Auditory adaptation in real and virtual rooms

FLORIAN KLEIN[1,*], STEPHAN WERNER[1], GEORG GÖTZ[1,2] AND KARLHEINZ BRANDENBURG[1]

[1] *Electronic Media Technology Group, Technische Universität Ilmenau, D-98693 Ilmenau, Germany*

[2] *Department of Signal Processing and Acoustics, Aalto University, FI-00076 Aalto, Finland*

Walking from room to room in real listening conditions is a natural process in our everyday life and there is no obvious challenge for our auditory system to cope with. However, in experiments with virtual acoustic environments switching the virtual room or switching from real to virtual rooms can result in auditory confusions which can lead to in-head localization. This effect is known as the room divergence effect. A series of listening tests were conducted to verify this effect under different conditions as well as experiments which studied the effect of prior sound exposure and the time variant behaviour of it. In this paper two of these experiments are described and discussed. The first experiment shows that the extent of the room divergence effect depends on the room acoustics we have just learned. That indicates, that the room divergence effect is diminished during ongoing exposure to a specific room acoustic condition. The second listening test shows further evidence of this time-variant effect and we show that it can be suppressed by interrupting with the adaptation process. These tests raise the question why switching virtual rooms leads to temporary confusions but doing so with real rooms is unproblematic. Different theories are discussed in this publication.

INTRODUCTION AND MOTIVATION

Auditory adaptation effects are well known in a broad range of research areas like neurosciences as outlined in King (2008) and hearing rehabilitation as described by Moore *et al.* (2009). Listening training is important for hearing impaired people to familiarize with their newly fitted hearing aid or cochlear-implants. However, such adaptation effects are rarely taken into account during the evaluation of binaural synthesis systems or other spatial audio reproduction techniques. Previous research in this field has shown, that auditory adaptation to altered localization cues as shown by Mendonça (2014) can improve localisation. Also, adaptation to changing room acoustic situations were observed by Keen and Freyman (2009) as well as Seeber *et al.* (2016).

The ability for spatial hearing is not only based on signal driven processing but also

*Corresponding author: florian.klein@tu-ilmenau.de

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 7: Auditory Learning in Biological and Artificial Systems, August 2019, Nyborg, Denmark. Edited by A. Kressner, J. Regev, J. C.-Dalsgaard, L. Tranebjærg, S. Santurette, and T. Dau. The Danavox Jubilee Foundation, 2019. ISSN: 2596-5522.

on listener experience and expectations. Research in the domain of quality perception suggests that expectations about sound serve as an internal reference for the listener in order to rate the perceived quality (see Raake and Blauert (2013)). These expectations can change depending on prior sound exposure. Based on this concept, it might be possible that listeners learn how to interpret spatial cues and room reflections for localization tasks.
Virtual acoustic environments aim to place the listeners in different acoustical environments and therefore forces them to adapt to these situations. This publication presents research which shows such adaptive processes in different listening tests and it discusses differences between the perception of real and virtual rooms.

STATE OF THE ART

The auditory precedence effect describes the prioritization of the first sound waves of a sound event arriving at a listener in the perception of the sound event as described by Wallach *et al.* (1949). Sound waves arriving after this are assigned to the first sound until an echo threshold is reached. This effect refers in particular to the localization and directional assignment of auditory events in an environment affected by sound reflections. The temporal range of the precedence or fusion depends on the spectral composition of the sound waves arriving later in relation to direct sound and on adaptation to the spatial temporal patterns of direct sound and sound arriving later.

The temporal order of magnitude of the echo thresholds underlies a build-up process. The build-up of the precedence effect has been intensively investigated in experiments by Clifton and Freyman (1989); Freyman *et al.* (1991) as well as Clifton *et al.* (1994). A repetition of the same patterns of direct sound and reflections led to an increase in the echo threshold. This indicates an adaptation and learning process that is less dependent on the length of time and more on the number of comparable reflection patterns (see Djelani and Blauert (2001)). In conclusion, this means that in a changing acoustic environment there is no sudden collapse (or rebuilding) of the precedent effect. A significant extension of the precedence effect is the one proposed by Clifton (1987) and Litovsky *et al.* (1999) on the spatial variation of the pattern of direct sound and reflection. A change of the pattern leads to a reduction of the echo threshold and thus to a collapse of the precedence effect. After the change a new precedence effect is established. This effect is commonly referred to as the Clifton effect.

The room divergence effect (RDE, see also Werner *et al.* (2016)) describes the influence of the acoustical differences between an virtually auralized room (for example via binaural synthesis) and real room. If such a divergence is present, the perceived externalization of the auditory event is reduced. The reason for this effect lies in a cognitive disproportion between the expected auditory event and the actual perceived auditory event. A basic approach to the explanation can be found in the precedence effect and its extension, the Clifton effect. The patterns stored in the auditory system for recognizing the room and the audio scene do not correspond to those derived from the synthesis. If these deviations are sufficiently large, the brain is no longer able to produce perceptive fusion between virtual and real room. The assimilation of the currently perceived event to a stored schema/pattern fails. The term externalization describes the perception of the location of an auditory event outside the head. The counterpart to this is the in-head localization (IHL). The perception of auditory events outside the head is regarded as a mandatory quality feature of a binaural headphone system for the generation of a plausible auditory illusion. In studies by Toole (1970) and Plenge (1972) it becomes clear that the effect of the IHL is not necessarily dependent on the use of a headphone system. In his experiments, Toole (1970) was able to show that IHL also occurs when loudspeakers are used in environments with low reverberation. In a further study on the emergence of the IHL through Plenge (1972), the hypothesis is put forward that the IHL arises through a lack of adaptation or an inadequate learning process. The learning process includes the short-term learning of properties of the sound source and the listening room. Further experiments show that a smooth transition between an out-of-head localization and an in-head localization in loudspeaker reproduction in a low-reflection room cannot be clearly established. Even small changes in the test signals lead either to IHL or to the perception of externalization. The perception of externalization can be understood as a dichotomous quality feature based on the results of this experiment. Plenge (1972) states that an in or at the head localization occurs when there is a "missing, inadequate or incorrect sound source and sound field knowledge and/or the signals and thus the stimuli are of such a nature that they cannot be assigned to any stimulus pattern contained in the long-term memory". The results suggest that the quality feature externalization is influenced by the context of the playback and the listening situation. To acquire a deeper understanding how prior-listening experience alters the perception of externalization the following listening tests were conducted. This way it may be possible to find methods to estimate the role of context parameters in quality perception. The results could help to design listening tests which are closer to real-life experiences of auditory augmented or virtual realities.

LISTENING TESTS AND RESULTS

This section outlines two studies which indicate an effect of prior room exposure on the perception of externalization.

Externalization rating

The evaluation of externalization of an auditory event is performed by selecting a inner, middle or outer region on rating sheet similar to Figure 1. The following definitions are used for the individual areas: a) mid-point: "The auditory event is completely in my head and very diffuse."; b) inner circle: "The auditory event is completely in my head and easy to locate."; c) mid circle: "The auditory event is external but very close to my ears or head."; d) outer circle: "The auditory event is external and easy to locate."; e) outer point: "The auditory event is external and very diffuse." According to the definition, externalization occurs when the auditory event is perceived outside the geometric extension of the head. The position of the perceptive decision point varies from person to person. For example, one person may already have an unambiguous degree of externalization if the auditory event is located very close to the head, while another person may still have an unambiguous degree of in-the-head localization. On the basis of the individual decision, externalization is evaluated on a multi-point scale. In the current study, the externalization index is calculated by dividing the number of external ratings d) and e) by the number of all ratings.

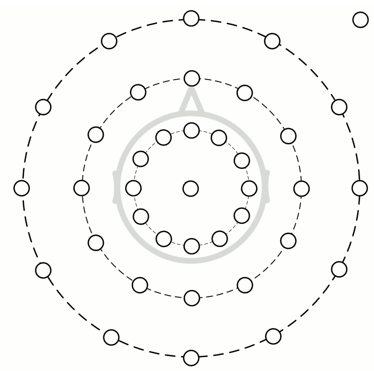

Fig. 1: Graphical user interface (GUI) for externalization and localization rating. Concentric rings are declared as in-head localization (1st ring), near but outside the head localization (2nd ring), and outside the head localization (3rd ring).

Listening test design

In order to provoke the room-divergence effect in the experiments discussed here, two rooms of similar size but strongly differing reverberation time and direct-to-reverberant ratios were chosen. The seminar room (SR) has a reverberation time at 1 kHz of 2 s and the listening lab (LL) has a reverberation time at 1 kHz of 0.339 s. For both listening tests, individual binaural room impulse responses (BRIRs) and headphone compensation filters were recorded in both rooms. The first listening test was conducted with a static binaural synthesis system. For the second test a dynamic binaural synthesis was realized by using the Smyth Realizer (Smyth *et al.*, 2008). For both studies speech and saxophone signals were used.

Excerpt from study no. 1

For the first listening test 31 participants were randomly separated into two groups, each trained to one of the rooms. The “convergent group” was trained to the SR room by listening to real loudspeakers in this room (LS) and a binaural synthesized stimuli of these loudspeakers (Synth SR). The “divergent group” was trained by listening to binaural synthesized stimuli of loudspeakers from the LL room (Synth LL). After the training, both groups were faced with the familiar and unfamiliar room condition to measure how the training sessions would influence the externalization ratings. The training was designed as a simple localization task accompanied with a judgment on the perceived level of externalization. Next to the rating task, visual feedback was provided at the correct source position by visually highlighting a loudspeaker model. After training, the listeners had to rate externalization of stimuli from the familiar and unfamiliar room.

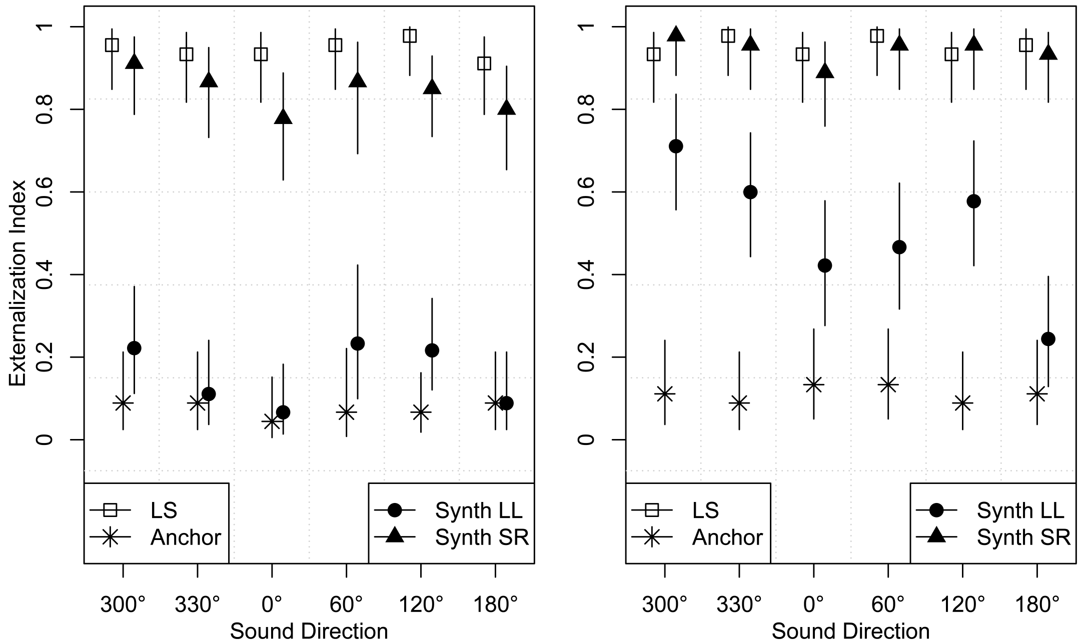

Fig. 2: Results for the externalization ratings of **left**, the convergent group and, **right**, the divergent group. Ratings are separated according to the direction of the presented sound. **Synth SR**, synthesis of the actual listening room (SR), **LS**, real loudspeakers in the listening room (SR), **Synth LL**, synthesis of acoustically dry listening room (LL) (Klein and Werner, 2017).

Figure 2 shows an excerpt of the test results. Ratings are separated according to the direction of the presented sound, because externalization is often direction dependent (see Werner *et al.* (2016)). The listening tests were conducted in the SR room. The low-quality anchor was measured with an omnidirectional microphone in both rooms and is aimed to provoke in-head localization. The results show high ratings for the actual room (LS and Synth SR) regarding to the externalization. The ratings for Synth SR of the convergent group is in tendency a bit lower than for the divergent group. Since this group listened to the Synth SR signals during training they might have discovered flaws of the binaural synthesis system. The rating of room LL is very different between the groups. A difference between Synth LL and Synth SR is clearly visible for both groups and relates to the room-divergence effect. It is particularly strong when the synthesis of an acoustically dry room is presented in a reverberant room as it was the case here. The ratings of Synth LL are significant higher for the divergent group than for the convergent group. This shows that the perceived externalization can shift according to the previous training session. In other words, the room-divergence effect highly depends on the listeners' acoustic experience.

Excerpt from study no. 2

The main aim of this study was to measure effects of head movements on the externalization (Werner *et al.*, 2017). In the original study several playlists were presented subsequently with and without head tracking enabled. Because of this test design, adaptation effects regarding externalization were expected. To avoid a mixup between the effects of head tracking and adaptation on the externalization rating, the room related adaptation was interrupted on purpose. Overall 36 participants were divided into a room convergent group which rated the synthesis of the actual listening test room (SR room), and a divergent group which rated the LL room while sitting in the SR room.

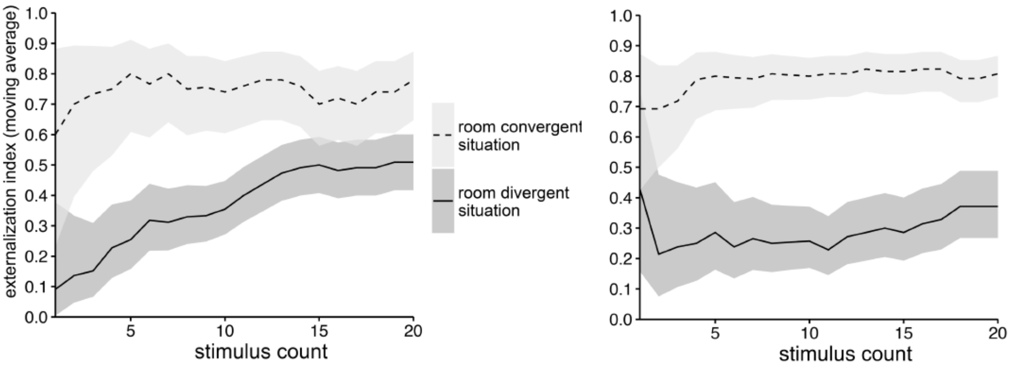

Fig. 3: Moving average for the externalization index over stimulus count and for two room situations. Gray area corresponds to 95% binomial confidence interval. Between first test (left) and second test (right) adaptation was interrupted by the presentation of 36 stimuli from loudspeakers.

In the first test each group had to rate the externalization of 20 test stimuli of Synth LL or Synth SR depending on which group they belong to. The ratings from the first test are shown on the left side of Figure 3. During this test session the ratings of the room divergent situation increases clearly with stimulus count. Before the second test run was conducted, 36 stimuli from the real loudspeakers in the SR room were presented in order to interrupt this adaptation behaviour. The second test run was identical to first one and the results are shown on the right hand side of Figure 3. Until stimulus five there are again some variations, but the ratings of the divergent situation never reach the values of the first test. Also, the increase over stimulus count is much less then in the first test. The results of this study show that a rapid room related adaptation can occur in such tests (at least regarding externalization) and that it is important to control this adaptation. It may be necessary to suppress this time-variant effect in order to study other quality elements such as the effect of specific system components (for example the benefit of head-tracking) on the perception of externalization.

CONCLUSION AND QUESTIONS

The listening tests show the effect of training to a specific room situation. The results indicate that externalization is influenced by prior knowledge and expectations about room acoustics. These results fit to the early findings of Toole (1970) and Plenge (1972). The experiments have shown that altering prior experience affects the perception of externalization. Listening in virtual acoustic environments and quality ratings thereof are strongly influenced by auditory adaptation effects. At this point the question arises whether there is a difference between the perception of real and virtual rooms. The studies on the precedence and Clifton effect indicate that the human auditory system needs time to interpret direct sound and its reflections in order to perceive a distinct position of a sound source. The relevant time frames in which an adaptation occurs is in order of a few milliseconds to seconds in these studies while in our studies the adaptation happens over the time span of several minutes.

Based on the listening experience in real rooms most people have probably never experienced a familiarization phase accompanied with in-head localization. So why is it the case in virtual acoustics environments? Experiments like those presented, create situations which normally do not exist. For example, rooms are switched with the press of a button while in reality there is always a transition phase when switching rooms. Furthermore, we hypothesize that an adaptation to new acoustic environments happens all the time. In reality a vast amount of acoustic information is available because mostly there are several sound sources at once and the listener also emits or creates sounds by walking in the room for example. In addition, all other senses also provide coherent information: visual and acoustic sound source positions match, the listener movement is translated into a change of the acoustic signal and so on. In laboratory experiments the amount of information which is provided to understand an acoustic scene is mostly limited. These limitations possibly require a longer time for our brain to understand the scene and as long the scene is not understood, perceptual errors like in-head localization are likely.

REFERENCES

Clifton, R. K. (**1987**), “Breakdown of echo suppression in the precedence effect,” J. Acoust. Soc. Am., **82**(5), 1834–1835.

Clifton, R. K. and Freyman, R. L. (**1989**), “Effect of click rate and delay on breakdown of the precedence effect,” Percept. Psychophys., **2**(49), 139–145.

Clifton, R. K., Freyman, R. L., Litovsky, R. Y., and McCall, D. (**1994**), “Listeners’ expectations about echos can raise or lower echo threshold,” J. Acoust. Soc. Am., **95**(3), 1525–1533.

Djelani, T. and Blauert, J. (**2001**), "Investigations into the Build-up and Breakdown of the Precedence Effect," Acta Acust united Ac., **87**(2), 253–261.

Freyman, R. L., Clifton, R. K., and Litovsky, R. Y. (**1991**), "Dynamic Processes in the Precedence Effect," J. Acoust. Soc. Am., **90**(2), 874–884.

Keen, R. and Freyman, R. L. (**2009**), "Release and re-buildup of listeners' models of auditory space," J. Acoust. Soc. Am., **125**(5), 3243 – 3252.

King, A. J. (**2008**), "Visual influences on auditory spatial learning," Philos. Trans. R. Soc. Lond., B, Biol. Sci., **364**(1515), 331–339.

Klein, F. and Werner, S. (**2017**), "Influences of training on externalization in binaural synthesis in situations of room divergence," J. Audio Eng. Soc., **65**(3).

Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. (**1999**), "The precedence effect," J. Acoust. Soc. Am., **106**(4), 1633–1654.

Mendonça, C. (**2014**), "A review on auditory space adaptations to altered head-related cues," Front. Neurosci., **8**(219), 1–14.

Moore, D. R., Halliday, L. F., and Amitay, S. (**2009**), "Use of auditory learning to manage listening problems in children," Philos. Trans. R. Soc. Lond., B, Biol. Sci., **364**, 409 – 420.

Plenge, G. (**1972**), "Über das Problem der Im-Kopf-Lokalisation [On the Problem of In Head Localization]," Acustica, **26**(5), 241–252.

Raake, A. and Blauert, J. (**2013**), "Comprehensive modeling of the formation process of sound-quality," Proc. Quality of Multimedia Experience (QoMEX).

Seeber, B. U., Müller, M., and Menzer, F. (**2016**), "Does learning a room's reflections aid spatial hearing?" Proc. ICA.

Smyth, S., Smyth, M., and Cheung, S. (**2008**), "Smyth SVS headphone surround monitoring for studios," 23rd UK Conference of the Audio Engineering Society.

Toole, F. E. (**1970**), "In-Head Localization of Acoustic Images," J. Acoust. Soc. Am., **48**(4B), 943–949.

Wallach, H., Newman, E. B., and Rosenzweig, M. R. (**1949**), "The Precedence Effect in Sound Localization," Am. J. Psychol., **62**(3), 315–336.

Werner, S., Götz, G., and Klein, F. (**2017**), "Influence of Head Tracking on the Externalization of Auditory Events at Divergence between Synthesized and Listening Room Using a Binaural Headphone System," Audio Engineering Society Convention 142.

Werner, S., Klein, F., Mayenfels, T., and Brandenburg, K. (**2016**), "A Summary on Acoustic Room Divergence and its Effect on Externalization of Auditory Events," Proc. Quality of Multimedia Experience (QoMEX).