euronoise

Minimum BRIR grid resolution for dynamic binaural synthesis

A. Lindau, H.-J. Maempel and S. Weinzierl

Department of Audio Communication, Technical University of Berlin, Sekr. EN-08, Einsteinufer 17c, 10587 Berlin, Germany
alexander.lindau@tu-berlin.de

The binaural synthesis of acoustical environments is based on binaural room impulse responses (BRIRs) measured with a dummy head for discrete head positions and angular resolutions of typically between 1° and 15°. The resolution of the BRIR grid defines the size of the BRIR database as well as the duration of its measurement. To determine the minimum grid resolution required for dynamic binaural synthesis a listening test was performed. Following an adaptive 3AFC procedure, a spatial grid of BRIR data was gradually coarsened from a maximum resolution of 1° / 1° until audible artefacts were introduced. Thresholds of audibility were tested for a sound source located at 0° / 0° with dynamical auralization in two rotational degrees of freedom. The datasets used were acquired in an anechoic environment and in two rooms of different size and reverberation time. Pink noise and acoustical guitar were used as stimuli. A third octave band filter bank analysis of the data sets, using a 1dB-deviation-in-a-band criterion for the audibility of spectral differences, was in good accordance with the listening test results.

1 Introduction

The dynamic binaural synthesis of acoustical environments based on measured binaural room impulse responses (BRIRs) can provide a high degree of perceptual realism [1]. One important parameter for the quality of binaural auralizations is the degree of interaction allowed between the listener's head movements and the synthesized sound field, i.e. whether it reacts only to horizontal or also to vertical movements, and how fine the spatial resolution of the BRIR dataset is.

The horizontal and vertical resolutions of HRTF resp. BRIR data sets used in today's virtual auditory displays cover a wide range of between 1-2° / 1-2° [2] and 10-22.5° / 10-11.5° [3, 4, 5]. Very often, BRIRs originally acquired with lower resolution are numerically interpolated [6, 7]. With regard to the size of the BRIR database as well as the duration of its measurement it would be desirable to know an auditory threshold above which audible artefacts are introduced by the granularity of the BRIR grid. This problem is related to quantities such as the minimum audible angle (MAA, [8]) and the minimum audible movement angle (MAMA, [9], [10]). In anechoic environments, MAAs of 1° for frontal sound incidents have been reported [9]. But, especially the MAMA, referring to situations with a moving source (or listener) is relevant in this context. MAMA values of 5-20° have been reported for moving sources, with values increasing with the velocity of the source and decreasing with the bandwidth of the signal [10].

Whereas the quality of localization was shown to depend very much on the direction of sound incidence, this factor is neither predictable nor stable for a dynamic binaural system. Therefore, the present investigation has chosen a setup with a sound source in the area around frontal incidence, where the sensitivity to changes in localisation and/or timbre is expected to be highest [8, 10], and for a listener free to move his head in any direction. In contrast to studies investigating changes in ITD/ILD and minimum phase spectrum separately that can be revealing with regard to fundamental localization cues [11, 12], the present approach seems closer to the practical application of binaural simulations. Moreover, the influence of different stimuli (pink noise and acoustical guitar) and the influence of the diffuse sound field has been investigated by using binaural data acquired in an anechoic environment and in two rooms of different size and reverberation time. Finally, a standard model for the audibility of spectral differences, using a 1 dB-deviation-in-a-band criterion based on a third octave filter bank analysis, was compared to the listening test results.

2 Experimental setup

2.1 Binaural Room Impulse Responses

Sets of binaural room impulse responses where measured in a studio (V = 260 m³, RT = 0.7 s, r_{crit} = 1.4 m) and in a large lecture hall (V = 8600 m³, RT = 2.1 s, r_{crit} = 4m). For the measurement the FABIAN head and torso simulator, developed at the TU Berlin was used [1]. In the studio, FABIAN was seated at the mixing desk with the frontal speaker serving as source at a distance of 2.8 m. In the lecture hall FABIAN was seated in the fourth row at 8 m distance from the source on stage. Hence, in both rooms the listening position was at approximately twice the critical distance r_{crit}, providing a predominantly diffuse sound field. In both cases the same loudspeaker was used. The BRIRs were measured for a frontal sphere of head movements of ±80° horizontally and ±35° vertically with 1° resolution, resulting in 11431 BRIRs for each room. Hence, 1° was the finest resolution presented in the listening test and supposed to be inaudible [7]. An additional set of HRTFs for a frontal source was provided by earlier measurements [14], providing 1° resolution only for the horizontal plane, which is why the HRTFs were not tested for vertical grid resolution. A standard system of spherical coordinates was chosen with head movements given by azimuth and elevation angles, so that (+45°, +45°) means "half-right, half-up".

2.2 Auralization

A software for fast partitioned convolution was used to auralize various BRIR sets at a time with a sampling rate of 44.1 kHz. Only the initial 2^{14} samples of the BRIR were updated dynamically, since changes in the diffuse reverberation tail due to different head and source positions were shown to be inaudible [1]. The initial block size was set to 256 samples, whereas for the diffuse tail a block size of 8192 samples was chosen. The result of a filter exchange is available one audio block after recognizing a trigger event; the resulting latency of one audio block is already introduced by the underlying jack audio server architecture [www.jackaudio.org]. The time-domain cross fade results are then output consecutively. This approach ensures a minimum response time to head movements, while

stretching out the fade process in time (see [15]). The inaudibility of the cross fading between BRIRs was approved by a pre-test crossfading white noise and sines between identical HRTFs. Since no switching was audible, it was concluded that all artefacts heard later should be due to differences in the BRIRs themselves. A Polhemus Fastrack head tracker was used, providing an update rate of 120 Hz, i.e. new head positions every 8 ms. STAX SR202 Lambda headphones were used, equalized using a linear phase inverse filter optimized by a least squares criterion and based on the magnitude average of 10 measurements carried out while repositioning the headphones on the dummy head after each measurement. As the 3AFC listening test design (see below) requires instantaneous switching between the full resolution and reduced resolution data sets, the complete set of BRIRs was held in random access memory (ca. 22 GB).

2.3 Listening test procedure

An adaptive three alternative forced choice (3AFC) test procedure was chosen. Three stimuli were presented, including the reference situation (1° / 1° resolution) twice and a reduced grid resolution once. Since the threshold of just audible grid granularity should be tested independently for horizontal and vertical grid resolution, the resolution was changed only for one direction, while kept constant for the other direction at maximum resolution. The presentation order was randomized for each trial. Following a separate training phase, each test run started with a test stimulus in the middle of the range of provided grid resolutions. These ranges (20° for HRTFs; studio & lecture hall: 30° horizontal, 35° vertical) were derived from pretests. The BRIR resolution was changed according to a maximum likelihood adaption rule ("Best-Pest", see [16]). According to pretests a stop criterion of 8 trials was assumed to be sufficient. As can be seen from figure 1, the test procedure converges very fast, resulting in test durations ranging from 54 min. to 1:30 hours, involving 10 test conditions for each subject (see Table 1).
Due to the 1° step size of BRIR resolutions while using a 3AFC stimulus presentation comparing the 1° reference to resolutions >1°, the finest audible grid resolution threshold measurable was 2°. As will be shown below, this value was never reached for the tested conditions and subjects.

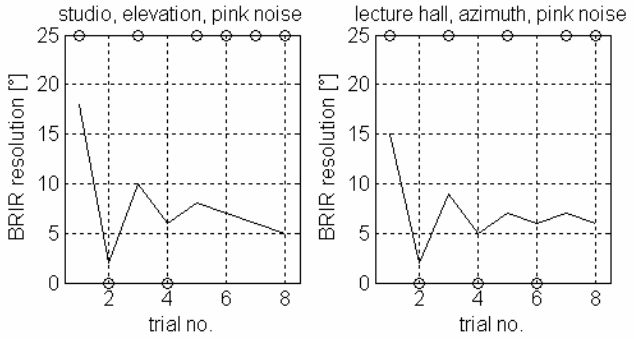

Fig. 1 Progression of the listening test procedure for two conditions (same subject), showing 8 consecutive trials. The BRIR resolution was automatically adapted according to the answers of the subject (a circle below means "difference not recognized", a circle above "recognized").

2.4 Stimuli

Two kinds of stimuli where chosen: pink noise with a duration of 5 seconds (20 ms fade in/out), and a 5 s excerpt from an acoustic guitar piece (bourée by J. S. Bach) which has turned out to be particularly sensitive to simulation artefacts [1]. Pink noise was expected to be particularly sensitive to spectral changes (and also used in [7], [11], [12], [13]). Furthermore, the results of the wide band filter bank analysis discussed below were expected to correlate best with listening test results of a wide band test stimulus. The guitar sample was selected to represent a musical audio content with transient as well as static harmonic signal parts. Due to the band pass used as headphone compensation target the bandwidth of all presentations was constricted to 50 Hz – 21 kHz.

2.5 Experimental design

The thresholds determined were dependent on three factors: room type (HRTF, studio, lecture hall), content (noise, guitar), and direction of grid reduction (horizontally, vertically). 21 subjects (age 24 – 35, 19 male, 2 female) participated in the test. All subjects were experienced in listening tests; most had musical education. Following the repeated measures design, thresholds were collected for each subject in each situation, resulting in 10 x 21 = 210 threshold values. The missing vertical data in the HRTFs produce two empty cells in the full factorial test design.

Room	Stimulus	Direction of reduction
HRTF (anechoic)	pink noise	horizontally
		-
	ac. guitar	horizontally
		-
studio	pink noise	horizontally
		vertically
	ac. guitar	horizontally
		vertically
lecture hall	pink noise	horizontally
		vertically
	ac. guitar	horizontally
		vertically

Table 1 Conditions tested in threshold experiment

3 Results

The following results for each test condition are labelled with three letters indicating room (**H**RTF/**S**tudio/**L**ecture hall), direction of grid reduction (**H**orizontal/**V**ertical) and content (**N**oise/**G**uitar). The distribution of thresholds over all 21 subjects is shown in figure 2. Basic descriptive values are summarized in table 2. Due to the repeated measures design the derived threshold values show a high reliability between subjects (Cronbachs Alpha = 0.934).

condition	mean	median	std.-dev.	5%-perc.
HHN	4,29	4	0,96	3,0
HHG	6	5	2,51	3,1
SHN	6,1	6	1,55	4.0
SVN	6,05	5	1,88	4,0
SHG	11,14	11	4,7	6,0
SVG	13,67	13	5,82	9,0
LHN	5,95	6	1,77	3,1
LVN	4,52	4	1,83	3,0
LHG	11,29	10	5,49	6,0
LVG	14,95	11	11,13	3,1

Table 2 Descriptive results for test conditions

While medians are within a range of 4° to 13° grid resolution, none of the subjects could reliably detect a grid resolution below 3°. Threshold values and variances were higher for the conditions presented with the guitar sample compared to those presented with noise. This effect was shown to be clearly significant by a later ANOVA. For the vertically reduced grid resolution some subjects were even unable to detect differences between the highest (35°) and the lowest (1°) resolution when presented with the guitar sample. Due to the skewed distributions medians and interquartile ranges seem to be more appropriate descriptors of the empirical data.

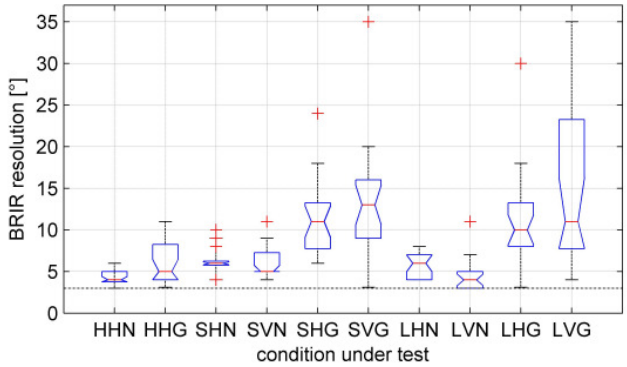

Fig. 2 Just audible BRIR grid resolutions derived for all 10 conditions. Box plots show medians and interquartile ranges, whiskers show 5 % resp. 95 % percentiles. Non-overlapping box-constrictions indicate a significantly different median. The dotted line shows a 3° margin that has never been exceeded.

The integration of threshold distributions for every condition provides cumulated detection frequencies which can be interpreted as estimates of the psychometric functions of the sample (fig. 3).

For comparability, functions related to the noise stimulus (upper row) have been plotted on a 12° range; functions related to the guitar stimulus (lower row) have been plotted on a 36° range. The sensitivity to reductions in horizontal and vertical resolution depends on stimulus type: for noise a vertical grid reduction becomes audible earlier than for the horizontal case, for guitar this is reversed. Particularly remarkable is the mean threshold value for vertical resolution in the lecture hall (noise stimulus), which is significantly lower than the horizontal thresholds, as shown by a non-parametric Friedman Test for multiple related samples.

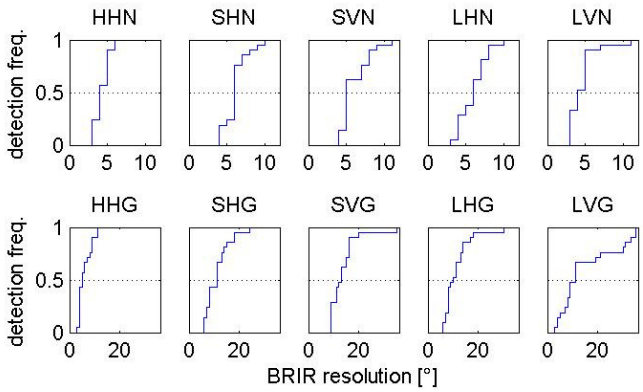

Fig. 3 Cumulated distributions of thresholds derived for all 10 conditions

An ANOVA was conducted for the data related to horizontal grid resolution, taking into account the factors of room (HRTF, studio, lecture hall) and stimulus content (noise, guitar). The 3x2 ANOVA for repeated measures showed significant main effects of content and room. A post hoc test proved the HRTF thresholds to be significantly different from both studio and lecture hall, but not the latter being different from each other. Additionally a significant ordinal interaction between room and stimulus could be found: For horizontal grid resolution, the observed thresholds for the guitar sample increased significantly stronger when played in the studio or lecture hall (as opposed to the HRTF case) than for the noise stimulus.

For the guitar stimulus, significant differences could be found neither between the two rooms nor for the directions of grid reduction. Hence, the minimum audible grid resolution for natural stimuli and both vertical and horizontal head movements in the diffuse field of different rooms seems to be given by single a value (see discussion below).

4 Modeling of results

In [7] the audibility of differences between HRTFs was corresponding rather well to a mean absolute difference in magnitude for 94 logarithmically distributed frequencies from 100 to 20 kHz (summed for left and right ear) of more than of 1 dB. A similar investigation was applied here to the obtained BRIR data.

A third octave filter bank analysis from 63 Hz – 16 kHz was conducted on the first 2^{14} samples of the BRIRs and averaged for both ears. For studio and lecture hall the results are shown in figs. 4 and 5, separately for differences in horizontal (i.e. left-right movements of the head) and vertical (i.e. up-down movements of the head) direction.

The position of the source is marked by a cross at (0° / 0°). Spectral differences are coded in greyscale for each head position at its azimuth and elevation (±80°, ±35°). Therefore the (-80° / 35°) spot is related to a left and upward orientation of the head, while the source always remains at (0° / 0°). 3 – 4° degrees on the greyscale indicate that the next audible BRIR is 3 – 4° away in horizontal or vertical direction depending on the diagram. Hence, the

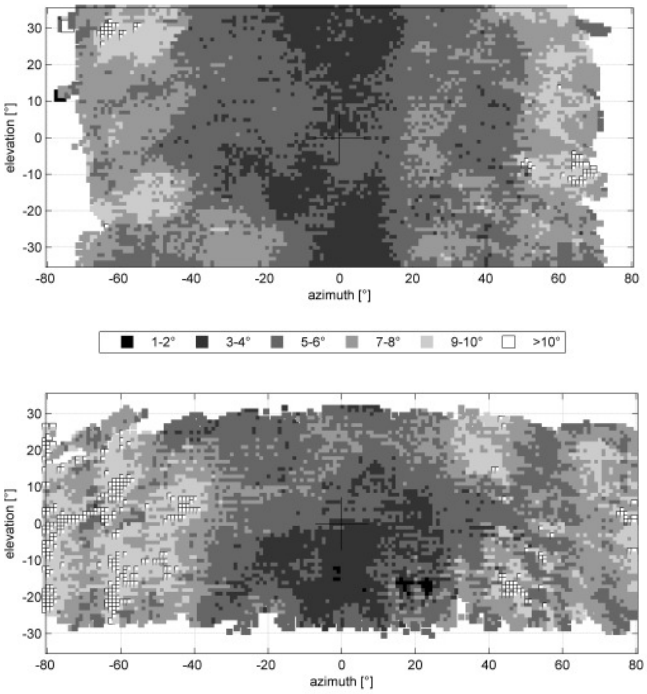

Fig. 4 Just audible grid resolution as estimated from spectral differences for all measured head positions: Studio, evaluated in horizontal (above) and vertical (below) direction only

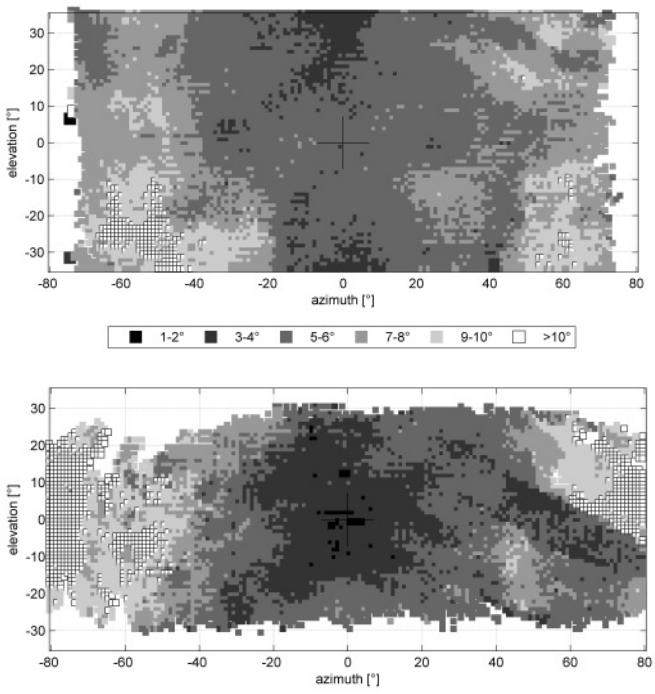

Fig. 5 Just audible grid resolution as estimated from spectral differences for all measured head positions: Lecture hall, evaluated in horizontal (above) and vertical (below) direction only

brighter the area on the plotted angular region, the smaller the spectral difference between adjacent BRIRs is. The white spaces at the boundary of the diagrams indicate that within the data range no audible neighbour could be found at least in one direction of head movement. The results for HRTFs are not shown because of the missing data in vertical direction.

Whereas the values obtained for both rooms (S, L) are rather similar, HRTF values were found to be lower, as can be seen in figure 7 and in table 3. The highest spectral differences occur for head movements in the range of ±20°. Differences due to movements by less than 3° are predicted to be inaudible by the model. For horizontal head movements differences are continuously decreasing for larger excursions of the head to the sides. For horizontal movements of the head there might be a domination of spectral cues due to alternating shadowing because this would occur independently from vertical angle of the head, as receivers (ears) and obstacle (head) don't change their constellation towards the sources. Small differences found would then be mainly due to changes of the pinnae's angle of approach. For vertical head movements sensitivity seems to be highest in a frontal hot spot, decreasing for increasing vertical and horizontal excursions of the head. When moving in the median plane spectral changes are caused mainly by the pinnae, these seem to be largest in a frontal range. Also, they differ between the rooms as for the studio they seem larger for head positions pointing downwards.

5 Comparing results to simulation

In figure 6 the 5%-percentiles of angles predicted to produce spectral differences as derived by the filter bank analysis for the whole frontal sphere is compared to 5%-percentiles derived from the distribution of thresholds from the listening tests.

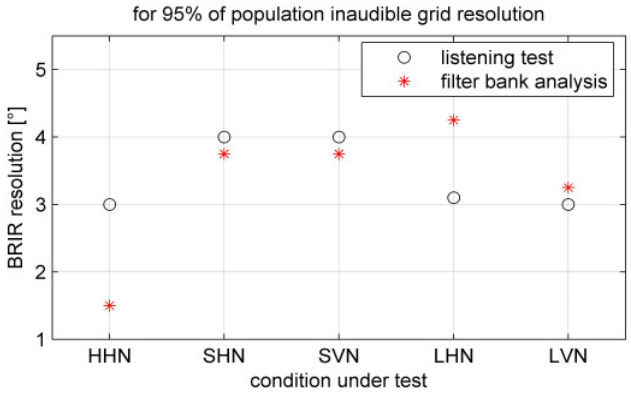

Fig. 6 5%-percentiles of just audible grid BRIR grid resolution derived from listening test and filter bank analysis

Table 3 shows means and standard deviations from the listening test (pink noise situation) compared to those predicted by the filter bank analysis. Distributions (fig. 7, left) and cumulated distributions (fig. 7, right) of a) thresholds from the listening test and b) predicted minimum grid resolution values according to the filter bank analysis for each dataset are in good accordance with each other, suggesting the filter bank analysis as an efficient model for the prediction of just audible grid resolutions of BRIR data.

conditions	Mean	Std.
HVN	4,29	0,96
HRTF: filter bank, horizontal	3,98	1,37
SHN	6,10	1,55
Studio: filter bank, horizontal	6,02	1,58

SVN	6,05	1,88
Studio: filter bank, vertical	6,47	1,88
LHN	5,95	1,77
L-hall: filter bank, horizontal	6,47	1,64
LVN	4,52	1,83
L-hall: filter bank ,vertical	6,33	2,72

Table 3 Comparison of some descriptive values from listening test and filter bank analysis

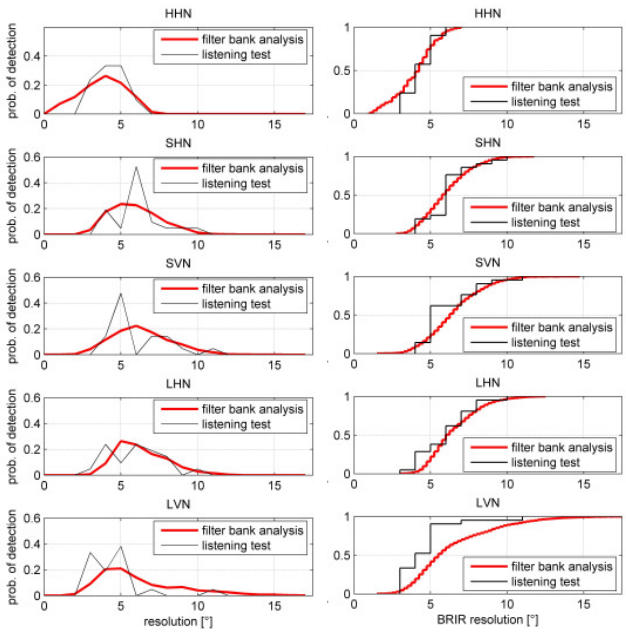

Fig. 7 Distribution of thresholds (left) and cumulated distribution (right) of just audible grid resolution obtained from listening tests and values derived from filter bank analysis

6 Conclusion

The minimum audible grid resolution of measured BRIR data has been studied. For three different acoustic environments, two audio contents and two degrees of rotational head movement, thresholds have been derived in an adaptive 3AFC listening test. For musical content it was considerably more difficult to detect a reduction in grid resolution than for pink noise. Likewise, thresholds increased significantly for BRIRs measured in a diffuse sound field compared to HRTFs measured in an anechoic environment. Somewhat surprisingly, the obtained thresholds for vertical resolution were lower than for horizontal resolution in the case of a broadband excitation such as by pink noise.

For none of the presented conditions a BRIR grid resolution of less than 3° could be detected by the 21 test subjects. Hence, a BRIR data with a grid of 2° x 2° for horizontal and vertical head movements should provide a binaural synthesis of acoustical environments free from resolution artefacts. For musical stimuli and diffuse acoustic environments, the 5%-percentile from pooled threshold values indicates a resolution of 5° x 5° to be sufficient, i.e. only in 4 out of 84 test runs resolution artefacts could be detected.

7 References

[1] Lindau, A.; Hohn, T.; Weinzierl, S.: "Binaural resynthesis for comparative studies of acoustical environments". Presented at the 122nd Convention of the AES, Vienna: 2007, preprint no. 7032

[2] Lentz, T.; Schröder, D.; Vorländer, M.; Assenmacher, I.: "Virtual Reality System with Integrated Sound Field Simulation and Reproduction." In: *EURASIP J. of Advances in Signal Processing* (Article ID 70540), No. Volume 2007

[3] Wenzel, E. M.: "What Perception Implies About Implementation of Interactive Virtual Acoustic Environments". Presented at the 101st Convention of the AES, Los Angeles: 1996, preprint no. 4353

[4] Takala et. al.: "An Integrated System for Virtual Audio Reality". Presented at the 100th Convention of the AES, Kopenhagen: 1996, preprint no. 4229

[5] Sandvad, J.: "Dynamic Aspects of Auditory Virtual Environments". Presented at the 100th Convention of the AES, Kopenhagen: 1996, preprint no. 4226

[6] Hartung, K.; Braasch, J.; Sterbing, S.J.: "Comparison of different methods for the interpolation of head-related transfer functions." In: Proc. of the AES 16th International Conference. Rovaniemi, 1999

[7] Minnaar, P.; Plogsties, J.; Christensen, F.: "Directional Resolution of Head-Related Transfer Functions Required in Binaural Synthesis." In: *J. Audio Eng. Soc.*, Vol. 53 (2005), No. 10, pp. 919-929 7

[8] Mills, A.W.: "On the Minimum Audible Angle." In: *J. Acoust. Soc. Am.* (1958), Vol. 30, pp. 237-246

[9] Perrott, D. R.; Musicant, A. D.: "Minimum audible movement angle: Binaural localization of moving sound sources." In: *J. Acoust. Soc. Am.*, Vol. 62 (1977), No. 6, pp. 1463-1466

[10] Chandler, D. W.; Grantham, D.W.: "Minimum audible movement angle in the horizontal plane as a function of stimulus frequency and bandwidth, source azimuth, and velocity." In: *J. Acoust. Soc. Am.*, Vol. 91, No. 3, pp. 1624-1636 10

[11] Hoffmann, P. F.; Møller, H.: "Audibility of Time Switching in Dynamic Binaural Synthesis". Presented at the 118th Convention of the AES, Barcelona: 2005, preprint no. 6326

[12] Hoffmann, P. F.; Møller, H.: "Audibility of Spectral Switching in Head-Related Transfer Functions". Presented at the 119th Convention of the AES, New York: 2005, preprint no. 6537

[13] Hoffmann, P. F.; Møller, H.: "Audibility of Spectral Differences in Head-Related Transfer Functions". Presented at the 120th Convention of the AES, Paris: 2006, preprint no. 6652

[14] Moldrzyk, C. et al.: "Head-Tracked Auralization of Acoustical Simulation". Presented at the 117th Convention of the AES, San Francisco: 2004, preprint no. 6275

[15] Müller-Tomfelde, C.: "Time varying Filters in non-uniform Block Convolution." In: Proc. of the COST G-6 Conference on Digital Audio Effects. Limerick, 2001, Vol. 2001

[16] Pentland, A.: "Maximum likelihood estimation: The best PEST." In: *Perception & Psychophysics*, Vol. 28 (1980), No. 4, pp. 377-379