TECNIACUSTICA VALENCIA · 2015

46º CONGRESO ESPAÑOL DE ACÚSTICA
ENCUENTRO IBÉRICO DE ACÚSTICA
EUROPEAN SYMPOSIUM ON VIRTUAL ACOUSTICS AND AMBISONICS

COMBINING SPATIAL REPRODUCTION TECHNIQUES FOR ROOM ACOUSTIC AURALIZATION

PACS: 43.6c, 43.38Md

Kohnen[1], Michael; Pelzer[1], Sönke; Aspöck[1], Lukas; Vorländer[1], Michael;
[1]Institute of Technical Acoustics, RWTH Aachen University, Kopernikusstr. 5, 52074 Aachen, Germany
mko@akustik.rwth-aachen.de

ABSTRACT

3-D audio reproduction techniques differ in their performance of delivering cues needed for spatial perception of a sound field. For room acoustics situations, human perception is based on different cues during the different phases of a room impulse response. In case of spatial reproduction based on simulated room impulse responses, the reproduction quality can be optimized by choosing different reproduction techniques for each part of the impulse response. The aim is to compensate the weaknesses and to benefit from merging their strengths. Listening tests were performed to evaluate aspects of the applied reproduction methods.

RESÚMEN

Técnicas de reproducción de audio 3-D difieren en su desempeño de la entrega de las señales necesarias para la percepción espacial de un campo de sonido. Para situaciones de acústica de salas, la percepción humana se basa en diferentes señales durante las diferentes fases de una respuesta al impulso de una habitación. En el caso de la reproducción espacial basada en las respuestas de impulso de la sala simuladas, la calidad de reproducción se puede optimizar mediante la elección de diferentes técnicas de reproducción para cada parte de la respuesta de impulso. El objetivo es compensar las debilidades y beneficiarse de la fusión de sus puntos fuertes. Se realizaron pruebas de escucha para evaluar los aspectos de los métodos de reproducción aplicadas.

1. INTRODUCTION

Nowadays visual **V**irtual **R**eality systems (VR-Systems) have entered consumer markets in form of **H**ead **M**ounted **D**isplays (HMDs) and 3-D cinema technology. In research & development CAVE systems exist that provide a realistic holographic environment. The acoustic feedback in these VR systems should be spatially rendered and reproduced in a similar realistic quality as the stereoscopic visual feedback. Users should be able to freely move and interact with the environment which requires an in-time calculation of the sound propagation using room acoustic simulation algorithms to determine the transfer path from the source to the receiver, the **R**oom **I**mpulse **R**esponse (RIR), and to allow for a realistic experience. The reproduction of the RIR can be done using binaural technology played backed over headphones. Achieving a good channel separation this technology lacks in providing bone-structure sound, especially for low frequencies. Furthermore the attachment of a device to the listener's body might decrease the feeling of immersion as the listener feels constricted. Besides this for multiuser application it is a visual disturbance and lastly wearing headphones might become uncomfortable over time. Considering

these aspects a loudspeaker based solution is investigated. This paper proposes an idea to use the limited human perception of different time sections of a RIR to symbiotically combine different reproduction techniques to benefit of their strengths while compensating their weaknesses.

2. STATE OF THE ART

Nowadays RIRs for virtual reality applications are synthesized using computer generated models. Two different methods can be applied to calculate the data. The first one is the deterministic image source model that delivers exact information about time, magnitude and direction of incoming reflections. The time needed to calculate reflections with this method increases exponentially with the number of reflections and the complexity of the room geometry [1]. A stochastic and less accurate method is the ray tracing approach. It allows for a stochastic determination of the incoming reflections in time and direction and requires less calculation time. Modern room acoustic simulation software uses a hybrid approach of both techniques, the image source model for early reflections and the ray tracing approach for later arriving reflections.

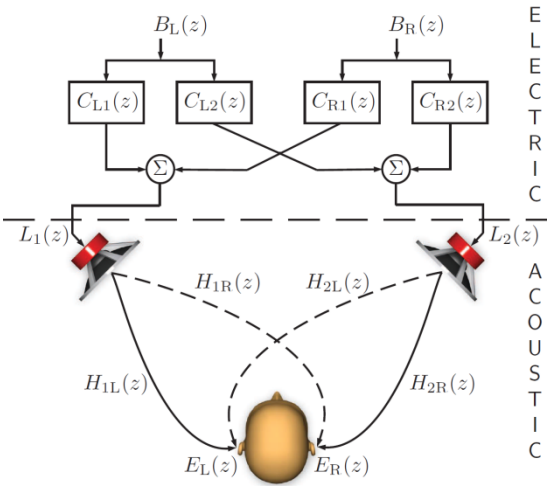

*Figure 1: Concept of CTC. The binaural input **B** is processed with filters **C** to the loudspeaker signals **L**. The signal then arrives the listeners ears over the transfer paths **H** where dotted lines should be cancelled out. For ideal conditions the ear signals **E** are time shifted version of the binaural input **B**. Figure taken from [5].*

Crosstalk Cancellation is a binaural reproduction technique invented by Schroeder and Atal [2]. Input signals are synthesized or recorded using the ear canal signals of an artificial head. Binaural reproduction then aims at presenting a stimulus at the listeners' ears and therefore requires a good channel separation of the two input signals between the left and the right ear. To avoid crosstalk, loudspeakers near the contralateral ear will play cancellation signals that remove the crosstalk however producing new crosstalk at a lower level at the ipsilateral ear. The aim is to eliminate all crosstalk and to produce a so called virtual headphone that allows for sufficient channel separation between both ears to present the binaural signal. This technique is able to reproduce sources in any direction and at any distance.

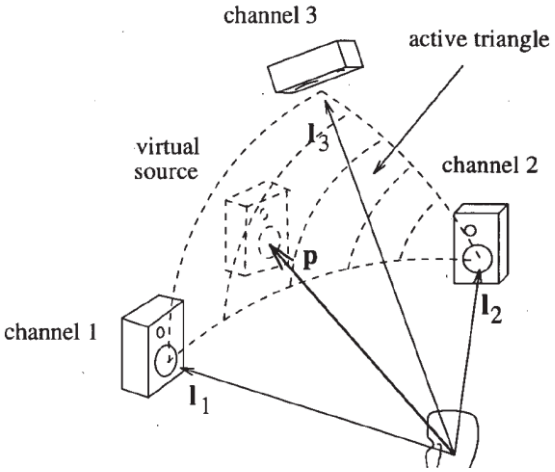

*Figure 2: Principle of VBAP. The loudspeaker vectors **l** are used to create the source vector **p**.*

Vector-base amplitude panning (VBAP) is the extension of the two loudspeaker stereo principle to a triangle and building a sphere with these triangles. Pulkki [4] invented VBAP on the principle of source summation using intensity panning. Virtual sources in any directions can be presented to a listener seated in this sphere. Information about distance can only be indicated by changes in sound pressure levels.

Ambisonics is a physical sound field reproduction technique invented by Gerzon [6]. The sound field at a specific point is either recorded with a special microphone or synthesized. It is then approximated using real-valued spherical harmonics. The spherical harmonics coefficients are then broadcasted and can be played back using the inverse of the spherical harmonics decomposition of the loudspeaker set-up. To enhance the sweet spot and to compensate for the listener that interferes with the sound field, different decoding strategies exist [7] which bring Ambisonics closer to intensity panning techniques. For the so called plane wave Ambisonics different source distances can only be indicated by a change of sound pressure level.

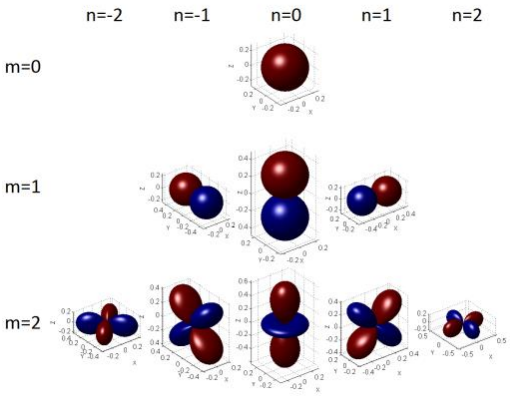

Figure 3: Real valued spherical harmonics basefunctions up to a truncation order of 2 as part of 2nd order Ambisonics.

Wave-field synthesis is based on the Huygens-Fresnel principle which states that each point of a propagating wave is the origin of new point source. These point sources are represented by loudspeakers which then reproduce the original wave front. The idea was depicted by Berkhout and de Vries [14]. While this technique has the advantage of not having a sweet spot, it requires a large number of loudspeakers to avoid spatial aliasing. Even for single user experiences a set-up that covers all spatial directions is unpractical and expensive. Therefore this technique is not considered to be part of the hybridization at this point of investigation.

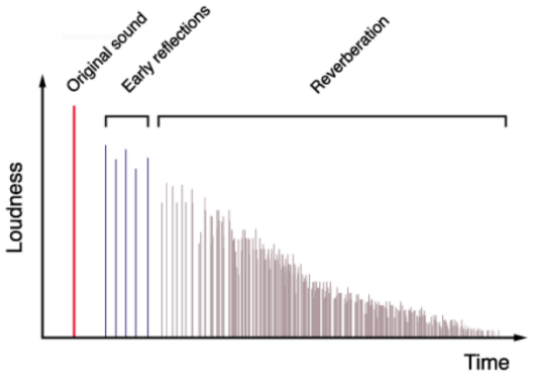

Figure 4: Separation of a RIR in time by different phases of perception

3. COMBINING REPRODUCTION MODULES

The human auditory system is limited in its perception of different cues of a RIR. The direct sound delivers information about position of the source. Early reflections contribute to the impression of loudness and size of the source [1]. The perception of reverberance and envelopment is mostly provided by reflections arriving after a certain transition time after which the reflections are considered to be diffuse [10] [11].

Reproduction technique	Strengths	Weaknesses
Transaural	Precise and easy localization Good readability	Poor realism and lack of immersion/envelopment
Ambisonics	Strong immersion and envelopment	Poor localization readability
Stereo / Panpot	Very precise localization	Lack of immersion/envelopment

Figure 5: Characteristics of different reproduction techniques. Taken from [13]

Spatial Audio reproduction techniques have different strengths and weaknesses. Gustavino et al. [13] showed that the weaknesses and strengths between Ambisonics and the other two

techniques are inverted. The findings shown in *Figure 5* illustrate that Ambisonics can provide a feeling of envelopment whereas VBAP and CTC are able to provide good localization. To achieve an optimal reproduction of a RIR the idea is to separate the RIR in time and to find the best suitable reproduction technique for each part of the RIR. As shown in *Figure 5* CTC and VBAP are suggested to provide more exact information of position and size of the sound source than Ambisonics does. Therefore CTC and VBAP are intended to be used for the presentation of direct sound and early reflections to deliver a good impression of position and size of the sound source whereas Ambisonics should be used for the late reverberant tail delivering a more authentic impression of listener envelopment and reverberance. The weaknesses of the reproduction techniques can be suppressed as they are not used for the other RIR parts. For this purpose, it is also convenient that the transition time between early reflections and late reverberation matches the time of the last audible image sources (e.g., for order 2) and the beginning of ray tracing results, as shown by Pelzer [3]. The results gathered by the used room acoustic software RAVEN [17] can be fetched separately and then directly be post processed. To ensure an inaudible transition between different reproduction techniques the time of arrival and the loudness of the different reproduction techniques have to be matched.

4. LISTENING TEST

The idea of combining different reproduction techniques aims at preserving the good localization properties of one reproduction technique while enhancing the impression of envelopment and reverberance using another technique. To investigate if localization performance and envelopment changes by combining techniques a listening test was conducted. CTC was implemented as a two loudspeaker CTC with simple regularization and windowing as described in [5]. Plane wave Ambisonics with a truncation order of four was implemented and decoded with $|r_e|$-*max* [7] decoding. Additionally two virtual loudspeakers below and above the listener where added to improve stability of the Ambisonics system. VBAP was implemented in its original formulation. Further detailed information about the listening test can be found in [8].

4.1 SET-UP

The listening test was held in an anechoic chamber with 24 loudspeakers which were set up in three horizontal rings with eight loudspeakers each. The approximate distance of each loudspeaker to the listener was about $1.7m$, with loudspeakers pointing towards the listener, one loudspeaker in direct frontal direction (azimuth $= 0°$) for each ring, spacing of $45°$ in azimuth and $30°$ in elevation, as shown in *Figure 6*. To optimize the pointing accuracy for the localization test a head mounted display was used, as suggested by Majdak [12]. The view inside the HMD (Oculus Rift DK1) was a sphere with a

Figure 6: Listening test set-up and view in the HMD

grid, illustrated in the bottom right corner of *Figure 6*. The own view direction was indicated by a red dot and the head rotation by four grey dots around the red dot. The front direction was indicated by a yellow dot in the grid. Playback was only enabled when the test subject was looking in front direction with no head rotation, i.e. the red dot was congruent with the yellow dot and the grey dots where on the grid. To avoid motion sickness and to ease orientation in the real world a fixed chair with arm-rests was used. The chair set-up and the loudspeaker set-up limited the virtual source positioning to the frontal directions as the subjects could not easily turn their head further than $\pm 80°$ in azimuth and as the loudspeaker set-up became unfavorable for source positions elevated further than $\pm 30°$ for Ambisonics and VBAP.

As virtual room a model of the Concertgebouw Amsterdam was used. The listener was positioned in the middle of the room, near the first row of the audience. To ensure an influence of the late reverberation tail on the perceived sound a source distance of $5.5m$ was chosen which matches the critical distance in the room. The four selected directions of the sources can be found in *Figure 7*. This figure also illustrates the source position in relation to the loudspeaker array. As Signal pink noise bursts were used. The playback could be repeated as often as desired.

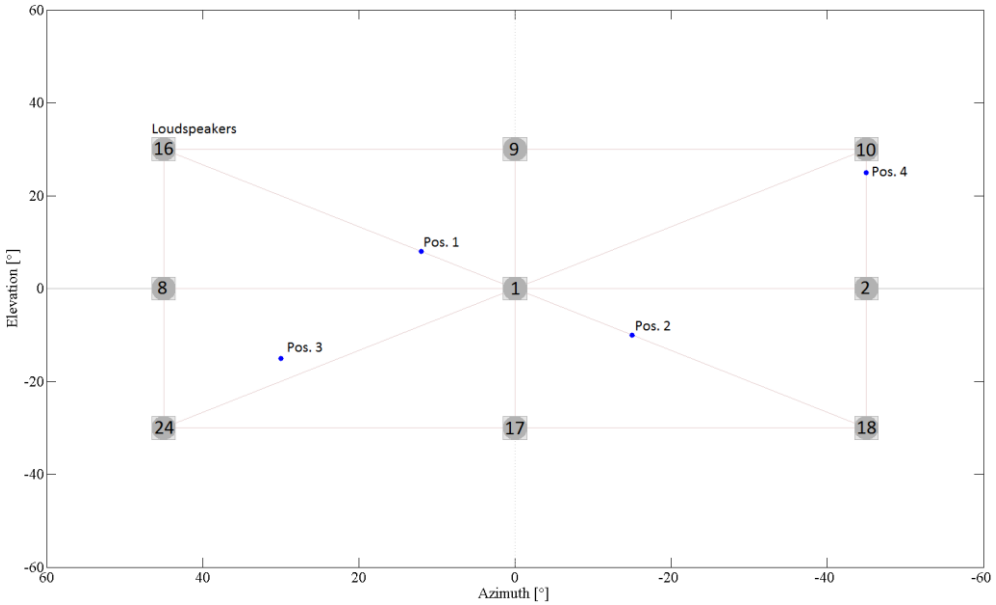

Figure 7: Directions of the source positions and frontal loudspeakers from the view of the listener

4.2 PROCEDURE

18 persons, including 7 acousticians, participated in the listening test. The five different reproduction techniques tested were Ambisonics, VBAP and CTC and a combination of CTC for early reflections (image source order of three) with Ambisonics for the late reverberant tail, and in the same way a combination of VBAP with Ambisonics. At first the participants were asked to rate the immersion of the three pure reproduction techniques in a 2-AFC. The definition of immersion was given as “feeling surrounded by the sound and being able to feel like ‘you are in a different room’ than the one you are seated in”. This part was done without the HMD.

Next the participants were set up with the HMD and had to proceed through a training phase in which single loudspeaker played pink noise bursts and were visualized in the HMD. The participants then had to point at the visualized loudspeaker and push a button to confirm their

selected direction. This way the participants could get used to the situation of being in the virtual world and familiarize with the operation. After the training phase the stimuli were played back using the five different reproduction techniques. In comparison to the training phase, no visualization of the sound source loudspeaker was used. Due to problems with the horizontal tracking of the HMD the subjects had to adjust the front direction before every sample by pointing the HMD to the loudspeaker in direct frontal direction which was playing pink noise bursts.

4.3 RESULTS AND DISCUSSION

The immersion test showed no significant difference between the reproduction techniques. Most participants stated that they had no specific idea of what immersion means. The formulation clearly was too vague and the assumption holds that participants changed the criteria they used to evaluate immersion during the test.

To investigate the immersion of the reproduction techniques more profoundly, a questionnaire is currently being developed and validated. This questionnaire contains a set of items addressing different aspects of the spatial audio reproduction such as envelopment, source localization or the room perception. The selection of items aspects is based on existing inventories of spatial audio attributes [15, 16]. The usage of this questionnaire will lead to a more detailed measurement of the immersion of the different reproduction methods.

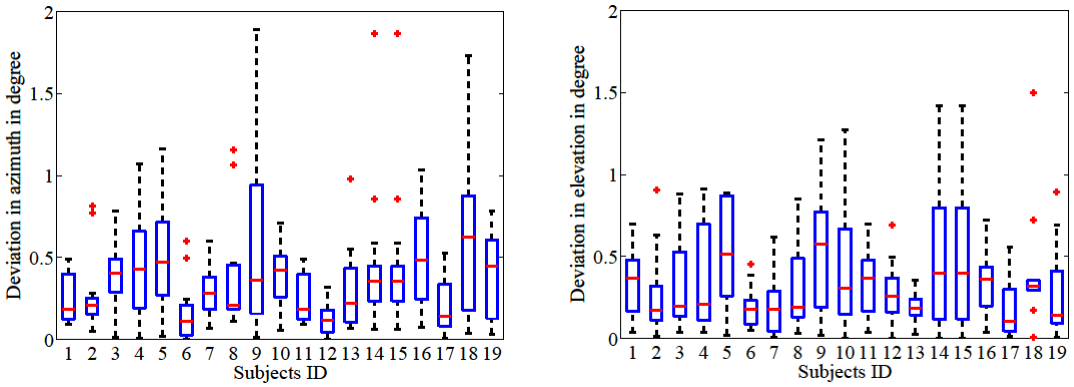

Figure 8: Pointing accuracy during training phase

In the localization experiment, the HMD as the pointing device resulted in a good performance. Both in azimuth and elevation the deviation was below 1°. Only two subjects complained about a slight feeling of motion sickness after finishing the listening test.

The localization test showed that the performance of the single reproduction techniques depends on the position of the virtual source in relation to the loudspeaker array as shown in *Figure 9*. The overall performance between the systems results no significant difference. For the CTC technique a more homogeneous localization can be found whereas Ambisonics and VBAP provide localization based on the relation of virtual source position to loudspeaker array.

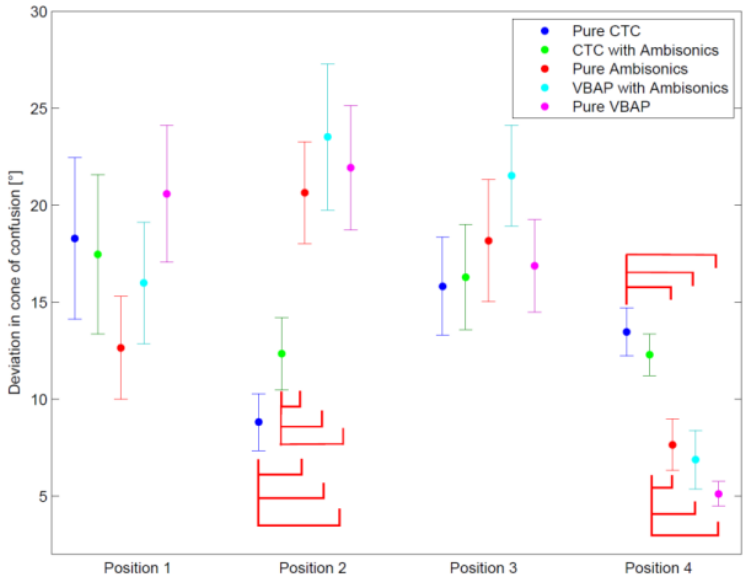

Figure 9: SPSS analysis of the localization test. The results depend on the position of the source in relation to the loudspeakers. No significant difference between pure and combined reproduction techniques can be found.

5. CONCLUSIONS AND OUTLOOK

The presented idea is a promising way to improve reproduction of spatial audio for room acoustic simulations and virtual reality purposes by including knowledge of the difference in perception of different sections in a RIR.

The listening test showed that no significant differences between pure reproduction methods and combined ones exist regarding localization performance. Proving the enhancement failed which is assumed to be founded in the listening test design. Further listening tests are currently being prepared to investigate the impression of envelopment and reverberance and to find a suitable way of measuring the immersion of spatial reproduction techniques.

The results of the listening test indicate that the localization performance depends on the loudspeaker array and the position of the virtual source in relation to the loudspeaker array. This relation should be investigated to find optimal solutions for the hybridization process.

Furthermore the listening test showed that a HMD can improve pointing accuracy and that participants can handle the device without problems. It should be noted that the used HMD, Oculus Rift DK1, is an outdated version and that motion sickness, pointing accuracy and usability can be expected to be enhanced in current versions of the HMD.

6. REFERENCES

[1] VORLÄNDER, M. Auralization, 2008, Springer Verlag
[2] Schroeder, M. R., Atal, B. S. "Computer Simulation of Sound Transmission in Rooms," IEEE Conv. Rec., 150-155 (1963)
[3] PELZER, S., MASIERO, B., AND VORLÄNDER, M., 3D Reproduction of Room Acoustics using a Hybrid System of Combined Crosstalk Cancellation and Ambisonics Playback. 2011.
[4] PULKKI, V. Virtual Sound Source Positioning Using Vector Base Amplitude Panning. Journal of the AES, 45(6):456–466, 1997.
[5] MASIERO, B. Individualized Binaural Technology: Measurement, Equalization and Perceptual Evaluation. Ph.D. thesis, RWTH
[6] GERZON, M. A. Periphony: With-height sound reproduction. Journal of the Audio Engineering Society, 21(1):2–10, 1973.
[7] DANIEL, J., "Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia," PhD Thesis, Ph. D. Thesis, University of Paris VI, France, 2000
[8] KOHNEN, M. 3-D Reproduction of Room Auralizations by Combining Intensity Panning, Crosstalk Cancellation and Ambisonics. Master Thesis, RWTH Aachen University, 2014.
[9] REILLY, A., MCGRATH, D., AND DALENBÄCK, B.-I. Using Auralisation for Creating Animated 3-D Sound Fields Across Multiple Speakers.
[10] MEESAWAT, K. AND HAMMERSHOI, D., editors. The Time When the Reverberation Tail in a Binaural Room Impulse Response Begins, 2003.
[11] LINDAU, A., editor. Perceptual evaluation of physical predictors of the mixing time in binaural room impulse responses. Audio Engineering Society, 2010.
[12] MAJDAK, P., LABACK, B., GOUPELL, M., AND MIHOCIC, M. The Accuracy of Localizing Virtual Sound Sources: Effects of Pointing Method and Visual Environment. 2008.
[13] GUASTAVINO, C., LARCHER, V., CATUSSEAU, G., AND BOUSSARD, P. Spatial Audio Quality Evaluation: Comparing Transaural, Ambisonics and Stereo. 2007.
[14] BERKHOUT, A. J., DE VRIES, D. "Acoustic holography for sound control," in 86th AES Convention, 1989
[15] BERG, J. AND RUMSEY, F., Identification of quality attributes of spatial audio by repertory grid technique. J. Audio Eng. Soc., 54 (5). 365-379, 2006.
[16] LINDAU, A., Binaural resynthesis of acoustical environments: Technology and perceptual evaluation. PhD thesis, Technical University Berlin, 2014.
[17] SCHRÖDER, D. And VORLÄNDER, M., "RAVEN: A real-time framework for the auralization of interactive virtual environments" in Proceedings of Forum Acusticum 2011 : 27 June - 01 July, Aalborg, Denmark, 2011.