Acoustics'08 Paris
June 29-July 4, 2008
www.acoustics08-paris.org

# Externalization in binaural synthesis: effects of recording environment and measurement procedure

Florian Völk, Fabian Heinemann and Hugo Fastl

AG Technische Akustik, MMK, TU München, Arcisstr. 21, 80333 München, Germany
florian.voelk@mytum.de

Databases of head related impulse responses (HRIRs) for binaural synthesis can be measured either in anechoic or reflective environments. If high synthesis quality is needed, miniature microphone measurements are performed in the ear canals of each individual user (individual measurement). Sometimes impulse responses measured in the ear canals of one individual are used for synthesis for other persons (non-individual measurement). In most other cases, artificial head measurements are used. This paper considers the dependence of the perceived distance of auditory images (externalization) on the measurement procedure (individual, non-individual, or artificial head) and on the recording environment (anechoic or reflective). For each measurement, the same system and the same setup, especially the same geometric parameters, are used. Differences in the corresponding impulse response databases are determined and related to the subjective relative externalization differences in the front, in the back, and to both sides. For each direction, a seven point rating scale was used. Statistical analysis suggests that the applied measurement parameters influence the externalization of auditory images: reverberation in impulse responses increases externalization significantly if a human head is used for recording. If the considered artificial head (Neumann KU 80) is used, only a marginal increase in externalization occurs.

# 1 Introduction

In the past decade, auralization using binaural technique (cf. Møller, [1] or Hammershøi and Møller, [2]) has gained more and more attention in the context of virtual reality applications (e.g. Völk et al., [3], Blauert, [4]). Since sufficient processing power for real-time computation of well-known fast convolution algorithms is available, even complex virtual auditory scenes including moving sources and moving listeners as well as user interaction may be rendered.

A fundamental component of each binaural synthesis system is the HRIR library used. The term *head related impulse response* is widely used for impulse responses recorded under anechoic conditions. If the recording is carried out in a reflective environment, the resulting impulse responses are called *binaural room impulse responses* (BRIRs). To point out that the impulse responses contain information stemming from the room, the term *room* is included. Whenever in the following one term for both groups of impulse responses (with and without reflections) without an explicit distinction between them is necessary, the term *head related impulse response* will be used.

There are two common approaches to the collection of the HRIR library: a model- and a data-driven method (for an overview cf. Vorländer, [5]). The difference between these approaches is the method used for the room simulation.

The first approach is based on a HRIR library measured under anechoic conditions. These HRIRs are convolved with a room impulse response that may be measured or - under certain conditions - rendered in real-time (cf. Vorländer, [5]). The latter procedure allows maximum flexibility since changes of the room’s acoustical properties during system operation and even simulations of non-existent rooms are possible.

The second, more traditional and restrictive approach relies on BRIR measurements in the room of which a simulation is desired. This room may be an anechoic chamber. Here, the data-driven and the model-driven approach without room simulation are identical. If there are reflections in the recording room, the data based approach requires lots of measurements, making the synthesis of a reflective environment a time consuming and resources intensive task (e.g. memory requirements).

Each of the aforementioned approaches requires many HRIR measurements, which leads to the necessity of a quick and easy measurement procedure. Artificial heads allow fast and automated measurements, but it is well known that the perceptual quality of a synthesized scene strongly depends on the used head (cf. Møller et al., [6]). Additionally, it is always lower than the quality of synthesis with recordings made in human ears (cf. Minaar et al., [7] as well as Møller et al., [8]), especially for measurements in the subject's own ears (cf. Minaar et al., [9]). Because of the complexity and physical burden of an individual HRIR measurement, it is often desirable to use artificial head recordings or at least non-individualized recordings from a standard subject (cf. Møller et al., [10] and [11]), although some perceptual factors will therefore decrease, for example directional localization (Wenzel et al., [12]). For many practical applications, the reduced complexity is much more desirable than the highest possible synthesis quality. The perceptual impression created by a virtual auditory display based on binaural technique is dominated by the perceived distance of a sound event, the distance of the auditory image. If the synthesis is done with improper HRIRs, auditory images are very close to the head, even if they are not intended to be as close. In the worst case, they are located inside the head.

This paper deals with the questions what improper HRIRs are and especially which auditory image distance (which degree of externalization) can be achieved with a certain recording method. The differences, which are perceived between distances of auditory events created by binaural technique with different impulse response databases, will be quantified. After a consideration of the listening situation and especially the distance perception in virtual acoustics, a short literature review is given to motivate the present work. The aims of the current work, procedures and stimuli, as well as other conditions are defined and results are shown. A discussion of the results and a comparison to previous works conclude this paper.

# 2 Auditory distance perception and externalization in virtual acoustics

The main goal of a virtual auditory display could at first glance be defined as the synthesis of a sound scene that is (or at least might be) present in the recording (original) situation. After some moments of thought, it is obvious that there exists no pure sound scene at all. Objects in the perceptual space of humans always arise at least from a combination of inputs of all senses (other effects like previous knowledge and learning shall be neglected here for simplicity).

For that reason, a better definition of the goal of a virtual auditory display would be the proper synthesis of the auditory part of a real scene, presumed that the remaining parts of that scene (the visual and tactile components etc.) are synthesized in a proper way. To verify if a certain system reaches this goal, a comparison between the synthesized and the original scene is necessary. With nowadays technology, a proper synthesis of a real scene's non-auditory part is not possible; therefore, a comparison to verify the acoustical part is not practicable.

Because the definition given above is correct within theoretical consideration but practically not helpful, a different definition of the goal of virtual acoustics is needed. Another way to deal with the considered situation is trying to isolate the auditory part of the real and the virtual scene for comparison. As mentioned above, in reality there is no purely auditory scene. For that reason, the human perception mechanism expects in addition to auditory stimuli some more (non-auditory) stimuli. It then combines all of them for the generation of objects in the perceptual space (cf. Blauert and Jekosch, [13]). Among them, there might be one or more auditory events. Because it is not possible to block the inputs to the non-auditory senses, it is only applicable to give as little input as possible to them and to keep the conditions for the comparison as constant as possible.

A common method to reduce the input to the visual sense (also used in the present work) is to carry out the experiments in complete darkness. It should be mentioned that darkness does not mean that there is no visual stimulus at all. It can be ensured that darkness is the only visual stimulus present, but it is not possible to avoid an influence of the visual stimulus darkness or of physical effects like fibrillation, caused by darkness, on the auditory event. Additionally, in this way, comparable circumstances for all subjects can be assured and an influence of a visually perceptible sound source on the auditory event (cf. Seeber, [14]) is avoided.

Therefore, we define the goal of virtual acoustics as the creation of the auditory events that arise in the corresponding real scene in complete darkness. For that reason, the experiments reported in this paper were all conducted under dark circumstances. Inputs to other modalities (e.g. the tactile sense) were neglected. It is assumed that they play an inferior role, when subjects are seated in a dark room and sound levels remain in the range around 70 dB (A) for broadband stimuli as applied in the current work. In a straightforward manner, externalization is defined here as the perceived distance of an auditory event to the center of the head (following Kim and Choi, [15]), but with the additional requirement of dark circumstances.

# 3 Previous work

Externalization is a subjective perception and is generally not defined exactly. It is possible to create externalization in different ways, not only by trying to reproduce correct ear signals. Sakamoto et al. ([16]) for example created externalized auditory events using artificial reverberation. Begault et al. ([17], [18]) showed that reverberation in the used impulse responses might influence externalization in virtual auditory displays. They found nearly a doubling in externalization caused by reverberation.

Externalization can mean in the one extreme the perception of auditory events comparable to reality, in the other extreme auditory events perceived a little outside the head. Hartmann and Wittenberg ([19]) were able to continuously move the auditory event from inside the head to the outside (also described by Blauert, [20]). It has been shown by Hartmann and Wittenberg ([19]) that HRIRs measured with an artificial head lead to externalized auditory events that are often diffuse or localized at a wrong position, regarding direction and distance. Besides, many front/back confusions occur (see Wightman et al., [21]) and the auditory events are closer to the head than those created by real sources are. The situation gets better when using individual HRIRs (cf. Wenzel et al., [12]).

Hartmann and Wittenberg ([19]) showed that the correct spectrum at the ears is essential for the creation of externalized auditory events, whereas the correct reconstruction of interaural level differences is not sufficient. Toole ([22]) studied localization with real sound sources and recognized an influence of the signal bandwidth. Additionally, they mentioned that the source position plays an essential role for externalization. For these reasons, it seems plausible that individual HRIRs lead to the largest externalization, as they reproduce the most individual spectral cues. The externalization decreases when using HRIRs measured in the ear canals of another human being (cf. Wightman and Kistler, [23], [24]).

Kim and Choi ([15]) compared the degree of externalization for different HRIR-sets (recorded in an anechoic environment). Their results suggest that externalization can be reached with artificial-head HRIRs as well as with individual ones, but the latter lead to more externalization than the first. The sound stimuli used in [15] were white noise pulse trains (impulse duration 250 ms, 20 ms ramps) and the distance of the virtual source was 1.4 m. A virtual sound source was rotated in steps of 15° around the subjects' heads, starting in frontal direction. A similar procedure was used in the present study.

# 4 Stimuli and Procedure

All used impulse responses were measured with a well-known method using Maximum-Length-Sequences (MLS) as measurement signals (see Schroeder, [25] and Rife and Vanderkooy, [26]). As artificial head, a Neumann KU 80 (with torso) was used, which is known to produce many distance errors (cf. Møller et al., [27]). For the individual measurements, miniature microphones (Sennheiser KE 4-211-2) were inserted in the blocked ear canal (following Hammershøi and Møller, [28]) of a so called good listener (a person whose HRIRs have shown good localization results in previous studies, cf. Møller et al., [10], Seeber and Fastl, [29] and [30]).

With both measurement objects (the artificial head and the individual), two sets of HRIRs (one pair for every five degree in the horizontal plane) were recorded, one in an anechoic chamber and one in a laboratory with reflecting walls and ceiling as well as a carpet on the floor. The distance between the measurement loudspeaker and the center of the head was kept constant at 2 m for all recordings. After the measurement, a spatial interpolation was performed to reach the desired spatial step size of one degree in the horizontal plane. This procedure consisted of an appropriate temporal shifting of the measured impulse responses, a spline-interpolation of the responses in the time domain and of the time-shifting-vector and finally of a back-shifting step. The impulse responses were used as FIR-filters and cut to 256 samples (at 44.1 kHz sampling frequency) in the anechoic case and 2048 samples for the ones measured in the laboratory environment.

As sound stimulus, pulsed uniform exciting noise (UEN, cf. Fastl und Zwicker, [31]) was used. Because this stimulus contains the same intensity in each critical band, all spectral cues contained in the HRIRs are available with the same perceptual weight. Therefore, all possible spectral information is available to the hearing system, but no influence of the sound stimulus on the auditory event should be present. To add some temporal cues to the signal besides the random temporal structure of the noise, the UEN was pulsed with 200 ms pulse and pause duration. Following Blauert and Braasch ([32]), this is the minimal duration allowing dynamic localization cues. The pulses were modulated with 10 ms Gaussian gating signals to prevent audible clicks. A virtual source was rotated two times on a circle around the head of the listeners (starting at a randomly chosen direction) with a virtual acoustics system (cf. Völk et al., [3]), but with no respect to the orientation of the listener's head. That means, no dynamic localization cues evolving from head movements were present, but there should arise dynamic cues resulting from the source movement.

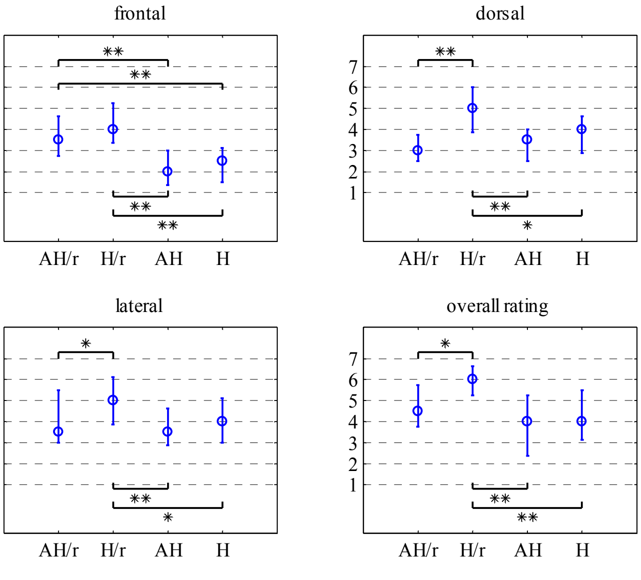

Fig. 1 Externalization differences between HRIR sets. Results for the artificial (AH) as well as the human head (H) recording. The index "/r" indicates recording in reflective environment, otherwise, recordings took place in an anechoic chamber. Three different directions and the overall quality of the intended circle of the auditory event were judged on a seven point rating scale, each stimulus four times. An asterisk indicates significant differences on a 5 % significance level, two asterisks on a 1 % significance level.

Each HRIR-set was presented four times. Thus, every subject had to perform 16 judgements, which led to trial durations of 11 to 17 minutes (mean value: 13 minutes). The presentation sequence was chosen randomly for each subject. A software tool, running on a consumer PC, automated the whole trial. The subjects were seated in front of a tablet-PC and had to answer by selecting a radio button corresponding to the intended answer with the computer mouse. Their task was to complete the following three sentences on a seven point rating scale (in German): "I heard the noise in the front / behind me / to the side". The answer scale ranged from "not at all" to "very far" for each judgement with no additional identifiers associated with the scale-steps. It was intended to ask for the distance of the auditory event, not for the position of the sound source (cf. Blauert, [20]). Additionally, the overall quality of the circle (of the auditory event) in the horizontal plane had to be judged again on a seven point rating scale ranging from "very badly" to "very well". The subjects were explicitly instructed not to avoid bad judgements, because it was known from previous studies (see for example Kim and Choi, [15]) that the results, especially with artificial head HRIRs, could be rather bad.

# 5 Results

Thirteen normal hearing subjects (two female and eleven male) aged between 21 and 55 years (mean value: 27 years) participated in the experiment. Four subjects had previous experience in listening tests, two of them were familiar with listening in virtual acoustical displays and experienced with localization experiments. No persons had participated in listening experiments before. From the median-values of each person for each stimulus (individual medians), the median-values and inter-quartile-ranges of the individual medians were computed. In addition, the data sets were checked for significant differences (ANOVA with post-hoc comparisons according to Bonferroni).

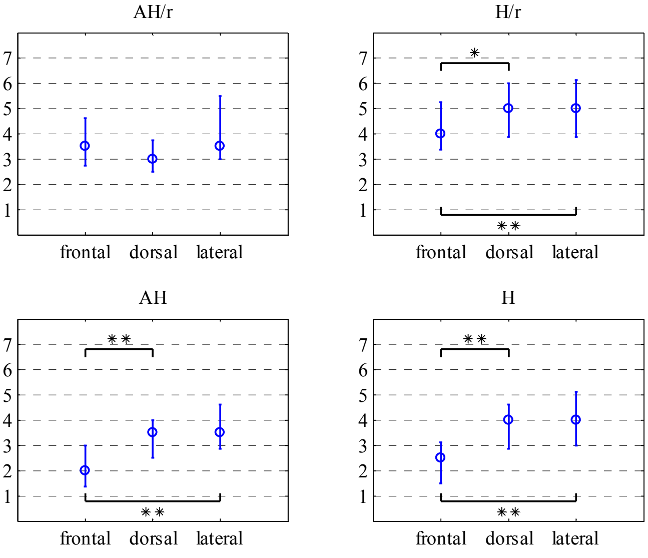

Fig. 2 Externalization differences in HRIR sets. Results of all subjects are shown for the artificial (AH) as well as the human head (H). Index "/r" indicates recording in reflective environment.

Medians are displayed as circles; inter-quartile ranges as lines with markers at the quartiles. Significant differences between the corresponding data sets are indicated by one asterisk on a 5 % significance level and by two asterisks on a 1 % significance level. Fig. 1 shows the results of all subjects for the frontal, the lateral, and the dorsal direction as well as the overall ratings of the quality of the intended circle of the auditory event. On the abscissa the different HRIR-sets, i.e. the artificial head (AH) and the human head (H) are displayed. An additional "/r" (e.g. AH/r) indicates recording in reflective environment. On the ordinate, the seven-point rating scale is shown. Fig. 2 shows the same results as Fig. 1, but grouped as externalization differences to the different directions for each of the used HRIR-sets.

# 6 Discussion

The results displayed in Fig. 1 show a significant difference in the degree of externalization and in the overall rating between the anechoic recordings and the individual recording in a reverberant room. This result is in accordance with Begault et al. ([17], [18]), who showed that adding reverberation to HRIRs measured on a human head leads to larger externalization. Our results suggest that more detailed spectral information at the ears, as they are contained in human head HRIRs compared to the used artificial head HRIRs, are a prerequisite for the mechanism described above, which might also be concluded from the results of Hartmann and Wittenberg ([19]).

Fig. 2 shows that for three out of the four used HRIR-sets, the externalization is significantly worse to the front than to the other directions. In this critical frontal direction, even the artificial-head-recording in a reflective environment creates significantly more externalization than the anechoic recordings, which is not the case in all other directions. The worst externalization happens in front of the persons if the recording contains no reflections, regardless which head is used. This may lead to the assumption that reverberation causes more externalization when the spectral cues are in line with the information contained in the reverberation. The latter should also be the case if very little spectral information is available, as for example in the frontal direction with the artificial head HRIRs.

The results displayed in Fig. 1 suggest that no significant difference occurs in any case between the two anechoic recordings, but there is a tendency that the auditory events created with the human head-recording are a little farther out, which has been shown also by Kim and Choi ([15]). The presented results suggest furthermore that the inclusion of reverberation in the impulse responses improves externalization more than the use of human head measurements instead of artificial head ones.

Apart from the critical frontal direction, the recordings within human ear canals in a reflective environment create significantly more externalized auditory events than the artificial-head ones. Together with the findings of Wightman and Kistler ([23], [24]), it may be concluded that the greatest externalization can be reached with individual recordings and only to a lesser extent with recordings from other human ears (as used here).

The least externalization is possible with artificial head recordings. It might be the case that measurements from other artificial heads than the one used here could create more externalization than human HRIRs from a so-called bad listener. On this account, the order mentioned here presumably holds for a good listener and an average artificial head. Fig. 3 summarizes the above-mentioned dependencies by showing the used HRIR-sets in sequence of the created degree of externalization for each considered direction.

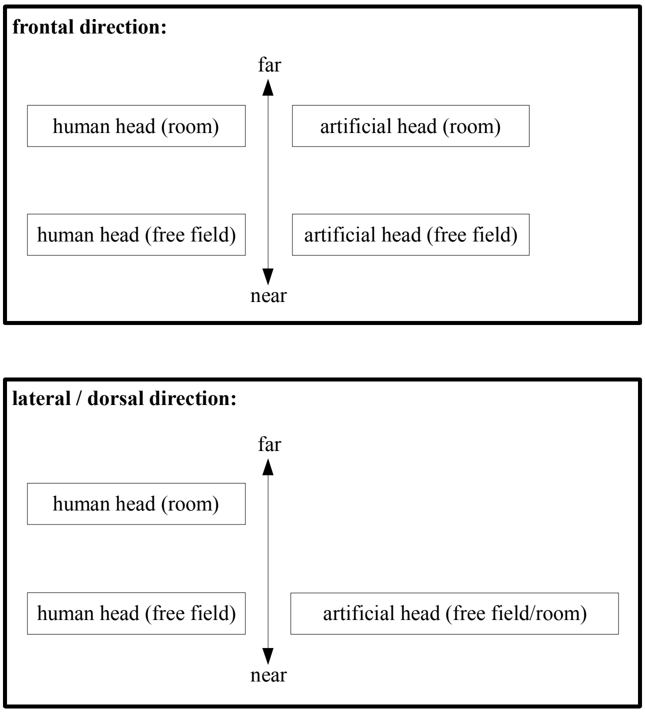

Fig. 3 Degree of externalization.
Significant differences of virtual auditory displays to the considered directions, dependent on the used HRIR-library. Significant differences were computed from the individual median values. A Neumann KU80 was used as artificial head.

A possible demonstrative explanation of the aforementioned constraints and especially of the fact that anechoic recordings create auditory events very close to the head might be the following:

Our hearing system acts like being in a comparative real-life situation although listening to a virtual auditory display. The little amount of diffuse energy contained in a HRIR measured under anechoic conditions occurs most likely in a situation with a sound source very close to the head or in a free field situation.

While a free field situation with as little reflections as are contained in a recording taken in an anechoic chamber is very unlikely to occur, a sound source very close to the head is a very common situation. For that reason, our hearing system might create the - under realistic conditions - more probable auditory event of a source being close to the head.

# 7 Summary

The results of the work presented in this paper are in accordance with the data presented by Begault et al. ([17]) as well as with the results of Kim and Choi ([15]). In addition, some quantitative values are presented. The findings might be summarized as follows:

Reverberation in impulse responses used for binaural technique increases the perceived sound source distance. For human heads, this effect is significant, for the used artificial head, only a tendency is visible. This may be because the used artificial head is known to produce a greater number of wrong distance perceptions than others do. This is most obvious regarding the critical frontal direction.

## Acknowledgments

Part of this work was supported by grant FA 140/4 of the Deutsche Forschungsgemeinschaft (DFG). The authors gratefully acknowledge the support of Sennheiser electronic GmbH & Co. KG, who provided the microphones for the HRIR measurements.

## References

[1] H. Møller, "Fundamentals of Binaural Technology", *Appl. Acoustics* 36, 171-218 (1992)

[2] D. Hammershøi, H. Møller, "Methods for Binaural Recording and Reproduction", *ACUSTICA - acta acustica* 88, 303-311 (2002)

[3] F. Völk, S. Kerber, H. Fastl, S. Reifinger, "Design und Realisierung von virtueller Akustik für ein Augmented-Reality-Labor", *Fortschritte der Akustik, DAGA '07*, DEGA e. V., Berlin (2007)

[4] J. Blauert, "3-D-Lautsprecher-Wiedergabemethoden", *Fortschritte der Akustik, DAGA '08*, DEGA e. V., Berlin (2008)

[5] M. Vorländer, "Auralization – Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality", *Springer*, Berlin, Heidelberg (2008)

[6] H. Møller, D. Hammershøi, C. B. Jensen, M. F. Sørensen, "Evaluation of Artificial Heads in Listening Tests", *J. Audio Eng. Soc.* 47, 83-100 (1999)

[7] P. Minaar, S. K. Olesen, F. Christensen, H. Møller, "Sound localization with binaural recordings made with artificial heads", *ICA 2004*, V3651-V3654 (2004)

[8] H. Møller, M. F. Sørensen, C. B. Jensen, D. Hammershøi, "Binaural Technique: Do We Need Individual Recordings?", *J. Audio Eng. Soc.* 44, 451-469 (1996)

[9] P. Minaar, S. K. Olesen, F. Christensen, H. Møller, "Localization with Binaural Recordings from Artificial and Human Heads?", *J. Audio Eng. Soc.* 49, 323-336 (2001)

[10] H. Møller, C. B. Jensen, D. Hammershøi, M. F. Sørensen, "Selection of a typical human subject for binaural recording", *ACUSTICA - acta acustica* 82, 215 (1996)

[11] H. Møller, C. B. Jensen, D. Hammershøi, M. F. Sørensen, "Using a Typical Human Subject for Binaural Recording", *100th AES Convention* (1996)

[12] E. Wenzel, M. Arruda, D. Kistler, F. Wightman, "Localization using nonindividualized head-related transfer functions", *J. Acoust. Soc. Am.* 94, 111-123 (1993)

[13] J. Blauert, U. Jekosch, "Sound-Quality Evaluation – A Multi-Layered Problem", *ACUSTICA - acta acustica*, 83, 747-753 (1997)

[14] B. U. Seeber, "Zum Ventriloquismus-Effekt in realer und virtueller Hörumgebung", *Fortschritte der Akustik, DAGA '02*, DEGA e. V., Oldenburg (2002)

[15] S. Kim, W. Choi, "On the externalization of virtual sound images in headphone reproduction: A Wiener filter approach", *J. Acoust. Soc. Am.* 117, 3657-3665 (2005)

[16] N. Sakamoto, T. Gotoh, Y. Kimura, "On "Out-of-Head Localization" in Headphone Listening", *J. Audio Eng. Soc.* 24, 710-716 (1976)

[17] D. R. Begault, E. M. Wenzel, A. S. Lee, M. R. Anderson, "Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source", 108th *AES Convention* (2000)

[18] D. R. Begault, "Perceptual effects of synthetic reverberation on three-dimensional audio systems", *J. Audio Eng. Soc.* 40, 895-904 (1992)

[19] W. M. Hartmann, A. Wittenberg, "On the externalization of sound images", *J. Acoust. Soc. Am.* 99, 3678-3688 (1996)

[20] J. Blauert, "Spatial Hearing – The Psychophysics of Human Sound Localization", *The MIT Press*, *Cambridge, Massachusetts, London, England*, Revised Edition (1997)

[21] F. Wightman, D. Kistler, M. Arruda, "Perceptual consequences of engineering compromises in synthesis of virtual auditory objects", *J. Acoust. Soc. Am.* 92, 2332 (1992)

[22] F. E. Toole, "In-head localization of acoustic images", *J. Acoust. Soc. Am.* 48, 943-949 (1970)

[23] F. L. Wightman, D. J. Kistler, "The Perceptual Relevance of Individual Differences in Head-Related Transfer Functions", *ACUSTICA - acta acustica* 82, S92 (1996)

[24] F. Wightman, D. Kistler, "Measurement and Validation of Human HRTFs for Use in Hearing Research", *ACUSTICA - acta acustica*, 91, 429-439 (2005)

[25] M. R. Schroeder, "Integrated-impulse method measuring sound decay without using impulses", *J. Acoust. Soc. Am.* 66, 497-500 (1979)

[26] D. D. Rife, J. Vanderkooy, "Transfer-Function Measurement with Maximum-Length Sequences", *J. Audio Eng. Soc.* 37, 419-444 (1989)

[27] H. Møller, C. B. Jensen, D. Hammershøi, M. F. Sørensen, "Evaluation of Artificial Heads in Listening Tests", 102nd *AES Convention* (1997)

[28] D. Hammershøi, H. Møller, "Sound transmission to and within the human ear canal", *J. Acoust. Soc. Am.* 100, 408-427 (1996)

[29] B. U. Seeber, H. Fastl, "Effiziente Auswahl der individuell-optimalen aus fremden Außenohrübertragungsfunktionen", *Fortschritte der Akustik, DAGA '01*, DEGA e. V., Oldenburg (2001)

[30] B. U. Seeber, H. Fastl, "Subjective Selection of Non-Individual Head-Related Transfer Functions", *Proc. of ICAD 2003* (2003)

[31] H. Fastl, E. Zwicker, "Psychoacoustics – Facts and Models", *Springer, Berlin, Heidelberg*, 3rd Edition (2007)

[32] J. Blauert, J. Braasch, "Räumliches Hören", Contribution for the Handbuch der Audiotechnik (Chapter 3, Stefan Weinzierl, Ed.), *Springer, Berlin, Heidelberg* (2007)