MADRID
inter.noise 2019
June 16 - 19
NOISE CONTROL FOR A BETTER ENVIRONMENT

Virtual scene adaption for compensation of the reproduction room

Kohnen, Michael[1]
Vorländer, Michael
Institute of Technical Acoustics, RWTH Aachen University
52056 Aachen, Germany

ABSTRACT

To study the effect of real-life acoustic scenes on human behavior and well-being, a repeatable scene in an acoustically controllable environment must be set up that can be perceived as naturally as possible. As headphone reproduction might distract the user from a natural immersion, loudspeaker-based solutions are considered. These are designed to work perfectly in anechoic rooms and neglect the influence of the reproduction room. A possible solution by substitution of reflections in the target scene with unavoidable reflections occurring in the reproduction room is proposed. Reflections which are close in time and direction of arrival are substituted. As these criteria are less important for the late reverberant part of a room impulse response (RIR) this work focuses on early reflections only. Reflection detection was done using a hybrid model of virtual acoustics and analysis of RIR measurements. As loudspeaker arrays are used according to a specific reproduction technique chosen the algorithm was adapted to three different reproduction techniques: higher order Ambisonics, crosstalk-cancellation and vector-base amplitude panning. For the performance evaluation, a listening test was conducted regarding the perceived width of a virtual source.

Keywords: Room-compensation, Spatial audio, Auralization
I-INCE Classification of Subject Number: 74

1. INTRODUCTION

The perception of sound and its effect on our performance in everyday tasks is subject to research in measures like productivity or well-being. The psychoacoustic and clinical investigations of these effects have to be done under controllable and repeatable conditions while at the same moment they should be as realistic, plausible and natural-sounding as possible. Loudspeaker-based reproduction methods are able to set up virtual acoustics environments in which subjects can be examined without attaching any additional devices

[1]mko@akustik.rwth-aachen.de

that may limit their actions or distract them from the task of a behavioral test. With the demand of authenticity of a reproduced scene comes its need for a correct reproduction. A virtual acoustic scene is usually rendered to the intended acoustic properties. Yet, the room in which this scene is reproduced superimposes the rendered presentation with its own acoustics. Reproduction techniques assume free-field environment, which usually can not be fulfilled, and do not compensate for the presence of the environing room.

Different compensation techniques have been proposed [1–4] which solely relate on the information only given on the reproduction side. The presented approach takes into account knowledge of the virtual scene and adapts it and therefore focuses on the use on room acoustically auralizations. This paper proposes a method for compensation of early reflections in which the virtual scene will be adapted to match, combined with the acoustics of the reproduction room, the intended acoustics of the virtual acoustic scene. Early reflections will be detected in both, the virtual scene and the reproduction room, and then be compared by means of time, direction of arrival and energy. The procedure can be found schematically in figure 1.

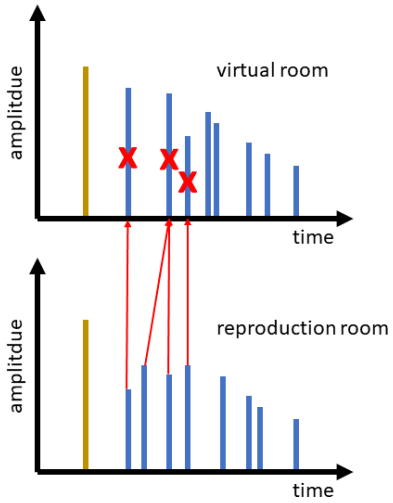

Figure 1: The principal idea is schematically illustrated. Reflections in the virtual room are erased as they will naturally occur in the reproduction room.

Depending on a further defined 'distance' (see section 2.3) of a reflection in the virtual scene to a reflection in the reproduction room, a reflection in the virtual scene will be suppressed as it will be substituted by an unavoidable reflection in the reproduction room. For the late reverberant field an adaptive algorithm is intended that changes the absorption material homogeneously in the virtual room to match the energy decay of the convolved rooms, reproduction and virtual, to the intended target room. This preserves time and direction of arrival of the reflections and their relative energy differences [5].

This paper presents a compensation algorithm by outlining reflection detection and distance estimation as well as implementation for different reproduction methods. A subjective evaluation in terms of a listening test is done that validates the narrowed perception of the apparent source width (ASW) of a virtual source.

2. COMPENSATION APPROACH

2.1. Example room

Figure 2: The figure shows the VRLab used to evaluate the proposed approach. The loudspeaker array is visible as well as the acoustics treatment in terms of curtains and an acoustic ceiling. The artificial head in the middle was used for measurements. During measurements the curtains covered the screen in front of the artificial head.

For the verification of the proposed approach the virtual reality laboratory (VRLab) of the Institute of Technical Acoustics at the RWTH Aachen University in Germany was used. It roughly has a ground plane of 5 meters times 8 meters and a height of 2.8 meters. The ceiling is acoustically treated and the corners contain low frequency resonance absorbers. The walls are covered by curtains. The reverberation time is about 0.15 seconds. The curtains can also be used to divide the room into a listening area of about $5m$ x $5m$ and a control space. On the $5m$ x $5m$ area a 12 loudspeaker array is set up which allows for different reproduction methods as illustrated in figure 2. Four loudspeakers in the horizontal plane are placed at an azimuth angle of 45°, 135°, 225° and 315°. Two loudspeaker rings elevated by ±30° place four loudspeaker each at an azimuth angle of 0°, 90°, 180° and 270°. The distance of each loudspeaker is 2.3 meters.

The room is furthermore equipped with an optical motion tracking system (red cameras in the ceiling) and stereoscopic beamers for 3-D viewing. The visual screen was covered by curtains for the purposes in this paper.

2.2. Reflection detection

Simulation models require carefully chosen acoustic characteristics of the materials in the room model [6]. To get more precise information about the reproduction room and to avoid complex gathering of the material data a hybrid model is proposed. For each loudspeaker a room impulse response (RIR) with a simple omni-directional microphone is measured. A reflection detection algorithm, as described below, is then searching for early reflections in the measured RIR. After that, the arrival time of these reflections is compared to the early reflections in a RIR calculated by an image source method based on a geometrically accurate model that only contains rough estimation of the acoustic

properties of the room boundaries. Each reflection in the measured RIR can then finally be mapped to a certain direction.

Simulation methods, especially geometrical acoustics, often assume a reflection that influences the incident sound only in terms of energy. Yet, the reality shows a complex behavior including frequency depending phase shifts and wave-based effects. Consequently, the algorithm is not searching for attenuated copies of the direct sound but for similarities of it. This is done cross-correlating the direct sound with the whole RIR [7]. The direct sound is selected using a five sample time-window starting from the peak value of a RIR. Once it detects a full window without any peak above -30dB it will stop and search in the other time direction. The extracted direct sound part is then cross-correlated to the RIR. A reflection will be detected if the cross-correlation exceeds a certain threshold (here 0.5), neglecting reflection from strongly damped directions. The measured RIR can be found in figure 5 and does not directly provide information about the time or direction of arrvial of early reflection.

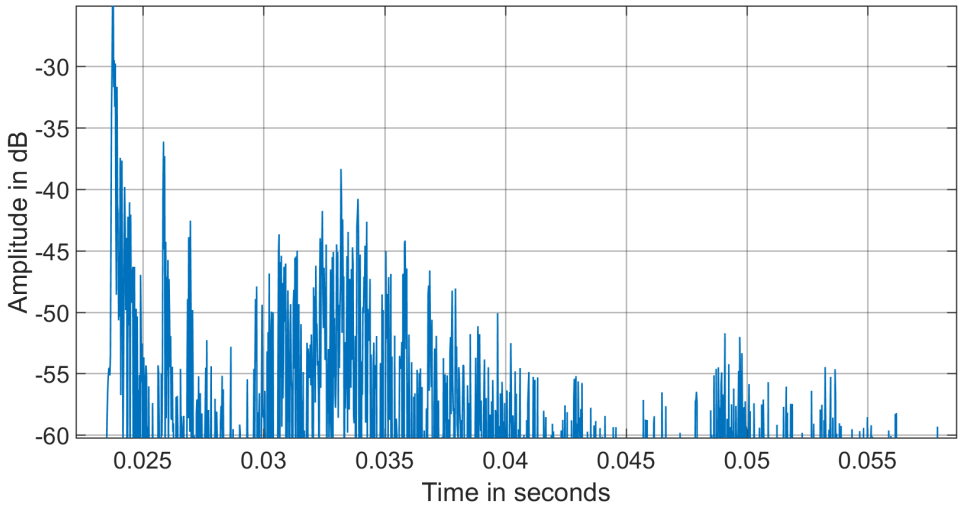

Figure 3: RIR in the reproduction room measured with an omni-directional microphone. Early reflections can not directly be separated from each other and no information of their direction of arrival is immanent.

The results of the cross-correlation can be seen in figure 4. Single reflections are revealed and can be mapped due to their time of arrival to the image sources calculated by an in-house software solution called RAVEN (Room Acoustics for Virtual ENvironments [8]). This combination enriches the measured RIR with information of the direction of arrival of the early reflections.

2.3. Reflection distance

The distance of a reflection in the target room (i.e. the virtual room) to a reflection in the reproduction room is a multi-dimensional problem. Additional to the difference in time there is the difference in angle of incidence, which might also have to be further divided in horizontal and elevation difference for differences in perceptual aspects. Furthermore, the intensity difference between the reflections is important. Figure 5 shows a three dimensional illustration of the problem, in which one room direction (the x-axis) is arbitrary chosen as time axis, to get an image of the constructed problem. The length

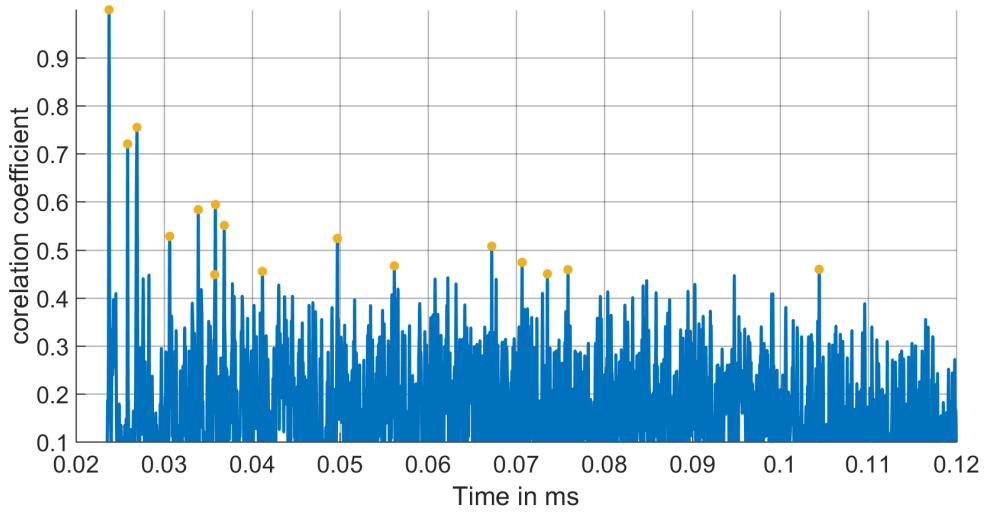

Figure 4: Cross-correlation function of the direct sound and the RIR. The yellow circles indicate a detected reflection.

of the vector is the intensity and the color indicates the order of reflection with black being direct sound, blue first order and green second order reflection.

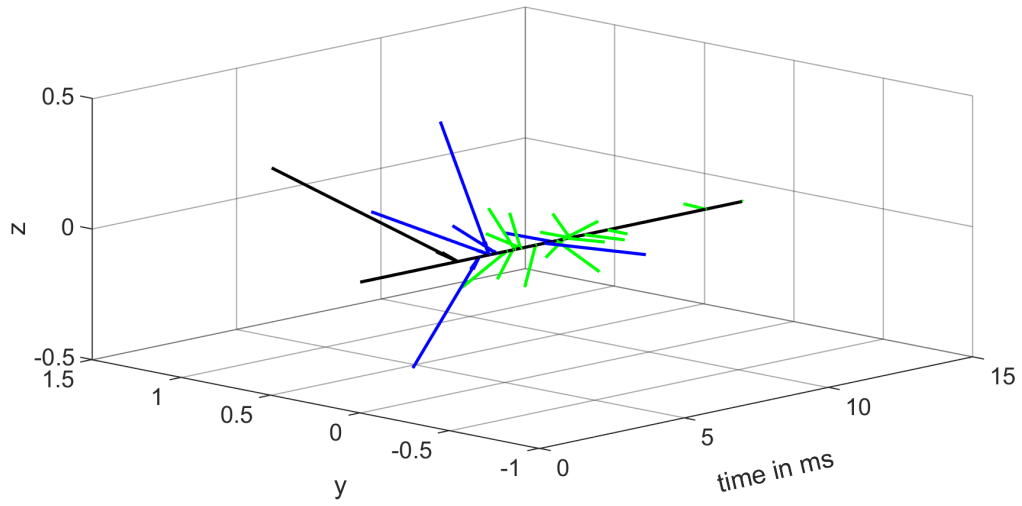

Figure 5: Visualization of the incoming reflections. The x-axis is arbitrary chosen as time axis to allow illustration of the multidimensional problem. Black indicates the direct sound, blue first order reflections and green second order reflections. The length of each vector is the energy of the reflection.

The euclidean distance between two multi-dimensional vectors can be expressed as:

$$d = \sqrt{\sum_{k=1}^{dim} (x_{k1} - x_{k2})^2} \tag{1}$$

As the difference of each aspects is perceived differently the absolute distance is a formulation of weighted parts time of arrival (w_t), direction of arrival (w_d) and energy

(w_I):

$$d = \sqrt{(w_t \cdot \Delta t)^2 + (w_d \cdot \Delta \overrightarrow{r})^2 + (w_I \cdot \overrightarrow{\Delta I})^2} \tag{2}$$

The weighting factors in d are perceptual motivated and need yet to be found. An overview to the perception of reflection can be found for example in [9]. The influence of a single side reflection on the perception when altered in time and intensity is stated. Late and strong reflections are perceived as echoes while image shifts occur for very early or loud reflections. Spatial impression decreases over time for one side reflection. The more complex problem of tone coloration is not quantitatively described. These factors are a good starting point for designing the weighting parameters for distance in time and intensity. Both seem to be time dependent and substituting very early reflections in the virtual scene has to be done very carefully to avoid image shifting. The echoing part seems to be less critical as early reflections are not considered to be late enough, especially not for one single reflection. Tone coloration should be decreased by having less reflections in the reproduction chain.

2.4. Integration of reproduction techniques

VBAP and encoded Ambisonics signals of a target acoustic scene can be generated using RAVEN. After decoding the Ambisonics signal using the loudspeaker set-up information in reproduction room the target loudspeaker signals are generated. These can directly be compared to the weighted RIR of each loudspeaker in the reproduction room and fed to the compensation algorithm described above. The implemented CTC system uses free-field HRTFs to compensate the crosstalk. Another way to compensate would be taking into account the reflections of the room which in previous work turned out to perform worse than the free-field approach [10]. Additional reflections in the reproduction room should be seen as reduction of the channel separation yet do have an perceptive effect of the perceived acoustics. Therefore, the target binaural impulse response of the virtual room is compared to the CTC-filtered BRIR of the loudspeaker in the reproduction room at the listener point.

3. SUBJECTIVE EVALUATION

The system at this stage is not realized as real-time implementation, especially due to the hybrid modeling of the reflections in the reproduction room which requires measurement of every point in the room to compensate listener movement. Consequently, artificial head recordings in the sweet-spot of the loudspeaker array were used for correct listening and controllable reproduction situations in the listening test. In contrast to simulations these procedure still contains all uncertainties like the loudspeakers frequency response, a non-perfect positioning of the loudspeakers and listener etc.. Anyhow, there is still the downside of using a non-individual HRTF set which might lead to less localization performance. As we are testing both systems with the same artificial head, the influence is assumed to be negligible. During the listening test, three different reproduction methods (Higher Order Ambisonics, crosstalk-cancellation (CTC) and Vector-Base Amplitude Panning (VBAP [11])) were tested. The compensation does only affect the early reflections and will neither influence the direct sound nor the late reverberant part of the reproduced RIR. To evaluate the perceptive part of the compensation a measure is needed to rate only the effect of early reflections. Additionally,

to distinguish between very subtle differences a direct comparison must be possible. These two factors led to the choice of the apparent source width as direct comparable measure. Additional reflections will de-correlate the signals at both ears and increase the ASW, therefore the compensation is supposed to decrease the perception.

3.1. Listening test design

The listening test was designed as direct comparison between two stimuli: With and without compensation. The compensation method is assumed to depend on the reproduction methods and the room impulse responses used. To reduce the duration of the listening test the reproduction scenario was set to be fixed to the investigated room in section 2.1. The target RIR was altered by using two different rooms and six different virtual sound source positions, the same in each room. One target room was the used reproduction room itself, the other one was a $12m$ x $10m$ x $4.5m$ room with a reverberation time of about 1.5s. The virtual sound source positions were all located on the right hemisphere due to right ear advantage [12]. Four of them in the horizontal plane at 180°, 240°, 300° and 340°. For the latter horizontal direction two additional sources elevated by 30° and 60° from the horizontal plane were used. The total number of different Stimuli is then 36 which were repeated in eight randomized blocks. The stimulus contained three $300ms$ white noise pulses divided by $200ms$ pauses.

For the binaural synthesis and the CTC reproduction free-field measured HRTFs of the artificial head are used, the CTC calculation was done using four loudspeaker and the calculation suggested by Masiero [13]. Ambisonics was decoded second order with r_e-max and two virtual loudspeaker in north and south pole to avoid an instable decoder matrix. The order of the playback of binaural synthesis and the recording of the loudspeaker reproduction was randomized. The listening test was conducted in an acoustically treated listening booth using a Sennheiser HD650 with a total number of 34 subjects. The average duration of the test was 33 minutes.

3.2. Results

The results of the test can be found in figure 6. The test only asked for a perceivable change of the ASW, not for the perceived ASW as such. The x-axis indicates the choices made with '0' being a decreased perceived ASW during the uncompensated playback and '1' for the compensated one and is therefore a measure of how often the compensation was perceived as decreasing the ASW. For each variable the mean value and the 95% confidence interval is plotted. The positions are indicated with elevation as first value (with 0° being the frontal direction) and the azimuth angle as second. The results show that for all cases the tendency is towards the compensated playback. The confidence interval is not intersecting the guessing rate of 0.5. A Shapiro-Wilk-Test results normally distributed data. For the significance analysis a two-sample t-test was used. The results for significances smaller than 0.05 are shown in table 1. It should be noted that only the performance of the compensation is analyzed not the overall perceived ASW. Consequently, two significantly different situations can be perceived with the same ASW but different effectiveness of the compensation.

Variable 1	Variable 2	p-value
CTC	VBAP	0.0162
Pos1	Pos2	0.0389
Pos1	Pos3	0.0301
Pos2	Pos5	0.0001
Pos2	Pos6	0.0005
Pos3	Pos5	0.0001
Pos3	Pos6	0.0005

Table 1: The table shows significant differences found in the tested variables in the listening test as shown in figure 6

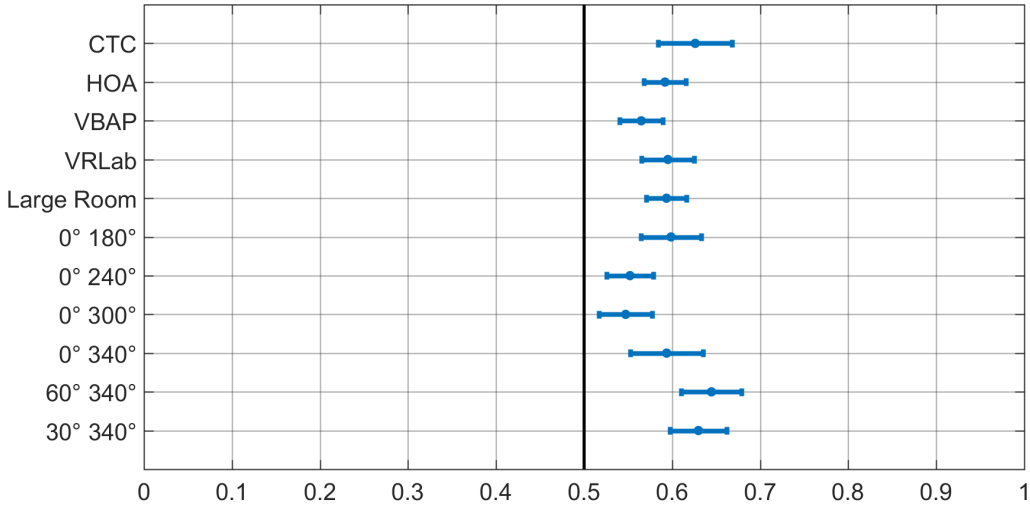

Figure 6: Mean values and 95% confidence intervals of the listening test results. On the y-axis the different variables are plotted. The positions are indicated with elevation angle first and then azimuth angle. On the x-axis '0' indicates all perceived ASW were narrower during the uncompensated playback, '1' during the compensated playback.

3.3. Discussion

The results indicate an effect established by the compensation. All mean values are above 0.5 and the confidence interval do not overlap this threshold. Yet, the results from the single variables differ from the expectations. For CTC the compensation method is theoretically not correct but seems to perform significantly better than for VBAP. Additionally, a better performance for a larger room would be expected which can not be seen in the results. A longer reverberation time would allow for more reflections to be canceled out and therefore a bigger change in the compensation. Yet, the matched room situation can lead to a total vanishing of all reflections and therefore results in the same findings. For further analysis a virtual room smaller than the reproduction room should be tested. One interesting aspect is that the elevation of the source at an azimuth angle of 340° does not have a significantly effect on the performance. For the virtual sources located more to the side the compensation performance seems to be weaker. As the perceived ASW is linked to the lateral energy this can be explained as the ratio from

frontal to side energy is not changed to much whereas for front and back direction the frontal energy should stay the same and the lateral energy be less.

4. CONCLUSION

The paper presents an approach for compensation of a reproduction room which novelty lies in the integration of the reproduction room into the virtual room. The implementation of reflection handling is presented and its proof of concept demonstrated using a real room. To evaluate the approach a listening test was conduct that shows that the approach does effect the perceived apparent source width of a virtual source. Yet, the differences of performance for the tested variables were unexpected as a larger target room does not necessarily lead to a better performance of the Algorithm.

The approach uses input parameters that are perceptual. More quantitative perceptual data is needed to optimize the algorithm and to evaluate whether the euclidean norm is the most suitable distance and weighting measure for the multi-dimension reflection spacing. Further evaluation is in progress for objective measures in terms of inter-aural cross-correlation (IACC) and lateral energy. Additionally, the effect of the correlation of acoustic parameters in the target room and the reproduction room on the performance of the compensation will be investigated in objective measurements and subjective listening tests using the parameters suggested in the SAQI [14] and RAQI [15].

5. ACKNOWLEDGEMENTS

The authors like to thank Markus Voth whose master thesis contributed to this paper and all participants who volunteered in the listening test. For the implementation, measurements and analysis of the data the ITA-Toolbox was used [16].

This research was financed by the Head Genuit Foundation under the project ID P-16/4-W.

6. REFERENCES

[1] J. Grosse and S. van de Par. Perceptually accurate reproduction of recorded sound fields in a reverberant room using spatially distributed loudspeakers. *IEEE Journal of Selected Topics in Signal Processing*, 9(5):867–880, 2015.

[2] S. Spors, H. Buchner, and R. Rabenstein. Eigenspace adaptive filtering for efficient pre-equalization of acoustic mimo systems. In *2006 14th European Signal Processing Conference*, pages 1–5. IEEE, 2006.

[3] J. J. López, A. González, and L. Fuster. Room compensation in wave field synthesis by means of multichannel inversion. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2005.*, pages 146–149. IEEE, 2005.

[4] D. S. Talagala, W. Zhang, and T. D. Abhayapala. Efficient multi-channel adaptive room compensation for spatial soundfield reproduction using a modal decomposition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22(10):1522–1532, 2014.

[5] S. Pelzer and M. Vorländer. Inversion of a room acoustics model for the determination of acoustical surface properties in enclosed spaces. In *Proceedings of Meetings on Acoustics ICA2013*, volume 19, page 015115. ASA, 2013.

[6] M. Vorländer. *Auralization: fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality*. Springer Science & Business Media, 2007.

[7] A. M. Noxon. Correlation detection of early reflections. In *Audio Engineering Society Conference: 11th International Conference: Test & Measurement*. Audio Engineering Society, 1992.

[8] D. Schröder. *Physically based real-time auralization of interactive virtual environments*, volume 11. Logos Verlag Berlin GmbH, 2011.

[9] M. Barron. *Auditorium acoustics and architectural design*. Routledge, 2009.

[10] M. Kohnen, J. Stienen, L. Aspöck, and M. Vorländer. Performance evaluation of a dynamic crosstalk-cancellation system with compensation of early reflections. In *Audio Engineering Society Conference: 2016 AES International Conference on Sound Field Control*. Audio Engineering Society, 2016.

[11] V. Pulkki. Virtual sound source positioning using vector base amplitude panning. *Journal of the audio engineering society*, 45(6):456–466, 1997.

[12] D. S. Emmerich, J. Harris, W. S. Brown, and S. P. Springer. The relationship between auditory sensitivity and ear asymmetry on a dichotic listening task. *Neuropsychologia*, 26(1):133–143, 1988.

[13] B. S. Masiero. *Individualized binaural technology: measurement, equalization and perceptual evaluation*, volume 13. Logos Verlag Berlin GmbH, 2012.

[14] A. Lindau, V. Erbes, S. Lepa, H. Maempel, F. Brinkman, and S. Weinzierl. A spatial audio quality inventory (saqi). *Acta Acustica united with Acustica*, 100(5):984–994, 2014.

[15] S. Weinzierl, S. Lepa, and D. Ackermann. A measuring instrument for the auditory perception of rooms: The room acoustical quality inventory (raqi). *The Journal of the Acoustical Society of America*, 144(3):1245–1257, 2018.

[16] P. Dietrich, M. Guski, M. Pollow, B. Masiero, M. Müller-Trapet, R. Scharrer, and M. Vorländer. Ita-toolbox–an open source matlab toolbox for acousticians. *Fortschritte der Akustik–DAGA*, pages 151–152, 2012.