Reduction of Dynamic Cues in Auralized Binaural Signals

Markus Fruhmann*, Philip Mackensen§, Günther Theile
Institut für Rundfunktechnik GmbH (IRT), Floriansmühlstr. 60, D-80939 München, Germany

PACS no. 43.66.Pn,43.66.Qp,43.66.Md

1. Introduction

Two ways of auralizing virtual or real listening environments via headphones are commonly in use: Model based and data based systems. The first group uses a more or less complex model to reproduce the acoustical specifications of a room. With higher complexity of the model used, the quality of the results rises, but the processing power needed increases rapidly. In contrast, a data based system will store either “pre-synthesized” impulse responses or measured data of real existing rooms. The ear signals are produced in the same way in both systems, namely a convolution of an impulse response with the incoming audio signal.

Experiments by Mackensen [1], Thurlow and Runge [2], Thurlow et al. [3] and Wallach [4] showed the importance of head rotations in the horizontal plane for accurate localization. To take these results into account, the convolution in an auralization system needs to be calculated in dependence of the actual head orientation. By doing so, front-back reversals and inside-the-head locatedness as described by Blauert [5] can be widely avoided. In dummy head recordings or static auralization systems these time variant interaural differences (ILDs and ITDs) are absent as dynamic cues in the binaural signal.

The convolution with impulse responses is memory and processing power intense. For each relevant head position and orientation the related transfer functions need to be stored when using a databased auralization system. Thus it would be useful to reduce the stored data to a minimum required set.

The aim of the present study was to investigate the dependence of auralization quality on dynamic cues in the binaural impulse responses. Thus the impulse response for a certain head orientation was simplified and audible changes in sound characteristics due to these changes were investigated. The simplification consisted of a crossfade from the orientation dependent impulse response to the impulse response for zero degrees head orientation after a certain time t_{dyn}. By means of this simplification, the impulse response is separated into a dynamic and a static part as shown in Figure 1.

2. Measurements

For the experiments, a studio listening room at the Bavaria Film Studios, Munich, was selected. The base dimensions were 11 by 15 meters and with a height of 6 meters the volume of the room was 990 m^3. The reverberation time has been determined to T_R=300 ms at frequencies around 500 Hz. This is substantially longer than the stored impulse response length of 85 ms implemented in the auralization system which was used in these experiments. Thus the auralization of the room can only be an approximation of the original room acoustics. Looking at the measured impulse response, the amplitude has declined more than 30 dB below the maximum after the measured time of 85 ms. Thus a difference to the real listening situation should only be perceived with extremely critical signals like clicks. With continuous signals, no problems should occur due to post-masking effects.

Received 1 December 2001,
accepted 1 May 2002.

* now: AG Technische Akustik, MMK, TU München, Arcisstr. 21, D-80333 München, Fruhmann@mmk.ei.tum.de

§ now: T-Systems debis Systemhaus GEI GmbH, München

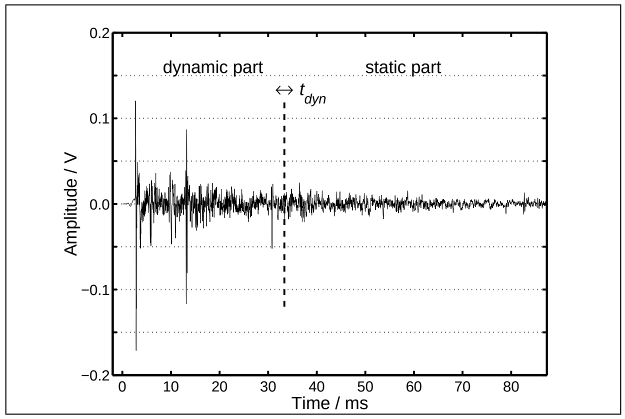

Figure 1. Impulse response of the used studio listening room, measured with the dummy head, for frontal sound incidence with the center speaker as the sound source. Convolution depends on head orientation up to t_{dyn} and afterwards the impulse response is faded to a static part. The value of t_{dyn} is varied in the experiment to find a threshold for audible artefacts due to reduced dynamic cues

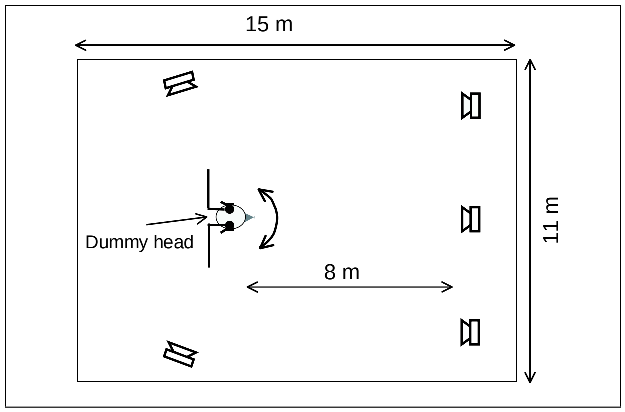

Figure 2. Sketch of the measured studio listening room

The loudspeaker setup as shown in Fig. 2 did not fit the needs described in [6] as the speaker setup in this room was intended to conform with a real cinema. The measured listening position was about 8 meters from the front speakers and equidistant to the side walls.

The impulse responses were measured for each speaker separately with a diffuse field equalized dummy head (Neumann KU100) connected to an Audio Precision System ONE. For each speaker and dummy head position, the binaural impulse response was measured using an MLS (Maximum Length Sequence) procedure. The impulse responses used in the experiments started at a threshold of 40 dB below the measured maximum and had an overall duration of 85 ms. For each speaker 15 impulse responses for dummy head orientations of -42 to +42 degrees azimuth relative to the front direction were measured with a horizontal resolution of 6 degrees. No elevation data was included as the measurements were only conducted within the horizontal plane. During the measurements, the dummy head was mounted on a motor which was controlled by a host system (Personal Computer) connected to the System ONE and the motor.

3. Psychoacoustic experiments

3.1. Stimuli

The intention of the experiment was to detect changes in the perceived room characteristics using simplified transfer functions. Thus no reverberation should be included in the stimulus used as input signal for the auralization process. A signal meeting these re-

quirements is track 49 of the EBU SQAM CD ([7], “to administer medicine to animals...”), which is a dry recording of female speech. The impulse responses stored in the auralization system were modified so that the part used for dynamic convolution had a length of t_{dyn}. The length of this dynamic part was varied within a range from 7 to 43 ms in steps of 4 ms and thus 10 different listening situations were realized.
The impulse responses were modified in a way that a cross-fade to the impulse response for a frontal head orientation (zero degrees) was calculated for each speaker. The time windows for this calculation had a length of 1.3 ms which, with a sampling frequency of 48 kHz, results to a length of 64 samples. The cross-fade part of the signal is calculated as follows:

$$IR_{mod}(n, \alpha) = \\ CR(n) \cdot IR(n, \alpha) + (1 - CR(n)) \cdot IR(n, 0),$$

where $IR_{mod}(n, \alpha)$ is the modified impulse response for a certain azimuth angle α, CR is the following cosine ramp

$$CR(n) = \cos^2(\frac{n \cdot \pi}{2 \cdot 64}), \qquad n = [1 \dots 64]$$

with a length of 64 samples and $IR(n, \alpha)$ is the original impulse response for an azimuth angle of α.
For the experiments only one front speaker (center) and one surround speaker (left-surround) were chosen. For each of the two speakers, separate experiments were conducted.

3.2. Auralization tool

The test signals were presented by means of a Studer BRS (Binaural Room Scanning) system ([8, 9]). This system is a data based auralization system with head tracking which is able to reproduce an existing studio listening situation.
To achieve this, a real listening situation is acoustically “scanned” as described in section 2 with a dummy head. The measured data is stored within the processor. For sound reproduction, a STAX Lambda Pro with a Polhemus Fastrak head tracker sensor mounted on top is used. When listening to signals using the BRS system the impulse responses, corresponding to the actual head orientation, are selected for convolution. These responses are interpolated from the measured data with a resolution of 0.5 degrees. This is a higher resolution than claimed by Mills [10] who found a minimum audible angle of 1 degree. Thus the interpolation accuracy should result in a proper room auralization.
As described in [8], the interpolation is computed in the frequency domain within each processing interval. According to Theile [11] the dummy head and headphones were diffuse field equalized to avoid orientation dependent errors.

3.3. Procedure

An A-B comparison procedure was chosen for testing. The reference sequence (A), which is convolved dynamically over the whole length of 85 ms, is followed by the modified sequence (B). In the modified sequence only the first part up to t_{dyn} is chosen dynamically according to the listener’s head orientation, whereas the later part ($t_{dyn} \dots 85$ ms) is identical for all head orientations.
As the stimulus was steadily repeated until the subject was able to draw a decision, in each pause between the repetition the listening situation was switched from A to B and vice versa. 12 expert listeners aged 23 to 49 took part in the investigation. Their task was to report if any difference was audible between the presented listening situations. To make active usage of dynamic cues that might help to detect the differences between the two listening situations, subjects were instructed to turn their head while listening to the test stimuli. For each value of t_{dyn} this procedure was repeated until the subject was able to reach a decision.
During the experiments, the subjects were surrounded by a curtain to eliminate optical influences which could disturb the plausibility of the auralized environment.

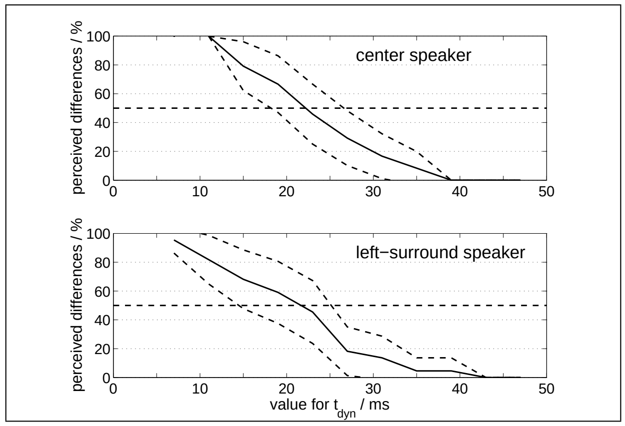

Figure 3. Experimental results for center (upper panel) and left-surround (lower panel) speaker. The solid line represents the percentage of perceived differences. With increasing dynamic cues due to a higher value for t_{dyn} less differences are evoked. Results are shown as arithmetic mean and 95 % confidence intervals over 12 subjects.

4. Results

4.1. Reported differences

Subjects were asked orally about the differences they used as a criterion to judge the presented pairs of signals in the experiment. They described several differences between the two listening situations. These differences depended on the length of the dynamic part of the impulse response and appeared as described in the following.
For very small values of t_{dyn} (7 – 11 ms), inside-the-head locatedness and front-back reversals occurred, like in the completely static case. This is comparable to a recording with a fixed dummy head or with static auralization systems. With increasing values of t_{dyn} disturbing additional echoes are perceived, which lead to an incorrect sensation of the room’s size. When t_{dyn} almost reaches the detection threshold, slight changes in sound characteristics like bass enhancements appear, which are very difficult to detect.

4.2. Minimum dynamic computation time

Figure 3 shows the results for the two different speakers that were selected for the experiment. In each graph the arithmetic mean and the 95 % confidence interval for the percentage of perceived artefacts are shown. The determined threshold for a procedure as used in these experiments is found as the value of t_{dyn} where differences between the two presented variations are audible with a probability of 50 %. According to the mean values, this threshold-value is 22 ms. To retain some more dynamic cues for very sensitive listeners, the upper limit of the 95 % confidence interval is taken into account which leads to a proposed value for t_{dyn} of about 27 ms for the used room.
In the case of the surround speaker the confidence interval is somewhat larger than for the center speaker and subjects reported difficulties to evaluate their listening impression. In that listening situation, one subject determined differences even with a dynamic computation time of 39 ms.

5. Discussion

As the results show, in principle a reduction of dynamic cues is possible without producing audible changes in sound quality or room impression. Taking only into account the mean values of the experimental data, the 50 % threshold is reached for a value of t_{dyn} of about 22 ms. This value coincides with the time of early reflections that arrive at the listener’s ears within the first 20 ms according to Hartmann [12] and their well-known importance for localization.

Later parts of the impulse response, i.e. reverberation, apparently contain less significant localization cues. Thus it seems plausible to find a threshold in the described experiments, close to 20 ms. As the limit for early reflections is dependent on the actual auralized room and the audibility of differences due to simplified impulse responses varies from listener to listener, a higher value is recommended for the dynamic computation time t_{dyn}. Thus the upper limit of the 95 % confidence interval is taken as a measure for the searched limit and a value for t_{dyn} of 27 ms is recommended.

Acknowledgement

The authors would like to thank Ulrich Horbach and Attila Karamustafaoglu at Studer for their support. Further thanks to Prof. Hugo Fastl at the Technical University Munich for encouraging the author's work.
The study is part of the development examinations for the BRS system in cooperation of the Institut für Rundfunktechnik, Munich, and Studer, Regensdorf (CH).

References

[1] P. Mackensen, U. Felderhoff, G. Theile, U. Horbach, R. Pellegrini: Binaural room scanning - a new tool for acoustic and psychoacoustic research. 137th Meeting of the Acoustical Society of America, conference transcript (CD-ROM) (1999).
[2] W. Thurlow, P. Runge: Effect of induced head movements on localization of sounds. J. Acoust. Soc. Am. **42** (1967) 480–488.
[3] W. Thurlow, J. Mangels, P. Runge: Head movements during sound localization. J. Acoust. Soc. Am. **42** (1967) 489–493.
[4] H. Wallach: The role of head movements and vestibular and visual cues in sound localization. J. Exp. Psychol. **27** (1940) 339–368.
[5] J. Blauert: Spatial hearing. The MIT Press, 1997.
[6] ITU: Multichannel stereophonic sound system with and without accompanying picture. ITU Rec. TU-R BS.775-1 (1998).
[7] EBU: Sound quality assessment material recordings for subjective tests. EBU SQAM CD, EBU Tech. 3253-E (1988).
[8] U. Horbach, A. Karamustafaoglu, R. Pellegrini, P. Mackensen, G. Theile: Design and applications of a data-based auralization system for surround sound. 106. AES Convention, Munich, preprint 4976 (1999).
[9] P. Mackensen, M. Fruhmann, G. Theile, U. Horbach, A. Karamustafaoglu: Head-tracker based auralization systems: additional consideration of vertical head movements. 108. AES Convention, Paris, preprint 5135 (2000).
[10] A. Mills: On the minimum audible angle. J. Acoust. Soc. Am. **30** (1958) 237–246.
[11] G. Theile: On the theory of optimal reproduction of stereophonic signals with loudspeakers and headphones. Rundfunktechnische Mitteilungen **4** (1981) 18–26.
[12] W. M. Hartmann: Listening in a room and the precedence effect. Binaural and Spatial Hearing in Real and Virtual Environments, 1994. R. H. Gilkey, T. R. Anderson (eds.). Lawrence Erlbaum Ass., 191–210.