutes was used between runs. After three runs were completed subjects were instructed to hold a break.

3. RESULTS

Psychometric functions of all tested conditions are shown in Fig. 3. The abscissa specifies the ΔITD in μs, and is given in a logarithmic scale. The ordinate specifies the proportion of correct responses, thus giving an estimate of subject's performance per condition. The chance level corresponds to the proportion 0.33 (average and standard deviation of "catch trials" across subjects and conditions was 0.31 $\pm$0.04).

4. DISCUSSION

With a few exceptions, the general tendency observed in the psychometric functions is a shallow slope. This indicates that subjects were not as sensitive to changes in ITD as would have been expected from results obtained in previous studies. Early experiments on just-noticeable differences in ITDs show that listeners' sensitivity seems to be 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 [16, 17]. For click-like stimuli, thresholds have been found to be in the range of 20-40 μs as the reference ITD increases from 0 μs to around 500 μs [18]. These values roughly apply to broadband stimuli. However, a rise in threshold values of up to 70-80 μs can be observed for subjects with little or no experience [19, 20]. But generally, results show that thresholds tend to remain below 100 μs. Note that this situation is commonly observed when the stimuli employed contain energy at low frequencies [21]. Much larger thresholds are observed when stimuli contain only high-frequency components.

Insensitivity to large ΔITDs constitutes another observation, which is surprising. It was expected that the point of the psychometric function associated to the largest ΔITD (333.3 μs) could reach 1 (perfect performance), or at least be within an acceptable range, say between 0.8 and 1, for all conditions and subjects. However, there were several subjects, and for various conditions, who did not reach the expected level of performance for this ΔITD. It is possible that subjects had difficulty in determining what to listen for in order to detect the deviating sound. In general, subjects reported that the task was very difficult, and only some of them reported that the cue they used to make their judgments was a shift in the apparent position of the sound image. These reports, along with the shallow slopes observed in the psychometric functions, suggest that the task might have been difficult during a major part of a run. Long periods of difficulty are likely to make subjects exhausted, and that could have affected performance at the largest ΔITD, for which the discrimination task was assumed to be simple.

In order to obtain threshold estimations, a threshold was defined as the ΔITD that would produce a proportion of correct discriminations equal to 0.66. The number of repetitions for each ΔITD was selected to be slightly above the minimum number of repetitions necessary to produce a statistically significant difference between 0.66 and chance level ($p < 0.01$). Individual thresholds were computed by finding the two measured points in the psychometric function that bracket this proportion, and linearly interpolating against ΔITD. Frequency distributions of the obtained thresholds are shown in Fig. 4. A substantial variability across subjects is observed. A small inter-individual variation had been reflected in distributions showing thresholds clustered around a common value. Here, subjects' sensitivity to changes in ITD spanned along a wide range. A large spread within virtually all conditions occur, and the range observed is from about 20 μs to more than 300 μs. This makes it difficult to evaluate on average whether there is a systematic variation in sensitivity as a function of

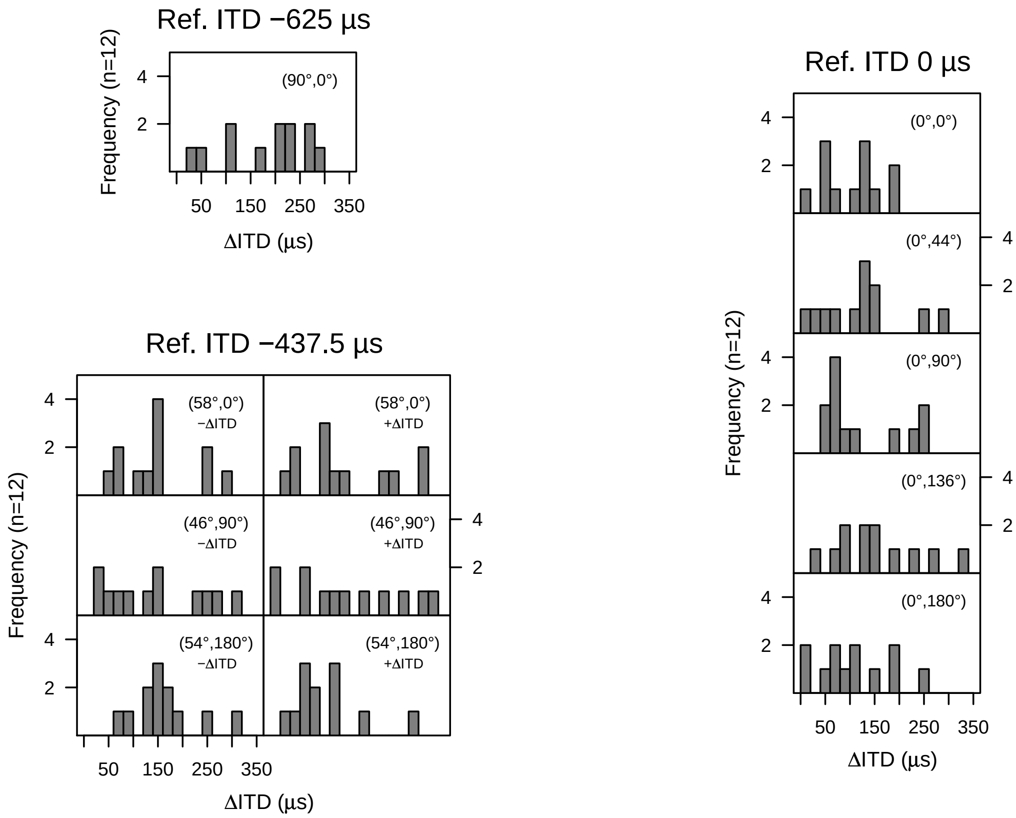

Fig. 4: Frequency distributions of the estimated thresholds for the different conditions employed in this experiment. The upper-left panel shows distribution of thresholds at 90°azimuth direction. Distribution of thresholds for directions in the iso-ITD contour are shown in the lower-left group of panels. Panels to the right show distributions for directions in the median plane.

reference ITD, and as function of spectral changes for directions in an iso-ITD contour.

From the thresholds obtained and the variation across conditions observed, subjects could roughly be classified into three groups of different sensitivity. There were two subjects who showed to be the most sensitive for almost all conditions. A group of five subjects showed to be fairly sensitive, mainly due to variations across conditions. The remaining subjects were generally insensitive. One possible cause of these variations could be attributed to the minimal practice given to the subjects. It is probable that with a more extensive practice period, or training, subjects would improve their performance. They might also produce more comparable results. However, the main purpose of this work was to study the performance of "average" subjects with a minimal amount of practice. The reason is that this condition might be more likely to exist on users of applications that may incorporate three-dimensional sound, e.g. teleconferencing, video-games.

In the context of dynamically varying ITD implementation, it seems worthy to compare results between time differences in HRTFs, and time switching between HRTFs. For time differences, the threshold estimated for the forward direction as the average between the two most sensitive subjects is 44.7 μs. For time switching, the threshold measured for the same direction is 5-6 μs [6]. The difference between both measurements has a factor of 8 approximately. Since thresholds of time switching are below 10 μs for all the other reference directions employed in this experiment, and assuming that the highest sensitivity to ITD changes is in the forward direction, the difference observed for this direction could, at a first approximation, be generalized to the other directions. Therefore, it appears that the requirements for time resolution in the implementation of ITD are significantly more demanding for time switching between HRTFs than for time differences in HRTFs.

As a summary, measured psychometric functions showed a shallow slope for almost all conditions and most of the subjects. There were large inter-individual variations in performance, making difficult to further analyze average tendencies. Only two subjects produced thresholds that are comparable to what is known from the literature on sensitivity to changes in ITD. A comparison between thresholds from these two subjects and thresholds on time switching suggests that, at a first approximation, time switching imposes higher demands on the needed spatial resolution for the implementation of ITD in three-dimensional sound systems.

5. ACKNOWLEDGMENTS

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

6. REFERENCES

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

[2] D. J. Kistler and F. L. Wightman. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *J. Acoust. Soc. Am.*, 91(3):1637–1647, 1992.

[3] P. Minnaar, J. Plogsties, S. K. Olesen, F. Christensen, and H. Møller. The interaural time difference in binaural synthesis. In *108th AES Convention*, Paris, France, February 19-22 2000. Preprint 5133.

[4] H. Strauss. Implementing Doppler shifts for virtual auditory environments. In *104th AES Convention*, Amsterdam, The Netherlands, May 16-19 1998. Preprint 4687.

[5] P. F. Hoffmann and H. Møller. Audibility of time switching in dynamic binaural synthesis. In *118th AES Convention*, Barcelona, Spain, May 27-31 2005. Preprint 6326.

[6] P. F. Hoffmann and H. Møller. Audibility of spectral switching in head-related transfer functions. In *119th AES Convention*, New York, USA, October 7-10 2005. Preprint 6537.

[7] T. I. Laakso, V. Välimäki, M. Karjalainen, and U. K. Laine. Splitting the unit delay - tools for fractional delay filter design. *IEEE Signal Processing Magazine*, 13(1), January 1996.

[8] P. F. Hoffmann and H. Møller. Audibility of spectral differences in head-related transfer functions. In *120th AES Convention*, Paris, France, May 20-23 2006. Preprint 6652.

[9] D. D. Rife and J. Vanderkooy. Transfer-function measurement with maximum-length sequence. *J. Audio Eng. Soc.*, 37(6):419–444, 1989.

[10] H. Møller, D. Hammershøi, C. B. Jensen, and M. F. Sørensen. Transfer characteristics of headphones measured on human ears. *J. Audio Eng. Soc.*, 43(4), April 1995.

[11] B. P. Bovbjerg, F. Christensen, P. Minnaar, and X. Chen. Measuring the head-related transfer functions of an artificial head with a high directional resolution. In *109th AES Convention*, Los Angeles, California, USA, September 22-25 2000. Preprint 5264.

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

[13] J. Sandvad and D. Hammershøi. What is the most efficient way of representing HTF filters?

In *Proceedings of Nordic Signal Processing Symposium*, NORSIG '94, Lesund, Norway, June 2-4 1994.

[14] D. Hammershøi and H. Møller. *Communication Acoustics*, chapter in Jens Blauert (edt.) : Binaural Technique, Basic Methods for Recording, Synthesis and Reproduction, pages 223–254. Springer Verlag, 2005. ISBN 3-540-22162-x.

[15] D. A. Oppenheim, R. W. Schafer, and J. R. Buck. *Discrete-Time Signal Processing*. Prentice Hall, 2nd edition, 1999.

[16] R. G. Klumpp and H. R. Eady. Some measurements on interaural time difference thresholds. *J. Acoust. Soc. Am.*, 28(5):859–860, September 1956.

[17] J. Zwislocki and R. S. Feldman. Just noticeable differences in dichotic phase. *J. Acoust. Soc. Am.*, 28(5):860–864, September 1956.

[18] E. R. Hafter and J. De Maio. Difference thresholds for interaural delay. *J. Acoust. Soc. Am.*, 57(1):181–187, 1975.

[19] J. Koehnke, C. P. Culotta, M. L. Hawley, and H. S. Colburn. Effects of reference interaural time and intensity differences on binaural performance in listeners with normal and impaired hearing. *Ear Hear.*, 6(4):331–353, 1995.

[20] B. A. Wright and M. B. Fitzgerald. Different patterns of human discrimination learning for two interaural cues to sound-source location. *Proc. Natl. Acad. Sci.*, 98(21):12307–12312, 2001.

[21] W. A. Yost, F. L. Wightman, and D. M. Green. Lateralization of filtered clicks. *J. Acoust. Soc. Am.*, 50(6):1526–1531, 1971.