THE VIRTUAL CONCERT HALL - A RESEARCH TOOL FOR THE EXPERIMENTAL INVESTIGATION OF AUDIOVISUAL ROOM PERCEPTION

HANS-JOACHIM MAEMPEL AND MICHAEL HORN
Department for Acoustics and Music Technology | Studio Facilities and IT
Staatliches Institut für Musikforschung Preußischer Kulturbesitz,
Germany

Abstract

The Virtual Concert Hall is a virtual environment that has been specifically designed for the optoacoustic simulation of performance rooms. As a tool for experimental research, its design is derived from particular methodological demands including harvesting comparable acoustical and optical information, dissociating the space and content of multisensory events, varying optical and acoustical room properties in a mutually independent manner, while also providing a full set of stimulus cues. The system features 3D sound and vision by applying dynamic binaural synthesis and a 161° stereoscopic projection on a cylindrical screen. Room simulation data were acquired in situ in the form of orientational binaural room impulse responses and stereoscopic panoramic images. Music and speech performances were recorded acoustically in an anechoic room and optically in a greenbox studio and inserted into the virtual rooms. The Virtual Concert Hall provides nearly all perceptually relevant acoustical and optical cues, enabling experiments on the audiovisual perception of optoacoustically conflicting rooms under rich-cue conditions.

Keywords: performance room, concert hall, architectural acoustics, simulation, stereoscopy, binaural synthesis, audio-visual perception

CONTACT
Hans-Joachim Maempe
Maempel@sim.spk-berlin.de

1 Introduction

In recent years, many virtual environments (VEs) have been developed which provide both sound and vision. Most VEs (Keyes, 2016; Weissig 2017) are used for the reproduction of documentary, artistic or entertaining content which is, as a rule, technically designed in accordance with the principles of media aesthetics. Other VEs are specifically designed for scientific purposes and are used for the investigation of human-machine interaction (Pfeiffer, 2011), quality assessments (Iljazovic et al., 2012; Permentier, 2013), media aesthetics (Hendrickx et al., 2015) and multimodal perception (Bolanos & Pulkki, 2012; McArthur, 2016), as well as for the visualisation of scientific data and models in the fields of engineering, archaeology and medicine (Barnes, 2013; Blewett, 2014), to mention just a few.

Within these VEs, pictures or videos are generally projected on screens or displayed on monitors and stereoscopic viewing is mostly supported. Planners strive for high spatial resolutions and large fields of view by combining several projectors or monitors. In most instances the sound is reproduced according to Vector Base Amplitude Panning (Pulkki, 1997) or various multichannel formats via loudspeakers.

Three of the above-mentioned VEs (McArthur, 2016; Parmentier, 2013; Weissig 2017) apply methods of virtual acoustics; which is to say, they reproduce natural sound fields or ear signals in a near-physical-correct manner. There is therefore no need to rely on psychoacoustic phenomena such as summing localisation. However, every loudspeaker-based sound reproduction system, even when capable of performing sound field synthesis, raises the problem of exciting the real room it is installed in, i.e. the problem of superimposed room acoustics. Only the VE described by McArthur (2016) avoids this shortcoming by using a headphone-based reproduction, and applies at the same time a method of virtual acoustics (dynamic binaural synthesis). Collectively, it is not always clear whether the reproduced acoustic content is data-based or numerically-modeled, to what extent it has been manipulated, and how much background noise is produced by the technical equipment in each case.

When addressing specific research questions, the properties of both the VE and the reproduced content should ideally be determined in accordance with methodological considerations. The experimental investigation of the perceptual effects of optoacoustic properties of performance rooms makes certain methodologically-founded demands on the rooms serving as test stimuli. This is particularly the case when testing the extent to which acoustical and optical information each contribute to auditory, visual, and audiovisual features (Maempel, 2012). To ensure coherence in such investigations, the acoustical and optical properties of each room must (a) be physically and experientially congruent, (b) be mutually independently variable, and (c) include all perceptually relevant cues without any bias (full cue condition) or almost all perceptually relevant cues without

major biases (rich cue condition). Moreover, (d) it may be required to reproduce the same artistic performances within different virtual rooms.

Applying real rooms as test stimuli does not meet (b) and (d) means that simulated rooms must be applied instead. Numerical simulations on the basis of computer-aided designed room models are, however, not appropriate, because they neither meet the criterion of empirical congruence (a) nor feature a high level of acoustical and optical detail in general (c). Hence, data-based models must be applied. Only by acquiring the acoustical and optical room properties in situ can we guarantee the optoacoustic congruence of the respective room simulation: which is to say, the room 'sounds as it looks' in an ecologically valid manner.

In order to maintain perceptually relevant cues, room data must be transferred into the laboratory using a transmission system that meets demanding technical criteria: a geometrically correct display, a large field of view (FOV) and of listening, three-dimensional acoustical and optical reproduction, spatial and temporal resolution below perceptual threshold, linear acoustical transmission, and near-total absence of interfering light and sound. Additionally, maintaining the ecological validity places certain requirements upon the performances used as stimuli. They must feature a high artistic quality, and respondents must consider it plausible that they would take place within the rooms reproduced.

As older VEs do not meet this specific combination of demands, the authors have designed a new and specialised VE, as initially proposed by Maempel & Lindau (2013). Furthermore, as this VE is primarily intended to simulate musical content and appropriate performance rooms, we call it the Virtual Concert Hall (VCH). The present article describes comprehensively the technical concept, the selection and acquisition of the real rooms to be simulated, the selection of the artistic content, the recording of its performance, the integration of rooms and performances, the properties of the optoacoustic reproduction system, the technical collection of the perceptual data, and the resulting features of the VCH.

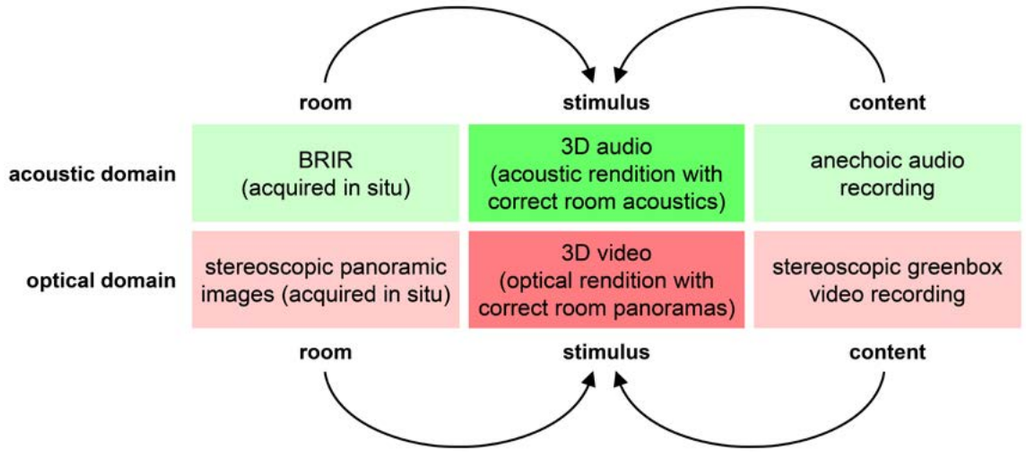


Figure 1: Technical concept. The simulation requires a mutual independence of 1st the acoustical and the optical domain, and 2nd the room and the artistic performance.

2 Technical Concept

Currently, there is no loudspeaker-based technique of virtual acoustics that demonstrably provides a sufficiently high plausibility of simulation. Sophisticated headphone-based techniques, on the other hand, are capable of simulating rooms with a considerably high degree of plausibility. That is to say, test subjects are not able to reliably identify simulated rooms as such, though most subjects notice minor timbre differences between real and simulated rooms (Lindau & Weinzierl, 2012). Hence, we decided in favour of a headphone-based technique utilising dynamic binaural synthesis (i.e. convolving anechoic sound recordings with head-orientational binaural room impulse responses). This has yielded three-dimensional acoustic renditions with correct room acoustics at a certain listener's position (figure 1, green). The synthesis system is interactive. Rather than utilising acoustical stimuli that have been completed in advance, it responds to the momentary head orientation of the listener, merging room impulse responses and performance information in real-time during playback.

Unfortunately, a highly-plausible optical simulation cannot currently be achieved due to the subjects' awareness of the screens or displays that must be used. A cost-efficient method for transmitting the detailed optical properties of performance rooms can be attained through the use of stereoscopic panoramic images. An invariant artistic performance can then be integrated into the imaged rooms using the procedure of chroma key compositing. This involves the recording of a stereoscopic performance video in a greenbox, after which the performer's images are extracted and graphically embedded in the room image (figure 1, red). In doing so, the optical stimuli can be completely prepared in advance. The concept allows for presenting identical renditions in different rooms, and varying their optical and acoustical properties independently .

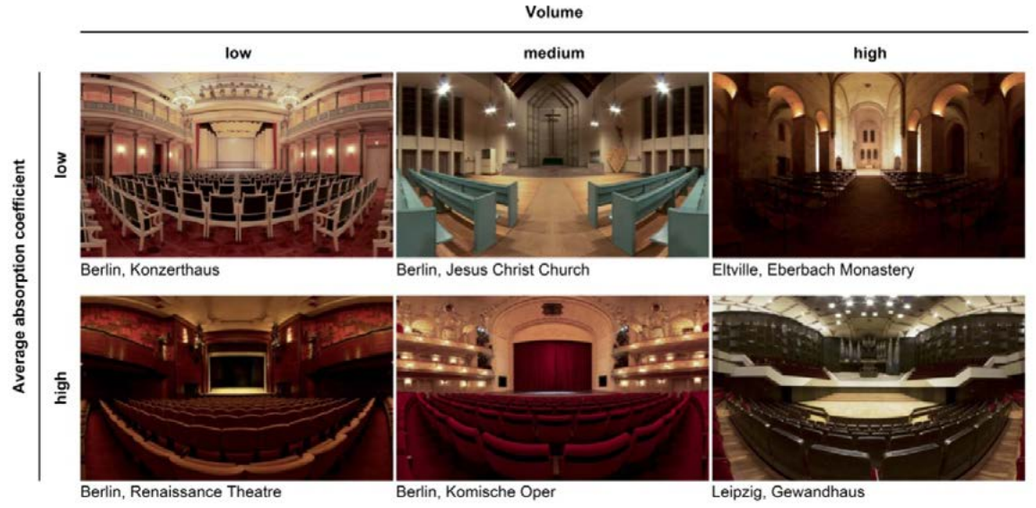


Figure 2. Selected performance spaces, differentiated by volume and average absorption coefficient (shown in equirectangular projection geometry).

3 Rooms

3.1 Room selection

Six concert halls adequate for musical and spoken performance were selected, each of which differs in volume and average acoustical absorption coefficient (figure 2). Taking geometrical measurements (distances, angles), acoustic measurements (reverberation time, ambient noise level) and listening professionally (however subjectively) to provisional performances during on-site visits allowed for the assessment of which concert halls best met the selection criteria.

3.2 Room models

Based on the geometrical measurements, interior space models were built using the software *SketchUp* (by Google), and these models' volumes and surface areas were calculated using the plugin *Volume Calculator* (by TGI). In conjunction with accurate measurements of the reverberation time (see 3.4), these values allowed for the assessment of the average absorption coefficient. In this way, the selection criteria were quantified and confirmed ex post.

3.3 Receiver and source positions

On location, an optimum receiver position was defined for each room. Criteria were good speech intelligibility, accurate perceptibility of acoustic room properties, and accurate perceptibility of optical room properties such as the visibility of the ceiling height. The principle of maintaining constant relative distances (in terms of multiples of the critical distance to the sound sources on stage) was attempted. However, it proved impossible to meet the above-mentioned criteria, so this principle was abandoned.

To simulate a string quartet, four artificial sound sources (loudspeakers) were placed on stage in a semi-circle which was 3 m in diameter. The speakers were arranged in accordance with common musical practice, with the sound source correlating to the 1st violin on the left, followed by the sources correlating to the 2nd violin, viola, and cello (in clockwise order). The sound source representing the vocalist was positioned in the center of the circle.

On the basis of anthropometrical data (Deutsches Institut für Normung e.V., 2005, p. 29), the average of common seat heights (43 cm), and the common posture of stringed bowed instruments the height of the acoustic center of the instruments was determined to be 151 cm for vocals, 108 cm for violins and viola, and 76 cm for violoncello (each valid for an unisex player). The height of the interaural center of a prototypical seating unisex listener was determined to be 120 cm.

3.4 Acoustical rooms: acquisition of BRIRs

The plausible representation of natural sound sources by electroacoustic means is still at an early stage of development (Pollow & Behler, 2009; Zotter, 2009). However, through analysis of Pollow & Masiero's (2009) measurements of musical instrument directivity, it was determined that – on average – vocalists and string instruments feature a rather small directivity index. Thus, as a convenient substitute for natural sound sources wide-angle reinforcement speakers of type QSC K8 (105° conical opening angle) were used (c.f. Meyer, 1992, for an extended discussion). The K8 speakers are capable of delivering large sound pressure levels (127 dB SPL peak) over a wide frequency range (61 Hz – 20 kHz, -10dB). All speakers were additionally equalized to a flat on-axis frequency response. In each hall, the loudspeakers were placed at the virtual artists' positions, and bass-emphasised linear sine sweeps with FFT order 17-20 (depending on the specific reverberation time of a room), 44.1 kHz sampling rate, and 24 Bit word length were played back (figure 3).

Figure 3: Playback of excitation signals and recording of binaural room impulse responses using the head and torso simulator ***FABIAN***.

At the receiver position, measurement signals were acquired by the in-ear-microphones (DPA 4060) of the automated, motion-controlled head-and-torso-simulator (HATS) *FABIAN* (Lindau & Weinzierl, 2006; Lindau, Hohn & Weinzierl, 2007). In order to allow for a realistic auralisation responding seamlessly to head movements, for each sound source BRIRs were measured for different horizontal head orientations with a fine angular resolution (±80° range, 1° steps). BRIRs were calculated via non-circular spectral deconvolution of the acquired measurement signals and a reference spectrum derived via loop-back measuring the electroacoustical transfer chain.

To enable further analysis, standard room acoustic measures were taken in dependence on DIN EN ISO 3382-1 (Deutsches Institut für Normung e.V., 2009) for several acoustic transfer paths, i.e. RT30, EDT, BR, C50, STI, TS, C80, G, H, IACC, IACCearly, IACClate, LF, and LFC. When applicable, frequency dependent measures were first calculated (octave bands), and then averaged spatially and spectrally later on.

3.5 Optical rooms: acquisition of stereoscopic panoramic images

The acquisition of stereoscopic panoramas demands the introduction of interocular separation (Bourke, 1999). Instead of simultaneously using two spaced cameras, a single camera was used, first taking a photograph corresponding to the left eye, then shifting the camera's position with an L bracket before taking a photograph corresponding to the right eye. However, since a static camera arrangement does not allow for a consistent parallax for all sight angles possible in a panorama (figure 4), a pair of pictures had to be taken for each head orientation. The central vertical strips of these directional images were stitched together afterwards.

In order to cover the front hemisphere (particularly from the nadir to the zenith), a circular fisheye lens was used. To maximize the resolution of the mosaiced panoramic images, the composite images' vertical resolution had to be maximized. Hence,

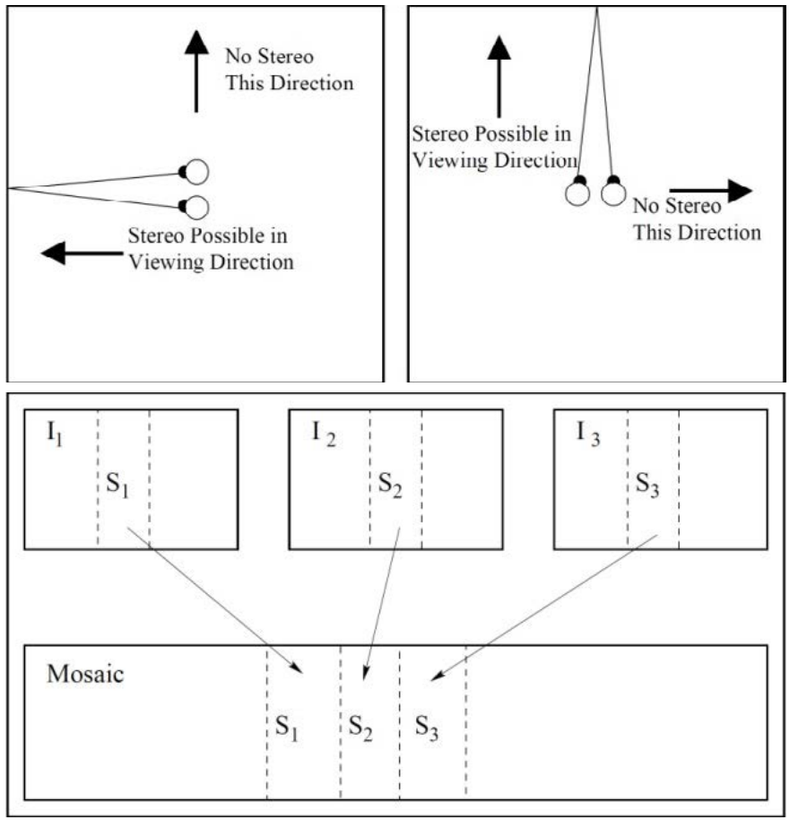


Figure 4: A consistent parallax requires the accordance of sight angle and head orientation, thus the "mosaicing" of the central vertical strips of directional images (Peleg & Ben-Ezra, 1999, p. 1395).

the upright format was chosen and mounted a full-frame fisheye lens (Sigma 8mm f3,5 *EX DG*) on a camera body featuring a smaller APS-C sensor (Canon *EOS 600D*). This yields a maximum vertical resolution of 5,184 pixels, thereby covering 176.6° vertical FOV. In order to rotate the optical recording system the rotatable neck hinge of the HATS *FABIAN* (cf. 3.4) was used. The nodal point of the camera lens system was adjusted using a slide bar (figure 5). After alignment by means of a laser pointer and water level, the setup automatically took images for all azimuthal head orientations with a step size of 1°.

Significant interocular errors with respect to size, vertical position, rotation and vergence did not occur due to the motor precision of the *FABIAN* robot. With regard to the effective FOV defined by the projection system (c.f. 6.2) the deviations of both the circular directional images and the stitched panoramas do not exceed the recommended value of 1/30 of the width of the picture (Herbig, n.d.).

The circular images were geometrically mapped onto an equirectangular projection geometry. Central vertical strips with 60×5400 pixels size (2°×180° equivalent) were cut and stitched together using the software *Panoramastudio 2* (by tshsoft). In

order to maintain the interocular offset consistent for all azimuthal angles, the stitching process was 'syn-located' by stacking the left eye strip and the right eye strip of the same direction in advance. The mosaicing produces a stereoscopic equirectangular full panorama of each room, arranged from top to bottom. In post-production, the left and right parts of the stereoscopic images were disconnected, each part was manually vertically aligned with reference to the image horizon, picture errors (e.g. annoying chairs) were retouched, and picture sizes were normalized. The resulting stereoscopic full panoramic images provide a consistent resolution of 10800×5400 pixels, i.e. 30 pixels per degree. Compared to the panoramic images produced by Baumbach (2009) who developed and applied the basic procedure, the images described here provide a higher resolution (factor 2.38), an increased sharpness, a better exposure with regard to the system dynamic range, and the absence of visible stitching artifacts.

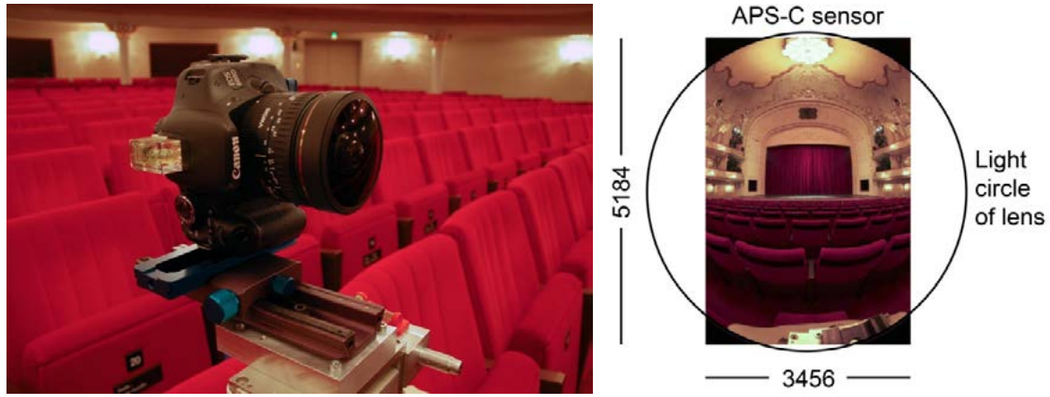


Figure 5: Canon ***EOS 600D*** with Sigma 8mm circular fisheye lens on L-bracket, slide bar and rotatable neck hinge of ***FABIAN*** in the Komische Oper Berlin. A laser pointer and a water level facilitated the alignment.

4 Content and performance

4.1 Selection of Content

The artistic content selected for experimental use comprised a musical work and a text appropriate for stage performance and capable of supporting the perceptibility of specific properties of the performing rooms. In the acoustic domain, this is generally achieved by a high dynamic range, a wide frequency spectrum, impulsivity, and sufficient pauses.

Works were sought which did not demand too much musical or literary expertise on the part of experiment participants. Accordingly, only European works of the 19th-20th century were considered. Among the performing musicians' concert repertoire the 2nd movement of Claude Debussy's String Quartet in g minor, op. 10 (composed 1893) met the criteria best. For the spoken performance, the first paragraph of the first elegy of Rainer Maria Rilke's *Duino Elegies* (published 1912) was chosen.

4.2 Acoustical performance: anechoic audio recording

To maximize ecological validity during future experiments, stimuli of high artistic quality were required. To this end, the music piece was played by a professional quartet, the text was performed by a professional actress, and the performances were professionally produced according to artistic criteria and recorded in the anechoic chamber (figure 6) of the Technische Universität Berlin (TUB). Each instrument and the voice were recorded individually and by not too closely applying one cardioid microphone. The four signals of the quartet were not summed; rather, each was recorded monophonically. Direct sound paths between the instruments were obstructed by acoustic partition walls. As a substitute for natural mutual listening, headphone monitoring with artificial reverberation was provided. The musicians performed simultaneously, i.e. no overdubbing was applied. The musical execution was not cleaned up unduly by editing in order to meet a degree of precision expected of live performance. The audio production yielded a four-track (music) and a one-track (speech) wave file with 44.1 kHz sampling frequency and 24 Bit word length.

4.3 Optical performance: Stereoscopic greenbox video recording

The optical content was recorded stereoscopically in a greenbox studio (figure 7) applying a stereoscopic television camera type Panasonic *AG-3DA1*. The video format was set to 1080/30p (Full HD) and the MPEG-4 AVC/H.264 codec was selected. The musicians and the actress, respectively, were placed in accordance to the previously defined source positions (c.f. 3.3) and performed with adequate expression while listening to their own anechoic recording (full playback).

The quartet realised a very good degree of optoacoustic synchrony using this technique. The nature of poetic recitation proved less amenable to capture, and although the actress coped masterfully with the problem, some video resampling had to be applied ex post at certain parts in order to ensure synchrony below perceptual level without changing the audio.

Three principal perspectives (specifically, vertical sight angle and distance) were derived from the room-specific source receiver arrangements (c.f. 3.3), and then arranged in view of a plausible integration of performances and rooms. Thus, the optical performances differ slightly between some halls. However, the acoustic performance is invariant.

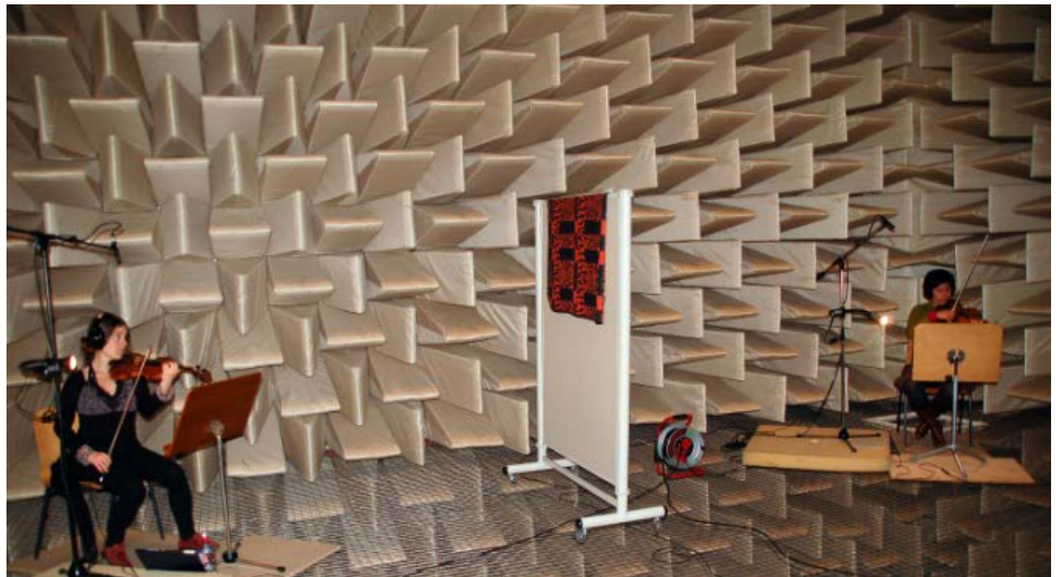
5 Stimuli

5.1 Acoustical Stimuli: dynamic binaural synthesis

The procedure of dynamic binaural synthesis was introduced as *Binaural Room Scanning* (*BRS*) by Horbach et al. (1999) using time variant convolution – i.e. exchanging BRIRs in response to head movements observed by a head tracker (figure 8).

Further improvements have subsequently been made in order to (1) provide sufficient spatial resolution (Lindau et al., 2008; Schultz et al., 2009), (2) compensate for spectral coloration (Erbes et al., 2012; Lindau & Brinkmann, 2012), (3) minimise system latency to a level below perceptual threshold (Lindau, 2009), and (4) reduce cross-fade artifacts and allow for individual adaption to the subjects' interaural time differences (Lindau et al., 2010). This has resulted in systems with an amazing degree of perceptual plausibility (Lindau & Weinzierl, 2012). In the present project, the respective real-time process was implemented in the reproduction system (c.f. 6.1).

The loudness differences between the rooms were preserved by adapting the respective level of auralisation based on measurements in each room with a SPL calibrated sound source.



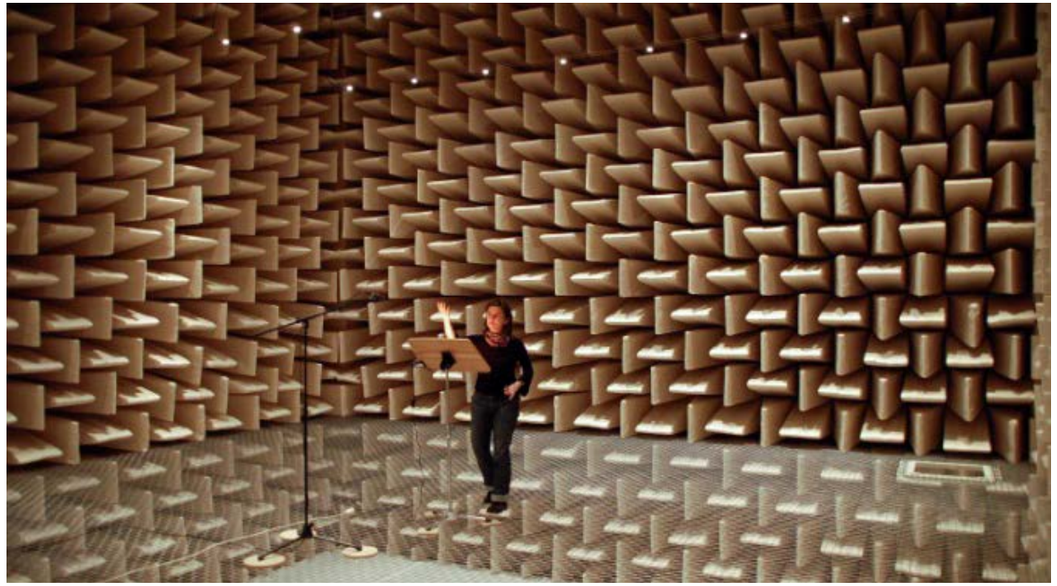


Figure 6: Anechoic audio recording of string quartet and voice.

Figure 7: Stereoscopic greenbox video recording of string quartet and voice.

5.2 Optical Stimuli: chroma key compositing

Since the stage of each hall was covered by an image area of 1920×1080 pixels, full HD images were cut out of the center of the panoramas. As the greenbox material provides a rectilinear geometry due to the Panasonic *AG-3DA1* camera's gnomonic lens, the cutout was mapped onto a rectilinear projection geometry. After adding deletable red markers at the stages indicating the positions of the quartet and the actress according to the previously defined positions (c.f. 3.3 and 3.4), the images served as backgrounds for image compositing.

By means of the software *After effects* (by Adobe) and the plug in *Keylight* (by The Foundry), the quartet and the actress were chroma-keyed, scaled, and added to the background images frame by frame. Since the objects are correctly sized, chromatically decontaminated, and cast shadows, the rendered compositions look amazingly plausible.

The composed frames were remapped onto an equirectangular projection geometry and set back into the panorama seamlessly (figure 9). The panoramic scene frames were mapped onto both a cylindrical and a rectilinear projection geometry. They were then cropped and rescaled in order to fit the horizontal and vertical FOV of each projection system (c.f. 6.2). These single frames were then joined and – together with the synchronous audio streams (or stream) – coded, applying the H.264 codec, thereby completing the optical stimuli.

6 Stimulus reproduction

6.1 Acoustic stimulus reproduction: BKsystem – system for binaural headphone playback

As indicated in 5.1, there are numerous requirements for the transparent binaural reproduction of virtual acoustic environments (Lindau & Weinzierl, 2012). Specific requirements attach to headphone-based reproduction. A suitable headphone system will require a frequency response which is tolerant towards repositioning and morphology of test participants. Furthermore, the headset should comply with free air equivalent coupling (FEC) criterion (Møller et al., 1995) as far as possible, i.e. approach the acoustic impedance of free air as seen from the ear canal entrances. Finally, the headset should be combinable with other devices such as insert microphones, 3D-glasses or head tracking sensors.

The system which currently best matches these criteria is the *BKsystem* of the TUB's audio communication group, an internal development of the research unit *SEACEN* (Erbes et al., 2012). The system comprises the DSP-driven power amplifier *BKamp* and the extraaural headset *BK211*, which is prepared for mounting a head tracking sensor. Its noise is below auditory threshold, crosstalk attenuation is greater than 23dB, and total harmonic distortion is good above 200Hz. The DSP provides IIR filtering which is applied for headphone linearisation. The rendering computer compensates the Headphone Transfer Function (HpTF) based on FIR filters. The HpTF may be fitted quite well to a linear target response applying parametrical filter regularised LMS-inversion. Compared to alternative extraaural and circumaural headphones, the overall irregularity of the linearised HpTFs is considerably reduced.

In view of different experimental purposes, two simulators were installed (cf. 6.2): A large VCH at the TUB, and a small one at Staatliches Institut für Musikforschung (SIM). The dynamic binaural synthesis is performed at the TUB using the software *fwonder* (Lindau et al., 2007) and a Polhemus Fastrak® head tracking system, at the SIM using the software *SoundScape Renderer* (Geier & Spors, 2013) and a Polhemus Patriot™ head tracking system. The total background noise of the system at the

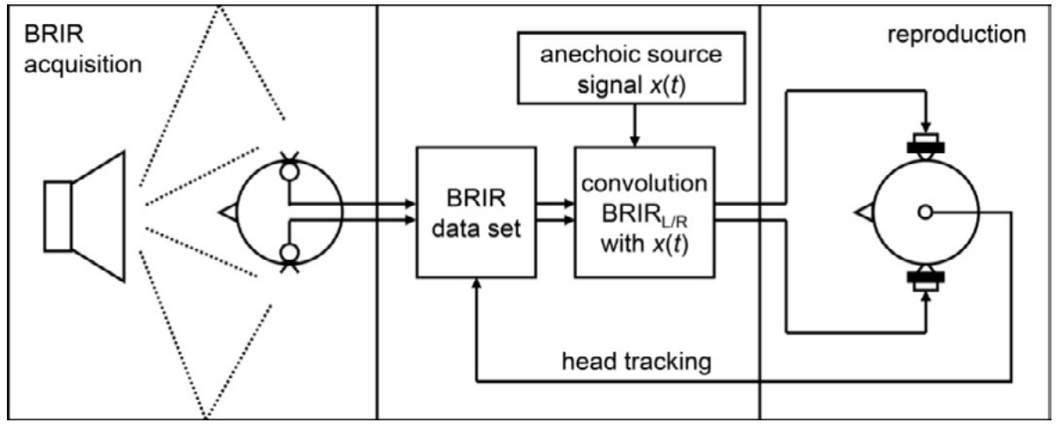


Figure 8: Principle of dynamic binaural synthesis.

Figure 9: String quartet in the Gewandhaus as a result of chroma key compositing (equirectangular projection geometry).

listener's position accounts just for 32 dB(A) at TUB and 21 dB(A) at SIM, respectively, and this could be attributed to the basic passive acoustic measures and the location of computers in separate machine rooms.

6.2 Optical stimulus reproduction: projection systems

In order to maximise immersive potential, a projection system providing stereoscopy, a large horizontal and vertical FOV, and a high resolution was required (cf. 1). Due to a limited budget, active stereoscopy was applied: Five home-cinema 3D Full HD projectors (EPSON *EH-TW9000W*) were installed providing overlapping images in upright format. Geometric warping and edge-blending were done using the software *Pixelwarp* (by Pixelwix). The panoramic 3D videos are projected on a 180° cylindrical hard screen with 5 m diameter and 3,22 m overall height which is permanently installed at the TUB.

Figure 10: *BKsystem*, comprising the extraaural headset *BK211* and the DSP-driven power amplifier *BKamp* (Erbes et al., 2012).

The projection provides a field of view of 161° horizontally and 56° vertically (figure 11, above), thus enabling test participants to look at both the side walls and the ceiling of the displayed concert halls (figure 12). It also provides an effective physical resolution (without loss by warping and overlap) of 4812 pixels horizontally and 1800 pixels vertically. This exceeds the resolution of a 4k projection and corresponds to 2.67 times the average human eye's angular resolution which, however, depends on the brightness and the contrast of the optical stimulus as well as on the age of the subjects (Adams et al., 1988), and which is about 0.8 arc minutes (equivalent to -0.1 logMAR and 6/4.8 acuity) for the age class 50-59 years under optimum conditions (Elliott et al., 1995). However, due to limitations in the capacity of the video signal chain, for the time being the vertical resolution mentioned above must be halved. The projection system is controlled by a computer based on an Intel *Core i7* 970 3.2GHz 6 core processor, 4 GB RAM, and a Powercolour *HD 7870 Eyefinity 6* graphics board. The playback of the audio-video files is done by means of the *VLC media player* (by VideoLAN organisation).

Experiments on distance perception in particular do not necessarily require a large field of view (Creem-Regehr et al., 2005; Knapp & Loomis, 2005). A better angular resolution near perceptual threshold might, however, be required. Experimentation has eliminated resolution as a potential factor in distance underestimation as it occurs in VEs (Bruder et al., 2016, pp. 9-14). Nevertheless, regarding stereopsis, "*the angular* resolution of pixels on a screen acts as an artificial cutoff to the theoretical capabilities of human vision" in principle (Bruder et al., 2016, p. 3). Therefore the second, small VCH was installed at the SIM (cf. 6.1), wherein the optical stimuli are reproduced by means of an 85" plane monitor (Samsung *UE85JU7090*) with a spatial resolution of 3840 horizontally and 2160 pixels vertically (ultra high definition) that features also active stereoscopy. Given a viewing distance of 1.4 m, the angular resolution amounts to 1.2 arc minutes corresponding to 1.5 times the average human eye's angular resolution (see above). However, because the system under consideration displays photographs of real rooms instead of artificial graphics for clinical test purposes, the contrast between adjacent pixels is low. Thus, the human eye's minimum angular resolution does not really take effect, and single pixels were not reported to be visible.

For reasons stated above, the vertical resolution of the VCH at the SIM has to be halved for the time being. The display system is controlled by a computer based on an Intel *Core i7* 5820k 6x 3.30 GHz 6 core processor, 32 GB RAM, and a 4 GB MSI *GeForce GTX 960* graphics board.

7 Procedure control and data collecting

The Virtual Concert Hall is not only set up for the reproduction of the experimental stimuli but also (a) to the control of the presentation sequence according to the combinations of levels of the independent variables (Maempel, 2012), and (b) to the collection of the test participant's response data. In each VCH, this is done by a control computer applying *Pure Data* (*Pd*, core by Miller Puckette). The *Pd* patch selects the correct stimuli according to a prepared presentation sequence, sends Open Sound Control (OSC) messages to the player computer, and sends/receives OSC messages to/from the test participant's *iPad* (by Apple) running *TouchOSC* (by hexler.net). The questionnaire displayed on the *iPad* screen names item sets, item qualities and item labels. The test participant is asked to rate the stimuli using touch-sensitive sliders. The rating values are sent back to the control computer as OSC messages. The digital questionnaire covers auditory, visual and audiovisual qualities. In compliance with Berg & Rumsey (2001) and Rumsey (2002) the qualities are explicitly referred to specified objects: the artist(s), the room or the overall scene. Ratings of the overall loudness, the visual brightness of the ensemble and of the room, the distance of the ensemble, the room dimensions, the audiovisual matching, and the overall aesthetic impression are collected.

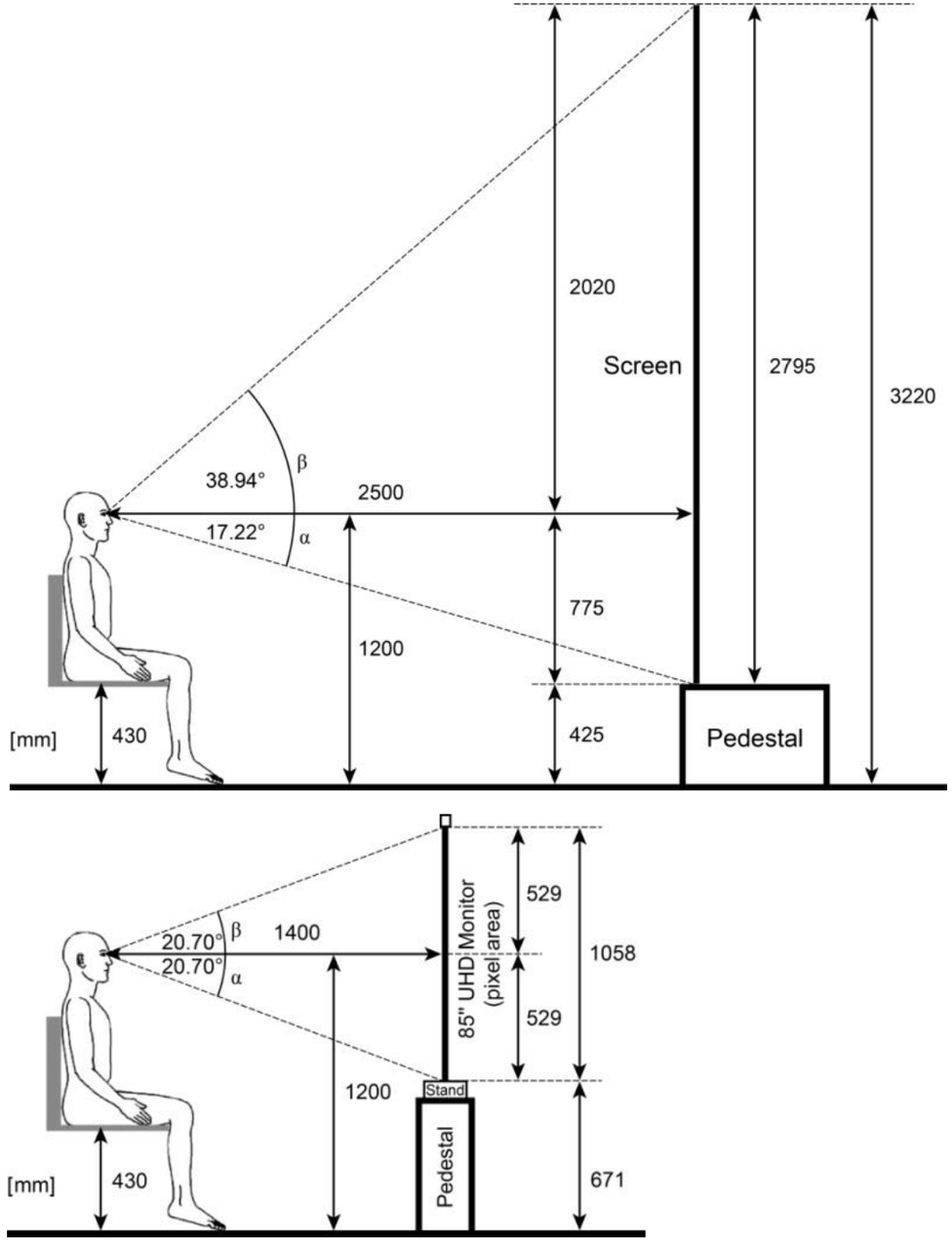


Figure 11: Side view of the optical projection setups at TUB (above) and SIM (below).

Figure 12: VCH at the TUB.

Figure 13: VCH at the SIM.

8 Features of the Virtual Concert Hall

As the VCH is based on professional artistic performances, room properties acquired in situ, and a state-of-the-art reproduction system, it provides a geometrically correct three-dimensional reproduction of concert and theatre performances under rich cue conditions. Thus, optoacoustically balanced experimental designs may be applied. The recipient may actively explore the virtual scene because the environment does not turn with the recipient's head movements. The optical simulation features a large FOV, and the acoustical simulation is fully spherical. Interfering optical and acoustical stimuli are minimal, and there are no superimposed room acoustics. Hence, the VCH may be regarded as an immersive environment. The six optical and acoustical performance rooms may be changed independently of each other and independently of the artistic performance, allowing for the presentation of optoacoustically conflicting stimuli, e.g. of the acoustical Gewandhaus and the optical Konzerthaus. The integration of both a procedure control and a digital questionnaire facilitates the application of the VCH as a research tool.

9 Outlook

Applying the VCH as a replacement for real concert halls requires the knowledge of potential perceptual biases caused by virtualisation. Comparing a real concert hall with its simulation will reveal respective perceptual effects and might validate the stimuli reproduced by means of the VCH to some extent. As a research tool, the VCH allows for the quantification of the proportional contribution of hearing and sight to auditory, visual, and audiovisual features. Additional rooms and content may be acquired in order to further investigate the audiovisual perception of egocentric distance, the role of room-acoustical parameters under optoacoustic conditions, the audiovisual perception of room shapes, the memorising and matching of acoustical and optical rooms, and the influence of recipients' expertise on audiovisual room perception. In a popular science context, smaller variants of the VCH might demonstrate both the relevance of the performance room and experimental methods of its investigation.

Acknowledgements

This work was carried out as a part of subproject 9, "Audio-visual perception of acoustical environments", within the framework of the *SEACEN* project, and funded by the German Research Foundation (DFG MA 4343/1-1). We would like to thank the directors and the technical staff of the named concert halls for their friendly cooperation, the *Berlin Budapest Quartet* (Dea Szücs, Éva Csermák, Itamar Ringel, Ditta Rohmann) and Ilka Teichmüller for their expressive performances, Annika Natus and Alexander Haßkerl for a perfect video shoot, and Shamir Ali-Khan for his creativity and patience during 3D compositing.

References

Adams, A. J., Wong, L. S., Wong, L., & Gould B. (1988). Visual Acuity Changes with Age: Some New Perspectives. *American Journal of Optometry and & Physiological Optics*, 65(5), 403-406.

Barnes, D. (2013). CAVE2. *Monash Immersive Visualisation Platform*, http://www.monash.edu/mivp/index.php?option=com_content&view=article&id=3&Itemid=104, access: 2017-02-07.

Baumbach, R. (2009). *Implementierung einer netzwerkfähigen und interaktiven stereoskopischen Visualisierungsumgebung.* Master's thesis. Berlin: Techn. Univ., audio comm. group.

Berg J., & Rumsey, F. (2001). Verification and correlation of attributes used for describing the spatial quality of reproduced sound. *AES 19th International Conference*, Article No. 1932.

Blewett, M., Pinkl. J., & Molle, B. D. (2014). Dynamic Audio Imaging in Radial Virtual Reality Environments. *137th AES Convention, Los Angeles, CA, USA*, Convention e-Brief 162.

Bolanos, J. G., & Pulkki, V. (2012). Immersive Audiovisual Environment with 3D audio playback. *132nd AES Convention, Budapest, Hungary*, Convention Paper 8604.

Bourke, P. (1999). *Calculating Stereo Pairs*, http://read.pudn.com/downloads30/sourcecode/windows/opengl/94939/stereo/Calculating%20Stereo%20Pairs.pdf, access: 2017-02-04.

Bruder, G., Argelaguet, F., Olivier, A.-H., & Lécuyer, A. (2016). CAVE Size Matters: Effects of Screen Distance and Parallax on Distance Estimation in Large Immersive Display Setups. *Presence – Teleoperators and Virtual Environments*, 25(1), 1-16.

Cree-Regehr, S. H., Willemsen, P., Gooch, A. A., & Thompson, W. B. (2005). The Influence of Restricted Viewing Conditions on Egocentric Distance Perception: Implications for Real and Virtual Indoor Environments. *Perception*, 34(2), 191-204.

Deutsches Institut für Normung e.V. (2005). *DIN 33402-2 – Ergonomie – Körpermaße des Menschen – Teil 2: Werte.* Berlin: Beuth.

Deutsches Institut für Normung e.V. (2009). *DIN EN ISO 3382-1 Akustik – Messung von Parametern der Raumakustik – Teil 1: Aufführungsräume.* Berlin: Beuth.

Elliott, D. B., Yang, K. C. H., & Whitaker, D. (1995). Visual Acuity Changes Throughout Adulthood in Normal, Healthy Eyes: Seeing Beyond 6/6. *Optometry and Vision Science*, 72(3), 186-191.

Erbes, V., Schultz, F., Lindau, A., & Weinzierl, S. (2012). An extraaural headphone system for optimized binaural reproduction. *DAGA 2012, Darmstadt*, 313-314.

Geier, M., Spors, S. (2013). Spatial Audio with the SoundScape Renderer. *27th Tonmeistertagung – VDT International Convention, November, 2012*.

Hendrickx, E., Paquier, M., & Koehl, V. (2015). Audiovisual Spatial Coherence for 2D and Stereoscopic-3D Movies. *Journal of the Audio Engineering Society*, 63(11), 889-899.

Herbig, G. P. (n.d.). *Leitlinien zur Herstellung von Stereobildern und Stereofilmen*, http://www.cosima-3d.de/download/leitlinien_stereofotografie.pdf, access: 2014-02-04.

Horbach, U., Karamustafaoglu, A., Pellegrini, R., Mackensen, P., & Theile, G. (1999). Design and applications of a data-based auralization system for surround sound. *106th AES Convention, München*, Preprint 4976.

Iljazovic, A., Leschka, F., Neugebauer, B., & Plogsties, J. (2012). The Influence of 2D and 3D Video Playback on the Perceived Quality of Spatial Audio Rendering for Headphones. *133rd AES Convention, San Franciso, CA, USA*, Convention Paper 8735.

Keyes, C. (2016). Designing a Laboratory for Immersive Arts. *140th AES Convention, Paris, France*, Convention e-Brief 263.

Knapp, J. M., & Loomis, J. M. (2005). Limited Field of View of Head-Mounted Displays Is Not the Cause of Distance Underestimation in Virtual Environments. *Presence – Teleoperators and Virtual Environments*, 13(5), 572-577.

Lindau, A. (2009). The Perception of System Latency in Dynamic Binaural Synthesis. *DAGA 2009, Rotterdam*, 1063-1066.

Lindau, A., & Brinkmann, F. (2012). Perceptual Evaluation of Headphone Compensation in Binaural Synthesis Based on Non-Individual Recordings. *Journal of the Audio Engineering Society*, 60(1/2), 54-62.

Lindau, A., Estrella, J., & Weinzierl, S. (2010). Individualization of dynamic binaural synthesis by real time manipulation of the ITD. *128th AES Convention, London*, Preprint 8088.

Lindau, A., Hohn, T., & Weinzierl, S. (2007). Binaural resynthesis for comparative studies of acoustical environments. *122th AES Convention*, *Vienna*, Preprint 7032.

Lindau, A., Maempel, H.-J., & Weinzierl, S. (2008). Minimum BRIR grid resolution for dynamic binaural synthesis. *Acoustics '08, Paris*, 3851-3856.

Lindau, A., & Weinzierl, S. (2006). FABIAN – An instrument for software-based measurement of binaural room impulse responses in multiple degrees of freedom. *24. Tonmeistertagung, Leipzig*.

Lindau, A., & Weinzierl, S. (2012). Assessing the Plausibility of Virtual Acoustic Environments. *Acta Acustica united with Acustica*, 98(5), 804-810.

Maempel, H.-J. (2012). Experiments on audio-visual room perception: a methodological discussion. *DAGA 2012, Darmstadt*, 783-784.

Maempel, H.-J., & Lindau, A. (2013). Opto-acoustic simulation of concert halls – a data-based approach. *27th VDT International Convention, Köln, 2012*, 293-309.

McArthur, A. (2016). Disparity in horizontal correspondence of sound and source positioning: The impact on spatial presence for cinematic VR. *AES Conference on Audio for Virtual and Augmented Reality, Los Angeles, CA, USA*.

Meyer, J. (1992). Klangliche Unterschiede zwischen realem Orchester und der Einspielung nachhallfreier Musik in Sälen. *17. Tonmeistertagung, Karlsruhe*, 144-154.

Møller, H., Jensen, C. B., Hammershøi, D., & Sørensen, M. F. (1995). Design Criteria for Headphones. *Journal of the Audio Engineering Society*, 43(4), 218-232.

Parmentier, M. (2013). A Special Room for 3D Audio and Ultra High Definition Video for Quality Assessment of Future TV. *135th AES Convention, New York, USA*, Convention eBrief 112.

Peleg, S., & Ben-Ezra, M. (1999). Stereo Panorama with a Single Camera. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'99)*, Vol. 1, 1395-1402. DOI: 10.1109/CVPR.1999.786969.

Pfeiffer, T. (2011). Virtual Reality Laboratory. *Arbeitsgruppe Wissensbasierte Systeme*, https://www.techfak.uni-bielefeld.de/ags/wbski/labor.html, access: 2017-02-07.

Pollow, M., & Behler, G. (2009). Variable Directivity for Platonic Sound Sources Based on Spherical Harmonics Optimization. *Acta Acustica united with Acustica*, 95(6), 1082-1092.

Pollow, M., & Masiero, B. (2009). Measuring Directivities of Natural Sound Sources with a Spherical Microphone Array. *Ambisonics Symposium 2009, Graz*.

Pulkki, V. (1997). Virtual source positioning using vector base amplitude panning. *Journal of the Audio Engineering Society*, 45(6), 456-466.

Rumsey, F. (2002). Spatial quality evaluation for reproduced sound: terminology, meaning, and a scene-based paradigm. *Journal of the Audio Engineering Society*, 50(9), 651-666.

Schultz, F., Lindau, A., & Weinzierl, S. (2009). Just Noticeable BRIR Grid Resolution for Lateral Head Movements. *DAGA 2009, Rotterdam*, 200-201.

Weissig, C. (2017). TiME Lab. *Heinrich Hertz Institute*, https://www.hhi.fraunhofer.de/index.php?id=199&L=1, access: 2017-02-08.

Zotter, F. (2009). *Analysis and Synthesis of Sound-Radiation with Spherical Arrays.* Doct. diss., Graz, Austria: Univ. of Music and Performing Arts.