Context-dependent quality parameters and perception of auditory illusions

STEPHAN WERNER[1,*], FLORIAN KLEIN[1], AND TAMÁS HARCZOS[2]

[1] *Technische Universität Ilmenau, Institute for Media Technology, Ilmenau, Germany*

[2] *Fraunhofer Institute for Digital Media Technology, Ilmenau, Germany*

This contribution introduces context-dependent quality elements, which have significant influence on perception of an auditory illusion. Binaural synthesis of an acoustic scene via a personalized headphone system is used. The investigated elements are divergent between synthesized scene and listening room, visibility of the scene, and personalization of the system. Two rooms with different acoustic parameters are used as recording and listening room. The test persons listen either to the same room as the listening room or to the other room. The plausibility of the perceived auditory scene is described by the probands with the help of the parameter perceived externality of the auditory event. Because it is unknown if the relevant quality elements are acoustically or visually based, two groups of test persons are used. The first group has no visual cues (dark room), while the second group sees the synthesized source positions and listening room. We have found significant differences in perceived externality depending on the synthesized and listening room, on the two groups, and on personalization of the system.

MOTIVATION

The development of audio systems is motivated by the purpose to create perfect auditory illusions with a high degree of immersion and plausibility (Heeter, 1992; Lindau and Weinzierl, 2011). A lot of work is done to increase the technical quality of such systems. Systems which use the principles of binaural synthesis are one possibility to achieve auditory illusion. Binaural synthesis takes the underlying perceptual processes conditioned by the direct synthesis of the corresponding sound pressure at the ear drums of a listener into account. The technical parameters are therefore well understood and controllable (see, e.g., Hess, 2006; Silzle, 2007). Sound sources in rooms can be described by binaural room impulse responses (BRIRs). The BRIRs can be derived from acoustic room simulations or from measurements of real sound sources in real rooms. A personalization of the binaural system is achievable by using individual BRIRs and individual headphone equalization for example. In addition to the technical realization of the correct binaural synthesis and signals, many psychoacoustic effects in perception of auditory scenes and their interconnections are not completely understood until now.

*Corresponding author: stephan.werner@tu-ilmenau.de

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

Such effects cover for example multimodal interactions between acoustical and visual stimuli like the McGurk-effect (McGurk and MacDonald, 1976) or the ventriloquism-effect (Bertelson and Radeau, 1981; Seeber and Fastl, 2004; Werner *et al.*, 2012). Other perceptual effects depending on the congruence or divergence between the synthesized scene (including room) and the listening situation also seem to have a not neglectable influence on perception (Werner and Siegel, 2011). The quality of experience of an audio reproduction system depends on technical quality elements of the system but also on context-dependent quality parameters. To contribute to the improvement of binaural synthesis this paper focuses on investigations on acoustic divergence between listening room and synthesized room, visibility of the listening room and simulated source positions, and on personalization of the binaural synthesis system. The quality of experience is measured with listening experiments. The ratings of perceived externalization of the auditory event are shown. However, this quality feature is only one possible feature that has an influence on a plausible perception of an auditory illusion (Raake and Blauert, 2013).

BINAURAL SYNTHESIS VIA HEADPHONES

For generating test stimuli, binaural recordings of individual and ‘mean’ (manikin KEMAR) BRIRs for the used rooms and sound source positions and the auralization via headphones were prepared. The binaural system was customized for each participant to avoid within-cone and out-of-cone of confusion errors (Møller *et al.*, 1996) and to increase the simulation’s similarity compared with the real loudspeakers (Begault and Wenzel, 2001). A listening lab and seminar rooms with defined room acoustics and an adequate source-receiver distance were chosen to include reverberation. Reverberation encourages the perception of externalization of an auditory illusion and the impression of distance (Laws, 1973). The headphones were equalized using individual headphone transfer functions (HPTFs) if individual BRIRs were used. HPTFs from the head-and-torso simulator (KEMAR) were used if ‘mean’ BRIRs were used. In-ear microphones were used to measure individual BRIRs and HPTFs at the entrance of the blocked ear canal of each subject. The microphones are not removed between the BRIR and HPTF measurements. The measurements of the HPTFs were averaged over five recordings, repositioning the headphones for each recording. The inverse of a HPTF was calculated by a least-square method with minimum phase inversion (Schärer and Lindau, 2009). The measurements of the BRIRs were averaged over three recordings. Stax Lambda Pro headphones were used for playback.

OBJECTS OF INVESTIGATION

The listening experiments were focused on the evaluation of context-dependent quality parameters and their influence on the perception of externality of the auditory event. Two listening tests were conducted. Both tests investigated the combinations of listening room and synthesized room. Additional context-dependent quality parameters like visibility of the listening room and personalization of the binaural synthesis were investigated in the first test. The second test was focused on perceived externalization depending on different distances of the synthesized sound source. Binaural recordings of non-individual BRIRs (KEMAR head-and-torso simulator) were prepared for the used rooms and sound source positions to generate the test stimuli in the second test.

Acoustic divergence between rooms

A listening lab (Rec. ITU-R BS.1116, V = 179 m³, RT60distance (2m) = 0.16 s), a depleted seminar room (V = 182 m³, RT60reference distance (2m) = 1.4 s), and another seminar room (V = 182 m³, RT60distance (2m) = 0.9 s) with different room acoustic characteristics were used for the listening tests and the measurement of the BRIRs at a distance of 2.2 m. The tests were conducted in the same listening lab (HL) and the same seminar rooms (SR) to evaluate the influence of the listening situation. The left part of Fig. 1 shows the combinations of listening room and synthesized room used in the tests.

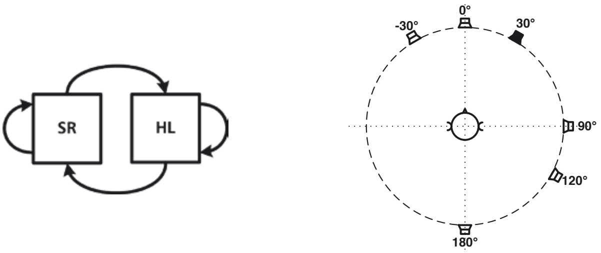

Fig. 1: Left: Combinations of listening room and synthesized room used in the listening tests; SR = seminar room, HL = listening lab; Right: Positions of the binaural synthesized sound sources for playback via headphones; distance of the sources to the listener (midpoint of the figure) approx. 2.2 m; the filled position (30°) was used in test two.

Visibility of the listening room

The test persons were randomly divided into two groups depending on the presence of visual cues within the tests. For the first group the illumination of the listening rooms was minimized (nearly complete darkness) and a sound-transparent black curtain with a distance of 2.2 m was arranged around the test persons. The test persons should have no visual impression or visual cues of the listening room. The test persons in the second group were placed in the illuminated listening rooms and dummy loudspeakers were placed at each hour position on a clock-like circle to provide additional visual cues. This situation was also used in the second test.

Sound source positions

Five sound source directions were checked for test one and one direction was used in test two. A Genelec 1030A loudspeaker was used to measure the BRIRs for each position. The right part of Fig. 1 shows the different positions. The distance from the loudspeaker to the listening point was approx. 2.2 m for test one and two. The height of the source position was approx. 1.3 m (ear position of a sitting person). The BRIRs for each position and for each test person were recorded in the two rooms. The recording position was the same as the listening position in the test.

Personalization of the binaural synthesis system

The individual BRIRs of the test persons from the two rooms and source directions and the individual headphone transfer function were recorded in a preceding session. Furthermore, the BRIRs and HPTFs of a KEMAR head-and-torso simulator (45BA) were recorded. Both the individual and 'mean' BRIRs were used to create the binaural test stimuli for test one. For test two only the 'mean' BRIRs were used.

LISTENING TESTS

Test one: The listening test was conducted in the listening lab and the seminar room separately in two sessions at different days. In every session every test person listened to individually synthesized and dummy-head synthesized source positions of both recording rooms. The stimuli were presented two times in a random order. The perceived incidence angle could be rated by choosing the respective direction on a top-down view. Externalization could be rated by choosing the midpoint, inner circle, or outer circle. The attribute externalization was oriented to definitions given by Hartmann and Wittenberg (1996). The following definitions were used in the test: a) midpoint: "The sound event is entirely in my head or it is very diffuse."; b) inner circle: "The sound event is external but it is next to my ears or head."; c) outer circle: "The sound event is external and good locatable." Note that the definitions were given in German.

Test two: The test persons rated the externalization in the listening test. The test persons indicated the externality of the auditory event by pressing one of three buttons on a graphical user interface. The same scale as in test one was used. The synthesized BRIRs of several distances from the listening lab and the seminar room were used as stimuli. A more detailed description about the BRIR synthesis and the test design can be found in Werner and Sass (2013).

Twenty-one test persons participated in the first and 16 test persons in the second listening test. The test persons were well experienced with listening tests and were trained before each test. For the first test the training consisted of an oral and written introduction and a definition of the used attributes localization and externalization. Each subject had to listen to all different test items. The test persons could compare each item with the others and could listen to each item several times. The test persons had to rate each test item on the same rating sheet as in the main test session. For the second test a presentation of non-binaural stereo panned signals, a playback via the reference loudspeaker, and a binaural synthesis of the reference loudspeaker were used as training. The test persons should build up an own internal reference and had to define differences between the items for the attributes localization and externalization.

RESULTS

The ratings of the test persons for externalization were counted as frequencies. The frequencies showed no significant dependency from the used sound signal. Both signals were put together for analysis. An externalization index was calculated as ratio between the ratings of extern (outer circle on the rating sheet) and all ratings within the test. An index of 0 indicates in-head localization, while an index of 1 indicates out of the head localization of the auditory event.

Figure 2 shows the rating of perceived externalization depending on the presence of visual cues, personalization method, and combinations of listening room and synthesized room. The midpoints of the polar plots represent an externalization index of 0 while the outer circle represents an index of 1 (linear scale in between). Wilcoxon signed rank tests at the 5% confidence level were conducted for statistical testing. The upper row of Fig. 2 shows the externalization indexes for the reverberant seminar room as listening room, while the lower row shows the ratings for the less reverberant listening lab as listening room.

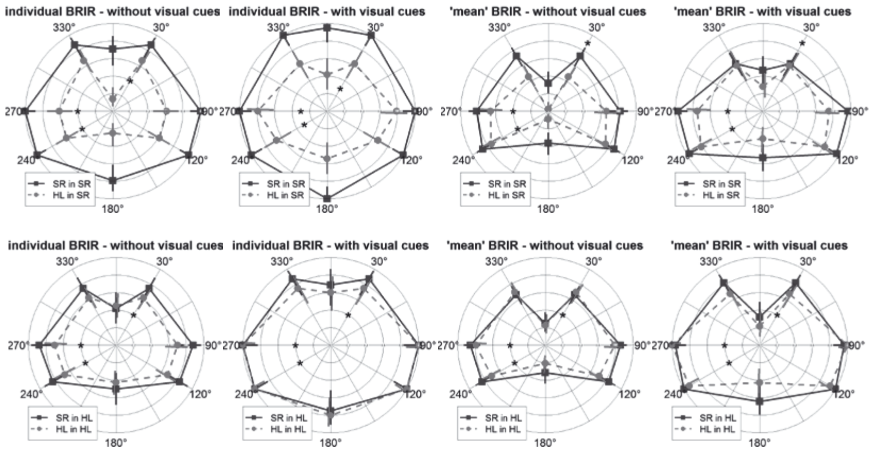

Fig. 2: Ratings for perceived externalization as externalization indexes depending on the combinations of listening room and synthesized room, personalization of the binaural synthesis, and presence of visual cues with 95% confidence intervals from test 1; SR = seminar room, HL = listening lab, * are the mirrored ratings at the 0° to 180° axis.

A general lower externalization index is achieved for binaural synthesis using 'mean' BRIRs compared to individual BRIRs. Very low indexes are visible especially for the direct front and back directions. The usage of an individualized synthesis increases the perceived externalization of the auditory event significantly for the direct front and back directions. Furthermore, a higher index is visible for congruence between the listening room and synthesized room (SR in SR) related to divergence between the rooms (HL in SR). This effect is mostly significant if individual BRIRs are used. The ratings show no significant differences if 'mean' BRIRs are used for the synthesis. However, the magnitude of the indexes is decreased compared to individual BRIRs at congruence between listening and synthesized room (SR in SR). The room effect is maybe covered by the effect caused by the personalization of the binaural synthesis. Further research is needed to determine the interconnection between these two context-dependent quality elements. The effect caused by room divergences seems to be independent of the visibility of the listening room. However, the visibility of the room increases the indexes especially for the front and back directions. The lower row of Fig. 2 shows the ratings of test one for the less reverberant listening lab as listening room. Significant differences depending on divergence or congruence between the listening and synthesized room are visible in contrast to the seminar room as listening room for the direct front and back directions. The visibility of the room also increases the externalization indexes for all conditions. The room effect seems to be much more present for synthesis of a less reverberant scene in a more reverberant room.

Figure 3 shows the rating as externalization index for different combinations of listening room and synthesized room and additionally for different distances of the auditory event. A similar effect of dependencies of the rooms is visible as in test one. Clearly higher ratings are reached if the synthesized room is the same as the listening room especially for the more reverberant seminar room (SR in SR compared to HL in SR). The source distance of one meter is rated with the lowest externalization indexes while the more far away distances are rated with higher values. Saturation is visible for the synthesis of the seminar room but not for the less reverberant listening lab. An increase of the externalization index is visible for synthesis of the listening lab in the listening lab (HL in HL) compared to the synthesis of the listening lab in the seminar room (HL in SR) for the distance of 5 m. The ratings of test two are consistent with the ratings of test one for the 2.2-m distance, 30° direction, and 'mean' personalization of the binaural synthesis.

CONCLUSIONS

The ratings from two listening tests to evaluate the perceived externalization of an auditory event using a binaural auralization via headphones were reported. Five source positions, four combinations of listening room and synthesized room, and two personalization methods were investigated. A dependency of the perceived externalization of an auditory event from the used personalization method was shown. Higher externalization indexes are reached especially for the direct front and back direction and for the frontal lateral direction. This is in contrast to own former investigations (Werner and Siegel, 2011). It would be insightful to investigate the correlation between externalization and errors in perception of direction.

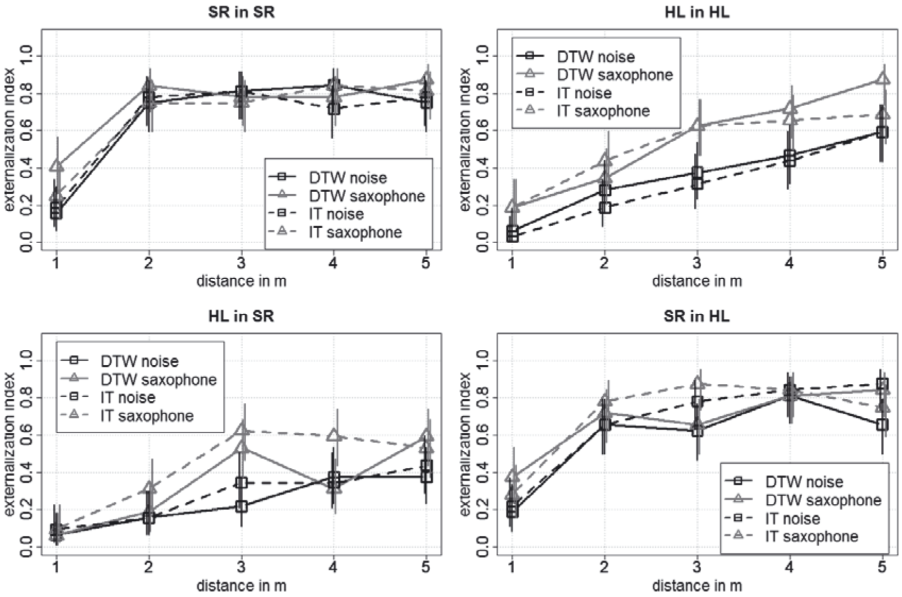

Fig. 3: Perceived externalization depending on distance of synthesized sound sources from test 2; BRIRs of different distances using interpolation methods (Werner and Sass, 2013) with 95% confidence intervals; azimuth of source direction = +30°; IT = interpolation in time domain; DTW = interpolation + dynamic time warping; 1 m = measured start-BRIR; 5 m = measured target BRIR; HL = listening lab, SR = seminar room.

Furthermore, low externalization indexes were found for synthesis of the less reverberant room in the more reverberant room. The highest externalization indexes were found for playback of test signals from the reverberant room in the same room. The personalization method maybe covers the room effect. The interconnection between personalization and room divergences is not well-known until now. The presence of visual cues has a supporting effect on the perceived externalization independent of the personalization method and combination of listening and synthesized room. The effect of perceived externalization depending on room divergences seems to be an acoustically based context-dependent quality element. Further investigations in evaluation of detailed quality elements based on a variety of plausibility features are meaningful.

REFERENCES

Begault, D.R., and Wenzel, E.M. (**2001**). "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," J. Audio Eng. Soc., **49**, 904-916.

Bertelson, P., and Radeau, M. (**1981**). "Cross-modal bias and perceptual fusion with auditory-visual spatial discordance," Percept. Psychophys., **29**, 578-584.

Hartmann, W.M., and Wittenberg, A. (**1996**). "On the externalization of sound images," J. Acoust. Soc. Am., **99**, 3678-3688.

Heeter, C. (**1992**). "Being there: The subjective experience of presence," in *Presence: Teleoperators and Virtual Environments* (MIT Press).

Hess, W. (**2006**). *Time-Variant Binaural-Activity Characteristics as Indicator of Auditory Spatial Attributes*, PhD Thesis, Ruhr-Universität Bochum, Bochum, Germany.

Laws, P. (**1973**). "Entfernungshören und das Problem der Im-Kopf-Lokalisiertheit von Hörereignissen [Auditory distance perception and the problem of 'in-head localization' of sound images]," Acustica, **29**, 243-259 (NASA Technical Translation TT-20833).

Lindau, A., and Weinzierl, S. (**2011**). "Assessing the plausibility of virtual acoustic environments," Forum Acusticum, European Acoustic Association, Aalborg, Denmark, pp. 1187-1192.

McGurk, H., and MacDonald, J. (**1976**). "Hearing lips and seeing voices," *Nature*, **264**, 746-748.

Møller, H., Sørensen, M.F., Jensen, C.B., and Hammershøi, D. (**1996**). "Binaural technique: Do we need individual recordings?" J. Audio Eng. Soc, **44**, 451-469.

Raake, A., and Blauert, J. (**2013**). "Comprehensive modeling of the formation process of sound quality," 5th Int. Workshop on Quality of Multimedia Experience (QoMEX), Klagenfurt, Austria, pp.76-81.

Schärer, Z., and Lindau, A. (**2009**). "Evaluation of equalisation methods for binaural signals," Proc. of the 126th AES Conv., preprint 7721.

Seeber, B., and Fastl, H. (**2004**). "On auditory-visual interaction in real and virtual environments," Proc. ICA 2004, 18th Int. Congress on Acoustics, Kyoto, Japan, volume III, Int. Commission on Acoustics, pp. 2293–2296.

Silzle, A. (**2007**). *Generation of Quality Taxonomies for Auditory Virtual Environments by Means of Systematic Expert Survey*, PhD Thesis, Ruhr-Universität Bochum, Bochum, Germany.

Werner, S., and Siegel, A. (**2011**). "Effects of binaural auralization via headphones on the perception of acoustic scenes," Proc. of the 3rd International Symposium on Auditory and Audiological Research, ISAAR, Denmark, pp.215-222.

Werner, S., Liebetrau, J., and Sporer, T. (**2012**). "Audio-visual discrepancy and the influence on vertical sound source localization," Quality of Multimedia Experience (QoMEX), Fourth International Workshop, Australia, pp.133-139.

Werner, S., and Sass, R. (**2013**). "Synthesis of binaural room impulse responses," AIA-DAGA Merano 2013, pp. 572-575.