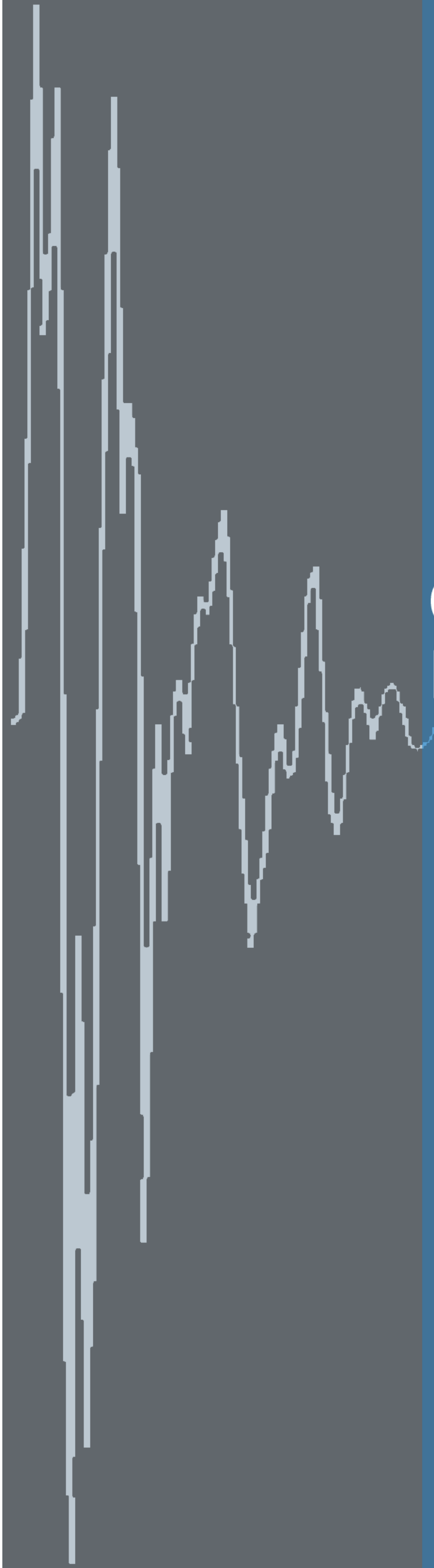
Characteristics of head-related transfer functions

discrimination and switching between adjacent directions

Hoffmann, Pablo F.



Characteristics of Head-Related Transfer Functions

Discrimination and Switching between Adjacent Directions

Pablo F. Hoffmann

Characteristics of Head-Related Transfer Functions (HRTFs) — Discrimination and Switching between Adjacent Directions

Pablo Faundez Hoffmann

August, 2007

Preface

This thesis has been submitted to the Faculty of Engineering, Science and Medicine at Aalborg University for partial fulfillment of the requirements for the award of the PhD degree. The research was conducted as part of the research framework at the university's Section (former Department) of Acoustics in the period from November 2003 to January 2007.

I want to thank all the section acoustics for such fine and inspiring working atmosphere. I want to thank Søren L. Buhl for his help in statistical methods. I also want to thank Flemming Christensen for helping me in the preparation of the resumé. Special thanks to my supervisor Henrik Møller for his valuable advice and guidance. Economic support from the Danish Technical Research Council and the Research Council for Technology and Production Science is greatly acknowledged.

I am most indebted to my family Mabel and Ivana. To Mabel for her constant support and comprehension during some very stressing periods in the preparation of this thesis. To Ivana for being the joy of my life.

Pablo Faundez Hoffmann
Aalborg, 2007

Summary

The sound transmission from the free field to the ears is described by the *head-related transfer function*, or HRTF. Measurement of the HRTF typically results in a time-domain representation, i.e. an impulse response, from which the frequency response can be computed using the Fourier transform. It is well known that the HRTF depends on frequency, listener and sound direction. Particularly because of the dependency on direction, a significant application of the HRTF is in three-dimensional sound synthesis. This is accomplished by convolving the HRTF in the time domain with an anechoic recording, and then delivering the result typically over headphones. If done properly, the sound is perceived as coming from the direction corresponding to the HRTF.

In several three-dimensional sound applications a dynamic scenario may be desired. That is, synthesis of moving sound such as moving sources, compensation for listeners (or listener's head) movements in interactive systems, or both. When sound moves, its position relative to the listener changes over time, and thereby the HRTF needs to be constantly updated. It is clear that in such dynamic scenario some switching between HRTFs must take place, and because the spatial resolution of measured HRTFs is limited, one can only switch between discrete directions. In this context, there are basically two perceptual criteria that should be considered. First, one should ensure that HRTFs are close enough so that differences between adjacent directions are not audible. Second, switching between discrete directions produces artifacts, and these should also be inaudible. This Ph.D. thesis investigates audibility thresholds for differences and switching in HRTFs. Thresholds have been estimated using an experimental paradigm in which discrimination was based on any possible cue.

Characteristics of the HRTF were grouped into two categories, one corresponding to time characteristics and the other to spectral characteristics. Time characteristics relate to the interaural time difference, or ITD, whose importance in sound localization has long been acknowledged. Spectral characteristics relate to the frequency-domain HRTF's magnitude to both ears, and hence to the spectral differences between the inputs to the two ears as well. The HRTF was modeled as a minimum-phase filter with a frequency-independent delay as ITD. In this way, the minimum-phase filter controlled the spectral characteristics and the delay controlled the time characteristics. In this way, thresholds for these characteristics were measured separately.

Audibility thresholds for differences in adjacent HRTFs were measured at several directions and for differences in elevation and differences in azimuth. Elevation and azimuth were described using an interaural-polar coordinate system. In one experiment naive listeners were required to discriminate differences in the spectral shape of the sound. Differences were produced by changes in the minimum-phase filters of HRTFs, while the ITD remained constant. In a second experiment listeners had to discriminate changes in ITD while the spectral information remained unchanged. Audibility of spectral differences was highly dependent on direction and also on the mode of change, i.e. azimuth and elevation. Large thresholds were observed at high elevations. Listeners showed a relatively low sensitivity to differences in ITD, and this was not significantly dependent on sound direction. Implications of these findings for three-dimensional sound systems aiming at a large population are (1) timing information in HRTFs do not require a very high resolution; and (2) spectral information requires different resolutions depending on sound direction.

Similar to the experimental design for HRTF differences, two experiments were conducted for switching between adjacent HRTFs. One experiment for the audibility of time switching and another for the audibility of spectral switching. The thresholds estimated from these two experiments were defined as the *minimum audible time switching* (MATS), and the *minimum audible spectral switching* (MASS) respectively. Listeners were surprisingly sensitive to time switching with mean MATS thresholds more than ten times smaller than those observed for time differences. Moreover, a comparison between MATSs and MASSs revealed that in general time switching becomes audible at smaller angular shifts than those for spectral switching. MASS thresholds were comparable to those from spectral differences suggesting that switching does not produce audible artifacts. Therefore, there are no additional requirements from spectral switching to the spatial resolution of the spectral information in HRTFs.

In short, when audibility of differences in HRTFs is based on any available cue, sensitive to spectral differences is higher than sensitivity to differences in the timing characteristics. The strategy of direct switching between HRTFs is viable for spectral switching, i.e., switching between the minimum-phase filters. Spectral switching does not require a denser representation of space than the one required to make spectral differences inaudible. In contrast, time switching requires a higher spatial resolution than the resolution required for time differences.

Resumé (Summary in Danish)

Lydtransmissionen fra et frit felt til en lytters ører å beskrives ved head-related transfer functions, eller HRTFer. Måling af HRTFer giver typisk resultatet i tidsdomænet, dvs. i form af en impulsrespons, hvorfra repræsentationen i frekvensdomænet kan findes ved en Fourier transformation. Det er velkendt at HRTFer afhænger af frekvens, lytter, og lydens retning. Specielt på grund af retningsafhængigheden er en vigtig anvendelse af HRTFer ved syntese af tredimensional lyd. Dette udføres ved at folde en HRTF givet i tidsdomænet med et lyddødt kildesignal. Resultatet reproduceres typisk over hovedtelefoner. Hvis det er gjort korrekt, opfattes lyden som kommende fra retningen svarende til HRTFen.

I adskillige anvendelser af tredimensional lyd kan et dynamisk scenarium være at foretrække. Det kan være syntese af bevægelige lydkilder, kompensation for lytterens (eller lytterens hoveds) bevægelser i interaktive systemer, eller begge dele. Når en lydkilde bevæger sig, ændres dens placering i forhold til lytteren over tid, hvorfor HRTFer skal opdateres løbende. Det er klart, at under sådanne dynamiske omstændigheder er det nødvendigt at foretage udskiftning af HRTFer, og fordi den rumlige opløsning af HRTFer er begrænset, kan der kun skiftes mellem diskrete retninger. I denne sammenhæng er der grundlæggende to perceptuelle kriterier at tage i betragtning. For det første skal det sikres, at HRTFerne er så tætte på hinanden, at forskellen mellem nærliggende retninger ikke er hørbar. For det andet skaber skift mellem diskrete retninger fejl, som ikke må være hørbare. I denne PhD afhandling undersøges tærsklen for hørbarhed af forskelle og skift mellem HRTFer. Der er estimeret tærskler gennem et eksperimentelt paradigme, i hvilket diskriminationen kan være baseret på enhver hørbar forskel.

Egenskaberne af HRTF blev klassificeret i to kategorier. En kategori for temporale egenskaber og en anden kategori for spektrale egenskaber. Temporale egenskaber relaterer sig til interaural time difference, eller ITD, hvis vigtighed i forhold til lokalisation af lydkilder længe har været kendt. Spektrale egenskaber relaterer sig til HRTFers amplitude i begge ører, og omfatter derfor også spektrale forskelle mellem de to ører. Tærskler for disse to egenskaber blev målt separat. HRTFerne blev modelleret med minimumfase filtre plus en frekvensuafhængig tidsforsinkelse som ITD. Minimumfasefiltrene kontrollerede de spektrale egenskaber, og tidsforsinkelsen kontrollerede de tidsmæssige karakteristika.

Hørbarhed af forskelle i nærliggende HRTFer blev målt for adskillige retninger og for forskelle i elevation og azimut ved hjælp af et interaural-polært koordinatsystem. I ét eksperiment skulle utrænede lyttere diskriminere ændringer i det spektrale indhold af lyd. Ændringerne blev skabt ved ændring af HRTFernes minimumfase filtre, mens ITD forblev

konstant. I et andet eksperiment skulle lytterne diskriminere ændringer i ITD, mens den spektrale information var konstant. Hørbarheden af forskelle i de spektrale egenskaber var meget afhængig af lydretning og også af retningsændring. Store tærskler blev observeret ved høje elevationer. Lytterne var relativt ufølsomme overfor forandringer i ITD, og dette afhang ikke signifikant af lydretning. Betydningen af disse resultater, i forhold til tredimensionale lydsystemer beregnet til en bred brugergruppe er, at (1) tidsinformation i HRTFer behøver ikke forelægge med særlig stor rumlig opløsning, og at (2) spektral information kræver forskellig opløsning afhængig af lydretning.

Tilsvarende det eksperimentelle paradigme for forskelle i HRTFer, blev der foretaget to eksperimenter med skift mellem HRTFer. Et eksperiment vedrørende tærskler for hørbarhed af temporale skift og et andet eksperiment om spektrale skift. Tærskler blev defineret som minimum audible time switching (MATS), og minimum audible spectral switching (MASS). Lytterne var overraskende følsomme overfor temporale skift, hvor i gennemsnit MATSer var over ti gange mindre end værdierne for temporale forskelle. Ydermere, afslørede en sammenligning mellem MATSer og MASSer at temporale skift generelt er hørbare ved mindre skiftevinkler end spektrale skift. MASS tærskler var sammenlignelige med tærskler for spektrale forskelle, hvilket antyder at skifteprocessen i sig selv ikke introducerer hørbare fejl. Derfor er der ikke yderligere krav til rumlig opløsning stammende fra spektrale skift i forhold til krav stammende fra spektral information i HRTFer.

Sammenfattet er følsomheden — målt i vinkel — for spektrale forskelle højere end følsomheden for forskelle i temporale egenskaber, når enhver hørbar forskel i HRTFerne ligger til grund. Strategien med direkte skift mellem HRTFer er brugbar for spektrale skift, det vil sige, skift mellem minimumfase filtre. Spektrale skift kræver ikke yderligere rumlig opløsning end hvad der skal til for at undgå hørbare spektrale forskelle. Derimod kræver temporale skift en signifikant forhøjelse af den krævede rumlige opløsning af temporal information.

Contents

Chapter 1

Introduction

The sense of hearing endow us with the extraordinary ability to transform those small pressure variations captured by our ears into a complete and coherent auditory image of the world around. But, how do we do that?; and what type of information does the auditory system use in order to provide us with this ability of *spatial hearing.* At the beginning of the 20th century a British physicist named John William Strutt — most widely known as *Lord Raleigh* — showed that time and intensity differences between the input to each ear were essential for the localization of sounds. These two binaural cues are identified as the *interaural time difference* (ITD) and *interaural level difference* (ILD). Further insight into the understanding of auditory spatial perception has acknowledged the importance of the spectral transformation produced by the external ears. These spectral transformations are regarded as monaural cues because they are extracted from sound received at one ear only. There is now a general view that binaural cues primarily contribute to the perception of the horizontal angle of the sound, and monaural cues provide the basis for determining the elevation of a sound in addition to front-back perception. If one is able to reproduce an identical copy of the sound at the ears produced by a real source, one would expect to provide the listeners with all these spatial cues, and thus to elicit the same spatial percept as the one produced by the real source. This constitutes the fundamental idea of *binaural technology.*

1.1 Binaural Technology

Although not perfect, a compelling illusion of spatial sound can simply be realized by listening to the binaural signals captured by microphones placed at the ears of an artificial head that is located on a different environment. The reason for not being perfect is that the artificial head is not a replica of yourself. Binaural technology is based on the assumption that equivalent sound stimuli generate equivalent percepts. If one can record the sound that occur at the two eardrums correctly, transmit them through an equalized chain and faithfully reproduce them, one would expect that the same auditory percept as the one produced by the original real sound field is elicited (Møller, 1992; Blauert, 1997) (for a recent review the reader is referred to Hammershøi and Møller, 2005). In this context, binaural recordings are measured at the ears of a listener and commonly reproduced over headphones since they offer the advantage of complete channel separation as compared to loudspeakers. Furthermore, in terms of studying auditory perception the use of headphones provides complete and accurate control over the acoustic stimulus delivered to the listener's ears. As an alternative to binaural recordings, the transfer function describing the acoustic transformation from the free field to the listener's ears can be measured and used to synthesize binaural signals.

1.2 The Head-Related Transfer Function

As a sound wave reaches the ears of a listener the acoustical properties of the sound are modified by the torso, head, pinna and ear canal before it hits the listener's eardrums. Although these modifications certainly change the original source's spectrum, they are generally not perceived as changes in the quality of the sound but actually as changes related to the direction of the source relative to the listener's head. The direction-dependant acoustic transformation from the free field to the eardrums is completely described by the *Head-Related Transfer Function (HRTF)*. Thus, the HRTF contains all binaural and monaural cues involved in spatial hearing.

The HRTF is computed as the ratio between the complex sound pressure measured at the eardrum and the complex sound pressure at the center of the head with the head absent. By doing this division and assuming that exactly the same system was used to do both sound-pressure measurements, the transfer function of the measuring system — which generally constitutes microphones, amplifiers, analog-to-digital and digital-to-analog converters, and loudspeakers — is eliminated. For distances of about one meter or greater, the HRTF is mainly function of direction and not of distance because the incidence wave is close to a plane wave. For distances closer than roughly one meter the HRTF is also function of distance, or range, and it has been referred to as near-field HRTF (Brungart and Rabinowitz, 1999). HRTFs are commonly specified in terms of spherical coordinates (azimuth θ, elevation ϕ) using a head-related system with its origin at the center of the head. For example, (0°,0°) would correspond to the direction in front of the listener. In this system three planes that intersect at the origin of the head are defined: the horizontal plane, that separates up/down directions; the median plane, that separates left/right directions; and the frontal plane, that separates front/back directions. The azimuth indicates the angular distance from the median plane along the horizontal plane. The elevation indicates the angular distance from the horizontal plane along a vertical plane that intersects with the vertical axis.

Although HRTFs have been defined as measured at the eardrums, it has been demonstrated that the effect of the ear canal is mostly independent of the direction of sound (Middlebrooks *et al.*, 1989; Hammershøi and Møller, 1996), and thereby HRTFs can be successfully measured at the entrance of the blocked-ear canal (Hammershøi and Møller, 1996). Here, the word “successfully" is used to emphasize the fact that all directional information is preserved. In addition, inter-subject variation is reduced as compared to measurements with the open ear canal (Møller *et al.*, 1995b). Figure 1.1 shows an example of an HRTF for 30° of azimuth and 0° of elevation. The left-hand panel presents the time representation

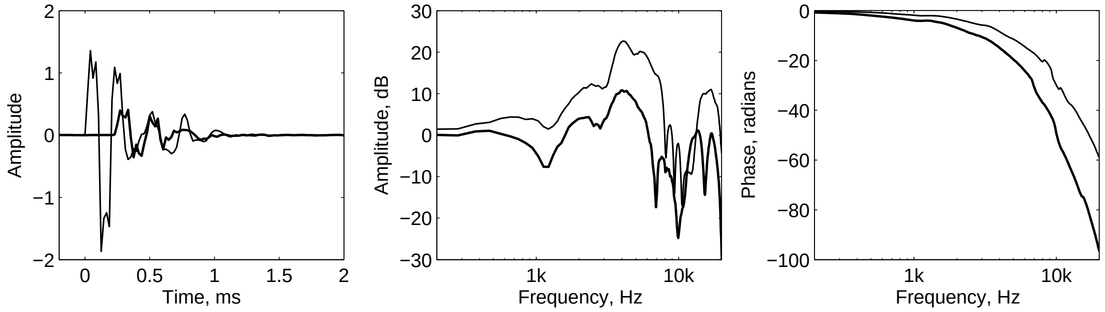


Figure 1.1: Measured HRTF corresponding to a direction of 30° azimuth and 0° elevation. Head-related impulse response, HRTF magnitude response and HRTF phase response are shown on different panels from left to right. The thin line represents the ear closer to the source (ipsilateral component of the HRTF). The starting point of the ipsilateral component is set as the time origin. The thick line represents the ear farther from the source (contralateral component of the HRTF).

of the HRTF or head-related impulse response (HRIR), the center panel shows the magnitude spectrum, and the right-hand panel shows the phase spectrum. ITDs can be observed from the differences in starting point of the impulse responses and from the differences in phase. Note also the complex spectral pattern characterized by peaks and notches. Because HRTFs capture these spectral patterns on both ears, they also capture the interaural spectral difference, which are more commonly referred to as the frequency-dependent ILD.

1.2.1 Binaural synthesis

One of the branches of binaural technology referred to as *binaural synthesis* aims at generating spatial sound by means of digital filters representing HRTFs. The traditional assumption that the ensemble of the external ear, head and torso constitutes a linear system, makes possible to use HRTFs as linear filters for any audio signal. The synthesis is performed by convolving an “anechoic" sound with a pair of HRTF filters in the time domain — one filter for each ear — and delivering the result typically over headphones. In this context, it is crucial that the headphones' transfer functions are correctly equalized for an authentic auditory reproduction (Møller *et al.*, 1995a; Pralong and Carlile, 1996). Therefore, if done properly, the sound can be made to seem as coming from a source positioned somewhere in space outside the head. There is substantial evidence indicating that a high degree of realism can be achieved with binaural synthesis. The highest degree of realism is generated when HRTFs from the same listener are employed for the synthesis (Wightman *et al.*, 1987). In this way, exactly the same spatial cues to which a person has grown familiar with will be available. Psychophysical studies using a discrimination paradigm, have shown that for stimuli synthesized with the listener's own HRTFs, listeners cannot discern between real a virtual sources (Zahorik *et al.*, 1995; Langendijk and Bronkhorst, 2000). Likewise, in localization experiments, errors in localization for real and virtual sources presented from the same direction are not significantly different (Wightman and Kistler, 1989; Bronkhorst, 1995). Some of the applications of binaural synthesis can be found in areas such as video games, virtual auditory displays, warning systems in fighter aircrafts, teleconferencing and other immersive environments.

An important aspect in the design of HRTF-based systems is the degree to which the general population of listeners can be provided with adequate spatial cues to create a convincing auditory experience. In this respect, one obvious disadvantage in the use of the listeners' own HRTFs, is its lack of feasibility due to the very time consuming task of measuring a complete set of HRTFs for each potential user of the system. Although some partially successful attempts have been made to make HRTF measurements more efficient (see e.g. Zotkin *et al.*, 2006), a more practical and common approach is the use of generic sets of HRTFs. These HRTFs are generally obtained from artificial heads (Gardner and Martin, 1995; Bovbjerg *et al.*, 2000; Kim and Kim, 2005), or, selected from the ears of a "representative listener" (Wenzel *et al.*, 1993). However, since these HRTFs will differ from the listeners own HRTFs distortions to the spatial cues are introduced. The most typical problems related to the use of generic HRTFs are lack of externalization of the sound image (Hartmann and Wittenberg, 1996), signal miscoloration (Silzle, 2002), increasing errors in elevation judgments, and confusion between sound sources in front and behind — also referred to as front-back confusions (Møller *et al.*, 1995c; Wenzel *et al.*, 1993).

Notwithstanding the problems stemmed from the use of generic HRTFs there are some strategies that can help to ameliorate them. Adding early reflections or reverberation improve externalization. There is also evidence that when dynamic cues derived from head movements are available, and provided that stimuli duration is long enough, a considerable reduction in the number of front-back confusions is observed (Perrett and Noble, 1997; Wightman and Kistler, 1999; Begault *et al.*, 2001). The advantage of allowing head movements may be attributed to the fact that binaural cues appear to be slightly affected by the use of generic HRTFs. Thus, dynamic changes in binaural cues can used to resolve front-back confusions. For example, if a sound source is located in a left cone and the subject rotates the head to the left, if the source is in the front the interaural differences would decrease whereas if it is behind they would increase. The issue of head movements becomes particularly relevant in the design of interactive three-dimensional sound systems. After a short review on how HRTFs are modeled we will return to this issue.

1.2.2 Modeling HRTFs

Several techniques have been proposed for the modeling of HRTFs (Huopaniemi *et al.*, 1999). Probably, the most common motivation is the efficient implementation of the digital filters used to control the directional characteristics. Thus, techniques are generally focused on deriving simplified representations of HRTFs (Kistler and Wightman, 1992; Grantham *et al.*, 2003). Reduced-order approximations may be derived as finite impulse response (FIR) filters (Kulkarni and Colburn, 1995; Sandvad and Hammershøi, 1994), or as infinite impulse response (IIR) filters (Kulkarni and Colburn, 2004). Since empirical HRTFs are directly obtained as FIR filters the most straightforward technique is to truncate the impulse response by applying a rectangular window. In Sandvad and Hammershøi (1994) FIR HRTFs of 12, 24, 48, 72 and 128 coefficients were compared with the original 256-coefficients HRTF using a forced-choice paradigm. Results showed that for reductions below 72 coefficients differences were audible, suggesting that an FIR filter of 72 coefficients is the shortest possible filter that preserves the necessary spatial information without introducing audible effects of the truncation.

One important aspect in the processing of HRTFs is the control of the response at low frequencies. This can be done by using a large number of coefficients for the HRIR. However, due to the impossibility of current systems to provide accurate measure at very low frequencies, the signal to noise ratio is bad. An alternative strategy is to adjust the DC value of HRTFs. From a physical point of view, a listener will become more acoustically transparent at lower frequencies, and thus the magnitude response of the HRTF will approach unity gain as the frequency gets closer to DC. On this basis, the DC value is typically adjusted to unity gain.

One could think of sound localization, at least to a first approximation, as a process in which interaural differences determine the lateral placement of the cone wherein the source is, and the spectral information provide the basis for finding the position in the cone. A model of HRTF that to some extent resembles this view, is based on the use of minimum-phase filters to control the variations in the magnitude spectrum, and a pure delay as a simplified approximation of the ITD. ILDs can be thought as being incorporated into the filters, namely, as interaural spectral differences. A minimum-phase filter (or system) has a unique correspondence between magnitude and phase. This is because the minimum-phase filter ensures the fastest possible energy release for a particular magnitude without violating causality. This characteristic is what makes minimum-phase filters so attractive since for the same magnitude shorter filters can be employed. However, because minimum-phase filters introduce phase distortions one legitimate concern is the perceptual validity of approximating the frequency-dependent ITD, found in empirical HRTFs, by a pure delay cascaded with the phase of a minimum-phase filter. From the study by Kulkarni *et al.* (1995) results indicate that in a discrimination paradigm modeled HRTFs were indistinguishable from empirical HRTFs. Furthermore, it has been shown that all-pass sections typically found in HRTFs can be replaced by a pure delay without audible effects (Plogsties *et al.*, 2000). In terms of localization performance the model has also found to be perceptually adequate (Kistler and Wightman, 1992; Wightman and Kistler, 2005). What is of great importance is that the empirical low-frequency ITD is calculated correctly.

1.2.3 Dynamic Binaural Synthesis

When virtual spatial sound becomes interactive dynamic aspects must be carefully considered. For example, in situations where the listener is free to move, the simulated sound field must remain constant relative to the listener's movements. To achieve this, the listener's position and head orientation is typically tracked and this information is passed to the synthesis engine which updates the HRTFs accordingly so as the apparent location of sound sources is fixed relative to the listener's movements. Several such systems have been reported in the literature (Savioja *et al.*, 1999; Blauert *et al.*, 2000; Miller and Wenzel, 2002; Silzle *et al.*, 2004; Pedersen and Minnaar, 2006).

In a similar way as how animated movies are produced by sequences of still images, apparent moving sound can be thought as synthesized by sequentially presenting sound filtered with adjacent HRTFs. Each HRTF is used to render a single spatial location corresponding to one point in the trajectory of the moving sound. It is important that the rendered sound is perceived as being updated without delays and changing smoothly. In a perceptual quality context these requirements are usually referred to as the responsiveness and smoothness of the system (Pellegrini, 2001; Novo, 2005). Engineering aspects related to these requirements correspond to the system latency (responsiveness), update rate (smoothness), and also the spatial resolution of the available HRTFs.

Latency is defined as the time between a change in the acoustic properties of the sound field and the reflected change in the system parameters at the output, e.g. time elapsed from a head movement to the corresponding adjustment in the audio signal presented to the listener's ears. It has been shown that localization accuracy is degraded by large latencies. Sandvad (1996) found that system latencies larger than 96 ms significantly affect localization accuracy. Brungart *et al.* (2006) reported that latencies should be below 70–80 ms. Wenzel (1999) reported thresholds of around 500 ms which are considerably long and argued that the main difference is because quite long stimulus duration was employed.

Update rate relates to the frequency at which new information is retrieved. This information usually comes as coordinates of the position and orientation of the listener and is used to refresh the virtual scenario in order to adjust for these changes. If the update rate is too slow the auditory experience is sluggish and flickering. Sandvad (1996) showed that reducing the update rate of the system from 60 Hz to 20 Hz did not have a significant effect on localization performance. However, at 20 Hz the switching between HRTF filters was clearly audible. In virtual spatial audio systems nominal update rates for the directional filters are in the range from about 60 Hz (Blauert *et al.*, 2000) up to 690 Hz (Miller and Wenzel, 2002).

Spatial resolution corresponds to the accuracy to which the continuous space is sampled. Due to the obvious constraints that make impossible to measure HRTFs for all directions the issue of interpolation rises (Wenzel and Foster, 1993; Hartung and Braasch, 1999). To address this issue a few studies have attempted to obtain analytical expressions of HRTFs (Chen *et al.*, 1995; Evans *et al.*, 1998). Some relatively sophisticated methods with the goal of efficient interpolation have also been proposed (Freeland *et al.*, 2004; Keyrouz and Diepold, 2006). It appears, however, that due to its simplicity (at least to the author) linear interpolation techniques are still a very attractive approach. In a recent study by Minnaar *et al.* (2005) the question of what is the required spatial resolution such that measured and linearly interpolated HRTFs were indistinguishable was addressed. Interpolation was performed on minimum-phase representation of HRTFs in the time domain. The major results show that for stationary sources, a resolution of about 24° appears to be sufficient for locations above the head. For lower elevations a higher resolution is necessary (4–8°), with the highest resolution for directions below the horizontal plane (less than 4°). For moving sources a similar pattern was observed. A simple model was used to generalize the required resolution for directions around the whole sphere. It was suggested that a set of 1130 measured HRTFs seems to be adequate such that interpolation errors are below the audible threshold. Assuming 72-coefficients minimum-phase filters, and coefficients represented with 16-bit resolution, this number of HRTFs would require approximately 325 kbytes of memory.

In conjunction with strategies to switch between HRTFs, the issue of HRTF resolution is of special relevance for the present study and further discussion will be given ahead.

1.3 Auditory Spatial Resolution

Independent of whether binaural material is produced from binaural recordings or from binaural synthesis, it is clear that the main goal of binaural technology is to produce a replica of the sound pressure at the ears that would be produced by a real sound field. If for a particular reason, this is not possible, then, knowledge about which characteristics of the directional information are more important, and/or what the hearing system cannot perceive, may be useful to maintain a perceptually adequate generation of virtual spatial sound.

1.3.1 Spatial resolution in static conditions

Auditory spatial resolution in static conditions is concerned with the ability of listeners to discriminate a change in the position of a sound. Experiments of this kind attempt to measure the smallest angular distance between two identical sources that can reliably be discriminated. This perceptual measure has been defined as the minimum audible angle (MAA) (Mills, 1958; Perrott and Pacheco, 1989; Perrott and Saberi, 1990; Saberi *et al.*, 1991; Strybel and Fujimoto, 2000; Grantham *et al.*, 2003). Typically, two sounds are presented in sequence and the listener has to judge the change in direction of the second sound relative to the first; for example, in case of horizontal MAAs the listener's task would be to decide whether the second stimulus was to the left or to the right of the first. Mills (1958) measured a horizontal MAA of about 1° for a 500-Hz tone presented from the forward direction. MAAs remain relatively constant for tones up to 1000 Hz, for higher frequencies MAAs increase. The sensitivity observed at low-frequency is preserved for broadband stimuli (Perrott and Pacheco, 1989). It has also been shown that spatial resolution is better for sounds presented directly in front of the listener, and thus diminishes as sources move to more lateral locations (Grantham, 1995, ch. 9 pp. 316–317).

Alternatives to the standard design of MAA experiments have been proposed by Hartmann and Rakerd (1989), and adopted by McKinley *et al.* (1992) to measure MAAs using virtual sources. In that study McKinley *et al.* (1992) measured MAAs of about 5°, 5.5°, 8° and 15° for target positions in the horizontal plane at 0°, 30°, 60° and 90° azimuth respectively. MAAs are also dependent on the direction of change; for the location directly in front of the listener horizontal MAAs are lower than vertical MAAs, whereas for the most lateral location (±90°azimuth) the opposite is observed (Perrott and Saberi, 1990; Saberi *et al.*, 1991). In addition, vertical MAAs strongly depend on the spectrum of stimuli (Grantham *et al.*, 2003), and they also increase considerably with increasing reference elevation (Bronkhorst, 1993).

An additional measure of spatial resolution corresponds to the ability of the auditory system to discriminate the separation of sounds presented simultaneously. Perrott (1984) defined the concurrent minimum audible angle (CMAA) as the smallest angular separation required to discriminate between two concurrent sounds. It has been found that horizontal CMAAs are generally larger than MAAs. Furthermore, similar to MAA thresholds, horizontal CMAAs increase for more lateral positions (Perrott, 1984; Divenyi and Oliver, 1989) and the opposite is observed for vertical CMAAs (Best *et al.*, 2004). For pure-tone stimuli CMAAs increase if the frequency difference between the stimuli is reduced (Perrott, 1984), and the same tendency is observed for more complex stimuli (Divenyi and Oliver, 1989).

1.3.2 Spatial resolution in dynamic conditions

In our daily lives, sound in motion is perhaps a more common auditory experience than only stationary sound. The experimental question of how far a sound source needs to move in order to be discriminated from a stationary sound has been addressed by measuring the minimum audible movement angle (MAMA). Usually two different paradigms are employed to measure MAMA. In one paradigm the listener has to discriminate a moving sound from a stationary reference (Perrott and Musicant, 1977; Grantham, 1986; Chandler and Grantham, 1992). In the other paradigm, the listener has to detect the direction of motion (Perrott and Tucker, 1988; Grantham, 1985; Perrott and Marlborough, 1989; Saberi and Perrott, 1990), e.g. whether the sound appears to move to the left or to the right. In a pioneer experiment Harris and Sergeant (1971) found that for slowly moving pure-tone stimuli MAMAs were larger than MAAs by a modest amount. In the horizontal plane the thresholds were about 2 to 4° for a sound moving at 2.8°/s. Factors such as velocity and frequency are known to affect MAMA thresholds. MAMAs increase with increasing velocity (Perrott and Musicant, 1977; Perrott and Tucker, 1988) and are affected by frequency in a manner similar to which MAAs are (Perrott and Tucker, 1988). MAMAs appear to also depend on the direction of motion in a similar way as for their stationary counterpart (Saberi and Perrott, 1990). MAMAs are consistently larger than MAAs with a tendency to approach resolution for static conditions as velocity decreases.

There are mainly two competing theories that try to account for the mechanism underlying motion perception. On the one hand, the “snapshot" theory states that listeners make no use of velocity per se, but they infer it by taking a look at the onset and offset of the sound and compare their spatial position. If these positions are perceptually different given a sufficient time to resolve them, then motion occurred (Grantham, 1997). On the other hand, some evidence against this theory has been shown by comparing MAMAs for stimuli presented during the entire trajectory of movement, and stimuli presented only to the onset and offset (Perrott and Marlborough, 1989). Thresholds were about 50% larger for the second condition, suggesting a special motion-sensitive mechanism that makes use of information obtained throughout the trajectory of the sound. However, it is argued that the duration of the stimuli presented at onset and offset (10 ms) may not have been long enough for assessing the validity of the snapshot theory. This is because a large increase in sound localization error is observed when stimuli duration is shortened from 50 ms to 20 ms (Middlebrooks and Green, 1991). The fact that MAMAs have been found to be consistently larger than MAA s has an interesting implication for the requirements of interactive virtual sound. Assuming that the snapshot theory is correct, the required resolution of HRTFs for dynamic spatial sound could be estimated from MAA measurements. That is the spacing between adjacent HRTFs should be below the MAA.

1.4 Motivation of the present study

In this thesis we are concerned with aspects related to auditory spatial resolution and audibility of artifacts in dynamic binaural synthesis. With respect to the former aspect, a body of experiments is designed to measure discrimination between HRTFs from adjacent locations. The outcome from these experiments can provide information useful to assess the resolution that is just sufficient for a three-dimensional sound system. It can also serve as criterion to select a representative spatial discretization for HRTF measurements. In connection with audibility of artifacts in dynamic binaural synthesis, experiments concerned with the detection of audible discontinuities produced by dynamically changing the HRTFs are designed.

For each of the two aspects previously specified, the experimental conditions are divided such as in one experiment only time characteristics of HRTFs are changed while spectral characteristics remain constant. In a second experiment spectral characteristics of HRTFs are changed whereas time characteristics remain unchanged. We hypothesize that the required resolution and the audibility of artifacts may not be the same for changes in time and spectral characteristics of HRTFs.

The HRTFs employed during the course of the present study correspond to HRTFs measured with a resolution of 2° on an artificial head. The artificial head named “Valdemar" was built at the acoustic laboratory of Aalborg University (Christensen *et al.*, 2000). Its adequacy for recording and reproduction of binaural material has proved to be comparable to other commercial artificial heads (Minnaar *et al.*, 2001). Throughout the experiments reported in this study HRTFs were implemented using the minimum-phase model described in 1.2.2. For all purposes, HRTF is always referred to as a pair of filters. Thus, it contains an ipsilateral component indicating the ear closer to the source, and a contralateral component indicating the ear farther from the source. Sound direction was specified using an interaural-polar coordinate system. This system has its poles to the leftmost and rightmost positions in the horizontal plane. In this system azimuth can be related directly to ITD, and elevation tells the position on the iso-ITD contour. This system is described in more detail in appendix A.

1.4.1 Audibility of Differences in HRTFs

HRTFs, being in the complex frequency domain, differ in their magnitude and phase. The magnitude contains monaural cues and interaural spectral cues which correspond to the frequency-dependent ILDs. The phase contains information related to ITD primarily. It is generally agreed that monaural phase does not contribute to spatial hearing and is relatively inaudible. In terms of magnitude spectra there is substantial evidence demonstrating that spectral modifications caused by changes in the directional properties of a sound field, introduced by the pinna mainly (Wright *et al.*, 1974), are responsible for our ability to localize changes in the elevation of a sound (Hebrank and Wright, 1974; Musicant and Butler, 1984; Langendijk and Bronkhorst, 2002). Sensitivity to changes in the spectral attributes has been investigated to a lesser extent. It is not clear what would be the smallest change between the spectral characteristics of two neighboring HRTFs that can be discriminated. From experiments on auditory profile analysis (Green, 1988) — experiments that test the ability of listeners to discriminate changes in the spectral shape of sound — knowledge has been obtained on that detecting changes in the spectrum of a signal depends on comparing levels at different regions of the spectra. However, to the author's knowledge it is unclear how the differences across frequencies are integrated or whether some frequency regions are given higher weights than others.

1.4.2 Audibility of Switching in HRTFs

Here, switching in HRTFs refers to the action of updating HRTF filters, typically all coefficients at once. More generally, filter switching relates to the field of time-varying digital filters (Mourjopoulos *et al.*, 1990). Switching between digital filters is a desired operation for several audio applications other than dynamic spatial sound, e.g. digital mixing consoles, parametric equalizers. Whenever the acoustic characteristics of a system change as a function of time we desire to control this change so as its output is perceived smooth. Rapid changes in the parameters of digital filters (e.g. gain, center frequency, bandwidth) may cause artifacts that are audible, e.g. “clicks". How audible these artifacts are depends on how large the difference between the switched parameters is. Several strategies have been proposed in order to guarantee a gradual transition between filters (Zoelzer *et al.*, 1993). It has been found that, besides the magnitude of the difference in filter's parameters, filter topology may also affect the audibility of artifacts (Clark *et al.*, 2002). Most of these studies have been centered on the use of IIR filters. One possible reason for this tendency could be because switching between IIR filters causes transients whereas switching between FIR filters does not. A transient is observed if the state variables of the new filter contain intermediate results related to the initial filter (Välimäki and Laakso, 2001, chapter 20 p. 860). A direct switching between FIR filters will cause a discontinuity in the output signal. Even though there are solutions to ameliorate transients produced by time-varying IIR filter (Välimäki and Laakso, 1998) their advantage in terms of efficiency may not be really substantial. The fact, for example, that at low frequencies IIR filters may provide better resolution than FIR filters, is not of major advantage here because HRTFs are practically flat at low frequencies. The work in the present study is based exclusively on FIR filters.

To our knowledge there are a few studies directly addressing the issue of audibility of discontinuities created due to switching between directional filters. One study by Kudo *et al.* (2005) compared several switching strategies: direct switching, overlap-add method, weighted overlap-add method, and cross-fading using three different envelope functions (square root, cosine, and a Fourier Series). From an objective analysis based on the expansion of the effective frequency bandwidth (Cohen, 1995) that occurs at the moment of switching, Kudo *et al.* (2005) concluded that the weighted overlap-add method and the cross-fading method using a Fourier Series generated the less amount of discontinuity to the signal waveform. This analysis was supported by a listening experiment that evaluated how much discontinuities affect the subjective quality.

1.5 Thesis organization

The thesis investigates issues related to auditory perception that are of relevance for the design of three-dimensional sound. The experimental work is described throughout three manuscripts attached to this report.

Manuscript I — Hoffmann, P. F., Møller, H. (2007). Some observations on sensitivity to HRTF magnitude spectrum.
Part of this work has been presented in *Proceedings of the 120th Convention of the Audio Engineering Society*, Paris, France, 2006 May 20–23, preprint 6552.

Manuscript II — Hoffmann, P. F., Møller, H. (2006). Audibility of differences in adjacent head-related transfer functions (HRTFs). in preparation for submission.
Part of this work has been presented in *Proceedings of the 121st Convention of the Audio Engineering Society*, San Francisco, USA, 2006 October 5–8, preprint 6914.

Manuscript III — Hoffmann, P. F., Møller, H. (2006). Audibility of direct switching between head-related transfer functions (HRTFs). in preparation for submission.
Part of this work has been presented in *Proceedings of the 118th Convention of the Audio Engineering Society*, Barcelona, Spain, 2005 May 28–31, preprint 6326.
Part of this work has been presented in *Proceeding of the 119th Convention of the Audio Engineering Society*, New York, USA, 2005 October 7–10, preprint 6537.

1.5.1 Manuscript description

Manuscript I This study constitutes a preliminary experiment on the audibility of spectral differences in HRTFs. The spatial resolution at which HRTFs are available is an important aspect in the implementation of virtual spatial sound. How close HRTFs must be, depends directly on how much the characteristic of HRTFs for adjacent directions differ, and most important, when these differences become audible. The audibility of differences in the spectral characteristics of HRTFs as a function of directional changes was estimated. Four listeners had to discriminate between stimuli spectrally shaped with different HRTFs but whose ITD remained the same. Results showed that listeners were less sensitive to changes in azimuth than to changes in elevation. Azimuth thresholds ranged from 4.7 to 17.2° and elevation thresholds ranged from 2.8 to 8.4°. Elevation thresholds were lower than azimuth thresholds for all conditions. Since discrimination was based on any possible difference it is probable that listeners have used non-spatial attributes of the stimuli, e.g. timbre changes. In relation to implementation of virtual spatial sound these results can provide some guidelines for the spatial resolution required in the spectral representation of HRTFs. In addition, the results may help in the selection of representative locations for HRTF measurements.

Manuscript II This study investigates how well human listeners discriminate differences between HRTFs, and what the minimum directional difference is, for which listeners can perceive a change of any kind. The discrimination of differences in the spectral and time characteristics of the HRTFs is studied separately. In one experiment the smallest angular distance needed to discriminate between the magnitude spectrum of HRTFs was determined. In a second experiment the smallest ITD shift needed to detect a difference was estimated. Results showed a large inter-subject variation, particularly for discrimination of changes in ITD. For the conditions involving discrimination of magnitude spectra mean thresholds ranged from 2.7 to 11° and significant differences were found between changes along azimuth and changes along elevation. These results were comparable to those obtained in Manuscript I wherein differences were between HRTFs that spanned symmetrically about a fixed direction. Here, the comparison was performed by always presenting a reference fixed direction. For changes in ITDs mean thresholds ranged from 87.8 to 163 μs. Implications for virtual acoustic systems aiming at a general population suggest that the required resolution for ITDs, at least in stationary conditions, can be relaxed.

Manuscript III This study investigates aspects related to binaural synthesis of moving sound. Typically, in order to synthesize dynamic changes, HRTFs must be switched as a function of time. In this context, the strategy of using direct switching between HRTFs is studied. Due to the discrete nature of the spatial representation available from the HRTFs, the switching operation generates artifacts that can be audible. These artifacts should ideally be below the threshold of human perception, here denoted as the *minimum audible switch* (MAS). The audibility of these artifacts was measured for time and spectral switching in HRTFs separately. It was found that artifacts produced by time switching were more audible than artifacts produced by switching of the magnitude spectrum. When the sound source was presented in front of the listener thresholds were about 6 μs for time switching and 5° for spectral switching. This shows that implementation of time-varying delays require high resolutions and/or update rates in order to be free of artifacts.

Chapter 2

Manuscript I : Some observations on sensitivity to HRTF magnitude

Some observations on sensitivity to HRTF magnitude*

Pablo F. Hoffmann† and Henrik Møller

Section of Acoustics, Department of Electronic Systems
Aalborg University

Abstract

The spatial resolution at which head-related transfer functions (HRTFs) are available is an important aspect in the implementation of virtual spatial sound. How close HRTFs must be, depends on how much the characteristic of HRTFs for adjacent directions differ, and most important, when these differences become audible. Thresholds for the audibility of variations in the spectral characteristics of HRTFs as a function of angular separation were estimated. Four listeners had to discriminate between stimuli spectrally shaped with different HRTFs but whose ITD remained the same. Results showed that listeners were less sensitive to changes in azimuth than to changes in elevation for several directions. Azimuth thresholds ranged from 4.7 to 17.2° and elevation thresholds ranged from 2.8 to 8.4°. In connection with synthesis of virtual spatial sound, these results can provide guidelines for the spatial resolution required in the spectral representation of HRTFs. In addition, the results may help in the selection of representative locations for HRTF measurements.

Key words: HRTFs, Minimum-phase filters, spectral-shape discrimination

1. INTRODUCTION

When sound reaches our ears many reflections from the torso, shoulders, head and pinna interact with the direct sound path. This interaction introduces patterns to the magnitude spectrum of the sound commonly characterized by peaks and notches. These patterns vary in a complex way as a function of sound direction [1, 2, 3]. The head-related transfer function (HRTF) fully describes the directional dependency of these patterns, and thus HRTFs can be used as filters to synthesize virtual spatial sound. In this context, if the HRTF used to filter the sound is changed, the corresponding change in the output of the filtering is assumed to be perceived as a shift in the apparent location of that sound.

Because different HRTFs give different spectral shapes to the sound at the ear, differences between HRTFs may not only cause an apparent shift in direction but also a change in the perceived timbre. Evidence of this has been reported in a study conducted by Langendijk and Bronkhorst [4] who measured discrimination between interpolated and measured HRTFs. Their results showed that for broadband stimuli presented with third-octave-band levels randomized within ±3dB on every presentation, a spatial resolution of 10–15° was required so that interpolated HRTFs generate the same spatial percept as measured HRTFs. It was argued that because of the randomization, changes in timbre were unlikely to be used, and actually, listeners reported that the only reliable cue was source location. If the level of the stimuli was fixed the required resolution was 6° suggesting that for small spectral differences timbre-based cues appear to precede spatial cues.

The ability of the auditory system to detect variations in the spectral pattern of sound, or profile analysis [5], has been studied in terms of detectability of changes in the sign of spectral slopes [6], discrimination of broadband noise shaped with different speech-like spectra [7], detection of peaks and notches from a flat spectrum [8, 9], and detection of a level increment in a single component relative to others in a multi-component complex stimuli [10]. The motivation of this study is to explore the ability of listeners to discriminate changes in the spectral pattern of stimuli when HRTFs are used to produced these changes. Discrimination is compared in a selected number of spatial positions and for different directions of change. In this experiment, spatial positions and directional changes are indicated using a coordinate system with the poles to the left and right. In this coordinate system the azimuth relates di-

*Portions of these results presented at the 120th Convention of the Audio Engineering Society, Paris, France, May 20–23, 2006

† Correspondence to pfh@es.aau.dk

rectly to the ITD and elevation tells the position around a cone of confusion. General results show that sensitivity to HRTF magnitude is dependent upon both spatial position and direction of change.

An important aspect of this experiment is that differences in the magnitude spectrum must provide the basis for discrimination, and thus, differences in phase should not be audible. To this purpose HRTFs are implemented as minimum-phase filters. It has been shown that it is perceptually adequate to approximate HRTFs by a frequency-independent delay to control the interaural time difference (ITD) and minimum-phase filters to control the magnitude spectra [11].

2. METHOD

2.1. Subjects

Four paid subjects participated in the listening experiment, one female and three males. Their ages ranged from 23 to 28. Subjects had normal hearing and they were selected by means of a pure-tone audiometry screening at less than 10 dB HL from 250 to 4000 Hz in octave steps, and less than 15 dB HL for 8 kHz. All subjects had previous experience on listening experiments.

2.2. Stimuli and Apparatus

Broadband pink noise (20–16000 Hz) was used as source signal. The simulation of directional sound was based on HRTFs measured with a resolution of 2° on an artificial head [12]. Eight directions were selected in the left half of the upper hemisphere. These directions are referred to as the *nominal directions*. Directions are given as (azimuth θ, elevation ϕ) in a polar coordinate system with horizontal axis and left-right poles (also referred to as the interaural-polar coordinate system). 90° and -90° azimuth correspond to left and right sides, 0° and 180° elevation to the frontal and rear portions of the horizontal plane respectively, and 90° elevation to the upper portion of the frontal plane. Fig. 1 shows the eight selected directions. Four directions were selected in the median plane (0° azimuth; 0°,44°,136° and 180° elevation). Three directions were chosen on a cone of confusion ((58°,0°), (46°,90°) and (54°,180°)) and they were selected to have the same ITD rather than being on the same geometrical cone, thus the azimuth varies with elevation. The ITD for these directions corresponded to -437.5 μs and was calculated from (46°,90°). Finally, (90°,0°) was also included and its corresponding ITD was -625 μs. ITDs were derived from the interaural differences in group delay of the excess-phase components of the HRTFs evaluated at 0 Hz. This procedure has been shown to be adequate for the computation of ITDs [13].

The measured HRTFs — available as pairs of 256-coefficients impulse responses — were truncated using a 72-coefficients rectangular window. At a 48-kHz sampling rate the resulting impulse responses had a duration of 1.5 ms. This duration has been shown to be sufficient to avoid audible effects of the truncation when using noise-like stimuli [14]. To control the HRTFs at low frequencies the DC value of each new impulse response was set to unity gain as described in [15, section 5.2]. Minimum-phase representations of the impulse responses were then constructed using homomorphic filtering [16, ch. 12].

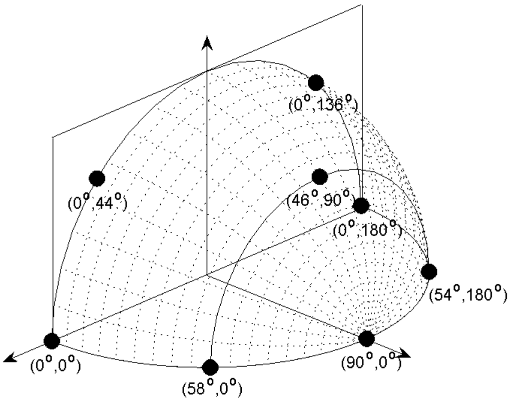


Fig. 1. Nominal directions selected in the left half of the upper hemisphere. Directions are specified in an interaural-polar coordinate system.

Stimuli were 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 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. To reduce the noise floor a custom-made 20-dB passive attenuator was connected to the output of the amplifier. The stereo output signal from the attenuator was delivered to the listener over equalized Beyerdynamic DT-990 circumaural headphones.

Two 256-coefficients minimum-phase FIR filters were employed in order to compensate for the left and right headphone transfer functions respectively. The equalization filters were based on measurements made at the entrance to the blocked ear canal on 23 subjects (none of them participated in this listening test). Five measurements were obtained from each ear and subject, and subjects were asked to reposition the headphones between measurements. Headphone responses were obtained using the maximum-length sequence technique (MLS) [17], and the results were in the form of 256-coefficients impulse responses for each ear. Impulse responses were then transformed to the frequency domain, and a representative transfer function was calculated by taking the mean of all transfer functions on a sound power basis. The equalization filter was designed based on the inverse of this mean response. To avoid excessive amplification at low frequencies due to the inversion, the DC value of the inverse was manually adjusted to unity gain. This value corresponded roughly to the observed gain at low frequencies. A 4th-order butterworth low-pass filter (19-kHz cut-off frequency) was applied to reduce the amplification that also occur at, and close to, the Nyquist frequency. Fig. 2 shows subjects' responses, mean response, and response of the equalization filter for the left ear. In order to obtain a time representation of the equalization filter, a minimum-phase approximation was computed for each ear using homomorphic filtering [16, ch. 12].

2.3. Psychophysical procedure

Each nominal direction, with the exception of (0°,0°), had an associated set of neighboring HRTFs symmetrically spaced about the nominal direction such that the absolute angular span was 2°, 4°, 16°, 24° and 32°. For the forward direction the selected angles were 2°, 4°, 8°, 12°, 20°. Given a nominal direction, changes between HRTFs could occur either in azimuth or in elevation. Fig. 3 shows a graphical representation of these changes when the selected nominal direction is (0°,0°). For changes in azimuth an arc is described along the horizontal plane (angle θ). For changes in elevation HRTFs along the median plane are used to describe the arc (angle ϕ). Note that the midpoints of the arcs always correspond to the nominal direction. For the particular case of (90°,0°), where in the strict sense changes in elevation cannot be applied, two azimuth modes were implemented. One mode for changes in the horizontal plane spanning the angle horizontally, and the other for changes in the frontal plane spanning the angle vertically.

Discrimination of HRTF magnitude was estimated in a three-interval, two-alternative forced-choice (3I 2AFC) task using the method of constant stimulus. Stimulus and inter-stimulus intervals were 300 ms of duration. The stimulus presented on either the second or third interval differed from the others two. Subjects were asked to identify the interval that differed (i.e. the target stimulus). They had to push one of two buttons to indicate a response. A feedback light was used to immediately show the subjects whether or not the response was correct. A 2-s silence interval was used between trials.

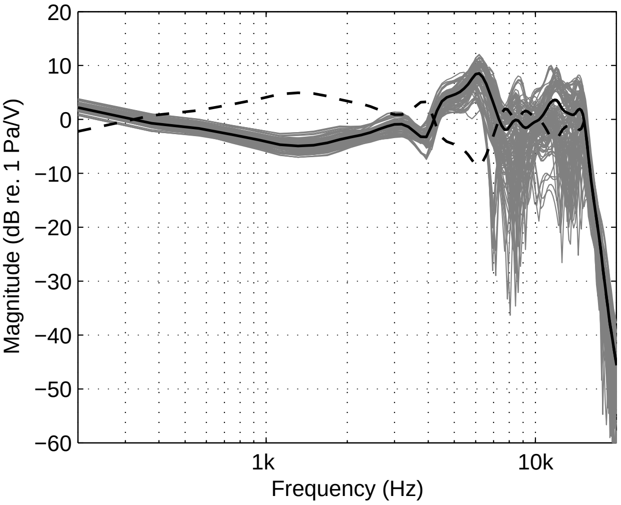


Fig. 2. Headphone transfer functions measured on the left ear of 23 subjects (grey lines). The solid line indicates the mean response. The response of the equalization filter, indicated by the dashed line, corresponds to the inverse of the mean response.

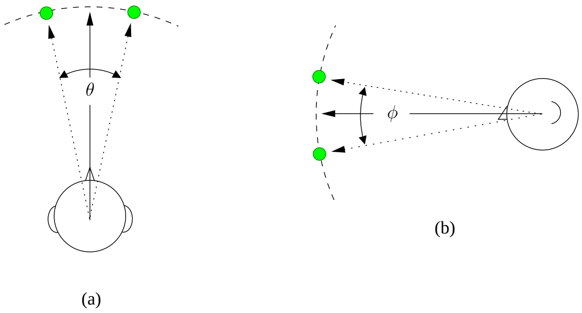


Fig. 3. Scheme of the two directional modes used in the experiment. The solid arrow represent the nominal direction (here (0°,0°)). One pair of adjacent HRTFs, indicated by the dotted arrows, describes an arc for; (a) changes in azimuth with angular separation θ, and (b) changes in elevation with angular separation ϕ.

On a single trial, a given nominal direction and angular separation were selected and the minimum-phase HRTF of one end of the arc (recall Fig. 3) was used to filter two of the three sound intervals (one corresponding to the first interval). This filter was regarded as the reference HRTF for that trial. The remainder sound interval was filtered with the minimum-phase HRTF that corresponded to the other end of the arc, i.e. the target HRTF. Note that the ITD of the nominal direction was used for both the reference and the target HRTFs. For changes in azimuth, the spatial configuration of the pair reference-target HRTF was randomly selected to be either left-right or right-left from the nominal direction. Similarly, for changes in elevation the pair reference-target HRTF could be either above-below or below-above the nominal direction.

2.4. Experimental Design

Subjects were in a sound-insulated cabin specially designed for psychoacoustical experiments. Blocks of sixteen trials were used for practice, and only the largest angular separation was presented. Since subjects had recently participated on similar listening tests and their performance was observed to be stable, no further practice was necessary.

On the main experiment, all angular separations were repeated 15 times within a block of trials. The order in which they were presented was random. Nominal direction and directional mode were held constant within a block of trials. Each combination of nominal direction, directional mode and angular separation was presented 30 times. A total of 2400 responses were obtained per subject (8 reference directions x 2 directional modes x 5 angular separations x 30 repetitions). Data were collected during three sessions that were held on different days. Blocks were distributed so that one session lasted from about one hour and half to two hours.

2.5. Psychometric functions

The proportion of correct responses p obtained at each angular separation were used to estimate psychometric functions for each subject and each condition. We assumed a logistic form of the psychometric function. Its mathematical expression is given by

$$\hat{p} = \lambda + (1 - \lambda)(1 + e^{-(x-\alpha)/\beta})^{-1} \qquad (1)$$

where $\hat{p}$ is the estimate of p, x is the angular separation, λ is the parameter that determines chance performance, α is the threshold parameter, and β is the slope parameter (shallow slopes correspond to large values of β). Here, chance performance is equal to 50%, and the threshold is defined as the angular separation that yields 75% of correct responses.

Psychometric functions were fitted using a least-square criterion based on the iterative Gauss-Newton algorithm. The fitting was performed on the log of the angular separation.

3. RESULTS

3.1. Individual results

Psychometric functions were obtained for the audibility of differences between minimum-phase HRTFs as a function of their angular separation. Fitted psychometric functions and proportion of correct responses for each condition and subject are shown in Fig. 4. The abscissa specifies the angular separation in degrees, and the ordinate specifies the proportion of correct responses. Each panel shows proportions and fitted functions for each nominal direction. Results were generally consistent across subjects and performance improved with increasing angular separation.

Using the bootstrapping technique we computed confidence limits for the estimated parameters of each psychometric function[1]. Fig. 5 shows the estimated confidence limits. The bootstrapping technique used to estimate confidence limits is described by [19]. From the psychometric function fitted to the empirical data we calculated the percent correct for each angular separation, and assuming they are binomially distributed, a simulated percent correct was randomly drawn for each angular separation and a new fitting was performed on the simulated percent correct. This operation was repeated 10000 times to provide 10000 estimates of the threshold and slope parameters. Then the 2.5% and 97.5% quantiles were taken as the 95% confidence limits. The purpose of these calculations was to give the reader an idea of the variation in the estimated parameters. We can observe from Figs. 5(a) and 5(b) that threshold and most slope estimates are reasonably good. Note that confidence limits for slope increases with increasing β. This may be due to the inverse relation between β and the slope of the psychometric function, which suggests that reliable estimates of β are difficult to obtain for shallow slopes.

3.2. Mean results

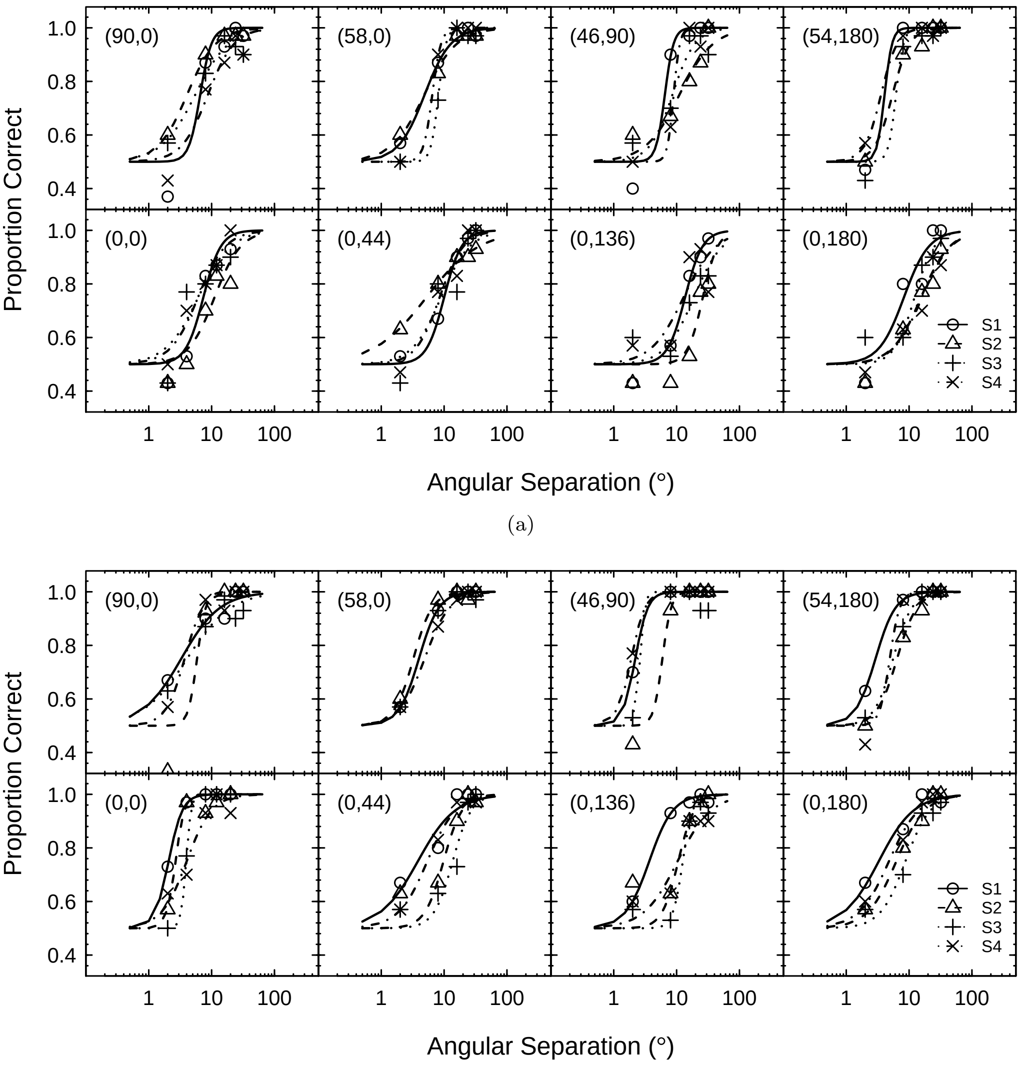
In order to observe group tendencies, the threshold parameter α and the slope parameter β were averaged across subjects. Geometric means were calculated for the threshold parameter and arithmetic means for the slope. Fig. 6 shows the calculated mean psychometric functions and the mean parameters are summarized in Table 1.

A two-factor within-subject analysis of variance with nominal direction and directional mode as factors was conducted on the logs of the thresholds. The analysis revealed a highly significant main effect of nominal direction (F(7,21) = 12.9, $p < 0.001$), and a significant main effect of directional mode (F(1,3) = 25.1, $p < 0.05$). That is, thresholds for changes in elevation were consistently lower than those for changes in azimuth. The interaction between nominal direction and directional mode was also significant (F(7,21) = 3.2, $p < 0.05$). This can be attributed to the fact that for some directions thresholds for changes in elevation were slightly lower than for changes in azimuth, whereas for others the difference was larger. A similar analysis on the logs of the slope revealed a significant main effect of nominal direction (F(7,21) = 4.7, $p < 0.01$). A post-hoc analysis (Tukey HSD) revealed that (46°,90°) was significantly different from (0°,44°) and (0°,180°), reflecting the very steep slope observed for changes in elevation around (46°,90°) as compared to the others.

4. DISCUSSION

Audibility of spectral differences in HRTFs was estimated by measuring how well subjects could discriminate between changes in the minimum-phase HRTFs while the ITD remained constant. For the directions used in this study, mean results are in the range of 2.8–17.2° depending on direction. This suggests that different resolutions may be required for minimum-phase filters depending upon the location of the virtual sources. This also apply to the selection of representative locations for HRTF measurements. Some spatial regions need a more dense set of measuring locations than others. In addition, the fact that thresholds for changes in elevation were consistently lower than thresholds for changes in azimuth implies that synthesis of moving sound, in which HRTFs need to be constantly updated, requires higher spatial resolution for trajectories that incorporate changes in elevation than those where azimuthal changes occur, i.e., trajectories where ITDs also need to be updated.

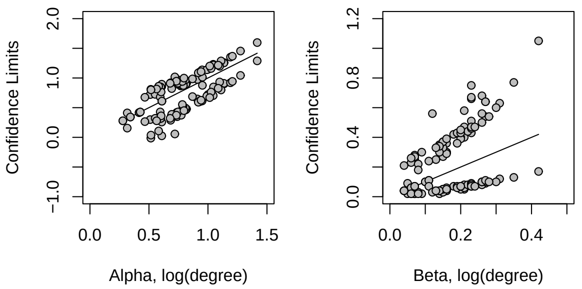
[1] We adopted this technique from the analysis done by [18] on psychometric functions for informational masking.



(a)

(b)

Fig. 4. Psychometric functions for the discrimination of HRTF magnitude. The eight panels in (a) show results for each nominal direction and changes in azimuth. The eight panels in (b) show results for changes in elevation. Individual proportions are represented by different symbols and fitted psychometric functions by different lines.



(a)

(b)

Fig. 5. Scatter plots of the estimated confidence limits obtained by "bootstrapping" technique (see text for details). The lines represent the estimated parameters. Each pair of points above and below the line represents upper and lower limits for each listener and condition.

To assess the possibility of specific frequency regions being more dominant as cues for discrimination, we computed spectral differences in third-octave bands between the HRTFs corresponding to the locations separated by the estimated thresholds. Fig. 7 shows these differences plotted for each nominal direction. They are given in absolute values in dB, meaning that higher values reflect larger differences for that frequency band. The left-hand column shows differences at threshold for changes in azimuth, and the right-hand column for changes in elevation. It can be observed that for most of the directions off midline, differences in the contralateral component of the HRTF (grey lines) are larger than those in the ipsilateral component.

It is clear that larger differences occur at high frequencies, and this is somewhat expected considering that the contribution of spectral cues to sound localization are more

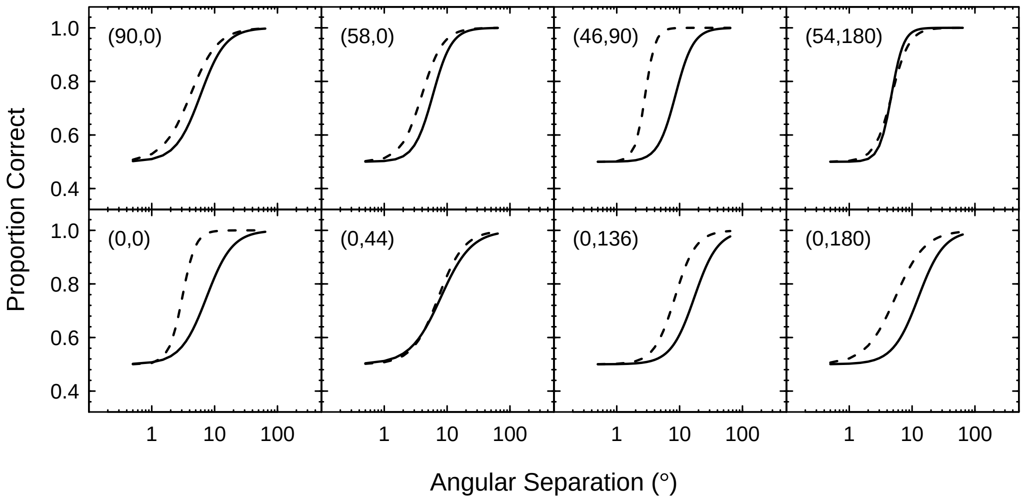


Fig. 6. Mean psychometric functions for the discrimination of HRTF magnitude. Solid line (–) indicates psychometric functions for changes in azimuth, and dashed line (- -) indicates psychometric functions for changes in elevation. Each panel shows psychometric functions for one nominal direction as indicated in the top-left part.

prominent at high frequencies [3]. The fact that HRTFs differ in complex ways makes difficult to confirm a specific criterion for discrimination. It seems more likely that discrimination could have been based upon the most reliable cue available, being this the integration of spectral differences over a wide frequency range or a prominent boost in a particular narrow frequency region. It is also important to note that, because of the smoothing effect caused by the third-octave analysis, if discrimination of notches was a reliable cue further analysis would be required to assess this possibility.

5. Spectral Distance Measure

In an attempt to represent the amount of spectral difference by a single number we computed the standard deviation (SD) of the difference (in dB) as a function of angular separation. This metric has been employed in modeling the contribution of spectral cues to localization [3], and in minimizing inter-subjects differences in HRTFs [20, in this study the variance was used]. One of the advantages of this metric is that overall level differences (differences that are constant across frequency) are eliminated, and thus only variations in spectral shape are emphasized.

Table 1
Mean threshold (Alpha) and slope (Beta) parameters across subjects for changes in azimuth and elevation on each nominal direction.

	Estimated parameters			
	Azimuth		Elevation	
Nominal Direction	Alpha	Beta	Alpha	Beta
(90°,0°)	6.0	0.20	4.1	0.22
(58°,0°)	5.9	0.15	4.0	0.17
(46°,90°)	8.6	0.14	2.8	0.08
(54°,180°)	4.7	0.09	4.8	0.14
(0°,0°)	7.4	0.21	3.1	0.11
(0°,44°)	7.9	0.25	7.2	0.21
(0°,136°)	17.2	0.19	8.4	0.17
(0°,180°)	12.6	0.21	5.4	0.24

Before computing the SDs the HRTFs were smoothed using a gammatone filterbank [21]. The motivation for using this smoothing technique is that this type of filters simulate the frequency analysis performed by the cochlea. Furthermore, since the frequency resolution of the cochlea is poorer at high frequencies, this procedure effectively smooth out some of the spectral details that may wrongly inflate the estimation of spectral differences (e.g. spectral notches). For anechoic HRTFs it has been shown that by selecting an appropriate filter order for the gammatone filters (3 or 4), the smoothing results in imperceptible differences as compared to the original ones [22]. Here, the order of the gammatone filters was set to 4 and the smoothing procedure was based on the procedure used by [22]. The mathematical equations used to derive the gammatone filters and to calculate the smoothed HRTFs are given in Appendix A.

SDs were computed on HRTFs differences produced by angular separations from 4 to 20° in steps of 4°. Recall that the resolution of the empirical HRTFs is 2°, and because in this experiment HRTFs were symmetrically spaced about nominal directions, then the smallest possible angular separation between measured HRTFs was 4°. To form a difference spectrum the difference between dB amplitudes was computed component by component in the frequency range of 1125–12000 Hz in 187.5-Hz steps. Then the SD across the components of the difference spectrum was computed.

Fig. 8 shows SDs for the directions in the median plane

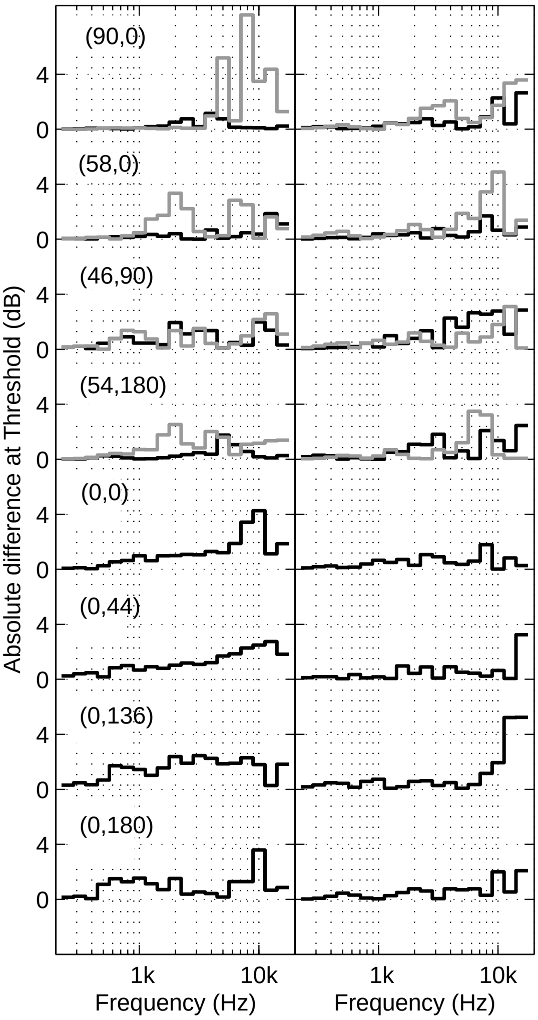


Fig. 7. Spectral differences at threshold in third-octave bands from 250 to 16000 Hz. Left-hand column presents differences for changes in azimuth and right-hand column presents differences for changes in elevation. Differences between right-ear HRTFs are indicated by gray lines, and differences between left-ear HRTFs are indicated by black lines.

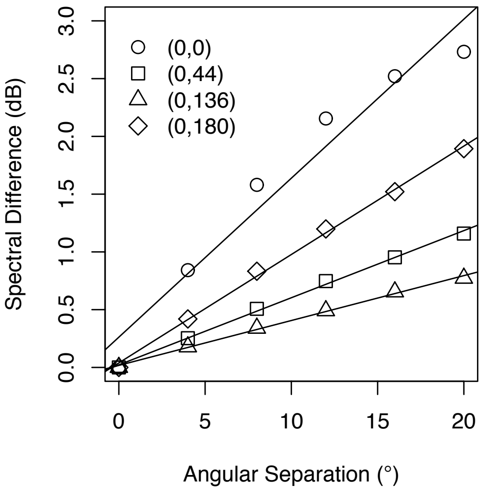


Fig. 8. Spectral differences between HRTFs symmetrically separated about the nominal direction for directions in the median plane (indicated by symbols). The angular separation is along elevation and lines represent linear fittings for each nominal direction.

and separations spanned along elevation. Note that spectral differences increase with increasing angular separation and this pattern is also observed for the other nominal directions. Furthermore, note that spectral differences increase more rapidly for nominal directions whose estimated thresholds were lower. By doing a linear fitting to the data, we can use the slope of the fitted line to establish a relation between the rate at which spectral differences increase and the observed thresholds. A simple model that describes this relation can be mathematically expressed by

$$\widehat{thr} = \frac{c}{slp} \qquad (2)$$

where $\widehat{thr}$ indicates a threshold estimate, slp is the slope of the fitted curve and c is a constant given in dB. By minimizing the error in the least-square sense between the measured thresholds and the estimated thresholds for the directions in the median plane and changes in elevation we found a value of 0.4 dB for c. In this calculation only one ear was used because in the median plane left and right HRTFs were identical.

Fig. 9 shows thresholds and the approximation obtained from eq.(2). The abscissa represents the elevation angle of the nominal directions, which are in turn grouped by common ITD. In the upper panel thresholds for changes in elevation are plotted together with the approximation for the left-ear HRTF and right-ear HRTF. The left-ear approximates well the thresholds obtained with the exception of the nominal direction (54°,180°). The approximation by the right-ear, which corresponds to the contralateral component, is less accurate. In the lower panel we observe an

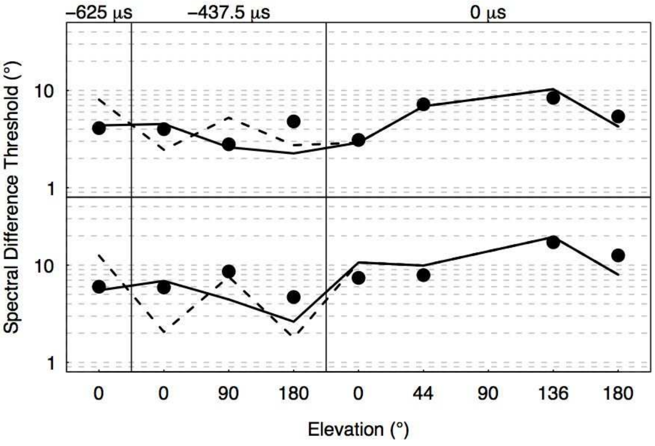


Fig. 9. Comparison between measured thresholds (circles) and predictions based on Eq. 2. Top and bottom panels show data for changes in elevation and azimuth respectively. Solid lines indicate approximations using spectral differences for the left-ear HRTFs (ipsilateral). Dashed lines indicate approximations for the right-ear HRTFs (contralateral).

almost identical pattern for changes in azimuth. Here, a good approximation is observed in the median plane. Approximations for the left-ear are also good for (58°,0°) and (90°,0°). Right-ear approximations are generally less accurate with the exception of (46°,90°). These observations suggest that for the more lateral positions discrimination was mostly based on differences in the ipsilateral component. It is also possible that overall interaural level differences might have been used for discrimination. These changes in overall level are not included in this simple model. However, in terms of predicting audibility of spectral differences along the median plane, results of this analysis are encouraging.

6. CONCLUSIONS

Listeners were able to discriminate spectral differences in HRTFs for angular separations in a range of 2.8–17.2° depending on direction. This relatively large range may be attributed to the fact that HRTFs change differently depending on the spatial location they are representing. Thresholds for changes in elevation were consistently lower than for changes in azimuth. A simple model for discrimination of spectral differences was proposed based on the standard deviation of a difference spectrum. It was possible to account for thresholds measured for locations in the median plane. For lateral directions approximations were less accurate, probably because some other cues not included in the model, such as overall interaural level differences, may have been used. Further investigation is necessary to evaluate which spectral features are more prominent for discrimination, and how the may depend on spatial location.

7. ACKNOWLEDGMENTS

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

APPENDIX A: Formulae for the smoothing of HRTF spectra with gammatone filters

Following the procedure described by [22], the smoothed magnitude $|Y(f_c)|$ of HRTF $X(f)$ was computed as

$$|Y(f_c)| = \sqrt{\frac{\int_0^\infty |X(f)|^2 |H(f,f_c)|^2 df}{\int_0^\infty |H(f,f_c)|^2 df}}$$

where $H(f,f_c)$ correspond to the frequency response of the gammatone filter with center frequency f_c. This transfer function is given by

$$H(f,f_c) = \Big(\frac{1}{1 + j(f-f_c)/b}\Big)^n$$

where n is the filter's order and b is the 3-dB bandwidth, which was set equal to the equivalent rectangular bandwidth (ERB) estimate of the human auditory system as derived by [23]. Its expression is given by

$$b(f_c) = \frac{24.7(0.00437 f_c + 1)}{2\sqrt{2^{1/n} - 1}}$$

References

[1] D. Wright, J. H. Hebrank, B. Wilson, Pinna reflections as cues for localization, J. Acoust. Soc. Am. 56 (3) (1974) 957–962.

[2] B. Rakerd, W. M. Hartmann, T. L. McCaskey, Identification and localization of sound sources in the median plane, J. Acoust. Soc. Am. 106 (5) (1999) 2812–2820.

[3] E. H. A. Langendijk, A. W. Bronkhorst, Contribution of spectral cues to human sound localization, J. Acoust. Soc. Am. 112 (4) (2002) 1583–1596.

[4] E. H. A. Langendijk, A. W. Bronkhorst, Fidelity of three-dimensional-sound reproduction using a virtual auditory display, J. Acoust. Soc. Am. 107 (1) (2000) 528–537.

[5] D. M. Green, Profile Analysis: Auditory Intensity Discrimination, Oxford University Press, New York, NY, USA, 1988.

[6] N. J. Versfeld, Discrimination of changes in the spectral shape of noise bands, J. Acoust. Soc. Am. 102 (4) (1997) 2264–2275.

[7] C. L. Farrar, C. M. Reed, Y. Ito, N. I. Durlach, L. A. Delhome, P. M. Zurek, L. D. Braida, Spectral-shape discrimination . I. results from normal-hearing listeners for stationary broadband noises, J. Acoust. Soc. Am. 81 (4) (1986) 1085–1092.

[8] B. C. J. Moore, S. R. Oldfield, G. J. Dooley, Detection and discrimination of spectral peaks and notches at 1 and 8 khz, J. Acoust. Soc. Am. 85 (2) (1989) 820–836.

[9] A. Alves-Pinto, E. A. Lopez-Poveda, Detection of high-frequency spectral notches as a function of level, J. Acoust. Soc. Am. 118 (4) (2005) 2458–2469.

[10] D. M. Green, Z. A. Onsan, T. G. Forrest, Frequency effects in profile analysis and detecting complex spectral changes, J. Acoust. Soc. Am. 81 (3) (1987) 692–699.

[11] A. Kulkarni, S. K. Isabelle, H. S. Colburn, On the minimum-phase approximation of head-related transfer functions, in: Proc. of the ASSP (IEEE) Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, USA, 1995, pp. 84–87.

[12] B. P. Bovbjerg, F. Christensen, P. Minnaar, X. Chen, Measuring the head-related transfer functions of an artificial head with a high directional resolution, in: 109th Convention of the Audio Engineering Society, Los Angeles, California, USA, 2000, convention paper 5264.

[13] P. Minnaar, J. Plogsties, S. K. Olesen, F. Christensen, H. Møller, The interaural time difference in binaural synthesis, in: 108th Convention of the Audio Engineering Society, Paris, France, 2000, convention paper 5133.

[14] J. Sandvad, D. Hammershøi, What is the most efficient way of representing HTF filters?, in: Proceedings of Nordic Signal Processing Symposium, NORSIG '94, Lesund, Norway, 1994, pp. 174–178.

[15] D. Hammershøi, H. Møller, Binaural technique, basic methods for recording, synthesis and reproduction, in: J. Blauert (Ed.), Communication Acoustics, Springer Verlag, Berlin, Germany, 2005, pp. 223–254.

[16] A. V. Oppenheim, R. W. Schafer, Discrete-Time Signal Processing, Prentice Hall, New Jersey, NJ, USA, 1989.

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

[18] R. A. Lutfi, D. J. Kistler, M. R. Callahan, F. L. Wightman, Psychometric functions for informational masking, J. Acoust. Soc. Am. 114 (6) (2003) 3273–3282.

[19] L. T. Maloney, Confidence intervals for the parameters of psychometric functions, Percept Psychophys 47 (1990) 127–134.

[20] J. C. Middlebrooks, Individual differences in external-ear transfer functions reduced by scaling in frequency, J. Acoust. Soc. Am. 106 (3) (1999) 1480–1492.

[21] R. D. Patterson, M. H. Allerhand, C. Giguère, Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform, J. Acoust. Soc. Am. 98 (4) (1995) 1890–1894.

[22] J. Breebaart, A. Kohlrausch, The perceptual (ir)relevance of HRTF magnitude and phase spectra, in: 110th Convention of the Audio Engineering Society, Amsterdam, The Netherlands, 2001, convention paper 5406.

[23] B. R. Glasberg, B. C. J. Moore, Derivation of auditory filter shapes from notched-noise data, Hearing Research 47 (1/2) (1990) 103–138.

Chapter 3

Manuscript II : Audibility of differences in adjacent head-related transfer functions

Audibility of differences in adjacent head-related transfer functions*

Pablo F. Hoffmann† and Henrik Møller

Section of Acoustics, Department of Electronic Systems
Aalborg University

Abstract

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 characteristic 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. Results from both experiments showed a large inter-subject variability, which was particularly pronounced for discrimination of changes in ITD. For the discrimination of spectral differences mean thresholds ranged from 2.4 to 11° depending on direction, and significant differences were found between changes in azimuth and changes in elevation. Mean thresholds for changes in ITD ranged from 87.8 to 163 μs. Results are discussed in the context of requirements for spatial resolution in the implementation of dynamic three-dimensional sound.

0. 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 consists in filtering a monophonic signal with a specific HRTF and reproducing the result typically over headphones. For dynamic virtual sound, in which sound source, listener, or both can move, directional information changes as a function of time. Therefore, HRTFs must be constantly updated in order to account for these changes. In real life the spatial characteristics of moving sound vary continuously, and this would in theory require an infinite number of HRTFs to be available. Because this is not physically realizable and only a discrete representation of the acoustic space is possible, strategies that exploit the perceptual limits of the auditory system must be evaluated in the design of such dynamic systems.

In order to define a given spatial resolution as perceptually adequate, one could assess the ability of listeners to differentiate the position of two sound sources irrespective of whether they can identify their location. 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 wether 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], and this spatial acuity has been found to be similar when using broadband stimuli [5]. Using stimuli reproduced over headphones (also 500-Hz tone), the horizontal MAA has been found to be about 5° for the forward direction[1]; increasing with lateral angle [6]. Vertical MAA is approximately 4° for the forward direction, and, in general, is larger than the horizontal MAA [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 the 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.

* Part of this work presented at the 121st Convention of the Audio Engineering Society, San Francisco, USA, October 5–8, 2006.

† Correspondence to pfh@es.aau.dk

0.1. 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 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 [9, 10]. 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 Fig. 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 [11, 12]. 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 [13], and sound localization tasks [14], have shown that empirical and modeled HRTFs are indistinguishable and that they generate the same spatial percept.

0.2. 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.

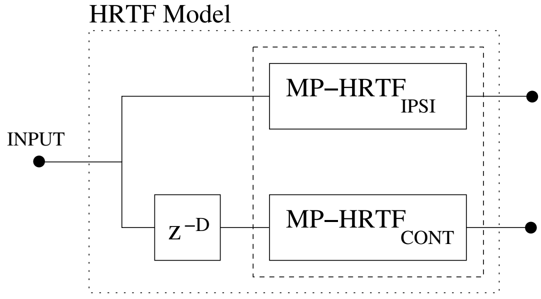


Fig. 1. Minimum-phase and frequency-independent ITD model of the HRTF. Minimum-phase filters are enclosed in the dashed box. The IPSI and CONT sub-indices indicate the ipsilateral and contralateral components respectively. The ITD is implemented by cascading the delay to the contralateral component of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

1. EXPERIMENT I: AUDIBILITY OF SPECTRAL DIFFERENCES

1.1. Method

1.1.1. *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.

1.1.2. *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 went 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. Details of the design of the headphone-equalization filters are given in [15].

1.1.3. *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 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 [16]. Nine positions were selected in the left half of the upper hemisphere. Directions are given as (azimuth, elevation) 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 azimuth, and elevation is used to specify source position around the cone determined by the ITD. This system has shown some advantages over the more conventional vertical coordinate system in explaining sound localization in the upper hemisphere [17]. The convention used here is that 90° and -90° azimuth correspond to left and right sides, 0° elevation to the anterior portion of the horizontal plane, 180° elevation to the posterior portion of the horizontal plane, and 90° elevation to the upper portion of the frontal plane. In this study, five positions were selected in the median plane (0° azimuth) at 0°, 44°, 90°, 136° and 180° elevation. Three positions were selected in an iso-ITD contour to the left ((58°, 0°), (46°, 90°) and (54°, 180°)). The positions at 0° and 180° elevation were chosen to match, or at least be the closest to, the ITD for (46°,90°). Because iso-ITD contours are not geometrically perfect, azimuth varied slightly with elevation. The position at 90° azimuth was also included. In the remainder of this article, these positions will be referred to as *nominal positions* and they are shown in Fig. 2.

The measured HRTFs were represented as minimum-phase FIR filters with the ITD calculated separately and inserted to the contralateral impulse response. Minimum-phase representations and ITDs were calculated using the same procedure as described in [15]. 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 [18, section 5.2]. Fig. 3 shows the HRTFs corresponding to the selected nominal positions.

1.1.4. *Psychophysical Method*

Audibility of spectral differences in HRTFs was determined in a three-interval, three-alternative forced-choice task using the method of constant stimulus. 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 in oder to immediately show the correct response. After a silence interval of 1 s a new trial was presented.

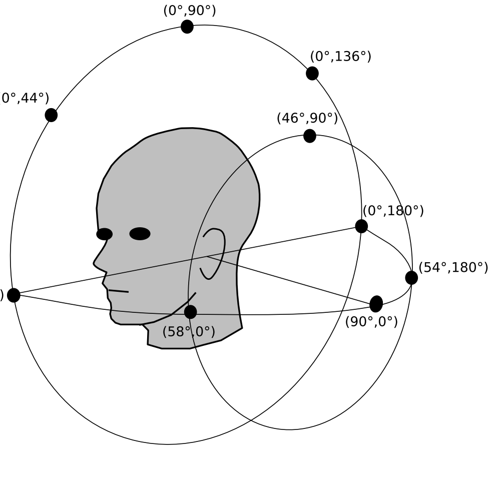


Fig. 2. Nominal positions employed in the listening experiment. These positions serve as reference in the experiment. Azimuth and elevation are indicated in an interaural-polar coordinate system.

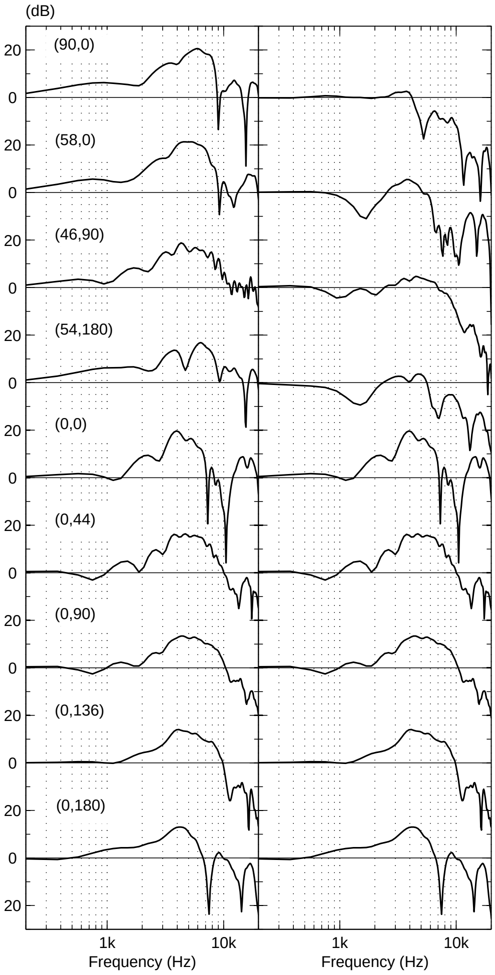


Fig. 3. HRTFs used for the nominal positions. Left and right columns represent HRTFs' components for the left- and right-ear respectively.

For the nominal position at 90° azimuth and the positions in the iso-ITD contour the directions in which HRTFs could change were: left, right, up and down. For the positions in the median plane changes to the right were not included because symmetry about the median plane was assumed. There are a few additional observations regarding the directional changes that shall be pointed out. If we specify the changes relative to the coordinate system, we observe that for 90° azimuth a left/right change would actually correspond to a backward/forward change. For the two positions at 90° elevation up/down would correspond to backward/forward, and for (46°,90°) left/right would correspond to down/up. In spite of these observations we decided, for clarity, to keep the left/right and up/down convention for all nominal positions.

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].

1.1.5. *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 were 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 direction of 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-hours sessions and each session was held on different days.

1.1.6. *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

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

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.

1.2. Results

Fig. 4 shows proportions of correct responses for each listener and each condition. Nominal positions are arranged in rows, and directional changes are separated in columns. 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). 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 higher elevations and this was more evident for discrimination along the vertical angle. Poorest performance was observed for (0°,90°) with upward/downward changes. 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 1. 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 upward/downward changes.

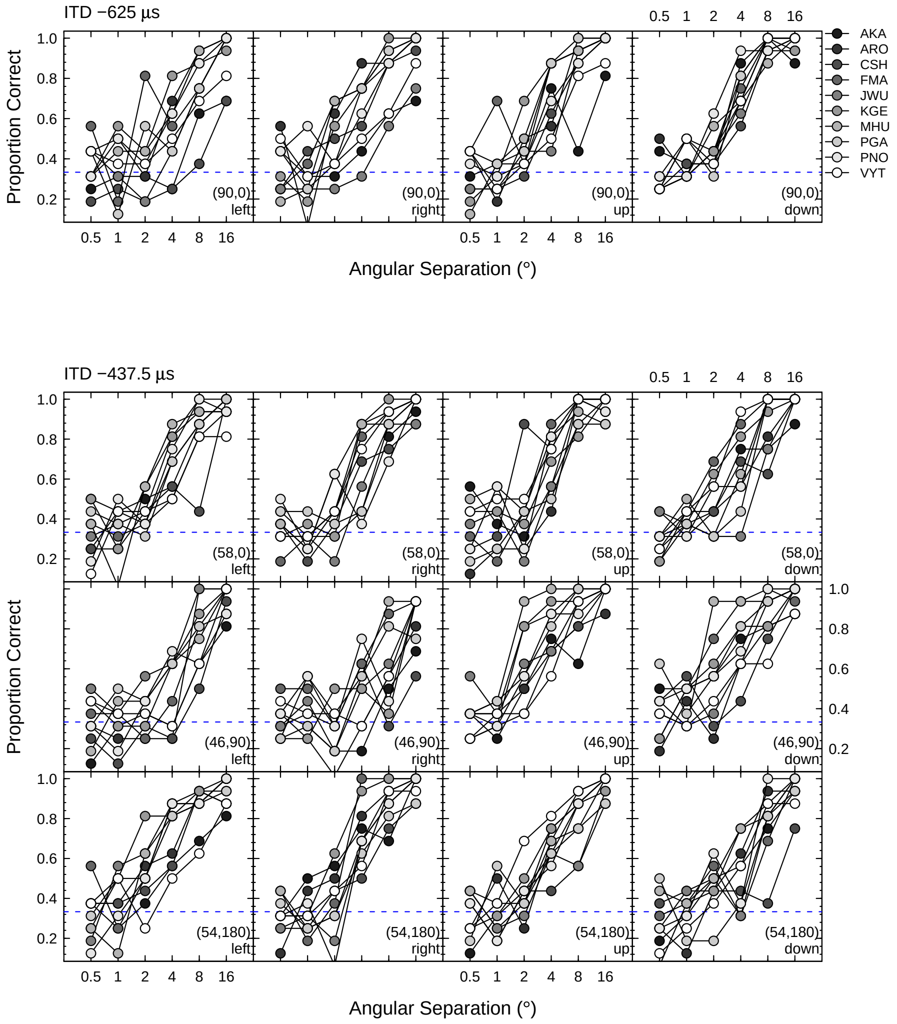


Fig. 4. Proportion of correct responses for spectral differences in HRTFs. Results for each combination of nominal position and directional change are plotted on single panels, and panels are grouped by common ITD. The four top panels show proportions for the position at 90° azimuth. The group of twelve panels at the bottom shows proportions for positions in the iso-ITD contour. The fifteen panels on next page show proportions for the positions in the median plane. The dashed line indicates chance performance.

1.3. 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 elevation (up/down) increase as elevation moves towards 90° for positions in the median plane, but they decrease for positions in the iso-ITD contour. For changes in azimuth, thresholds also increase with elevation and this is seen in both the median plane and iso-ITD contour. The direction dependency and range of thresholds observed in this study are comparable to those from a study conducted by [15], 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 elevation than in azimuth. In fact, at (0°,90°) (above the head) subjects were unable to perform above chance level for any of the elevation modes. The decrease in sensitivity to changes in magnitude as elevation moves towards 90° for upward/downward differences 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 angular modes. Fig. 5 shows differences in dB (expressed in absolute values) for the nominal direction (0°,90°) and for leftward, upward and downward changes. It is clear that when HRTFs are changed along azimuth (i.e. changes to the left) a small angular separation produces larger spectral differences than when the change is in elevation, being either upwards or downwards. Note that for downward changes, there are almost no differences in the frequency range 5–12 kHz.

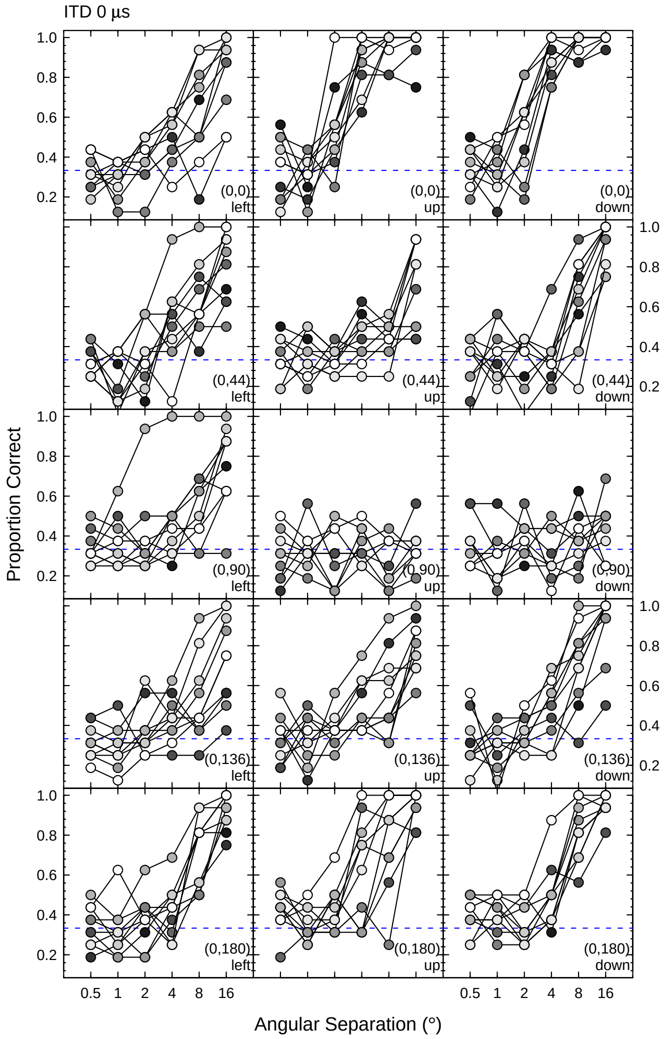


Fig. 4. Cont'd

For positions in the median plane the significance of the effects was evaluated in a two-way analysis of variance (nominal position x directional change). Because thresholds for (0°,90°) and changes in elevation could not be estimated, separate ANOVAs were done for lateral and vertical changes. For lateral changes main effects were not significant nor the interaction was significant. For vertical 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 down condition were lower 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 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°elevation whereas up/down thresholds decreased. Note that the effect of elevation on up/down changes was opposite to the effect observed in the median plane. This suggests that sensitivity to changes in elevation seems to increase as the sagittal plane moves to lateral positions.

Thresholds for (90°,0°) were significantly lower for up/down than left/right changes (p < 0.01). This result is consistent with vertical MAAs being generally smaller than horizontal MAAs for the 90° azimuth position in the horizontal plane [7, 21, 22]. 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 interaural level difference of

Table 1
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
-625	(90°,0°)	6.0	4.6	4.0	3.4
-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

roughly 15–20 dB makes unlikely that naive listeners could have made an effective use of spectral differences in the contralateral component.

2. EXPERIMENT II: AUDIBILITY OF TIME DIFFERENCES

2.1. 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°azimuth 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.

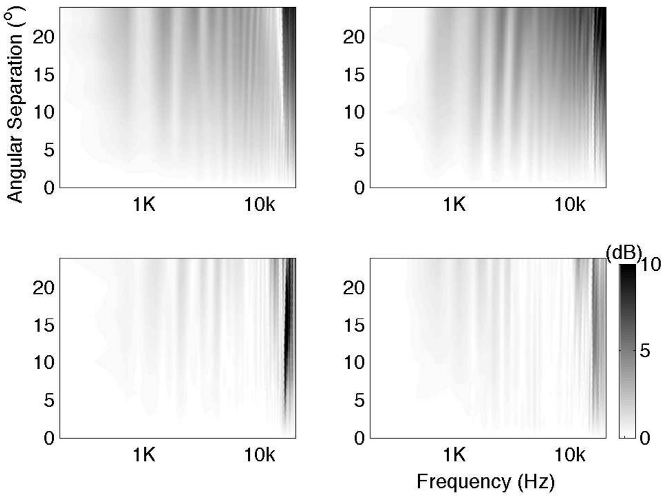


Fig. 5. Spectral magnitude differences as a function of angular separation for nominal direction (0°,90°). Top left and right panels show differences in azimuth for leftward changes produced in the left and right HRTFs respectively. Bottom panels show differences in elevation for upward changes (left panel) and downward changes (right panel). Differences are given in absolute dB values.

2.2. Results

Proportion of correct responses for the tested conditions are shown in Fig. 6. The abscissa specifies the ΔITD in μs, and is given in a logarithmic scale. Results for 90° azimuth refer to decrements from the -625-μs nominal ITD. For positions in the iso-ITD contour the left column represents increments in ITD and the right column represents decrements in ITD. For positions in the median plane 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.

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 2 and plotted on Fig. 7 along with individual thresholds. Data are grouped by common ITD and the abscissa represents elevation of the nominal position. For positions in the median plane mean thresholds ranged from 87.8 to 134.4 μs. There was not significant effect of nominal direction. For directions in the iso-ITD contour a two-way analysis of variance with sign of ΔITD and nominal direction as factors, showed that there was no significant difference between increments and decrements of ITDs nor was the difference between nominal direction signifi-

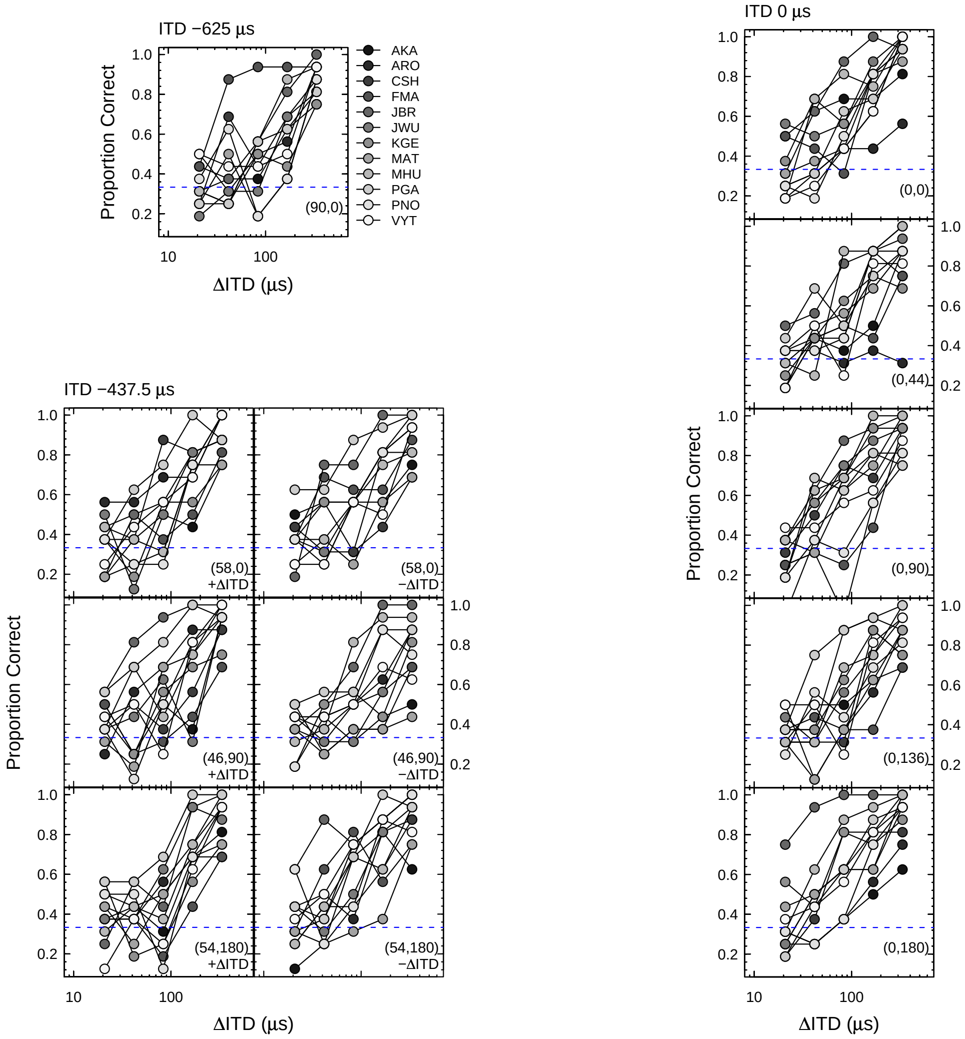


Fig. 6. Proportion of correct responses for discrimination of ITDs for all subjects and conditions. Results for each nominal position and ITD change are plotted on single panels, and panels are grouped by common ITD. The top-left panel shows subjects' performance for the 90° azimuth position. The group of six panels in the bottom-left shows performance for positions in the iso-ITD contour. The five panels to the right show performance for the positions in the median plane. The dashed line indicates chance level.

cant. 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.

2.3. 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 [23, 24]. 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 [25]. 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

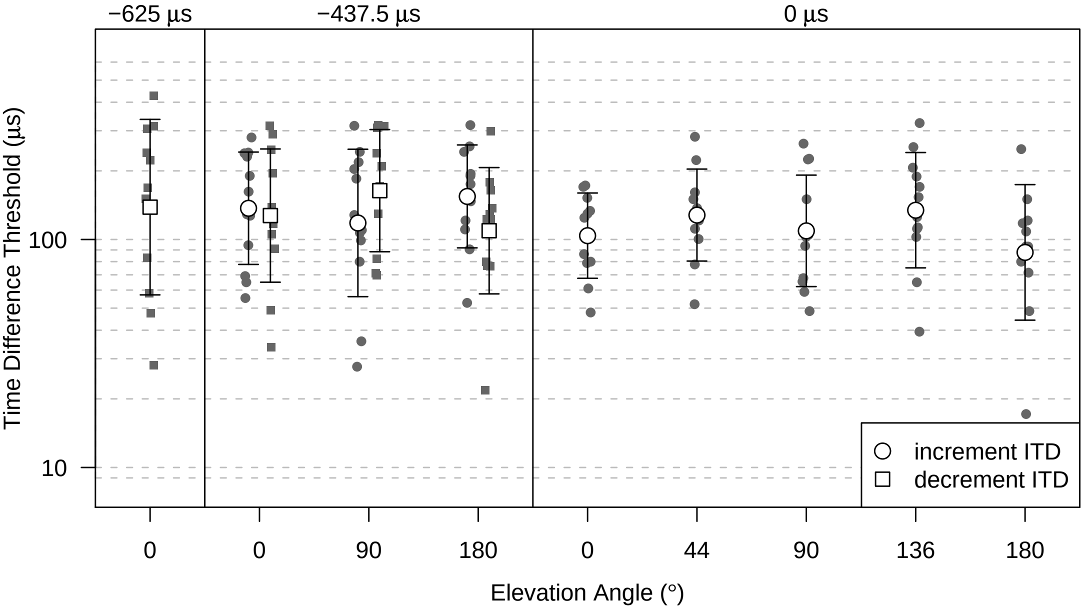


Fig. 7. Individual (grey color symbols) and mean thresholds on time differences in HRTFs. Data is grouped by nominal ITD and the abscissa describes the elevation for a given ITD. Thresholds for increments of ITD (circles) are observed for directions with ITDs 0 μs and -437.5 μs. Thresholds for decrements in ITD (squares) are observed for directions with ITD -437.5 μs and -625 μs. Error bars indicate 1 ± standard deviation.

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 [26, 27, 28]. Large differences between subjects have also been observed. In a study by [29] 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.

Table 2
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}	–

3. GENERAL DISCUSSION

3.1. 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. Approximating ITD thresholds to their corresponding change in degrees, and comparing them to those for spectral differences, indicate 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. It seems reasonable to think that changes in time differences would only offer a cue related to a shift in the apparent source position. Spectral differences on the other hand, 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 [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 were the stimuli 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°.

Here, the audibility of spectral and time differences has been tested separately. 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.

3.2. Implications in spatial resolution of HRTFs

In three-dimensional sound systems time-varying delays are commonly implemented with update rates equal to the sampling frequency. That is delay lines 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. 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. However, it is important to emphasize that at high elevations sensitivity to HRTF magnitude is more dependent on the direction in which HRTFs change than for lower directions.

4. 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.

5. ACKNOWLEDGMENTS

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

Notes

[1] A possible explanation for the difference between MAAs for real and virtual sources could be given in terms of in-head perception typically experienced when using headphones. For a source laterally displaced off midline with a given angle, the magnitude of the vector projected to the interaural axis would be greater for a well externalized source than for a source perceived inside the head. This means that if horizontal MAA for low-frequency tones are based on the ability to discriminate a change in lateral position (typically based on ITD changes at frequencies below 1.5 kHz), a stimulus reproduced over headphones would require a larger angular displacement to be as discriminable as if it was reproduced over loudspeakers.

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 (1) (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 (4) (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 (4) (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 (6) (1989) 2669–2672.

[6] R. L. McKinley, M. A. Ericson, D. R. Perrott, R. H. Gilkey, D. S. Brungart, F. L. Wightman, Minimum audible angles for synthesized location cues presented over headphones (a), J. Acoust. Soc. Am. 92 (4) (1992) 2297.

[7] D. R. Perrott, K. Saberi, Minimum audible angle thresholds for sources varying in elevation and azimuth, J. Acoust. Soc. Am. 87 (4) (1990) 1728–1731.

[8] R. M. Hershkowitz, N. I. Durlach, Interaural time and amplitude jnds for 500-Hz tone, J. Acoust. Soc. Am. 46 (6 (Part 2)) (1969) 1464–1467.

[9] J.-M. Jot, V. Larcher, O. Warusfel, Digital Signal Processing in the Context of Binaural and Transaural Stereophony, in: 98th Convention of the Audio Engineering Society, Paris, France, 1995, convention paper 3980.

[10] A. Kulkarni, S. K. Isabelle, H. S. Colburn, On the minimum-phase approximation of head-related transfer functions, in: Proc. of the ASSP (IEEE) Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, USA, 1995, pp. 84–87.

[11] P. Minnaar, H. Møller, J. Plogsties, S. K. Olesen, F. Christensen, Audibility of all-pass components in binaural synthesis, in: 106th Convention of the Audio Engineering Society, Munich, Germany, 1999, convention paper 4911.

[12] J. Plogsties, P. Minnaar, S. K. Olesen, F. Christensen, H. Møller, Audibility of all-pass components in head-related transfer functions, in: 108th Convention of the Audio Engineering Society, Paris, France, 2000, convention paper 5132.

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

[14] 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 (3) (1992) 1637–1647.

[15] P. F. Hoffmann, H. Møller, Some observations on sensitivity to HRTF magnitude, manuscript I. Submmited to J. Aud. Eng. Soc. (2007).

[16] B. P. Bovbjerg, F. Christensen, P. Minnaar, X. Chen, Measuring the head-related transfer functions of an artificial head with a high directional resolution, in: 109th Convention of the Audio Engineering Society, Los Angeles, California, USA, 2000, convention paper 5264.

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

[18] D. Hammershøi, H. Møller, Binaural technique, basic methods for recording, synthesis and reproduction, in: J. Blauert (Ed.), Communication Acoustics, Springer Verlag, Berlin, Germany, 2005, pp. 223–254.

[19] P. Minnaar, J. Plogsties, F. Christensen, Directional Resolution of Head-Related Transfer Functions Required in Binaural Synthesis, J. Audio Eng. Soc. 53 (10) (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] K. Saberi, L. Dostal, T. Sadralodabai, D. R. Perrott, Minimum audible angles for horizontal, vertical, and oblique orientations: Lateral and dorsal planes, Acustica 75 (1) (1991) 57–61.

[22] A. W. Bronkhorst, Horizontal and vertical MAAs for a wide range of sound source locations (A), J. Acoust. Soc. Am. 93 (4) (1993) 2351.

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

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

[25] E. R. Hafter, J. D. Maio, Difference thresholds for interaural delay, J. Acoust. Soc. Am. 57 (1) (1975) 181–187.

[26] 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 (4) (1995) 331–353.

[27] 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 (21) (2001) 12307–12312.

[28] 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 broadband noise, J. Acoust. Soc. Am. 103 (4) (1998) 2069–2078.

[29] J. Koehnke, H. S. Colburn, N. I. Durlach, Performance in several binaural-interaction experiments, J. Acoust. Soc. Am. 79 (5) (1986) 1558–1562.

Chapter 4

Manuscript III : Audibility of direct switching between head-related transfer functions

Audibility of direct switching between head-related transfer functions[*]

Pablo F. Hoffmann[†] and Henrik Møller

Section of Acoustics, Department of Electronic Systems
Aalborg University

Abstract

In binaural synthesis, signals are filtered with head-related transfer functions (HRTFs). In dynamic conditions HRTFs must be constantly updated, and thereby some switching between HRTFs must take place. For a smooth transition it is important that HRTFs are close enough so that differences between the filtered signals are inaudible. However, switching between HRTFs does not only change the apparent location of the sound but also generate artifacts that might be audible, e.g. clicks. Thresholds for the audibility of artifacts are defined as the smallest angular separation between switched HRTFs for which the artifacts is just audible. These thresholds were measured for temporal and spectral characteristics of HRTFs separately, and were denoted as the minimum audible time switching (MATS), and the minimum audible spectral switching (MASS). MATS thresholds were in the range of 5–9.4 μs, and MASSs were in the range of 4.1–48.2° being more dependent on the direction of sound than MATSs. Generally, for the implementation of dynamic binaural synthesis MATS impose higher demands in spatial resolution than MASSs.

0. INTRODUCTION

Similar to how animated movies are produced by sequences of still images, binaural synthesis of moving sound is typically done by sequentially presenting sound filtered with adjacent HRTFs. An inherent limitation of this technique is that moving sound, being a continuous phenomenon in real space, can only be synthesized using a discrete representation of space (HRTFs can only be measured for a finite number of directions). An intuitive criterion to evaluate whether a set of HRTFs provides the proper spatial resolution, is to make sure that switched positions are sufficiently close so that stimuli filtered with the corresponding HRTFs cannot be distinguished. In a study by [1] the audibility of differences in HRTFs was measured for changes in ITD and changes in spectrum separately. It was found that, using a discrimination paradigm, sensitivity to ITD was poorer than sensitivity to spectral differences, and that spectral differences require resolutions of 2.4–11° depending on direction.

Although this type of evaluation does not have a direct relation to the perception of moving sound, but rather to auditory spatial resolution for stationary sound sources, the approach is reasonable considering that human sensitivity to changes in the direction of sound is generally higher for stationary conditions than for dynamic conditions [see 2, for a review on dynamic and stationary spatial resolution]. However, note that this approach only evaluates our ability to detect differences in HRTFs. It does not take into account that due to the discrete nature of the spatial representation available, switching between HRTFs produces discontinuities in the waveform of the output signal. These discontinuities, if audible, are commonly heard as “clicks”, and their audibility is most probably proportional to the magnitude of the difference between HRTFs. Because HRTFs vary systematically with direction one would expect the audibility of clicks to depend on the spatial separation between the switched HRTFs. That is, the lower the spatial resolution of the switched HRTFs the higher the probability for the discontinuities to be perceived.

A common technique used to mitigate the problem of audible discontinuities is cross-fading. This technique is illustrated in Fig. 1(a) and is mathematically expressed by

$$y(n) = x(n) * h_i(n) \cdot \alpha + x(n) * h_j(n) \cdot (1 - \alpha) \qquad (1)$$

where $x(n)$ is the input signal, h_i and h_j the initial and target filters, α is a weighting factor, and $y(n)$ the output signal. Note that h_i and h_j represents head-related impulse responses. Here, $x(n)$ is convolved with h_i and h_j and the outputs of the convolution are weighted and summed up to yield $y(n)$. The cross-fading is controlled by α that gradu-

[*] Part of this work presented at the 118th Convention of the Audio Engineering Society, Barcelona, Spain, May 28–31, 2005. Part of this work presented at the 119th Convention of the Audio Engineering Society, New York, USA, October 7–10, 2005.

[†] Correspondence to pfh@es.aau.dk

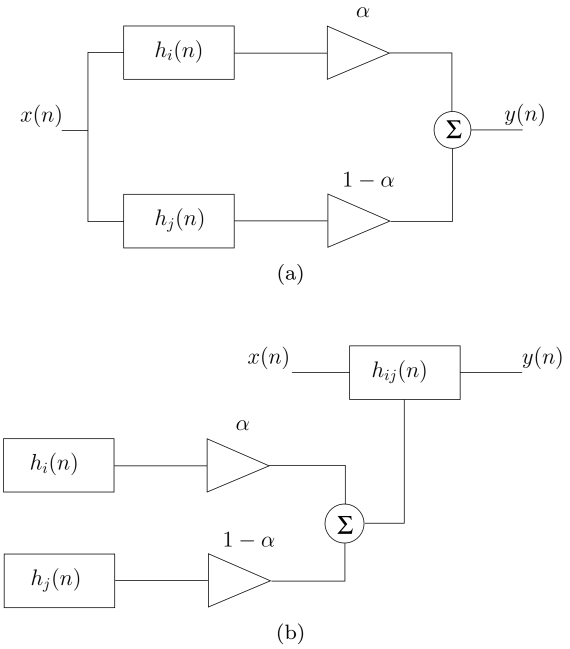


Fig. 1. Crossfade modalities corresponding to (a) cross-fading between filters' outputs and (b) cross-fading between filters.

ally change from 1 to 0 within a given time interval, thereby gradually changing from the initial filtering to the target filtering.

Observe that this cross-fading strategy requires at least two convolutions to run in parallel within the cross-fading interval. Also note that $x(n)$ is common to both convolutions, and using the distributive property of convolution, we can arrange Eq. (1) to

$$\begin{aligned} y(n) &= x(n) * [h_i(n) \cdot \alpha + h_j(n) \cdot (1-\alpha)] \\ y(n) &= x(n) * h_{ij} \end{aligned} \tag{2}$$

thus, the cross-fading is between filters (i.e. we obtain an interpolated filter h_{ij}) instead of between the filters' outputs. This scenario is depicted in Fig. 1(b). Note that when crossfading HRTFs the number of convolutions is reduced to one, hence the demand in computer power is also reduced. However, this reduction is just apparent since HRTF cross-fading requires α to be multiplied with all filter coefficients whereas output cross-fading only requires a number of multiplications equal to the number of HRTFs used in the cross-fading.

If we reduce the cross-fading time to its limit, this will converge to the sampling time, and thus the cross-fading will converge to a direct switching between HRTFs. Therefore, it appears that implementing HRTF cross-fading would only provide a substantial advantage over output cross-fading if we had previously computed and stored all the filters to be used. That is, measured HRTFs as well as interpolated HRTFs (all h_i, h_j and h_{ij} filters). This is a trade-off between computational power and memory requirements.

There are several aspects that may interact in dynamic systems. For example, depending on the velocity of sound source different resolutions may be needed. Furthermore, the update rate of the system will also play a role. This is because there are two aspects interacting. Say, for a given velocity and a fixed resolution, what would be the appropriate update rate?. Or, for a fixed update rate, what would be the appropriate resolution?. Thus to evaluate whether the strategy of direct switching is a viable alternative we should estimate the lower spatial resolution for which a direct switch does not produce audible artifacts.

0.1. Goal of the Study

In the present study two listening experiments are described. These experiments were conducted to measure the audibility of discontinuities produced by direct switching between HRTFs. The purpose is to estimate the largest possible angular separation for which switching between HRTFs does not produce audible artifacts. Timing and spectral characteristics of the HRTFs are separated in order to assess their individual effect. This is done by employing a model of the HRTF based on pure delays to control ITDs and minimum-phase filters to control the magnitude response. Experiment I measured audibility thresholds for dynamically changing delays and this threshold has been defined as the *minimum audible time switch* (MATS). Experiment II measured the *minimum audible spectral switch* (MASS), which is the threshold for direct switching between minimum-phase HRTFs. Audibility of HRTF switching in both experiments is estimated for several directions.

1. EXPERIMENT I: TIME SWITCHING IN HRTFs

1.1. Method

1.1.1. *Stimuli and Apparatus*

Broadband pink noise (20 – 9000 Hz) was used as source signal. HRTF switching was compared on thirteen directions distributed over the upper half of the sphere. The directions used in this study were the same as those used in [1] and they are shown in Fig 2. In addition, an iso-ITD contour to the right ((-56°,0°), (-46°,90°), (-54°,180°)) and the location at -90° azimuth in the horizontal plane were included. The HRTFs used to render directional sound were selected from a dataset measured with a directional resolution of 2° on an artificial head [3]. HRTFs were in the form of 72-coefficient finite-impulse-response (FIR) minimum-phase filters. Although not directly connected to binaural synthesis, it has been shown that the artificial head used to measure the HRTFs provides good localization cues for binaural recordings [4].

HRTF-filtered stimuli were played back over equalized Beyerdynamic DT-990 Pro circumaural headphones. The procedure used to compute the equalization filters is described in [5]. HRTF filtering and headphone equalization were done off-line and thirteen 5-s stimuli (one for each of the selected directions) were stored as 16-bit PCM stereo files. Stimuli were presented as a continuous sound and thus they were looped during playback. Raised-cosine ramps of 10 ms applied to the onset and offset of the stimuli were sufficient to avoid audible artifacts when looping.

An Intel-based personal computer (PC) equipped with a professional audio card RME DIGI96/8 PST was used to control the experiment. The rest of the equipment consisted of a 20-bit D/A converter (Big DAADi) set at a 48-kHz sampling frequency, and a headphone amplifier (Behringer HA4400). All the equipment was placed in the control room. The overall gain of the system was set so that the sound pressure at the ears produced by the source signal (unfiltered pink noise) was approximately equivalent to a free-field sound pressure level of 72 dB.

1.1.2. *Time Switching Implementation*

Time switching was implemented using digital delays and was produced by continuously alternating between a fixed reference delay and a variable delay. In this way, stimuli were presented as a continuously changing sequence like *off-on-off-on* and so forth. The *off* part corresponded to the signal with the fixed delay and the *on* part to the signal with the variable delay. Pilot experiments showed that audibility thresholds of time switching may be below one sample at a 48 kHz sampling frequency. Therefore, time switching was implemented by combining an integer variable-delay line with FIR fractional delay filters.

Fractional delay filters are capable of producing delays shorter than the sampling interval (for a thorough review of this topic refer to [6]). Typically, the cost of using FIR fractional delay filters is that the magnitude response is not flat over the entire frequency range. If a full-band flat response is a requirement, it is possible to design all-pass fractional delay filters. However, for time-varying filtering FIR filters are better suited than infinite impulse response (IIR) filters. This is because time-varying IIR filters produce transients at the output signal whereas FIR filters do not [7].

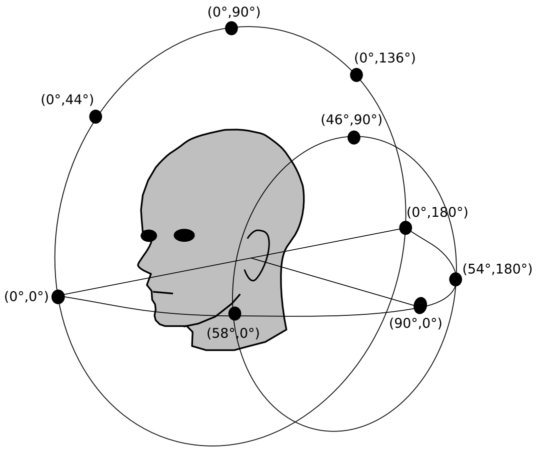


Fig. 2. Directions employed in the listening experiment. Azimuth and elevation are indicated in an interaural-polar coordinate system. Five directions were selected in the median plane, three directions in the iso-ITD contour of -437.5 μs, and the leftmost direction 90° azimuth with a calculated ITD of -625 μs. In addition, three directions in an iso-ITD contour to the right (437.5 μs) and the rightmost direction -90° azimuth (625 μs) were also included.

Coefficients of all fractional delay filters were calculated off-line, and a table-lookup method was used to switch between them. Filter coefficients were computed using Lagrange interpolation [6]. The simplicity of this design technique is that the coefficients are easily obtained using a closed analytical form given by

$$h(n) = \prod_{\substack{k=0 \\ k\neq n}}^{N} \frac{d-k}{n-k} \qquad \text{for } n = 0, 1, 2, ..., N \tag{3}$$

where d is the desired fractional delay in samples and N is the order of the filter. Here, we found that N = 11 was sufficient to ensure that filters had a flat frequency response and constant group delay in the effective bandwidth of the stimuli (20 – 9000 Hz). Fig. 3 shows examples of the filters implemented for 1/4, 1/2 and 3/4 sample delays. Note that the filters has an inherent integer delay corresponding to (N-1)/2. This delay was compensated for by delaying the signal by the same amount during the *off* part of the stimuli (it corresponded to the fixed reference delay).

During an *off–on* switching state the appropriate fractional delay filter was retrieved from memory and convolved with the signal. If delays larger than one sample were required, the additional integer part of the delay was introduced prior to the fractional delay filtering. All the operation was completed within one sampling interval, and thus for a sufficiently large delay shift a clear click was perceived. To test whether the switching rate had an effect on the audibility of time switching, two switching rates were used: 50 Hz and 100 Hz.

1.1.3. *Subjects*

Twenty-one paid subjects participated in this listening experiment. The panel consisted of 10 males and 11 females. Their ages ranged from 20 to 31. Subjects were selected by means of an audiometry screening at hearing levels $\leq$ 10 dB HL, at octave frequencies between 250 and 4 kHz, and a hearing level $\leq$ 15 dB HL at 8 kHz.

1.1.4. *Psychometric Method*

The listening experiment was conducted in a sound-insulated cabin specially designed for subjective experiments. Listeners were seated in front of a screen that displayed a graphic interface composed by a slider and a push button labeled “OK”. The slider could move along a vertical bar and was controlled via a mouse. The position of the slider determined the amount of delay introduced during the time switching. As the slider moved upwards and downwards the variable delay increased and decreased respectively.

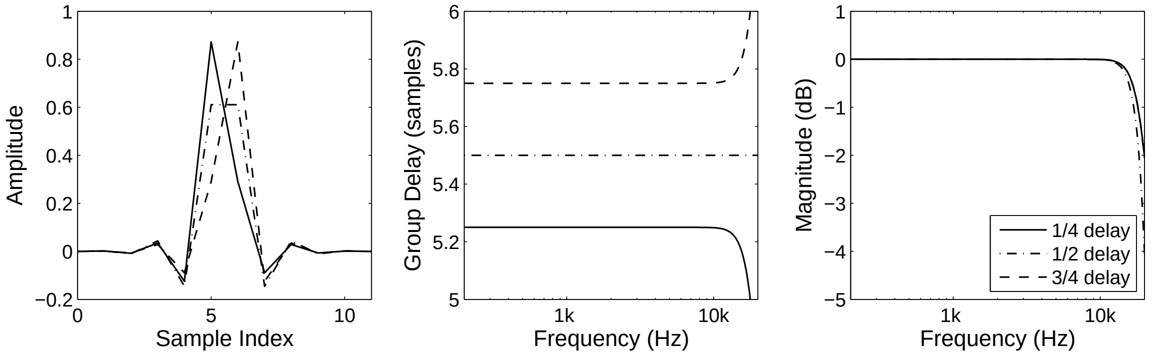


Fig. 3. Example of FIR fractional delay filters for delays of 1/4, 1/2 and 3/4 samples. Filters are of order 11th and were designed using Lagrange interpolation. The left panel shows the impulse responses. The center panel shows the group delay, where we observed that the filters' inherent delay is equal to (N-1)/2 with N being the filter's order. The right panel shows the magnitude responses of the filters.

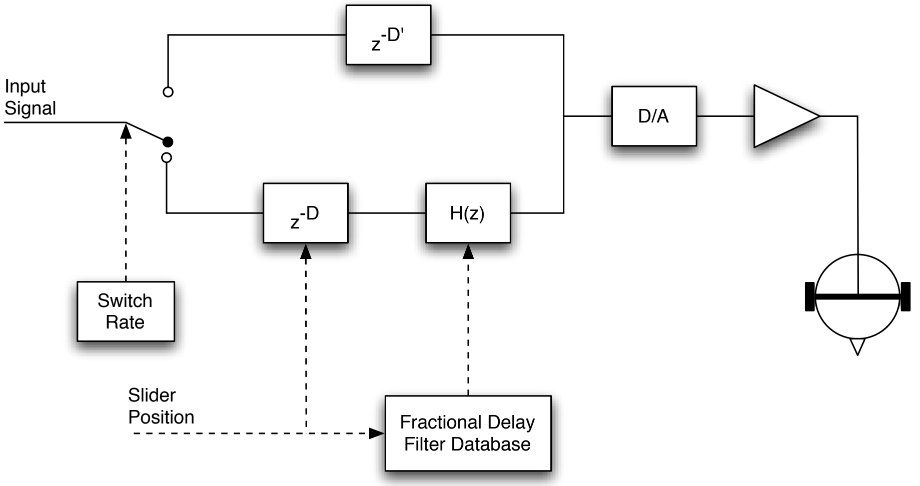


Fig. 4. Diagram of the dynamic time switching implementation. The input signal is the pink noise already filtered with an HRTF and the headphone equalization filters. Slider position data is constantly retrieved in order to select the appropriate fractional delay filter and integer delay. The fixed delay D' compensates for the extra integer delay introduced by the fractional delay filters.

A schematic of the system implemented to control the dynamic delay is shown in Fig. 4. The minimum delay was one tenth of a sample, which corresponded to about 2.1 μs; and the maximum delay was 4.2 ms. Delays were incremented logarithmically in 20 steps per decade, yielding a scale of 67 different delays. Time switching was presented diotically and the reason was because we wanted the audible artifacts to be the only cue, and avoid any confounding cue such as changes in the apparent direction of the stimuli. It is possible that this would have been the case if time switching had been applied to one ear only (dichotic presentation).

For estimating MATS thresholds the method of adjustment was employed [8]. Subjects were instructed to find the lowest position of the slider for which they just perceive a distortion (usually heard as a train of clicks). Listeners were encouraged to move the slider up and down several times, and to perform the task as fast as they could, but no limit was imposed to the response time. The scale of 67 delays was contained within a frame equal to half the length of the slider bar. Fig. 5 shows a representation of this. The position of the frame along the bar was randomized across trials, and this was done so as the position of the slider at threshold varied. In this way, we believe that a potential bias caused by threshold estimation based on visual cues was reduced, e.g. distance from the slider to the bottom. Below the lower end of the frame no switching was applied, and above the upper end of the frame the maximum delay was used for switching. The initial position of the slider was randomly located either to the bottom or to the top of the bar. This ensured that the slider position was at a clear distance from threshold at the beginning of each trial.

1.1.5. *Experimental Design*

Within an experimental block all nominal directions were presented once. Update rates were arranged so either time switching operated at 50 Hz for seven directions and at 100 Hz for the remaining six directions, or vice versa. Prior to the main experiment all subjects completed three blocks of practice. For the main experiment subjects participated in

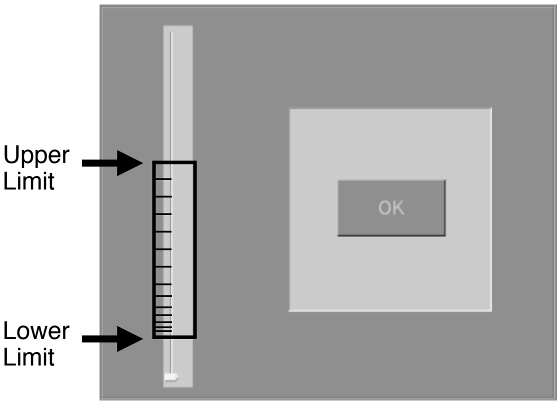


Fig. 5. Graphic interface presented to the subjects during the experiment. The graded box (not seen by the subjects) represents the scale of delays whose position along the slider-bar was randomized across trials.

two experimental sessions of 3 blocks each, and one session of 4 blocks. Experimental sessions were conducted in different days for the individual subjects. The 26 conditions (thirteen nominal directions x two switching rate) were repeated five times for each listener.

1.2. Results

A total of 130 responses were obtained per subject. None of the subjects gave responses equal to, or above, the maximum time switching (4.2 ms), and 0.11% (3 responses) of the total number of responses fell below the smallest time switching (2.1 μs). These three responses are not considered for further analysis. They were given for different conditions on two subjects, thus there were only 4 repetitions available for these conditions and subjects.

Since data appeared to better represent normal distribution on a logarithmic scale than on a linear scale, all statistics were done on the log domain. Individual thresholds were defined as the mean across repetitions for each condition. Figs. 6(a) and 6(b) show individual thresholds for each switching rate respectively. Directions are expressed in elevation angle and grouped by ITD. Individual thresholds are fairly consistent across directions. Also plotted are the responses of two subjects who represent extreme data. The most sensitive subject was capable to hear the artifacts produced by the minimum time switching for all conditions but one, which was also considerably lower than those of other subjects. The other subject showed the lowest sensitivity for all conditions and thus represents the upper bound of the data. Interestingly, the same situation is observed for both switching rates.

Fig. 6(c) shows mean thresholds calculated across subjects for each switching rate, and they are summarized in Table 1. The range of mean thresholds is 5.6–9.4 μs for 50-Hz switching rate, and 5.0–8.5 μs for 100-Hz switching rate. Both ranges correspond to thresholds obtained at (0°,44°)–(90°,0°). Thresholds increased as the nominal directions moved to the left side but this is not observed for nominal directions to the right. A two-way within-subject analysis of variance revealed highly significant main effect of direction (F(12,240) = 11.5, $p < 0.001$), and a highly significant main effect of switching rate (F(1,20) = 137.2, $p < 0.001$). Mean thresholds for 50-Hz switching rate were consistently greater than those for 100-Hz switching rate. The interaction between nominal direction and switching rate was not significant ($p = 0.61$).

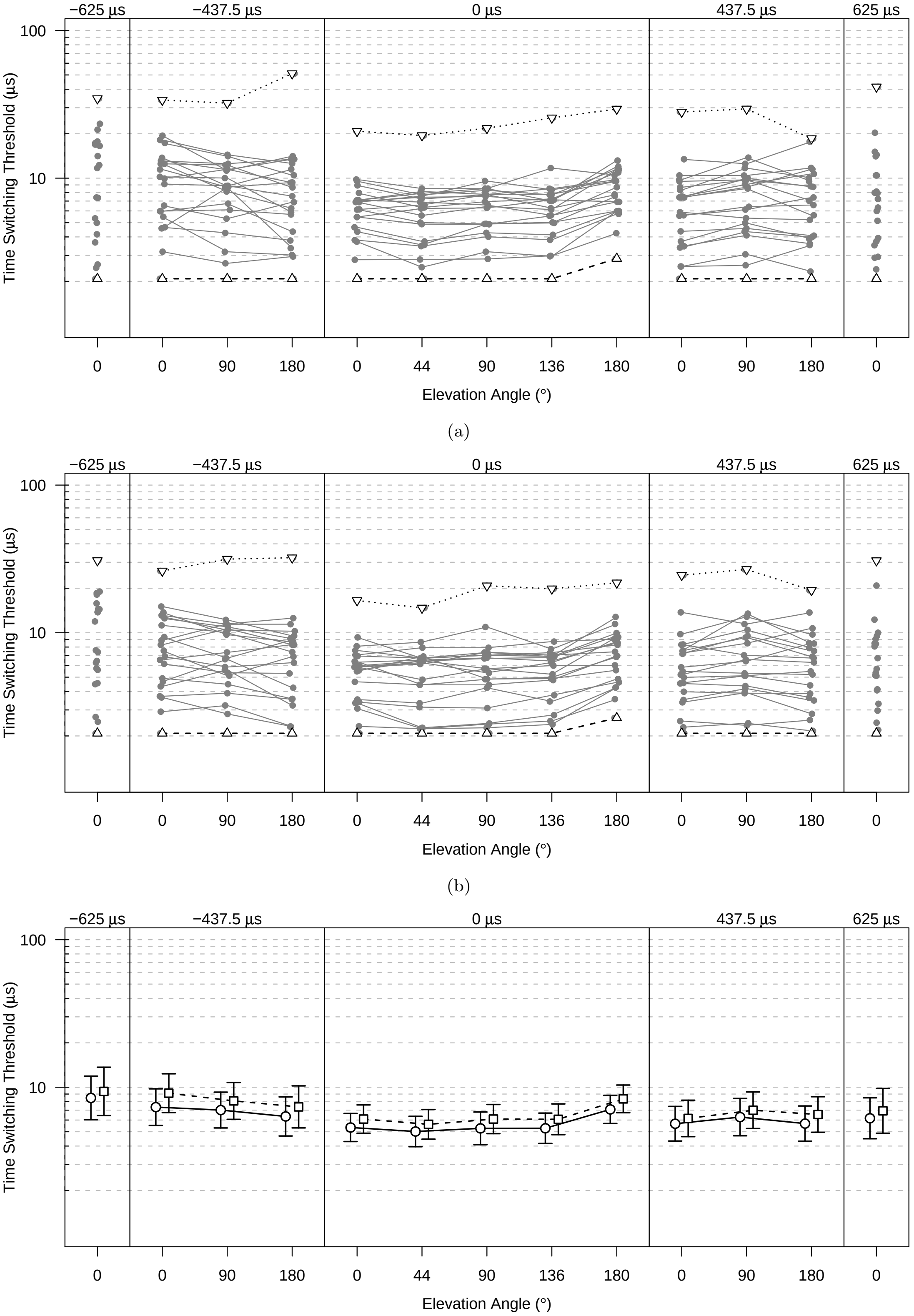
1.3. Discussion

Audible artifacts such as those introduced by dynamically changing delays are commonly perceived as clicks. This is because the energy of a click is in theory distributed all over the frequency with equal magnitude, and this energy is released within a very narrow time interval. A high switching rate would produce a larger number of clicks per time unit than a lower switching rate, and thus, the likelihood of the clicks to be audible is higher. Thresholds on the audibility of clicks have been reported to decrease as the click-presentation rate increases [9]. Our results are in agreement with this notion because subjects were significantly more sensitive to artifacts at a higher switching rate.

Even though MATSs were obtained for delays applied to both ears simultaneously, it seems worthy to compare these thresholds with sensitivity to dynamic changes in ITD. In a study by [10] discrimination between static ITDs and dynamically changing ITDs was examined. For low-rate fluctuations subjects could perceive lateral movements of the sound image. As the rate of fluctuation increases to values greater than 10 Hz, subjects could not longer track the

Table 1
Mean MATS for the tested directions. Thresholds are given in (μs).

Direction	Switching rate	
	50 Hz	100 Hz
(90°, 0°)	9.4	8.5
(58°, 0°)	9.1	7.3
(46°, 90°)	8.1	7.0
(54°, 180°)	7.4	6.4
(0°, 0°)	6.1	5.3
(0°, 44°)	5.6	5.0
(0°, 90°)	6.1	5.3
(0°, 136°)	6.1	5.3
(0°, 180°)	8.4	7.1
(-56°, 0°)	6.5	5.7
(-46°, 90°)	7.0	6.3
(-54°, 180°)	6.2	5.7
(-90°, 0°)	6.9	6.2



(a)

(b)

(c)

Fig. 6. MATS thresholds. Individual thresholds are shown in panel (a) for 50-Hz update rate and panel (b) for 100-Hz update rate. Lines connect direction with a common ITD. The dotted- and dashed-line represent extreme results (see text). Mean MATSs across subject are shown in panel (c) for 50-Hz (squares) and 100-Hz (circles) update rate. Error bars indicate 95% confidence intervals.

changes in source position but they started to perceive a wider intracranial image compared to the image produced by the fixed-ITD stimuli. This relatively poor ability of the binaural system to follow fluctuations in ITD has been called *binaural sluggishness* [11]. Therefore, it appears that in terms of synthesis of temporal changes, the generation of artifacts is the critical criterion for setting the minimum time interval required to change delays without audible artifacts.

Another factor that should be considered is the fastest velocity that one would like to simulate. Using the lowest threshold found for each switching rate we can estimate a limiting velocity in terms of amount of switched delay per second. In case of a 100-Hz switching rate we have a threshold of 5 μs, and in case of 50-Hz switching rate we have a threshold of 5.6 μs. By multiplying threshold with switching rate we can estimate the respective fastest velocities to be 500 μs/s and 280 μs/s. If we take these values as changes in ITD we could observe that simulating a source moving along the horizontal plane with a velocity of 90°/s (625 μs/s) would produce audible artifacts. This may be a problem if we consider that during localization it has been observed that listeners can move their heads with velocities of 175°/s [12]. In addition, even faster velocities are required in the use of propagation delays for incorporating Doppler effects [13]. As we will discuss in the following, current auralization systems update delays at much higher update rates than 50 and 100 Hz.

Findings from this experiment suggest that on average time switching should not exceed 5 μs. In interactive three-dimensional sound systems delay lines are commonly updated at every sample [14, 15, 16], and this applies to both propagation delays and ITDs. For example, the DIVA system [14] operates at a 20-Hz update rate and linearly interpolates between delays at every sample during a time interval of 50 ms. The interpolation is performed using a 1st-order FIR fractional delay filter. Because the system works at a 44.1-kHz sampling frequency the delay interpolation is performed in 2205 instances (50 ms x 44.1 kHz). Now, let us assume a sound moving at 250°/s (considered as a fast moving sound) and going from (0°,0°) to the side along the horizontal plane. Assuming that the first update captures the 0° azimuth direction, the second should return an azimuth value of approximately 12°. This directional change has an associated change in ITD of about 90 μs, meaning that intermediate ITDs are updated in successive steps of 40 ns (90μs/2205), which is two orders of magnitude below the threshold. This constitutes evidence that current applications are well within the range of time switching relative to MATS thresholds.

The fact that current computational power allows for update rates higher than 20 Hz implies that for a smooth delay transition the interpolation interval could be much shorter, and/or delays may not need to be updated at every sample. It is also possible that more accurate fractional delay filters (higher orders) can be implemented. Furthermore, these thresholds may be extended to dynamic varying delays for sounds moving close to the listener since it has been shown that ITDs for near-field HRTFs are similar to those measured from far-field HRTFs [17].

2. EXPERIMENT II: SPECTRAL SWITCHING IN HRTFs

The aim of this experiment is to estimate the ability of listeners to perceive artifacts when the magnitude spectrum of HRTFs is rapidly changed. The paradigm employed is a direct switching between minimum-phase HRTFs where the angular separation between the switched HRTFs is varied in order to find the just-audible switching.

2.1. Method

2.1.1. *Subjects*

Ten paid subjects participated in the listening experiment, nine males and four females. Their ages ranged from 22 to 31. Seven subjects had previously participated in the experiment on MATSs. All subjects fulfilled the hearing requirements corresponding to hearing levels $\leqslant$ 10 dB HL at octave frequencies from 250 Hz to 4 kHz and $\leqslant$ 15 dB HL for 8 kHz.

2.1.2. *Stimuli and Playback System*

Broadband pink noise (20–16000 Hz) was used as the source signal. The same thirteen directions employed in the previous experiment were used in this experiment. The playback system was almost identical to the one employed in the previous experiment. Here, the output from the D/A converter went 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 to inaudible levels. The stereo output from the attenuator was delivered to the listener through a pair of equalized Beyerdynamic DT-990 circumaural headphones.

2.1.3. *Spectral Switching*

Spectral switching was implemented by updating the minimum-phase component of the HRTFs while keeping the ITD unchanged. The switching was set to work at a rate of 100 Hz, and it was realized by changing all coefficients from one filter to another in a sample-to-sample operation. Angular separation between the switched HRTFs was the parameter that varied.

For each direction adjacent HRTFs were switched in two modes, and thus two sets of filters were computed. One mode corresponded to switching in azimuth, and the other mode corresponded to switching in elevation. For example let us assume that a sound is presented from (0°,0°) and switching is in azimuth. This particular scenario is depicted in Fig. 7(a). The switching operation takes place between two HRTFs in the horizontal plane, one spanning to the left of (0°,0°) and the other to the right at equal distance. For an angular separation of θ_1 switching would be between locations L1 and R1 by alternating between their corresponding minimum-phase HRTFs but keeping the ITD of (0°,0°). If the angular separation ϕ_2 is selected the switching would take place between locations L2 and R2. For switching in elevation, HRTFs corresponding to directions in the median plane (spanning up and down from the horizontal plane) would have been used instead. This scenario is illustrated in Fig. 7(b).

Note that for directions ±90° azimuth, switching in elevation cannot be applied. Instead, two azimuth-switching modes were implemented, one switching in the horizontal plane extending the angle horizontally (0°/180° elevation) and the other switching in the frontal plane extending the angle vertically (90°/270° elevation).

Angular separations ranged from 0.5° to 60° and they were incremented in steps of 0.5°. The resolution of the measured HRTFs was incremented using linear interpolation between the minimum-phase impulse responses. A total of 26 sets of adjacent HRTFs were constructed (13 nominal directions x 2 switching modes), and each set consisted of 120 pairs of impulse responses spanning ±30° from their respective directions. For directions (0°,90°) and (±46°,90°) the resolution was 1°, and thereby only 60 filters were effectively utilized. The use of less resolution on these nominal directions was based on results from preliminary experiments [18].

2.1.4. *Experimental Procedure*

The response protocol used for the estimation of MASSs was identical to the one used in the estimation of MATSs. A graphic interface composed of a slider and a push button was displayed on a screen. The slider could be moved along a vertical track-bar via a mouse. The position of the slider along the track-bar controlled the angular separation between the HRTFs used for the spectral switching. As the slider moved upwards or downwards the angular separation increased or decreased respectively.

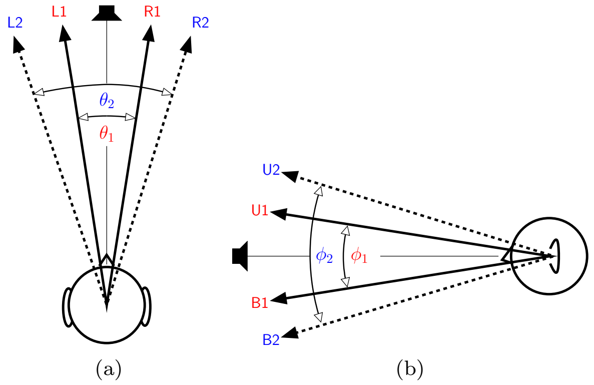


Fig. 7. Description of spectral switching for direction (0°,0°). In (a) switching is in azimuth. For an angular separation of θ_1 HRTFs are switched every 10 ms between positions L1 and R1 respectively. If the angular separation is θ_2, HRTFs for positions L2 and R2 are switched. In (b) switching is in elevation, and simmilarly, for angular separations ϕ_1 and ϕ_2 switching takes place between positions U1-B1 and U2-B2 respectively.

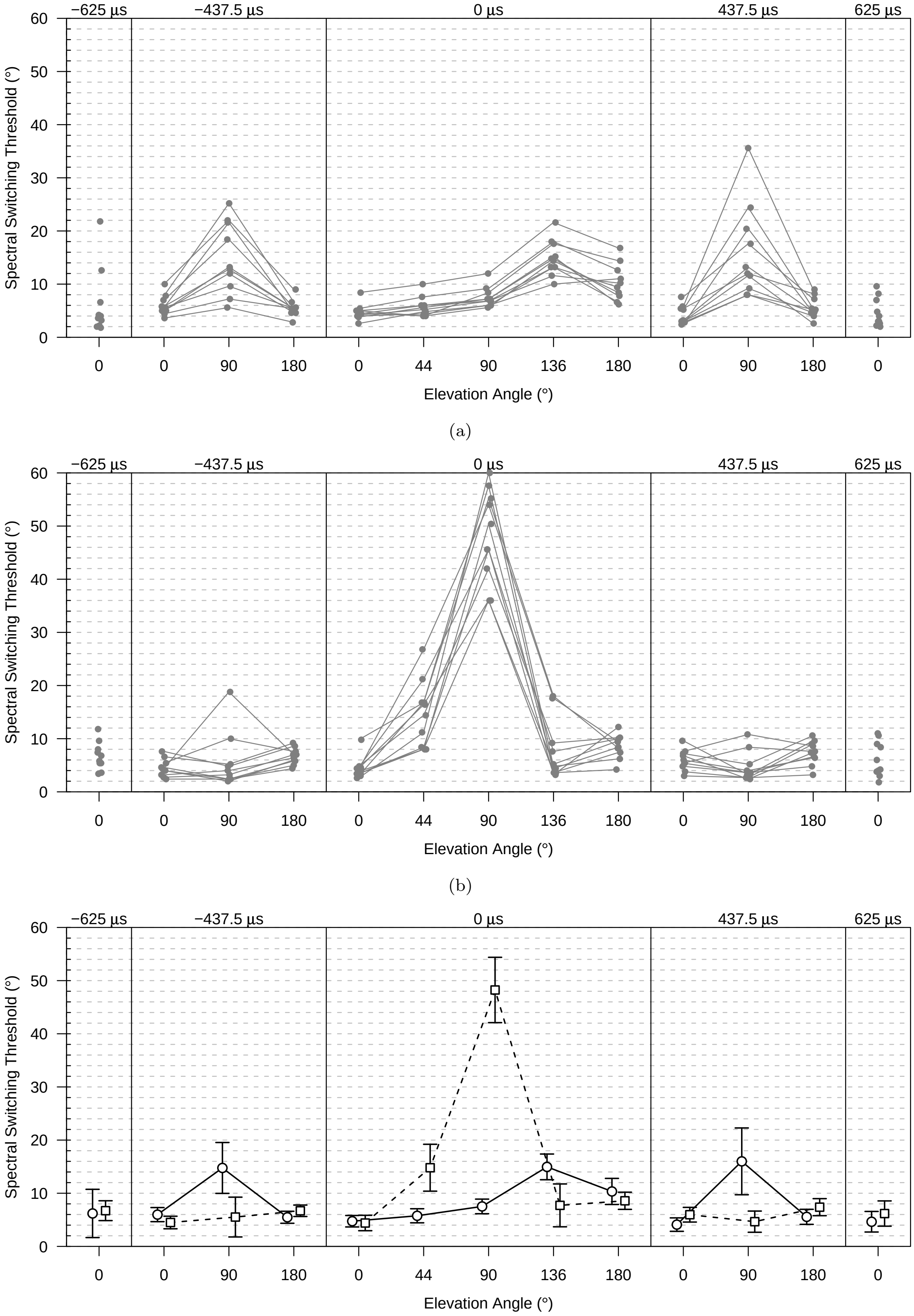
During a single MASS determination, a stimulus for a given nominal direction was presented as a continuous sound to the subject. The task of the subject was to find the lowest position of the slider where he/she could just perceive the presence of a "distortion" in the signal. The spectral switching effect became easier to perceive as the angular separation increased. Subjects were instructed to move the slider up and down several times before responding. Subjects were also encouraged to perform the task as fast as they could but no time limit was imposed. Once they had selected the slider position they entered a response by pressing the button. After a 2-s silence interval a new stimulus was presented. The position of the frame containing the array of angular separations was randomized along the track-bar. Below the lower end of the frame no switching was applied, and above the upper end of the frame the angular separation used for the switching was equal to the maximum (60°). The initial position of the slider was randomly selected at either the top or the bottom of the track-bar. This ensured that the slider position was at a clear distance from threshold at the beginning of each trial.

Subjects were seated in front of a screen that displayed the graphic interface. First, a few trials were presented in order to acquaint them with the task and the procedure. Posteriorly, they were presented with two or more blocks of stimuli for the practice sessions. One block consisted of thirteen trials. All nominal directions were presented in one block, and the switching modes, either seven times azimuth and six times elevation or vice versa, were randomly assigned. All subjects had at least two practice blocks and inexperienced subjects completed two or three additional blocks. Each combination of nominal direction and switching mode was repeated five times. Data were collected during two experimental sessions (5 blocks each) that were held on different days for each subject.

2.2. Results

A total of 130 responses were obtained per subject. Individual thresholds were computed as the arithmetic mean of the five repetitions for each condition. Twenty-one responses, corresponding to 1.6% of the total, were given at the largest angular separation and all of them were for the nominal direction (0°,90°) and switching in elevation. Eight responses (0.6%) were given below the smallest angular separation and they were not considered for further analysis. Therefore, for the conditions and subjects in which these responses were observed the computed mean was based on less than five repetitions (between three and four).

Figs. 8(a) and 8(b) show individual MASS thresholds for switching in azimuth and elevation respectively. For directions in the median plane and switching in azimuth, thresholds tended to increase with elevation from 0° to 136°,



(a)

(b)

(c)

Fig. 8. MASS thresholds. Individual thresholds are shown in panel (a) for switching in azimuth, and in panel (b) for switching in elevation. Lines connect responses for directions with a common ITD. Mean MASSs across subjects are shown in panel (c) where switching in azimuth and elevation are indicated by circles and squares respectively. Error bars indicate 95% confidence intervals.

Table 2
Mean MASS thresholds given in (°).

Direction	Switching mode	
	Azimuth	Elevation
(90°, 0°)	6.2	6.7
(58°, 0°)	6.0	4.5
(46°, 90°)	14.8	5.5
(54°, 180°)	5.5	6.7
(0°, 0°)	4.7	4.4
(0°, 44°)	5.8	14.8
(0°, 90°)	7.5	48.2
(0°, 136°)	15.0	7.7
(0°, 180°)	10.3	8.6
(-56°, 0°)	4.1	6.0
(-46°, 90°)	16.0	4.7
(-54°, 180°)	5.6	7.4
(-90°, 0°)	4.5	6.2

where the largest thresholds were observed for almost all subjects, and they decreased again for (0°,180°). A similar pattern was observed for switching in elevation but with largest threshold for (0°,90°) and a considerably fast increase/decrease in threshold as compared with switching in azimuth. Thresholds for the two cones also showed certain dependency with elevation, and this was more pronounced for switching in azimuth than in elevation.

Mean MASS thresholds across subject are shown in Fig. 8(c) and summarized in Table 2. Mean MASSs ranged from 4.1 to 16° for switching in azimuth, and from 4.4 to 48.2° for switching in elevation. A two-way within-subject analysis of variance revealed a highly significant main effect of direction (F(12,108) = 67.5, p <0.001), a significant main effect of switching mode (F(1,9) = 16, p<0.01), and a highly significant interaction between nominal direction and switching mode (F(12,108) = 70.7, p<0.001). This can be attributed to the fact that thresholds for 90° elevation in both left and right cones were higher for switching in azimuth than in elevation, whereas for 90° elevation in the median plane the opposite was observed. Considerable differences in mean thresholds between the two switching modes are only observed for elevations different from 0° and 180°.

2.3. Discussion

MASS thresholds are comparable to thresholds obtained from discrimination of spectral differences [1, 5]. This indicates that the audibility of switching between HRTF filters may stem from differences and not from artifacts. From a practical point of view this is encouraging, since it suggests that measurement on the ability of listeners to discriminate differences in HRTFs using stationary sources, may be sufficient to estimate an adequate resolution for the spectral characteristics.

If we use the lowest threshold to estimate the fastest velocity that we could implement without artifacts we obtain a value of about 400°/s. This velocity could be considered as a very high one if we relate it to how fast listeners, or listeners' heads, can move (approx. 180–200°/s).

MASS thresholds for directions in the median plane and the cones show a tendency to increase as a function of elevation. This tendency is in agreement with a study conducted by [19], who examined the required directional resolution for interpolated HRTFs such that they are indistinguishable from measured HRTFs. It was found that the requirements are less demanding for higher elevations. However, the increase in threshold at high elevations also depend on the switching direction. In the median plane thresholds increase for switching in elevation and in the cones thresholds increase for switching in azimuth.

The large difference in thresholds observed at higher elevations for the two switching conditions implies that the resolution becomes somewhat dependent upon the trajectory of the moving sound. The same tendency has been observed in thresholds for the audibility of spectral differences in HRTFs [1, 5]. In practical terms, one could simply select the lower threshold as the required spatial resolution, but the possibility of a system with a spatial resolution dependent of the direction of motion may also be considered.

3. GENERAL DISCUSSION

To our knowledge there only is one study that directly addresses the issue of audibility of discontinuities created by switching between directional filters. [20] compared several switching strategies: direct switching, overlap-add method, weighted overlap-add method, and cross-fading using three different envelope functions (square root, cosine, and a Fourier Series). From an objective analysis based on the expansion of the effective frequency bandwidth [21] that occur at the moment of switching, Kudo concluded that the weighted overlap-add method and the cross-fading using Fourier Series generated the less amount of discontinuity to the signal waveform. This analysis was supported by a listening experiment that evaluated how much discontinuities affect the subjective quality of virtual sound. Because cross-fading is based on intermediate filters it is not surprising that better weights were given to cross-fading methods than to a direct switching.

3.1. Comparison between MATS and MASS

Assuming MATSs as ITDs and computing their corresponding directional change in degrees results in thresholds of roughly 1° for directions in the median plane, 2–3° for directions in the cones, and 6–8° for directions at ±90° azimuth. Therefore MATSs are generally lower than MASSs, which supports the view that timing information should be updated at higher rates than those used to update the directional filters that control spectral information.

If we compare MATSs and MASSs with measures of static spatial resolution, such as the minimum audible angle (MAA), we observe that for the forward direction MATSs are comparable to MAAs for real sources (about 1°) [22]. MAAs have found to be about 5° for virtual sources based on generic HRTFs [23], and this is more comparable to the MASSs obtained in this study.

3.2. Implications for dynamic binaural synthesis

In the context of dynamically varying ITD implementation, it seems worthy to compare results between time differences in HRTFs, and time switching between HRTFs. In the study on time differences in HRTFs [1], the estimated threshold for the most sensitive subject for the forward direction was 48 μs. For time switching, the threshold measured for the same position is 5-6 μs. That is, MATS are at least 8–9 times lower than the minimum audible time difference. 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.

It is important to be cautious on how these thresholds can be generalized to other stimuli. We believe that for stimuli with broader bandwidths these thresholds may be applicable, but not for narrow-band stimuli. This is because the broader the bandwidth the more random is the nature of the sound, and thus, the less probable is for the switching to be audible. Essentially, for signals with broader bandwidth there is more masking of the switch by the signal.

4. ACKNOWLEDGMENTS

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

References

[1] P. F. Hoffmann, H. Møller, Audibility of differences in head-related transfer functions, in Preparation. Manuscript II (2007).

[2] D. W. Grantham, B. W. Y. Hornsby, E. A. Erpenbeck, Auditory spatial resolution in horizontal, vertical, and diagonal planes, J. Acoust. Soc. Am. 114 (2) (2003) 1009–1022.

[3] B. P. Bovbjerg, F. Christensen, P. Minnaar, X. Chen, Measuring the head-related transfer functions of an artificial head with a high directional resolution, in: 109th Convention of the Audio Engineering Society, Los Angeles, California, USA, 2000, convention paper 5264.

[4] P. Minnaar, S. K. Olesen, F. Christensen, H. Møller, Localization with binaural recordings from artificial and human heads, J. Audio Eng. Soc. 49 (5) (2001) 323–336.

[5] P. F. Hoffmann, H. Møller, Some observations on sensitivity to HRTF magnitude, manuscript I. Submmited to J. Aud. Eng. Soc. (2007).

[6] T. I. Laakso, V. Välimäki, M. Karjalainen, U. K. Laine, Splitting the unit delay - tools for fractional delay filter design, IEEE Signal Processing Magazine 13 (1) (1996) 30–60.

[7] V. Välimäki, T. I. Laakso, Fractional delay filters — design and applications, in: F. A. Marvasti (Ed.), Nonuniform Sampling Theory and Practice, Kluwer Academic/Plenum Publishers, New York, NY, 2001, pp. 835–895.

[8] S. A. Gelfland, Hearing An Introduction to Psychological and Physiological Acoustic, 3rd Edition, Marcel Dekker, Inc, 1998.

[9] D. R. Stapells, T. W. Picton, A. D. Smith, Normal hearing thresholds for clicks, J. Acoust. Soc. Am. 72 (1) (1982) 74–79.

[10] D. W. Grantham, F. L. Wightman, Detectability of varying interaural temporal differences, J. Acoust. Soc. Am. 63 (2) (1978) 511–523.

[11] D. W. Grantham, Spatial hearing and related phenomena, in: B. C. J. Moore (Ed.), Hearing, 2nd Edition, Academic Press Inc., 1995, pp. 297–345.

[12] E. M. Wenzel, The impact of system latency on dynamic performance in virtual acoustic environments, J. Acoust. Soc. Am. 103 (5) (1998) 3026.

[13] H. Strauss, Implementing Doppler shifts for virtual auditory environments, in: 104th AES Convention, Amsterdam, The Netherlands, 1998, convention paper 4687.

[14] L. Savioja, J. Huopaniemi, T. Lokki, R. Vaananen, Creating interactive virtual acoustic environments, J. Audio Eng. Soc. 47 (9) (1999) 675–705.

[15] E. M. Wenzel, J. D. Miller, J. S. Abel, Sound lab: A real-time, software-based system for the study of spatial hearing, in: 108th Convention of the Audio Engineering Society, 2000.

[16] A. Silzle, P. Novo, H. Strauss, IKA–SIM: A system to generate auditory virtual environments, in: 108th Convention of the Audio Engineering Society, Berlin, Germany, 2004, convention paper 6016.

[17] D. S. Brungart, W. M. Rabinowitz, Auditory localization of nearby sources. Head-related transfer functions, J. Acoust. Soc. Am. 106 (3) (1999) 1465–1479.

[18] P. F. Hoffmann, H. Møller, Audibility of spectral switching in head-related transfer functions, in: 119th AES Convention, New York, USA, 2005, convention paper 6537.

[19] P. Minnaar, J. Plogsties, F. Christensen, Directional Resolution of Head-Related Transfer Functions Required in Binaural Synthesis, J. Audio Eng. Soc. 53 (10) (2005) 919–929.

[20] A. Kudo, H. Hokari, S. Shimada, A study on switching of the transfer functions focusing on sound quality, Acoust. Sci. & Tech. 26 (3) (2005) 267–278.

[21] L. Cohen, Time-Frequency Analysis, Prentice Hall, 1995.

[22] A. W. Mills, On the minimum audible angle, J. Acoust. Soc. Am. 30 (4) (1958) 237–246.

[23] R. L. McKinley, M. A. Ericson, Flight demonstration of a 3-d auditory display, in: R. H. Gilkey, T. R. Anderson (Eds.), Binaural and Spatial Hearing in Real and Virtual Environments, Lawrence Erlbaum Associates, 1997, pp. 683–699.

Chapter 5

Conclusions

5.1 Summary of findings

This Ph.D. thesis investigated the audibility of differences and switching between anechoic HRTFs. The HRTFs were implemented as minimum-phase filters together with frequency-independent delays as ITDs. In this way, independent control over the spectral and time characteristics of the HRTFs was achieved.

Discrimination of differences in HRTFs was measured for the spectral characteristics and time characteristics separately. Thresholds for the audibility of spectral differences ranged from about 3 to 17° depending on direction. For the position directly above the head, thresholds for changes along the vertical angle could not be estimated. The reason is that for this position the spectral characteristics of HRTFs vary particularly smoothly when changes are along elevation. In the same way, the threshold for spectral switching was the greatest as compared to those measured for other directions.

A simple model was proposed to explain audibility of spectral differences in HRTFs. This model was based on the rate at which spectral differences increase with increasing angular separation. The model could successfully account for thresholds measured for directions in the median plane, but only to a modest extent for lateral directions. It is possible that for these directions discrimination of spectral differences could be mediated by changes in any of the two ears.

Discrimination of time differences was not dependent on sound direction, and a large inter-subject variability was observed. Results showed that in average naive listeners needed above 80 μs to be able to discriminate changes in ITD. Although this sensitivity to ITD may be considered surprisingly low, the fact that naive listeners do not perform as well as trained or selected listeners on tasks requiring binaural processing has been observed in previous investigations. It is important to note that for the same listeners, sensitivity to spectral differences was substantially higher than sensitivity to time differences when compared in terms of corresponding angular shifts. Apparently, monaural spectral differences in HRTFs were much easier to understand and use as a discrimination cue than differences in ITD.

Thresholds on spectral switching, or minimum audible spectral switching (MASS), were comparable to thresholds for discrimination of spectral differences. We consider this similarity as evidence indicating that audibility of switching between HRTFs stems from differences and not from switching artifacts. From a practical point of view this is encouraging because no extra requirements are imposed by spectral switching on the required spatial resolution for spectral information. In other words, there is no need for a denser representation of space than the one required for spectral differences in adjacent HRTFs.

In the study of time switching we did not focus on the perception of dynamic varying ITDs, but on the audibility of artifacts created when switching between delays that were presented diotically. Minimum audible time switching (MATS) thresholds were observed to depend on the switching rate. MATSs were consistently lower for higher switching rate. MATS thresholds were slightly dependent on sound direction, and listeners were able to just perceive, in average, time switching of 5–10 μs. MATS thresholds were more than ten times smaller than thresholds obtained for the audibility of time differences. An important implication of these results is that the spatial resolution required for differences in ITDs, which could also include propagation delays, should be increased considerably if time switching free of artifacts is desired.

5.2 Future Work

From this study the question naturally arises, of, what is the sensitivity to differences and switching in HRTFs when both time and spectral characteristics work together. This is how changes are produced in real life, and thereby corresponds to a more ecologically valid situation than changing one characteristic at time. In addition, results from such experimental paradigm could be compared more directly to measures of auditory spatial resolution such as the minimum audible angle (MAA). Moreover, a persisting challenge in the generation of virtual spatial audio is to find HRTFs that can work for a large population. Thus, it would also be of interest to examine how discrimination varies across different set of HRTFs.

It is important to emphasize that all the experiments conducted in this study refer to the use of HRTFs in anechoic conditions and for single source presentations. Therefore, one should be cautious in generalizing audibility of differences or switching in HRTFs for more reverberant environments, or for situations in which more than one source is active. This can certainly be a topic for further investigation.

Because HRTFs change in complex ways it is difficult to identify the exact spectral features that may be used in the discrimination of spectral differences. For example, spectral cues may arise from changes in narrowband frequency regions, e.g. discrimination of notches. Other possibility is the integration of spectral differences over a wide frequency range. Certainly, spectral cues that can be derived from HRTFs are diverse. Thus, further investigation could be directed towards the identification of spectral features that are more critical for the discrimination of changes in the spectral characteristics of HRTFs.

A limitation in the experiment of time switching was the use of broadband noise. This is because the broader the bandwidth the more random is the nature of the sound, and thus the less probable is for the switching to be audible. Essentially, for signals with broader bandwidth there is more masking of the switch by the signal. Therefore, it would be interesting to measure audibility thresholds for different types of stimuli, and particularly narrowband stimuli because they offer less masking to the switch.

Appendix A

Interaural polar coordinate system

Interaural polar coordinate system

This appendix describes the coordinate system employed in this Ph.D. study. Both source position and directional changes were specified using the *interaural-polar coordinate system* (also referred to as horizontal-polar coordinate system or lateral-polar coordinate system). This system has been employed for measurements of HRTFs (Brown and Duda, 1998; Algazi *et al.*, 2001; Martens, 2002), and to explain certain types of errors in sound localization (Morimoto and Aokata, 1984; Macpherson and Middlebrooks, 2002, 2003). One of the major advantages of this system is that sound localization cues can be directly associated to the coordinates. Interaural difference cues are related to the lateral angle and monaural/spectral cues are related to the polar angle.

Figure A.1 shows a graphical description of this coordinate system. The lateral angle, or azimuth, is represented by θ_{IP} and the polar angle, or elevation, by ϕ_{IP}. For comparison Figure A.2 shows the more commonly used vertical-polar coordinate system (azimuth θ_{VP}, elevation ϕ_{VP}). The coordinate θ_{IP} describes the lateral displacement of the source from the median sagittal plane. The range of θ_{IP} from the rightmost to the leftmost direction is -90 to 90°. Thus, here, positive values are associated to displacements to the left (some authors used positive values for displacements to the right). Observe that a constant θ_{IP} defines a parallel plane to the median plane, and therefore, an approximation to a cone of confusion. The polar angle defines the angle of rotation about the interaural axis. Thus, on a given sagittal plane, -90°/270° indicates the lowest point, 0° the front horizon, 90° the highest point, and 180° the rear horizon.

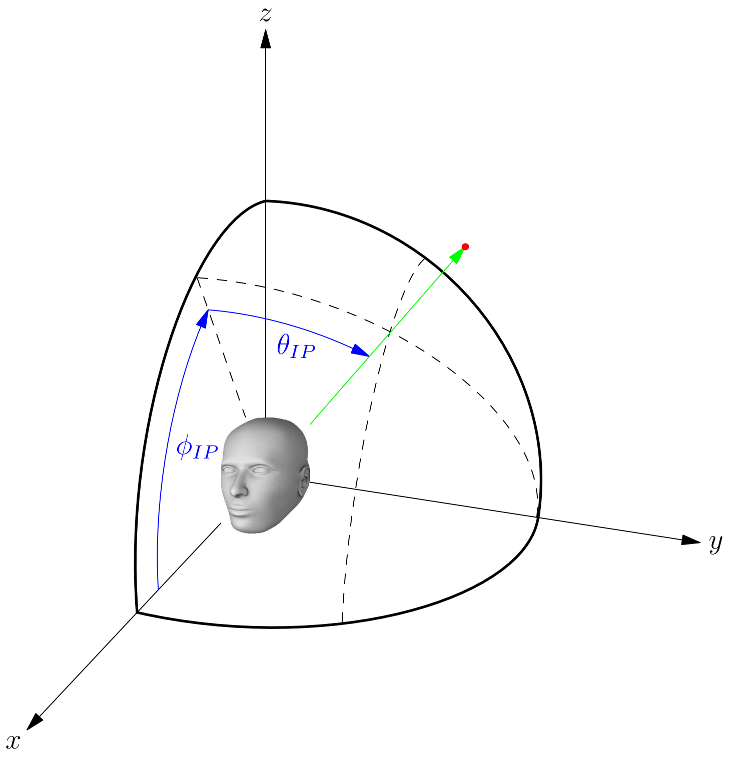


Figure A.1: Interaural-polar coordinate system. The red dot represents the source position.

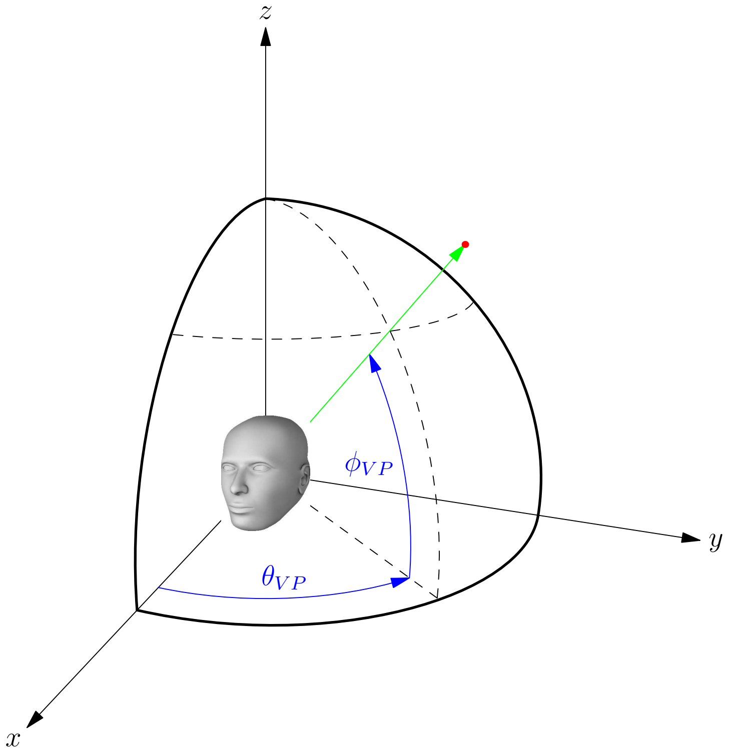


Figure A.2: Vertical-polar coordinate system.

MATLAB Code

```
[thetaIP phiIP] = vert2ip(theta,phi)

% function [thetaIP phiIP] = vert2ip(theta,phi)
%
% Transform coordinates from a vertical polar coordinate system
% to an interaural polar coordinate system. The coordinates theta and phi
% indicate azimuth and elevation respectively.
% The formulae have been modified from eq. (1) in:
% M. Morimoto and H. Aokata (1984), Localization cues of sound sources in the
% upper hemisphere, J. Acoust. So. Jpn. (E), Vol. 5, No. 3, pp. 165--173.
%
% Pablo F. Hoffmann, Aalborg University, March 2007

if(nargin ~= 2)
    error(’Number of input arguments must be two’);
end

thetaIP = asind(sind(theta)*cosd(phi)); % Transform azimuth

phiIP = asind(sind(phi)/sqrt(sind(phi)^2 ...
    +cosd(theta)^2*cosd(phi)^2)); % Transform elevation

if(theta > 90 && theta < 270) % adjust value for front-rear changes
    phiIP = 180-phiIP;
end
```

Bibliography

Algazi, V. R., Avendano, C., and Duda, R. O. (**2001**). Elevation localization and head-related transfer function analysis at low frequencies. *J. Acoust. Soc. Am.*, 109(3):1110–1122.

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

Best, V., van Schaik, A., and Carlile, S. (**2004**). Separation of concurrent broadband sound sources by human listeners. *J. Acoust. Soc. Am.*, 115(1):324–336.

Blauert, J. (**1997**). An Introduction to Binaural Technology. In Gilkey, R. H. and Anderson, T. R., editors, *Binaural and Spatial Hearing in Real and Virtual Environments*, pages 593–607. Lawrence Erlbaum Associates.

Blauert, J., Lehnert, H., Sahrhage, J., and Strauss, H. (**2000**). An Interactive Virtual–Environment Generator for Psychoacoustics Research. I: Architecture and Implementation. *Acustica*, 86:94–102.

Bovbjerg, B. P., Christensen, F., Minnaar, P., and Chen, X. (**2000**). Measuring the Head-Related Transfer Functions of an artificial head with a high directional resolution. In *109th Convention of the Audio Engineering Society*, Los Angeles, California, USA. convention paper 5264.

Bronkhorst, A. W. (**1993**). Horizontal and vertical MAAs for a wide range of sound source locations (A). *J. Acoust. Soc. Am.*, 93(4):2351.

Bronkhorst, A. W. (**1995**). Localization of real and virtual sources. *J. Acoust. Soc. Am.*, 98(5):2542–2553.

Brown, C. P. and Duda, R. O. (**1998**). A Structural Model for Binaural Sound Synthesis. *IEEE Trans. on Speech and Audio Processing*, 6(5):476–488.

Brungart, D. S., Kordik, A. J., and Simpson, B. D. (**2006**). Effects of Headtracker Latency in Virtual Audio Displays. *J. Audio Eng. Soc.*, 54(1/2):32–44.

Brungart, D. S. and Rabinowitz, W. M. (**1999**). Auditory localization of nearby sources. Head-related transfer functions. *J. Acoust. Soc. Am.*, 106(3):1465–1479.

Chandler, D. W. and Grantham, D. W. (**1992**). Minimum audible movement angle in the horizontal plane as a function of stimulus-frequency bandwidth, source azimuth, and velocity. *J. Acoust. Soc. Am.*, 91(3):1624–1636.

Chen, J., Veen, B. D. V., and Hecox, K. E. (**1995**). A spatial feature extraction and regularization model for the head-related transfer function. *J. Acoust. Soc. Am.*, 97(1):439–452.

Christensen, F., Jensen, C. B., and Møller, H. (**2000**). The Design of VALDEMAR–An Artificial Head for Binaural Recording Purposes. In *109th Convention of the Audio Engineering Society*, Los Angeles, California, USA. convention paper 5253.

Clark, R., Ifeachor, E., and Rogers, G. (**2002**). Filter Morphing - Topologies, signals and sampling rates. In *113th Convention of the Audio Engineering Society*, Los Angeles, California, USA. convention paper 5661.

Cohen, L. (**1995**). *Time-Frequency Analysis.* Prentice Hall.

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

Evans, M. J., Angus, J. A. S., and Tew, A. I. (**1998**). Analyzing head-related transfer function measurements using surface spherical harmonics. *J. Acoust. Soc. Am.*, 104(4):2400–2411.

Freeland, F. P., Biscainho, L. W. P., and Diniz, P. S. R. (**2004**). Interpositional Transfer Function for 3D–Sound Generation. *J. Audio Eng. Soc.*, 52(9):915–930.

Gardner, W. G. and Martin, K. D. (**1995**). HRTF Measurements of a KEMAR. *J. Acoust. Soc. Am.*, 97(6):3907–3908.

Grantham, D. W. (**1985**). Auditory spatial resolution under static and dynamic conditions. *J. Acoust. Soc. Am.*, 77(S1):S50.

Grantham, D. W. (**1986**). Detection and discrimination of simulated motion of auditory targets in the horizontal plane. *J. Acoust. Soc. Am.*, 79(6):1939–1949.

Grantham, D. W. (**1995**). Spatial Hearing and Related Phenomena. In Moore, B. C. J., editor, *Hearing*, pages 297–345. Academic Press Inc., 2nd edition.

Grantham, D. W. (**1997**). Auditory Motion Perception: Snapshots Revisited. In Gilkey, R. H. and Anderson, T. R., editors, *Binaural and Spatial Hearing in Real and Virtual Environments*, pages 295–313. Lawrence Erlbaum Associates.

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

Green, D. M. (**1988**). *Profile Analysis: Auditory Intensity Discrimination.* Oxford University Press, New York, NY, USA.

Hammershøi, D. and Møller, H. (**1996**). Sound transmission to and within the human ear canal. *J. Acoust. Soc. Am.*, 100(1):408–427.

Hammershøi, D. and Møller, H. (**2005**). Binaural Technique, Basic Methods for Recording, Synthesis and Reproduction. In Blauert, J., editor, *Communication Acoustics*, pages 223–254. Springer Verlag, Berlin, Germany.

Harris, J. D. and Sergeant, R. L. (**1971**). Monaural-binaural minimum audible angles for a moving sound source. *J. Speech Hear Res.*, 14(3):618–629.

Hartmann, W. M. and Rakerd, B. (**1989**). On the minimum audible angle — A decision theory approach. *J. Acoust. Soc. Am.*, 85(5):2031–2041.

Hartmann, W. M. and Wittenberg, A. (**1996**). On the externalization of sound images. *J. Acoust. Soc. Am.*, 99(6):3678–3687.

Hartung, K. and Braasch, J. (**1999**). Comparison of different methods for the interpolation of Head-related transfer functions. In *Proc. AES 16th Int. Conf.*, Rovaniemi, Finland.

Hebrank, J. and Wright, D. (**1974**). Spectral cues used in the localization of sound sources on the median plane. *J. Acoust. Soc. Am.*, 56(6):1829–1834.

Huopaniemi, J., Zacharov, N., and Karjalainen, M. (**1999**). Objective and Subjective Evaluation of Head-Related Transfer Function Filter Design. *J. Acoust. Soc. Am.*, 47(4):218–239.

Keyrouz, F. and Diepold, K. (**2006**). Efficient State-Space Interpolation of Head-Related Transfer Functions. In *Proc. AES 28th Int. Conference*, Piteå, Sweden.

Kim, Y. and Kim, J. (**2005**). New HRTFs (Head Related Transfer Functions) for 3D audio applications. In *118th Convention of the Audio Engineering Society*, Barcelona, Spain. convention paper 6495.

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

Kudo, A., Hokari, H., and Shimada, S. (**2005**). A study on switching of the transfer functions focusing on sound quality. *Acoust. Sci. & Tech.*, 26(3):267–278.

Kulkarni, A. and Colburn, H. S. (**1995**). Efficient finite-impulse-response filter models of the head-related transfer function. *J. Acoust. Soc. Am.*, 97(5):3278.

Kulkarni, A. and Colburn, H. S. (**2004**). Infinite-impulse-response models of the head-related transfer function. *J. Acoust. Soc. Am.*, 115(4):1714–1728.

Kulkarni, A., Isabelle, S. K., and Colburn, H. S. (**1995**). On the minimum-phase approximation of head-related transfer functions. In *Proc. of the ASSP (IEEE) Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 84–87, New Paltz, NY, USA.

Langendijk, E. H. A. and Bronkhorst, A. W. (**2000**). Fidelity of three-dimensional-sound reproduction using a virtual auditory display. *J. Acoust. Soc. Am.*, 107(1):528–537.

Langendijk, E. H. A. and Bronkhorst, A. W. (**2002**). Contribution of spectral cues to human sound localization. *J. Acoust. Soc. Am.*, 112(4):1583–1596.

Macpherson, E. A. and Middlebrooks, J. C. (**2002**). Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited. *J. Acoust. Soc. Am.*, 111(5):2219–2236.

Macpherson, E. A. and Middlebrooks, J. C. (**2003**). Vertical-plane sound localization probed with ripple-spectrum noise. *J. Acoust. Soc. Am.*, 114(1):430–445.

Martens, W. (**2002**). Rapid Psychophysical calibration using bisection scaling for individualized control of source elevation in auditory display. In *Proc. Int. Conf. on Auditory Display (ICAD 2002)*, Kyoto, Japan.

McKinley, R. L., Ericson, M. A., Perrott, D. R., Gilkey, R. H., Brungart, D. S., and Wightman, F. L. (**1992**). Minimum audible angles for synthesized location cues presented over headphones (A). *J. Acoust. Soc. Am.*, 92(4):2297.

Middlebrooks, J. C. and Green, D. M. (**1991**). Sound Localization by human listeners. *Annu. Rev. Psychol.*, 42(1):135–159.

Middlebrooks, J. C., Makous, J. C., and Green, D. M. (**1989**). Directional sensitivity of sound-pressure levels in the human ear canal. *J. Acoust. Soc. Am.*, 86(1):89–108.

Miller, J. D. and Wenzel, E. M. (**2002**). Recent Developments in SLAB: A software-based system for interactive spatial sound synthesis. In *Proc. Int. Conf. on Auditory Display (ICAD 2002)*, Kyoto, Japan.

Mills, A. W. (**1958**). On the Minimum Audible Angle. *J. Acoust. Soc. Am.*, 30(4):237–246.

Minnaar, P., Olesen, S. K., Christensen, F., and Møller, H. (**2001**). Localization with Binaural Recordings from Artificial and Human Heads. *J. Audio Eng. Soc.*, 49(5):323–336.

Minnaar, P., Plogsties, J., and Christensen, F. (**2005**). Directional Resolution of Head-Related Transfer Functions Required in Binaural Synthesis. *J. Audio Eng. Soc.*, 53(10):919–929.

Møller, H. (**1992**). Fundamentals of Binaural Technology. *Applied Acoustics*, 36:171–128.

Møller, H., Hammershøi, D., Jensen, C. B., and Sørensen, M. F. (**1995a**). Transfer Characteristics of Headphones Measured on Human Ears. *J. Audio Eng. Soc.*, 43(4):203–217.

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

Møller, H., Sørensen, M. F., Jensen, C. B., and Hammershøi, D. (**1995c**). Binaural Technique: Do We Need Individual Recordings? *J. Audio Eng. Soc.*, 44(6):451–469.

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

Mourjopoulos, J. N., Kyriakis-Bitzaros, E. D., and Goutis, C. E. (**1990**). Theory and Real-Time Implementation of Time-Varying Digital Filters. *J. Audio Eng. Soc.*, 38(7/8):523–536.

Musicant, A. D. and Butler, R. A. (**1984**). The influence of pinnae-based spectral cues on sound localization. *J. Acoust. Soc. Am.*, 75(4):1195–1200.

Novo, P. (**2005**). Auditory Vitual Environments. In Blauert, J., editor, *Communication Acoustics*, pages 277–297. Springer Verlag.

Pedersen, J. A. and Minnaar, P. (**2006**). Evaluation of a 3D-audio system with head tracking. In *Proc. of the 120th Convention of Audio Engineering Society*, Paris, France. convention paper 6654.

Pellegrini, R. S. (**2001**). Quality assessment of auditory virtual environments. In *Proc. Int. Conf. on Auditory Display (ICAD 2001)*, pages 161–168, Espoo, Finland.

Perrett, S. and Noble, W. (**1997**). The contribution of head motion cues to localization of low-pass noise. *Perception & Psychophysics*, 59(7):1018–1026.

Perrott, D. R. (**1984**). Concurrent minimum audible angle: A re-examination of the concept of auditory spatial acuity. *J. Acoust. Soc. Am.*, 75(4):1201–1206.

Perrott, D. R. and Marlborough, K. (**1989**). Minimum audible movement angle: Marking the end points of the path traveled by a moving sound source. *J. Acoust. Soc. Am.*, 85(4):1773–1775.

Perrott, D. R. and Musicant, A. D. (**1977**). Minimum auditory movement angle: Binaural localization of moving sound sources. *J. Acoust. Soc. Am.*, 62(6):1463–1466.

Perrott, D. R. and Pacheco, S. (**1989**). 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(6):2669–2672.

Perrott, D. R. and Saberi, K. (**1990**). Minimum audible angle thresholds for sources varying in elevation and azimuth. *J. Acoust. Soc. Am.*, 87(4):1728–1731.

Perrott, D. R. and Tucker, J. (**1988**). Minimum audible movement angle as a function of signal frequency and the velocity of the source. *J. Acoust. Soc. Am.*, 83(4):1522–1527.

Plogsties, J., Minnaar, P., Olesen, S. K., Christensen, F., and Møller, H. (**2000**). Audibility of All-pass Components in Head-Related Transfer Functions. In *108th Convention of the Audio Engineering Society*, Paris, France. convention paper 5132.

Pralong, D. and Carlile, S. (**1996**). The role of individualized headphone calibration for the generation of high fidelity virtual auditory space. *J. Acoust. Soc. Am.*, 100(6):3785–3793.

Saberi, K., Dostal, L., Sadralodabai, T., and Perrott, D. R. (**1991**). Minimum Audible Angles for Horizontal, Vertical, and Oblique Orientations: Lateral and Dorsal Planes. *Acustica*, 75(1):57–61.

Saberi, K. and Perrott, D. R. (**1990**). Minimum audible movement angles as a function of sound source trajectory. *J. Acoust. Soc. Am.*, 88(6):2639–2644.

Sandvad, J. (**1996**). Dynamic aspects of auditory virtual environments. In *100th Convention of the Audio Engineering Society*, Copenhagen, Denmark. convention paper 4226.

Sandvad, J. and Hammershøi, D. (**1994**). What is the Most Efficient Way of Representing HTF filters? In *Proceedings of Nordic Signal Processing Symposium*, pages 174–178, NORSIG '94, Lesund, Norway.

Savioja, L., Huopaniemi, J., Lokki, T., and Vaananen, R. (**1999**). Creating Interactive Virtual Acoustic Environments. *J. Audio Eng. Soc.*, 47(9):675–705.

Silzle, A. (**2002**). Selection and Tuning of HRTFs. In *112th Convention of the Audio Engineering Society*, Munich, Germany. convention paper 5595.

Silzle, A., Novo, P., and Strauss, H. (**2004**). IKA–SIM: A System to Generate Auditory Virtual Environments. In *108th Convention of the Audio Engineering Society*, Berlin, Germany. convention paper 6016.

Strybel, T. Z. and Fujimoto, K. (**2000**). Minimum audible angles in the horizontal and vertical planes: Effects of stimulus onset asynchrony and burst duration. *J. Acoust. Soc. Am.*, 108(6):3092–3096.

Välimäki, V. and Laakso, T. I. (**1998**). Suppression of transients in variable recursive digital filters with a novel and efficient cancellation method. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 46(12):3408–3414.

Välimäki, V. and Laakso, T. I. (**2001**). Fractional Delay Filters — Design and Applications. In Marvasti, F. A., editor, *Nonuniform Sampling Theory and Practice*, pages 835–895. Kluwer Academic/Plenum Publishers, New York, NY.

Wenzel, E. M. (**1999**). Effects of increasing system latency on localization of virtual sounds. In *Proc. AES 16th Int. Conf.*, pages 42–50.

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

Wenzel, E. M. and Foster, S. H. (**1993**). Perceptual consequences of interpolating head-related transfer functions during spatial synthesis. In *Proc. of the ASSP (IEEE) Workshop on Applications of Signal Processing to Audio & Acoustics*, pages 17–20, New Paltz, NY, USA.

Wightman, F. L. and Kistler, D. J. (**1989**). Headphone simulation of free-field listening. II: Psychophysical validation. *J. Acoust. Soc. Am.*, 85(2):868–878.

Wightman, F. L. and Kistler, D. J. (**1999**). Resolution of front-back ambiguity in spatial hearing by listener and source movement. *J. Acoust. Soc. Am.*, 105(5):2841–2853.

Wightman, F. L. and Kistler, D. J. (**2005**). Measurement and Validation of Human HRTFs for Use in Hearing Research. *Acta Acustica united with Acustica*, 91:429–439.

Wightman, F. L., Kistler, D. J., and Perkins, M. E. (**1987**). A New Approach to the Study of Human Sound Localization. In Yost, W. A. and Gourevitch, G., editors, *Directional Hearing*, pages 26–48. Springer.

Wright, D., Hebrank, J. H., and Wilson, B. (**1974**). Pinna reflections as cues for localization. *J. Acoust. Soc. Am.*, 56(3):957–962.

Zahorik, P., Wightman, F. L., and Kistler, D. J. (**1995**). On the discriminability of virtual and real sound sources. In *Proc. of the ASSP (IEEE) Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 76–79, New Paltz, NY, USA.

Zoelzer, U., Redmer, B., and Bucholtz, J. (**1993**). Strategies for Switching Digital Audio Filters. In *95th Convention of the Audio Engineering Society*, New York, USA. convention paper 3714.

Zotkin, D. N., Duraiswami, R., Grassi, E., and Gumerov, N. A. (**2006**). Fast head-related transfer function measurement via reciprocity. *J. Acoust. Soc. Am.*, 120(4):2202–2215.