Evaluation of Loudspeaker-based 3D Room Auralizations using Hybrid Reproduction Techniques

Sönke Pelzer, Michael Kohnen, Michael Vorländer

Institut für Technische Akustik, RWTH Aachen University, 52074 Aachen, Deutschland, Email: spe@akustik.rwth-aachen.de

Introduction

Room acoustic simulations use hybrid models for precise calculation of early specular reflections and stochastic algorithms for the late diffuse decay. Splitting the impulse response (IR) into early and late parts is also psychoacoustically reasonable. The early part is responsible for the localization and the spatial and spectral perception of sources, which makes the correct reproduction of its time-frequency structure important. In contrast the later part is responsible for the sense of spaciousness and envelopment, properties related to the room and its diffuse decay. There are reproduction systems better suited to coherent reproduction (important for the early arrivals of an impulse response) and others better suited for the reproduction of incoherent fields (the reverberant tail of an impulse response) [1]. A hybrid approach is presented which uses one common loudspeaker system for the simultaneous rendering of different reproduction methods. Strong localization cues should be provided for the direct sound and early reflections, while a method with higher immersion and envelopment can be used for the diffuse decay.

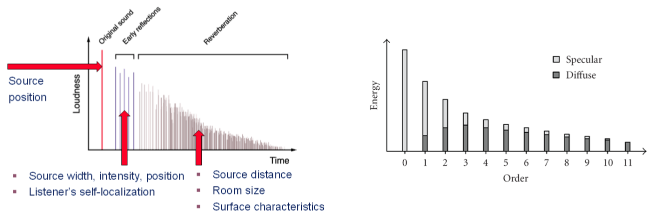

Figure 1: *Left*: Perceptual and physical division of the room impulse response. *Right*: Relation of specularly and diffusely reflected sound in a typical room.

Transition Time

The transition between the early and late part of the IR can be defined on a physical or perceptual basis. A recent study by Lindau reviewed existing definitions and added a perceptual evaluation [2]. A transition time t_m was found to be correlated to the mean free path length, as shown in Eq. 1, with V being the room volume and S the room's surface area. Using the image source method for early reflections, the image source filter length is proportional to the mean free path length. The product of reflection order and mean free travel time $\bar{t} = 4V/cS$ (c: speed of sound) should at least cover the transition time t_m, as shown in Eq. 2.

$$t_m = 20\frac{ms}{m} \cdot V/S + 12ms \qquad \text{[ms]} \qquad (1)$$

$$t_m = O_{IS} \cdot \bar{t} \qquad \text{[s]} \qquad (2)$$

$$O_{IS,min} = \frac{t_{m-12}}{\bar{t}} + 1 \approx 2.7 \qquad (3)$$

This results in a general estimation for a necessary image source order for many rooms (cf. tested rooms by Lindau with volumes from $182m^3$ up to $8500m^3$). Neglecting the additional 12 ms in the transition time formula in favor of a full additional order of image sources is a valid approximation for rooms with at least 4 m of mean free path. With this simplification, a general minimum image sources order criterion can be defined independently of reverberation time, volume or absorption to $O_{IS,min} \approx 3$, as shown in Eq. 3. Similar observations were made by Kuttruff, as shown in Fig. 1 (right).

Previous Work

Around 1980, the idea was mentioned in the Ambiophonics group to play CD recordings using a crosstalk canceled stereo-dipole for a wider stereo image and add optional ambiance loudspeakers for reverberation [3]. The aim was to enhance stereo or 5.1 recordings, while the optional ambiance channels were seen as an artificial effect. In 2010 Favrot proposed hybrid reproduction to account for different perception of room acoustics. He decoded simulated spatial impulse responses with variable Ambisonics orders for the early and late part and thereby reduced computation load for late reverberation and increased localization of the direct sound [4]. Guastavino et al. [1] compared different reproduction techniques (CTC, Ambisonics, Panning) and found differences in perception that are summarized in Table 1. It can be concluded that the reproduction method must also account for the psychoacoustics that define our hearing in rooms.

Table 1: Comparison of different reproduction techniques, as published by Guastavino [1], with additional comments.

Method	Advantages	Drawbacks
Binaural CTC	Precise localization, good readability, near field sources	Poor realism, lack of immersion/envelopment, needs individual HRTF
Ambisonics	Strong immersion and envelopment	Poor localization/readability
Stereo Panning	Precise localization	Lack of immersion/envelopment

Loudness Equalization

When mixing different reproduction methods their loudness must be accurately matched. This can be challenging due to the inhomogeneous direction-dependent loudness of individual systems and the assumption of an ideal constructive and destructive interference at the sweet spot. By applying Zwicker's loudness model to simulated dummy head responses of 900 sources from equally distributed directions, a sufficient loudness equalization could be achieved. [5]

Localization Performance Listening Tests

To measure the localization strength of different reproduction techniques, a listening test was conducted in an anechoic room with a 24-channel loudspeaker array (Fig. 2). A tracked head-mounted display provided an accurate and bias-free pointing method [6]. In a training phase an average pointing accuracy of 0.3 degrees was measured. Five reproduction methods (3 pure, 2 hybrid) were used to present four different source positions, as shown in Fig. 3. In total 18 subjects participated in the test with an average localization accuracy of 16 degrees (Fig. 4 left). Between the systems no significant differences were found in a one-way ANOVA test averaging all positions (Fig. 4 right). A two-way ANOVA revealed a dependency on the source position for VBAP and HOA, so that significant differences to the less sensitive CTC were found (Fig. 5).

Figure 2: *Left*: Listening tests were hosted in an anechoic chamber equipped with 24 loudspeakers for spatial reproduction. *Right*: Tracked head-mounted display and virtual environment for accurate pointing.

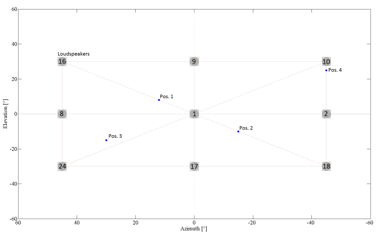

Figure 3: Presented 4 source positions in the listening test. Frontal 9 loudspeakers are shown for reference.

Conclusions

A powerful tool was developed for the preparation and reproduction of room auralizations using different reproduction techniques, including hybrid variants that combine different methods for an optimal reproduction of the room acoustics. For simultaneous hybrid playback the different methods have to be accurately calibrated to

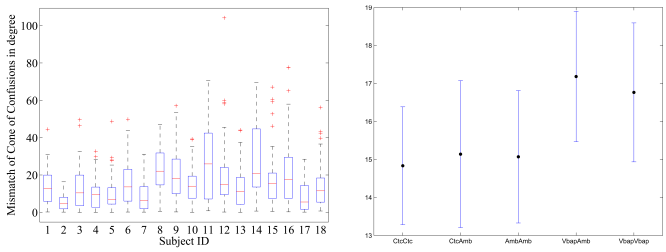

Figure 4: *Left*: Results of the localization performance tests for the 18 individual subjects. *Right*: Results of the ANOVA for different reproduction systems (CTC, CTC+HOA, HOA, VBAP+HOA, VBAP). No significant differences were found in localization performance.

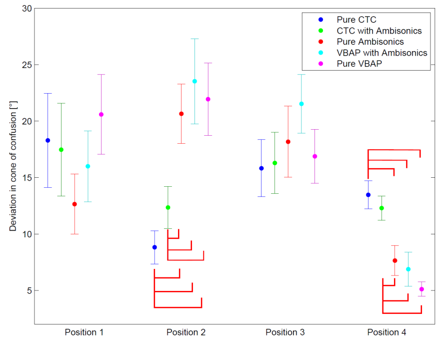

Figure 5: Results of two-way ANOVA grouped by presented source position. A high impact of the source position is visible. Red brackets indicate significant differences.

equal loudness. This was achieved by simulating array responses and calculating loudness for a virtual binaural receiver in the array. For the reproduction of the early part of the impulse response, a method with strong localization cues is preferred. Therefore different candidates were tested in a localization performance listening test. The test illustrated a strong dependency on the position of a virtual source resulting in different optimal reproduction systems for different positions. Overall no system was significantly more accurate than the others, while the CTC localization was in general more homogeneous across different positions.

References

[1] Guastavino et al.: Spatial audio quality evaluation: comparing transaural, ambisonics and stereo. Proc. 13th Internat. Conf. on Auditory Display, 2007

[2] Lindau, A.: Perceptual evaluation of physical predictors of the mixing time in binaural room impulse responses. Proc. AES 128th Convention, 2010

[3] Farina et al.: Ambiophonic principles for the recording and reproduction of surround sound for music. Proc. AES 19th Internat. Conf., 2001

[4] Favrot & Buchholz: Lora: A loudspeaker-based room auralization system. Acta Acustica, vol. 96, pp. 364–375, 2010

[5] Kohnen, M.: 3-D Reproduction of Room Auralizations by Combining Intensity Panning, Crosstalk Cancellation and Ambisonics. Master thesis, RWTH Aachen University, 2014

[6] Majdak et al.: 3-d localization of virtual sound sources: Effects of visual environment, pointing method, and training. Attention, Perception, Psychophysics, 72(2), 2010