7 - 12 SEPTEMBER 2014, KRAKÓW
FORUM ACUSTICUM

Does reverberation perception differ in virtual spaces with unrealistic sound reflections?

Fritz Menzer, Bernhard U. Seeber
Associated Institute Audio Information Processing, Technische Universität München, Germany.

Summary
Reverberation caused by sound reflected off surfaces and objects is present in abundance in everyday life. Yet, in most situations we are unaware of it. This is due to the ability of our auditory system to adapt to the incoming sound reflections, which results in increased echo thresholds, i.e. the minimum delay to hear a reflection as an echo is increased after the echo was heard repeatedly. Changes in the directional pattern of sound reflections disturb the adaptation process and for a short time after the change, reflections become audible. Besides a few notable exceptions, this process has been studied mainly with single reflections rather than with realistic reverberation. The present study addresses the question how changes in the reflection pattern of a reverberant environment are perceived. For the study, realistic simulations of reverberation in a room are created using a high order image source model. By altering parameters of individual image sources, in particular their direction, unrealistic reflection patterns are created which nevertheless maintain key properties of the room such as the direct-to-reverberant ratio and T60. This study tests the hypothesis that, compared to realistic reverberation patterns, differences can be observed in the release from adaptation for unrealistic reverberation patterns. We report the preliminary results from a listening test where the subjects first adapt to reverberant sound signals, and then indicate if a change in the reverberation pattern is perceived or not.

PACS no. 43.66.Pn, 43.66.Qp

1. Introduction

In everyday life, reflections caused by the surfaces of rooms and other acoustical environments are very common. However, these reflections are usually not perceived separately, and we are often unaware of them even in moderately reverberant environments. As was studied extensively in the context of the precedence effect, the human brain is able to suppress simple reflections, meaning that echos arriving before the so-called echoe threshold are suppressed. This echo threshold is not fixed, but can be increased by a repeated exposure to sound containing reflections [1]. Many studies related to the precedence effect use very simple stimuli such as clicks presented in free field or using headphones (without a realistic room simulation). An exception are the experiments presented in [2], where speech was used as a source signal in one experiment and a triangular room simulated up to the second reflection order in another experiment. In summary, it may be noted that phenomena related to the precedence effect such as build-up and release from adaptation can in principle also be observed with more than a single reflection or when speech is used as a source signal.

The first goal of the study presented here is to examine the release from adaptation that occurs when a reverberation pattern suddenly changes in a realistically simulated situation. Compared to other studies, here a very detailed simulation of a room was used (using an image source model up to the 20th reflection order) and instead of clicks, speech was used as the source signal, i.e. the signals used in this study are realistic with respect to both the modeled acoustic environment and the naturalness of the source signal.

The second goal was to examine if any differences in the release from adaptation can be observed when the directions of arrival of the reflections are artificially scrambled such that the resulting reflection pattern would be extremely unlikely to occur in reality. If the brain were to employ a geometrical model of the room for the echo suppression mechanism, the proposed scrambling would interfere with this mechanism, as the reflection pattern could not be interpreted as the result of a sound source emitting sound in a physically meaningful room.

In the following, the experimental setup as well as the method for generating the stimuli are explained, followed by subject-related information. The results and conclusion sections can be found at the end of this paper.

2. Experimental Setup

The experiment reported here was performed using headphone-based virtual acoustics. The necessary binaural room impulse responses were generated using the Simulated Open Field Environment [3], which was configured to use the "mer" HRTF set from the AUDIS catalogue [4].

The playback of source signals convolved with the simulated binaural room impulse responses uses Sennheiser HD600 headphones connected to a RME ADI-2 D/A converter, which was connected via S/PDIF to a PC equipped with an M-Audio Delta Audiophile card providing a 24 bit signal. The sampling rate was 44100 Hz.

Users could give feedback by pressing a button which was connected to the PC via a USB-to-serial-port-adaptor. A sound-proof booth was used to conduct the experiments.

3. Method

3.1. Overview

In this study, the test subjects listened to a speech signal convolved with a binaural room impulse response, which, after a randomly chosen time between 3 and 5.6 sec, was replaced by another binaural room impulse response. The subjects had to indicate that they heard a change in the reverberation characteristics of the room by pressing a button within 2 sec from the change of the impulse response.

The binaural room impulse responses were obtained by simulating a rectangular room using the image source model as implemented in SOFE [3] up to order 20 for a 5 m-by-9 m-by-2.3 m room, a listener position (1.500 m,1.800 m,1.600 m) and the source positions specified in Table I and illustrated in Figure 1. While in the first part of the stimulus, the source position was always position 1, in the second part it was any of the positions 1 to 8 (where a repeated use of position 1 means that no change in the impulse response occurred).

The source signal used to generate the stimuli was taken from track 50 of the EBU SQAM CD [5]. It consists of a recording of the sentence "In the course of a December tour in Yorkshire, I rode for a long distance in one of the public coaches, on the day preceding Christmas" spoken by a male voice. This sentence was split at a randomly chosen point between 3 and 5.6 sec after its beginning. The two resulting parts overlap by 2 ms with a sinusoidal crossfade, as illustrated in Figure 2. The first part was convolved with a binaural room impulse response for position 1,

Table I. Coordinates of source positions used and distances relative to the first position and the listener.

				Distance in [m] from:	
Pos.	x [m]	y [m]	z [m]	Pos. 1	Listener
1	2.200	3.060	1.600	0	1.441
2	2.225	3.105	1.600	0.052	1.493
3	2.250	3.150	1.600	0.103	1.544
4	2.300	3.240	1.600	0.206	1.647
5	2.400	3.420	1.600	0.412	1.853
6	2.500	3.600	1.600	0.618	2.059
7	2.600	3.780	1.600	0.824	2.265
8	2.700	3.960	1.600	1.030	2.471

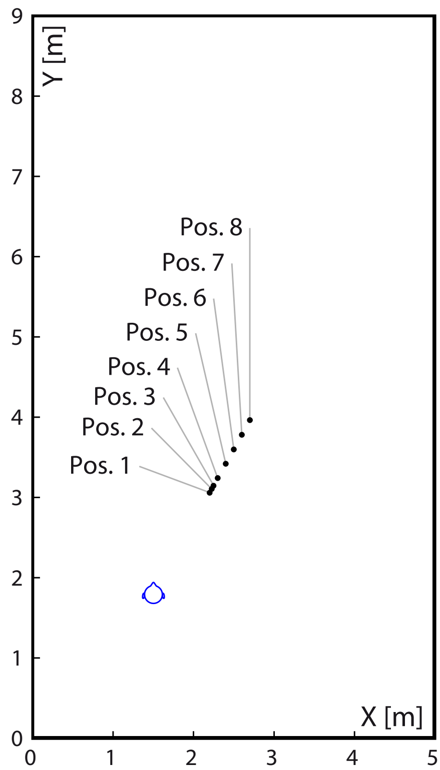

Figure 1. Source and listener positions used in this experiment. The simulated room was 5 m wide and 9 m long, with walls corresponding the the rectangular frame of the illustration.

while the second part was convolved with a binaural room impulse response for one of the positions 1 to 8. The signal was stopped 2 sec after the split point, i.e. the signal presented to the subject finished at the same time as the valid time interval for giving the response.

3.2. Impulse Response Scrambling

In order to compare a realistic room impulse response with a physically impossible one, scrambled versions of the binaural room impulse responses were obtained by randomly changing the directions of arrival of the individual reflections in the impulse response, as

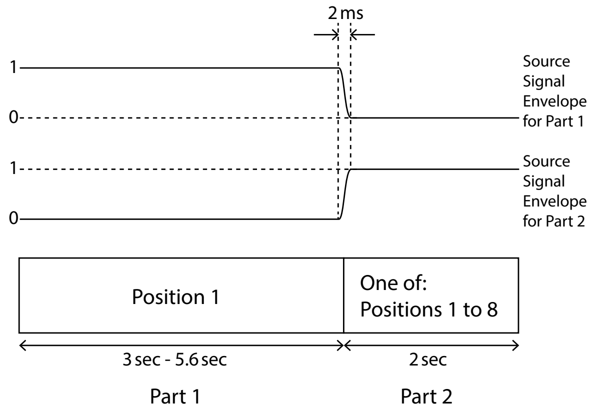

Figure 2. Envelopes used for the stimuli generation. By multplication with the two envelopes shown above, a speech signal was split into two parts overlapping by 2 ms with a sinusoidal crossfade. The first part was convolved with a binaural room impulse response corresponding to position 1, whereas the second part was convolved with a room impulse response for one of the positions 1 to 8. The stimulus was obtained by adding the two convolved signals.

shown in Figure 3. This means that the timing of the reflections, as well as their overall energy, is conserved. The binaural room impulse response is obtained by convolving the individual reflection responses (containing the wall absorption filters, etc.) by different head-related impulse responses than for the realistic version. This procedure preserves the reverberation time T60 and the direct-to-reverberant ratio. Furthermore, for each reflection, the angle of arrival was always modified by the same amount across all 8 different positions used in this experiment. Therefore, from one position to another the difference in the impulse response was comparable to the difference in the realistic simulation as is illustrated in Figure 4. The angle of arrival of the direct sound, as well as the first order floor and ceiling reflections were never changed, in order not to perturb the perception of the direct sound.

3.3. Tested Conditions

In total, 32 conditions were tested: 8 positions for the second part of the stimulus, each with original and scrambled image source positions, and also with normal and inversed channel assignments for the binaural output signal (where the inversed channel assignment corresponds to a situation where the room and source positions have been mirrored about the median plane of the listener).

3.4. Stimulus Presentation and Feedback Recording

The stimulus signals were divided into blocks of 92.87 ms (4096 samples at 44100 Hz sampling rate) and played using the playrec tool

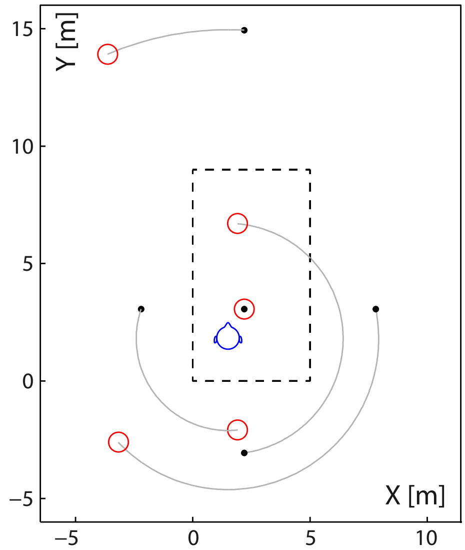

Figure 3. Comparison of original (realistic) and scrambled first order image source positions in the horizontal plane. The head symbol (not drawn to scale) indicates the listener position, the black dots indicate the original image source positions, while the circles indicate the scrambled positions. The grey arcs centered around the listener connect the original with the scrambled positions. The dashed line indicates the walls of the room used for the image source model.

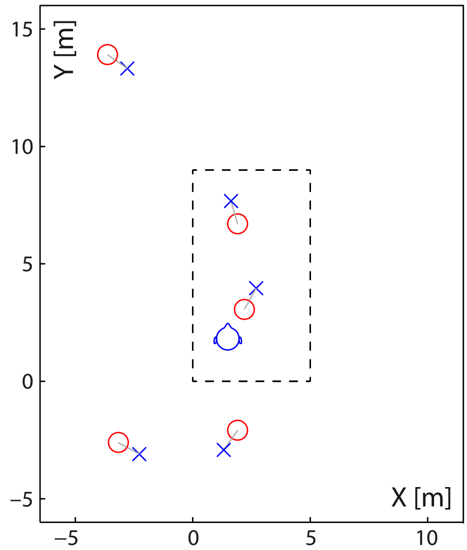

Figure 4. Comparison of scrambled image source positions for source positions 1 and 8. The head symbol (not drawn to scale) indicates the listener position, the circles correspond to position 1, and the x-marks correspond to position 8.

(http://www.playrec.co.uk/). Each time playrec indicated that a block had finished playing, the status of the push button was polled and this status (pressed / not pressed) was assigned to the block that had just finished.

If the subject pressed the button within the 2 seconds that followed the change in the reverberation characteristic, the result of the trial was considered as "change detected". Otherwise (i.e. if the subject did not press the button at all, or pressed it before the change happened), the result was considered as "no change detected".

The whole listening test consists of 640 trials split into 20 blocks of 32 trials each. Each block contains all 32 conditions in a randomized order and takes approximately 5 min to complete.

A single block could not be interrupted, but the test subject could choose when to start the next block. The subject whose results are presented here mostly did 2-3 blocks in one session, followed by a break of a few minutes. All sessions were completed within two days (10 blocks on one day and 10 blocks the next day).

4. Subjects

Here we present preliminary results for one subject (female, aged 26). Before starting the test, the subject could familiarize with the stimuli and the task by doing 3 blocks of 32 trials identical to those in the actual listening test. The results of these 3 blocks were not included in the statistical analysis. A significant change in responses could be observed between the first block and the two subsequent blocks, indicating that a familiarization session of at least one block is necessary before useful data can be recorded.

5. Results

Preliminary results for one subject are shown in Figure 5. The probability of detecting a change in the reverberation characteristic is shown as a function of the position of the source in the second part of the stimulus. For the statistical analysis, the trials for both normal and inversed channel assignments are merged, resulting in 40 trials per condition. As the non-overlapping confidence intervals for position 3 show, the subject was significantly more likely to hear a difference in reverberation characteristics for a step from position 1 to position 3 for the scrambled condition than for the realistic condition. This is likely to result also in a smaller just noticeable difference for the scrambled condition, but this could not be determined yet with a single subject having performed the test.

6. Conclusion

A statistically significant difference in reverberation perception between realistically simulated and physically implausible reverberant environments could be shown in the preliminary results for one subject. These results may be an indication that a geometrical model may be involved in the perception and suppression of reflections in reverberant environments. However, before drawing this conclusion, other causes for

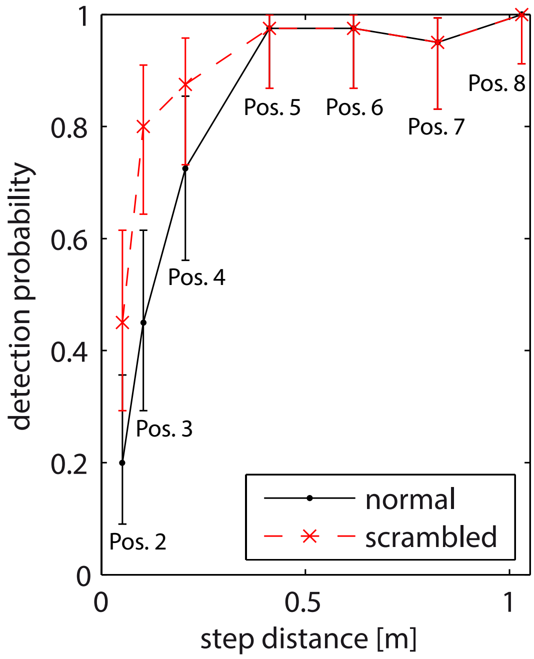

Figure 5. Preliminary results from one subject. The estimated probability for detecting the change in reverberation characteristic is shown as a function of the step distance, which is determined by the position of the source in the second part of the stimulus. The probability estimate is obtained by computing the average response over 40 trials. Error bars were calculated using the Clopper-Pearson method implemented in the MATLAB function "binofit" and indicate 95% confidence intervals. The solid lines and dots correspond to the realistic condition, while the dashed lines and x-marks correspond to the scrambled condition.

the observed phenomena should be ruled out in further experiments, e.g. monaural spectral effects in the left and right channels of the binaural signals presented to the listeners.

Acknowledgement

This project has been funded by BMBF projects 01 GQ 1004A and 01 GQ 1004B.

References

[1] R. Y. Litovsky, H. S. Colburn, W. A. Yost, S. J. Guzman: The precedence effect. J. Acoust. Soc. Am. **106** (1999) 1633-1654.

[2] T. Djelani, J. Blauert: Investigations into the build-up and breakdown of the precedence effect. acta acustica / ACUSTICA **87** (2001) 253-261.

[3] B. U. Seeber, S. Kerber, E. R. Hafter: A system to simulate and reproduce audio-visual environments for spatial hearing research. Hearing Research **260** (2010) 1-10.

[4] J. Blauert et al.: The AUDIS catalog of human HRTFs. 16th International Congress of Acoustics (1998).

[5] European Broadcasting Union: Sound Quality Assessment Material recordings for subjective tests - Users' handbook for the EBU SQAM CD. Geneva (2008).