Audibility of direct switching between head-related transfer functions

Hoffmann, Pablo F.; Møller, Henrik

Published in:
Acustica United with Acta Acustica

DOI (link to publication from Publisher):
10.3813/AAA.918112

Publication date:
2008

ACTA ACUSTICA UNITED WITH ACUSTICA
Vol. 94 (2008) 955 – 964

DOI 10.3813/AAA.918112

Audibility of Direct Switching Between Head-Related Transfer Functions

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

Summary
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 generates 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 are just audible. These thresholds were measured for temporal and spectral characteristics of HRTFs separately, and were defined as the minimum audible time switching (MATS), and the minimum audible spectral switching (MASS) respectively. MATS thresholds were in the range of 5–9.4 μs, and MASS thresholds were in the range of 4.1–48.2° being more dependent on the direction of sound than MATSs. Generally, results show that for dynamic binaural synthesis time switching imposes higher demands on spatial resolution than those imposed by spectral switching.

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

1. 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. That is, HRTFs can only be measured for a finite number of directions. An intuitive criterion to evaluate whether a set of HRTFs provides a 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 parallel study [1] the audibility of differences in HRTFs was measured for changes in interaural time difference (ITD) and changes in magnitude spectrum separately. Using a discrimination paradigm, it was found that sensitivity to ITD was poorer than sensitivity to spectral differences, and that spectral differences require resolutions of 2.4–11° depending on direction.

It has consistently been shown that human sensitivity to changes in sound direction is higher for stationary conditions than for dynamic conditions (for a review on stationary and dynamic spatial resolution see [2]). This suggests that information from auditory spatial resolution in stationary conditions may be sufficient to evaluate the required spatial resolution for dynamic binaural synthesis. However, this approach only evaluates our ability to detect differences in HRTFs. It does not consider that due to the discrete nature of the available spatial representation, switching between HRTFs produces discontinuities that can degrade the quality of the perceived sound [3, 4]. These discontinuities, if audible, are commonly heard as “clicks", and their audibility is thought to be proportional to the spatial separation between switched HRTFs. This study reports on experiments conducted to measure the just-audible switch for a direct switching between HRTF filters in the time domain.

A common technique used to mitigate the problem of audible discontinuities is crossfading. This technique is illustrated in Figure 1 and is mathematically expressed by

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

where $x(n)$ is the input signal, h_i and h_j are the current and target filters, $\alpha(n)$ is the crossfading function, and $y(n)$ the output signal. Note that h_i and h_j represent head-related impulse responses. Here, $x(n)$ is convolved with h_i and h_j and the outputs from both convolutions are weighted by $\alpha(n)$ and summed up to yield $y(n)$. The crossfading is controlled by $\alpha(n)$ that gradually and monotonically changes from unity to zero within a given interval, thereby gradually changing from the current filter to the target filter.

Observe that crossfading requires at least two convolutions to run in parallel within the crossfading interval. Also

Received 29 February 2008,
accepted 10 October 2008.

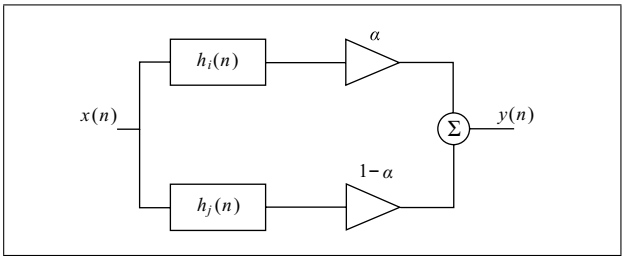

Figure 1. Schematic of crossfading strategy.

note that $x(n)$ is common to both convolutions, and using the distributive property of convolution, we can arrange equation (1) to

$$y(n) = x(n) * \left[h_i(n) \cdot \alpha(n) + h_j(n) \cdot \left(1 - \alpha(n)\right)\right],$$
$$y(n) = x(n) * h_{ij}(n), \qquad (2)$$

which reduces the number of convolutions to only one. However, the fact that in equation (2) $\alpha(n)$ needs to multiply all filter's coefficients instead of only the filter's outputs does not produce a real reduction in the number of operations. What we want to emphasize in the change from equation (1) to (2) is that for any new input sample, crossfading the filters' outputs is equivalent to first interpolate between the filters, and then switch to the new interpolated filter h_{ij}. In other words, from one output sample to the next one crossfading is equivalent to switching HRTFs that are intermediate between the current and target HRTFs. How close the intermediate HRTFs must be will depend upon the size of the switching step; it must be small enough so that switching artifacts are below the audible threshold, and large enough to avoid unnecessary switching. This suggests that knowledge about the just-audible switch can help in selecting better values for the parameters involved in dynamic binaural synthesis, e.g. update rate and spatial resolution.

In the present study, two experiments are conducted to measure the largest angular separation (lowest spatial resolution) for which switching between HRTFs does not produce audible artifacts. Experiment I measures audibility thresholds for dynamically changing delays and this threshold is defined as the *minimum audible time switch* (MATS). Experiment II measures the *minimum audible spectral switch* (MASS) defined as the threshold for direct switching between minimum-phase HRTFs. In both experiments, audibility of HRTF switching is estimated for a number of directions.

2. Experiment I: Time switching in HRTFs

2.1. Method

2.1.1. Stimuli and apparatus

Time switching was compared on thirteen selected directions distributed over the upper half of the sphere. These directions are referred to as nominal directions and they are summarized in Table I. Nominal directions are described using an interaural-polar coordinate system in the same way as in [1]. A sketch is shown in Figure 2. In this system, coordinates are given as (lateral angle, polar angle), where the lateral angle indicates the angle of incidence with respect to the median plane, and the polar angle indicates the angle around the interaural axis. From right to left the lateral angle goes from -90° to 90°; the polar angle goes from 0°, corresponding to the frontal part of the horizontal plane, to 180°, corresponding to the rear of the horizontal plane. Positive and negative values indicate positions above and below the horizontal plane respectively. The HRTFs used to render directional sound are from a dataset previously measured on an artificial head at a distance of 2 m and with a directional resolution of 2° [5]. This artificial head was designed so that HRTFs were representative of a typical human subject [6]. Its performance for recording and reproduction of binaural material has been shown to be comparable to other well-known artificial heads [7]. HRTFs were represented as 72-coefficient minimum-phase finite-impulse-response (FIR) filters with ITDs approximated as frequency-independent delays. ITDs were determined from the interaural group-delay differences of the excess-phase components of the HRTFs evaluated at 0 Hz [8]. Negative ITDs indicate that left ear is leading.

Table I. Nominal directions and ITD values of the HRTFs used in the listening experiment.

ITD [μs]				
-625.0	-437.5	0	437.5	625.0
Nominal direction (lateral angle, polar angle)				
(90°,0°)	(58°,0°) (46°, 90°) (54°,180°)	(0°,0°) (0°,44°) (0°,90°) (0°,136°) (0°,180°)	(-56°,0°) (-46°, 90°) (-54°,180°)	(−90°,0°)

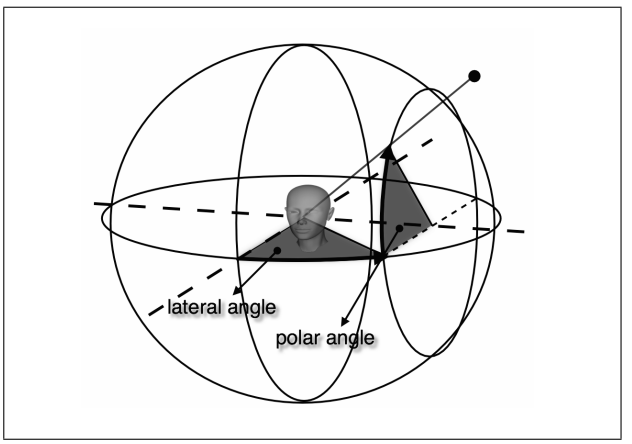

Figure 2. Interaural-polar coordinate system. The black dot represents an arbitrary position.

Broadband pink noise (20–9000 Hz) was used as source signal. Directional sound was synthesized by digitally filtering the pink noise with the HRTFs. HRTF-filtered stimuli were played back over equalized Beyerdynamic DT-990 Pro circumaural headphones. For the equalization of headphones 256-coefficient minimum-phase FIR filters were used. The procedure used to compute the equalization filters is described in a related work [9]. HRTF filtering and headphone equalization were done off-line and thirteen 5-s stimuli (one for each of the nominal directions) were constructed and 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.

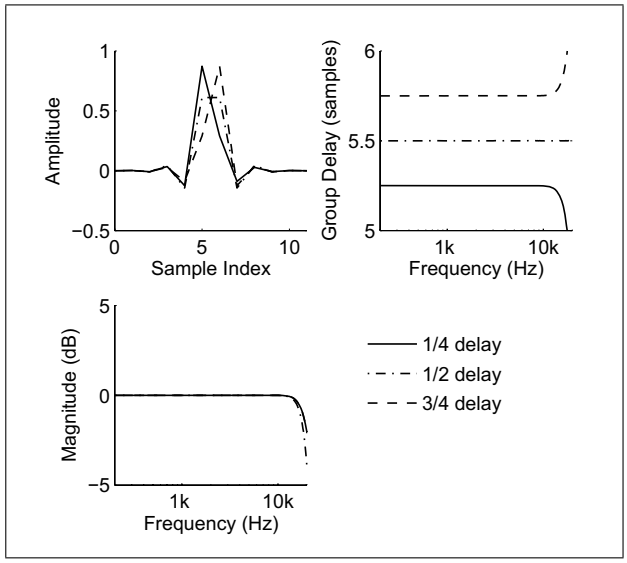

Figure 3. Example of FIR fractional delay filters for delays of 1/4, 1/2 and 3/4 samples. Filters are of order 11th and were designed using Lagrange interpolation. The upper-left panel shows the impulse responses. The upper-right panel shows the group delay, where we observed that the filters' inherent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

2.1.2. Time switching implementation

Time switching was implemented as a periodic delay shift applied to both left and right HRTF-filtered signals. Dynamic delay changes were produced by alternating continuously between a zero-delay state and a delayed state. Thus, there were two state transitions, one corresponding to switching from the zero-delay state to the delayed state and the other corresponding to switching back from the delayed state to the zero-delay state. Subjects controlled the amount of delay for the delayed state. 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. The switching operation was always completed on a sample-to-sample basis, and thus for a sufficiently large delay shift a clear click was perceived. The values for switching rates were selected to be within the range of those suggested and implemented for update of the processing in virtual spatial sound systems (20 Hz [10], 60 Hz [11], and 690 Hz [12]).

Pilot experiments have shown 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 [13]). 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 in the output signal whereas FIR filters do not [14].

Coefficients of all fractional delay filters were precalculated off-line and a table-lookup method was used to switch between them. Filter coefficients were computed using Lagrange interpolation [13]. The simplicity of this design technique is that the coefficients are easily obtained using a closed analytical form given by

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

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). Figure 3 shows examples of the filters implemented for 1/4, 1/2 and 3/4 sample delays. Note that the filters have an inherent integer delay corresponding to ceil$((N-1)/2)$.

A schematic of the system implemented to control time switching is shown in Figure 4. During switching from a zero-delay state to a delayed state, the amount of delay to be applied was read from the subject's response and the appropriate fractional delay filter was retrieved from memory and convolved with the stereo signal. If delays larger than one sample were required, the additional integer delay D was introduced prior to the fractional delay filtering. After 10 ms or 20 ms, depending on the operating switching rate (100 Hz or 50 Hz), the delayed state was switched back to the zero-delay state and the same operation was repeated.

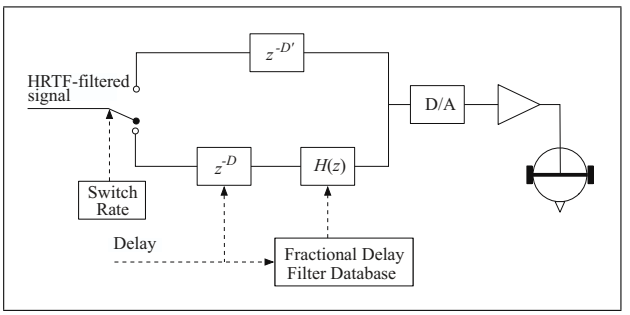

Figure 4. Diagram of the dynamic time switching implementation. The input signal is the pink noise already filtered with an HRTF and the headphone equalization filters. Delay is constantly retrieved in order to select the appropriate fractional delay filter and integer delay. The fixed delay D' compensates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

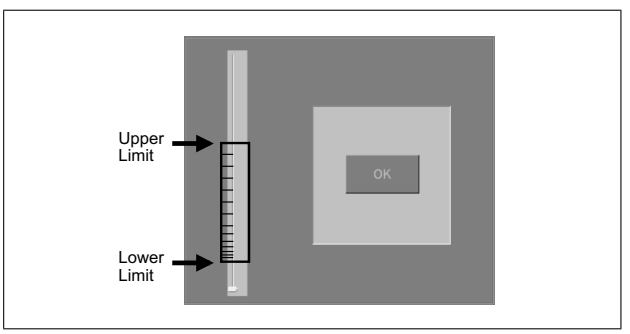

Figure 5. Graphic interface presented to the subjects during the experiment. The graded box (not seen by the subjects) represents the scale of delays whose position along the slider-bar was randomized across trials.

Note that the integer delay D' inherent to the fractional delay filters was compensated for during the zero-delay state.

2.1.3. 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 Hz and 4 kHz, and a hearing level $\leq$ 15 dB HL at 8 kHz.

2.1.4. 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. The minimum applicable delay was about 2.1 μs, and the maximum was 4.2 ms. Delays were incremented logarithmically in 20 steps per decade, yielding a scale of 67 different delays.

For estimating MATS thresholds the method of adjustment was employed [15]. 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. Figure 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 at the bottom or at the top of the bar. This ensured that the slider position was at a clear distance from threshold at the beginning of each trial.

2.1.5. Experimental design

This experimental design consists of a two-factor design with factors corresponding to the switching rate and nominal direction. Stimuli were grouped in blocks of thirteen trials so that each direction was presented once and in random order within an experimental block. Switching rates were arranged such that either time switching operated at 50 Hz for seven directions and at 100 Hz for the remaining six directions, or vice versa. This distribution was balanced across the experiment. Prior to the main experiment all subjects completed three blocks for familiarization and practice. For the main experiment subjects participated in 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 $\times$ two switching rates) were repeated five times for each listener.

2.2. 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. Figures 6a and 6b show individual thresholds for each switching rate respectively. Nominal directions are expressed in polar angle and grouped by ITD. Individual thresholds are fairly consistent across directions. The responses of two subjects who represent extreme data are

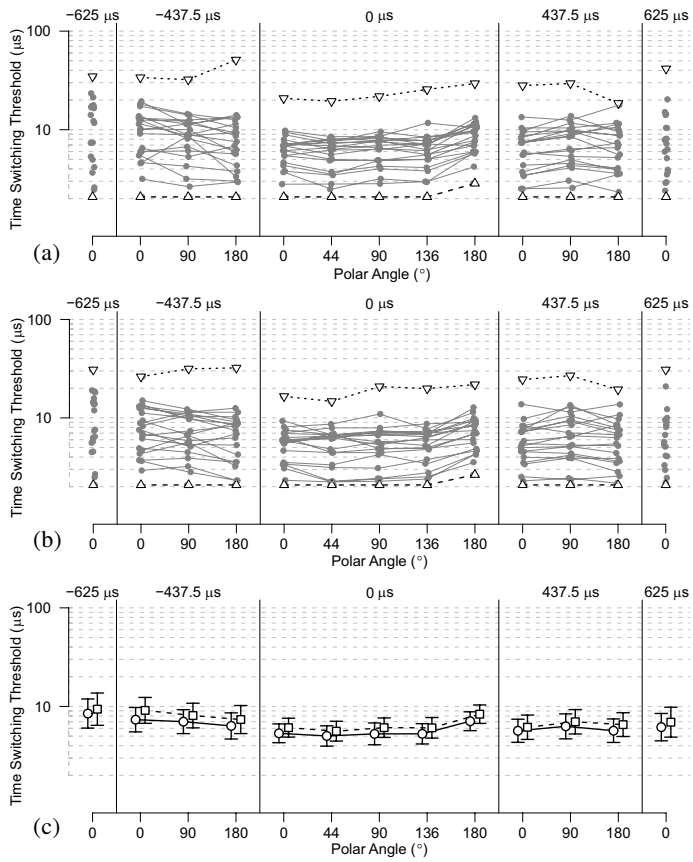

Figure 6. MATS thresholds. Individual thresholds are shown in panel (a) for 50-Hz update rate and panel (b) for 100-Hz update rate. Lines connect direction with a common ITD. The dotted- and dashed-line represent extreme results (see text). Mean MATSs across subject are shown in panel (c) for 50-Hz (squares)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

plotted with specific signatures. The most sensitive subject was capable of perceiving 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.

Figure 6c shows mean thresholds calculated across subjects for each switching rate, and they are summarized in Table II. The range of mean thresholds was 5.6–9.4 μs for 50-Hz switching rate, and 5.0–8.5 μs for 100-Hz switching rate. Smallest mean thresholds were found at (0°,44°) and largest mean thresholds at (90°,0°) for both switching rates. 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 nominal 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 ($p = 0.61$).

Table II. Mean MATS for the tested directions and switching rates of 50 Hz and 100 Hz. Thresholds are given in (μs).

Nominal direction	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

2.3. 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 that the clicks are audible is higher. Thresholds for the audibility of clicks have been reported to decrease as

the click-presentation rate increases [16]. Our results are in agreement with this notion because subjects were significantly more sensitive to artifacts produced at a higher switching rate.

Even though MATSs were obtained for delays applied to both ears simultaneously, it seems worth comparing these thresholds with sensitivity to dynamic changes in ITD. This comparison is justified by considering that ITDs can be implemented as a delay increment in one ear and a delay decrement in the other ear. The absolute delay being equal to half the ITD. In a study by Grantham and Wightman [17] 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 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* [18]. 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 assume that these values are changes in ITD, and considering that head movements during sound localization can easily reach velocities above 90°/s [19], corresponding to more than 630 μs/s, then, at these switching rates artifacts will be audible. In addition, even faster velocities are required in the use of propagation delays for incorporating Doppler effects [20]. 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 delays are commonly updated at every sample [12, 21, 22], and this applies to both propagation delays and ITDs. For example, the DIVA system [21] 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° lateral angle direction, the second should return a lateral angle 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.

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 dynamically 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 [23].

3. 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.

3.1. Method

3.1.1. 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.

3.1.2. Stimuli and playback system

Broadband pink noise (20–16000 Hz) was used as the source signal. The same thirteen nominal 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.

3.1.3. 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.

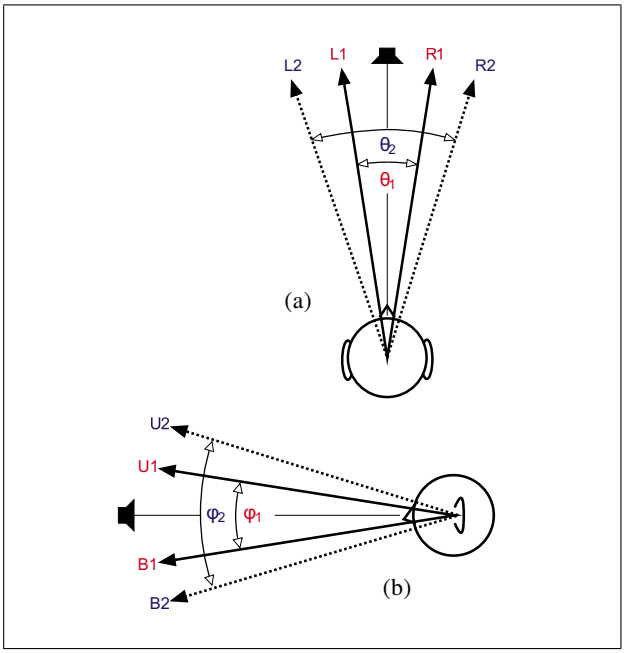

Figure 7. Description of spectral switching for direction (0°,0°). In (a) switching is in lateral angle. For an angular separation of θ_1 HRTFs are switched every 10 ms between positions L1 and R1 respectively. If the angular separation is θ_2, HRTFs for positions L2 and R2 are switched. In (b) switching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

For each nominal direction adjacent HRTFs were switched in two modes, and thus two sets of filters were computed. One mode corresponded to switching in lateral angle, and the other mode corresponded to switching in polar angle. For example let us assume that a sound is presented from (0°,0°) and switching is in lateral angle. This particular scenario is depicted in Figure 7a. 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 θ_2 is selected the switching would take place between locations L2 and R2. For switching in polar angle, 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 Figure 7b.

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

Angular separations ranged from 0.5° to 60° and they were incremented using linear steps of 0.5°. The resolution of the measured HRTFs was increased using linear interpolation between the minimum-phase impulse responses. A total of 26 sets of adjacent HRTFs were constructed (13 nominal directions × 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 [24].

3.1.4. 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.

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, followed by two or more blocks of stimuli for practice. One block consisted of thirteen trials. All nominal directions were presented in one block, and the switching modes, either seven times lateral angle and six times polar angle 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.

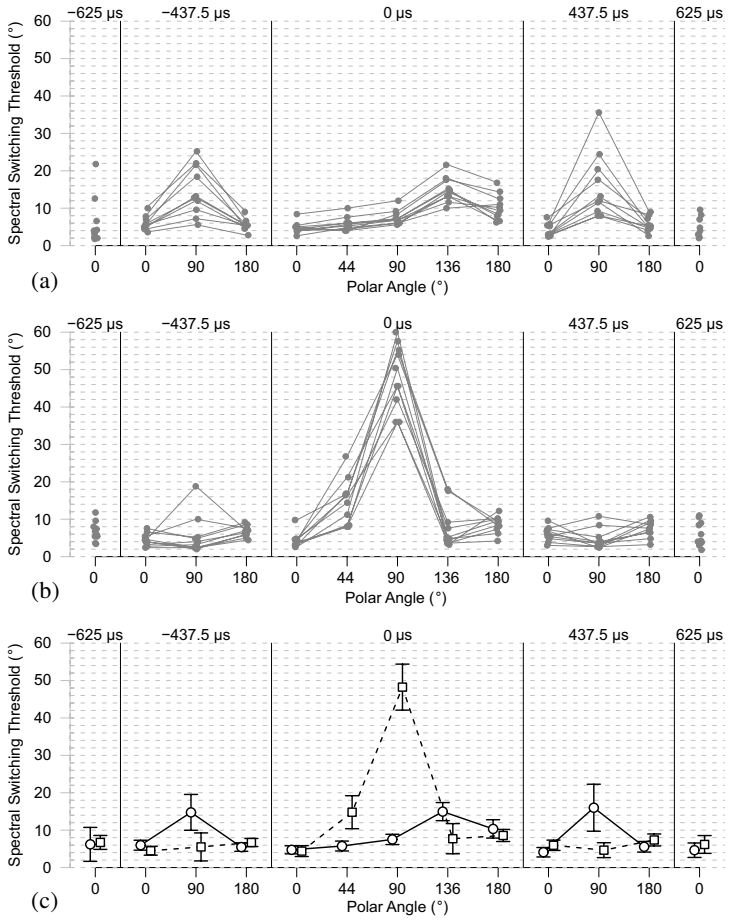

Figure 8. MASS thresholds. Individual thresholds are shown in panel (a) for switching in lateral angle, and in panel (b) for switching in polar angle. Lines connect responses for directions with a common ITD. Mean MASSs across subjects are shown in panel (c) where switching in lateral angle and polar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

3.2. 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 polar angle. 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).

Figures 8a and 8b show individual MASS thresholds for switching in lateral angle and polar angle respectively. For directions in the median plane and switching in lateral angle, thresholds tended to increase with polar angle from 0° to 136°, 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 polar angle but with largest threshold for (0°,90°), which has as consequence a faster increase/decrease in threshold as compared with switching in lateral angle. Thresholds for both left and right sagittal planes also showed certain dependency with polar angle, and this was more pronounced for switching in lateral angle than in polar angle.

Table III. Mean MASS thresholds given in (°) for lateral angle and polar angle switching mode.

Nominal direction	Lateral angle	Polar angle
(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

Mean MASS thresholds across subject are shown in Figure 8c and summarized in Table III. Mean MASSs ranged from 4.1 to 16° for switching in lateral angle, and from 4.4 to 48.2° for switching in polar angle. 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 sagittal planes were higher for switching in lateral angle than in polar angle, 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 polar angles different from 0° and 180°.

3.3. Discussion

MASS thresholds are comparable to thresholds obtained from discrimination of spectral differences [1, 9]. 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 in the left and right sagittal planes show a tendency to increase as a function of polar angle. This tendency is in agreement with a study conducted by Minnaar *et al.* [25], 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 depends on the direction of switching. In the median plane thresholds increase for switching in polar angle and in the lateral sagittal planes thresholds increase for switching in lateral angle.

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, 9]. In practical terms, one could simply select the lowest threshold as the required spatial resolution, one value for the lateral angle and one value for the polar angle.

4. General discussion

4.1. Related work

Some studies have addressed the issue of audibility of switching between directional filters by examining its effect on the perceived sound quality [3] and by evaluating the level of annoyance produced by the artifacts [4]. Kudo *et al.* [3] compared several switching strategies: direct switching, overlap-add method, weighted overlap-add method, and crossfading using three different crossfading functions (square root, cosine, and a Fourier Series). From an objective analysis based on the expansion of the effective frequency bandwidth [26] that occur at the moment of switching, Kudo et al. concluded that the weighted overlap-add method and the crossfading 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 crossfading is based on intermediate filters it is not surprising that better evaluations were given to crossfading methods than to a direct switching.

In [4] Otani and Hirahara found that annoyance indexes correlated with spectral differences produced by HRTF switching for a number of signals with different bandwidths. Broader bandwidths generated less annoying artifacts and smaller spectral differences than signals with narrower bandwidths.

4.2. 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° lateral angle. 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°) [27]. MAAs have been shown to be about 5° for virtual sources based on generic HRTFs [28], and this is more comparable to the MASSs obtained in this study.

4.3. Implications for dynamic binaural synthesis

In the context of dynamically varying ITD implementation, it seems worth comparing results of time differences in HRTFs with those of time switching in 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 switching by the signal.

Acknowledgements

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. Acta Acustica united with Acustica **94** (2008).

[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** (2003) 1009–1022.

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

[4] M. Otani, T. Hirahara: Auditory artifacts due to switching head-related transfer functions of a dynamic virtual auditory display. IEICE Trans. Fundamentals, **E91-A** (2008) 1320–1328.

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

[6] F. Christensen, C. Boje Jensen, H. Møller: The design of VALDEMAR - An artificial head for binaural recording purposes. 109th Convention of the Audio Engineering Society, Los Angeles, California, USA, 2000, convention paper 5253.

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

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

[9] P. F. Hoffmann, H. Møller: Some observations on sensitivity to HRTF magnitude. J. Aud. Eng. Soc. **2008** in print.

[10] J. Sandvad: Dynamic aspects of auditory virtual environments. 100th Convention of the Audio Engineering Society, Copenhagen, Denmark, 1996, convention paper 4226.

[11] J. Blauert, H. Lehnert, J. Sahrhage, H. Strauss: An interactive virtual-environment generator for psychoacoustics research. I: Architecture and implementation. Acustica united with Acts Acustica **86** (2000) 94–102.

[12] E. M. Wenzel, J. D. Miller, J. S. Abel: Sound lab: A real-time, software-based system for the study of spatial hearing. 108th Convention of the Audio Engineering Society, Paris, France, 2000. Convention paper 5140.

[13] 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** (1996) 30–60.

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

[15] S. A. Gelfand: Hearing an introduction to psychological and physiological acoustics, 3rd edition. Marcel Dekker, Inc., New York, 1998.

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

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

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

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

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

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

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

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

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

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

[26] L. Cohen: Time-frequency analysis. Prentice Hall, 1995.

[27] A. W. Mills: On the minimum audible angle. J. Acoust. Soc. Am. **30** (1958) 237–246.

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