## ACKNOWLEDGEMENT

The project was approved by The Danish National Committee on Biomedical Research Ethics. Financial support as been received from Danish Working Environment Research Fund no. 20070014504 and Ear Nose and Throat specialist Hans Skouby's and Hustru Emma Skouby's Foundation.

## REFERENCES

Atcherson, S. R., Martin, M. J., and Lintvedt, R. **(2008)**. "Contralateral noise has possible asymmetric frequency-sensitive effect on the 2F1-F2 otoacoustic emission in humans." Neurosci. Lett., **438**, 107-110.

Axelsson, A. and Lindgren, F. **(1978)**. "Temporary threshold shift after exposure to pop music" Scand. Audiol., **7**, 127-135.

Axelsson, A. and Lindgren, F. **(1981)**. "Pop music and hearing" Ear Hear., **2**,64-69.

Brashears, S. M., Morlet, T. G., Berlin, C. I., and Hood, L. J. **(2003)**. "Olivocochlear efferent suppression in classical musicians" J. Am. Acad. Audiol., **14**, 314-324.

Clark, W. W. **(1991)**. "Recent studies of temporary threshold shift (TTS) and permanent threshold shift (PTS) in animals" J. Acoust. Soc. Am., **90**, 155-163.

Darrow, K. N., Maison, S. F., and Liberman, M. C. **(2006)**. "Cochlear efferent feedback balances interaural sensitivity" Nat. Neurosci., **9**, 1474-1476.

Lindgren, F. and Axelsson, A. **(1983)**. "Temporary threshold shift after exposure to noise and music of equal energy" Ear Hear., **4**, 197-201.

Maison, S., Micheyl, C., and Collet, L. **(2001)**. "Influence of focused auditory attention on cochlear activity in humans" Psychophysiology, **38**, 35-40.

Micheyl, C., Carbonnel, O., and Collet, L. **(1995)**. "Medial olivocochlear system and loudness adaptation: differences between musicians and non-musicians" Brain Cogn., **29**, 127-136.

Schmidt, J. H., Pedersen, E. R., Juhl, P. M., Christensen-Dalsgaard, J., Andersen, T. D., Poulsen, T., and Baelum, J. **(2011)**. "Sound Exposure of Symphony Orchestra Musicians" Ann.Occup.Hyg.,[Doi:10.1093/annhyg/mer055,pp 1-13].

Strasser, H., Irle, H., and Legler, R. **(2003)**. "Temporary hearing threshold shifts and restitution after energy-equivalent exposures to industrial noise and classical music" Noise. Health, **5**, 75-84.

Swanson, S. J., Dengerink, H. A., Kondrick, P., and Miller, C. L. **(1987)**. "The influence of subjective factors on temporary threshold shifts after exposure to music and noise of equal energy" Ear Hear., **8**, 288-291.

Wagner, W., Heppelmann, G., Kuehn, M., Tisch, M., Vonthein, R., and Zenner, H. P. **(2005)**. "Olivocochlear activity and temporary threshold shift-susceptibility in humans" Laryngoscope, **115**, 2021-2028.



# Effects of binaural auralization via headphones on the perception of acoustic scenes

STEPHAN WERNER[1] AND ANDRÉ SIEGEL[1]

[1] *Institute for Media Technology, Ilmenau University of Technology, D-98693 Ilmenau, Germany*

The auralization of an acoustic scene can be realized with the presentation of binaural signals via headphones. One of the biggest challenges is the individualization of the headphone equalization and the generation of the binaural signals. A promising way is the use of probe microphones for equalization and recording. Very good results in terms of externalization and correct reproduction of the acoustic scene can be reached. However, former investigation indicates consistently that perceived acoustic illusion is much more plausible if the recording and the playback conditions are similar or even the same. Within this contribution we present a fully individualized binaural auralization system via headphones. Binaural recordings of sound sources on different representative positions in two real rooms with distinct different room acoustics are made. These recordings are presented via headphones to test persons. A series of listening tests show the expected influence of an accurate individualization on the correct localization of the synthesized sound source. Furthermore, a strong influence of congruence between the room acoustics of listening and recording room on the perception of the scene was observed. We can show that there is a significant decrease of perceived externalization if the listening rooms and the recording rooms are different.

## MOTIVATION

Observations from former investigations (e.g., Møller *et al.* (1999) and Klein and Werner (2011)) consistently show that perceived acoustic illusion is much more plausible if the recording conditions are similar or even the same as the listening conditions. At first, the influence of room acoustics is investigated. Other dependencies like audio-visual effects (Abou-Elleal (2003)) and adaptation effects are investigated consecutively. However, for this contribution we dispose the following hypothesis. H1: Perceived externalization of a single sound event synthesized by a binaural headphone system is less if the room acoustics of the listening room does not match the room acoustic of the recording room. This hypothesis is verified with listening tests. In this study, the term plausibility describes the perceived quality features of localization accuracy and externalization of the sound event regarding to different room conditions, customization methods, and sound source positions.

Proceedings of ISAAR 2011: Speech perception and auditory disorders. 3rd International Symposium on Auditory and Audiological Research. August 2011, Nyborg, Denmark. Edited by T. Dau, M. L. Jepsen, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN 87-990013-3-0. EAN 9788799001330. The Danavox Jubilee Foundation, 2012.

## BINAURAL AURALIZATION VIA HEADPHONES

The binaural system contains the recordings of individual and dummy head binaural room impulse responses (BRIRs) for the used rooms and sound source positions and the auralization via headphones. The system is customized for a dummy head or individually for the test persons. The headphones are equalized using individual or dummy head headphone transfer functions (HPTFs). Probe microphones are used to measure individual BRIRs and individual HPTFs next to the eardrum of each proband; see figure 1 (Sass et al., (2010)). Dummy head HPTFs are applied if dummy head BRIRs are used. The measurements of the HPTFs are averaged over five recordings with repositioning of the headphones. The inverse of a HPTF is calculated by a least-square method. The measurements of the BRIRs are averaged over three recordings. A Stax Lambda Pro headphone is used for playback.

**Fig. 1:** Methods for measurements of binaural transfer functions; left and middle: self-made probe microphones for individual customization, right: dummy head (Neumann Ku-80) customization

## EXPERIMENTAL DESIGN

The experimental design contains five different sound source positions, two different rooms, two test signals and dummy head or individual customization. A rating method for discrete positions is used.

### Sound source positions

Five sound source positions are checked. A Genelec 1030A loudspeaker is used to measure the BRIRs for each position. Figure 2 shows the different positions. The distance from the loudspeaker to the listening point is ca. 2.2 m. The height of the source position is ca. 1.3 m (ear position of a sitting person). The different positions are chosen to validate common errors of the binaural auralization system: front-back confusion (0° and 180°), within cone-of-confusion errors (-30° or 120°), and localization (without cone-of-confusion) errors. A clock style notation is used for the different positions during the listening test and presentation of the results. The BRIRs for each position and for each test person are recorded in the two rooms. The recording position is the same as the listening position in the test.





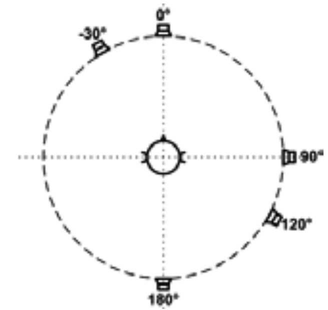


**Fig. 2:** Positions of the used sound sources for playback via headphones; distance of the sources to the listener (midpoint of the figure) = ca. 2.2 m.

### Conditions under test

Two rooms are used for recording and playback. Room 1 is a standardized listening lab at Ilmenau University of Technology. The lab fulfils the EBU Tech. 3276. In the following the room is named as dry room. Room 2 is an empty seminar room. The mean reverberation time at the listening position for room 2 is distinct longer than for room 1. Furthermore, room 2 shows strong early reflections from the walls. This room represents a wet room. Table 1 shows the mean reverberation time and the lateral energy fraction ($LF_{E4}$) for both rooms. The lateral energy fraction is an objective measurement parameter which indicates perceived spaciousness and is strongly correlated with the apparent source width. While in concert hall situations LF is measured with sources on stage and receivers on listening area, in this paper the mean LF was determined for the listening position over all source positions for four octaves (125 Hz, 250 Hz, 500 Hz, and 1 kHz). The measurements where performed according the standard ISO 3382-1:2009.

| | *Room 1* | *Room 2* |
|---|---|---|
| $T_{60}$ | ca. 0.3 s | ca. 1.4 s |
| $LF_{E4}$ | 0.08 | 0.15 |

**Table 1:** Room acoustic parameters for room 1 and room 2; mean reverberation time ($T_{60}$) and mean lateral energy fraction ($LF_{E4}$) over four octave bands.

The used signals for the listening tests are a female singing voice and a dry saxophone part. Both signals are broad band signals and faded at the beginning and the end. The duration of each signal is six seconds.

### Test persons and training

Three female and three male normal hearing persons with the age between 24 and 32 years participated in the listening tests. The test persons are well experienced in listening tests and are trained before each test. The training consists of an oral and written introduction and a definition of the used attributes localization and externalization. Each subject has to listen to all different test items in a random and unknown order. The proband can compare each item with the others and can listen to each item several times. The subject have to rate each test item on a discrete rating sheet. Diotic representations of dry test signals are given as known external reference for in-head localization. The test persons should build up an own internal reference and have to define differences between the items for the attributes localization and externalization.

## EXPERIMENTAL PROCEDURE

The listening tests are conducted in room 1 and room 2 separately in two sessions on different days. In every session every test person listens to individually synthesized and dummy head synthesized source positions of both recording rooms. Figure 3 shows the room conditions under test. The stimuli are represented six times in random order. Additionally, the unprocessed test signals are presented six times. These stimuli represent the reference for in-head localization of the sound event. Dummy loudspeakers are placed at the twelve possible source positions as visual foothold. The test persons are instructed to look straight forward and keep still during listening.

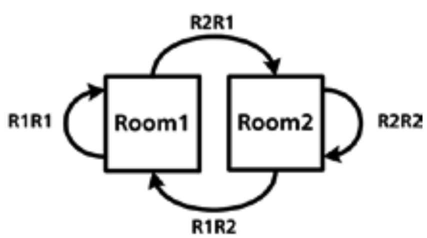


**Fig. 3:** Room conditions under test (e.g. R2R1 means playback of recordings from room 1 in room 2)

A session is divided into three parts. The first part contains the training of the test person. The second and third part includes three repetitions of the test stimuli respectively. A whole session takes ca. 100 minutes. The number of test stimuli is 252 per test person. Each stimulus is rated on the rating sheet shown in figure 4. The perceived incidence angle can be rated by chose the respective direction from one to twelve. Externalization can be rated by chose the midpoint, inner circle, or outer circle. The attribute externalization is oriented to definitions given by Hartmann and Wittenberg (1996). Following definitions are 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 extern but it is next to my ears or head.”, c) outer circle: “The sound event is extern and good locatable.” (note that the definitions are given in German). At first, the test persons have to listen to the whole stimulus and make their rating after listening.





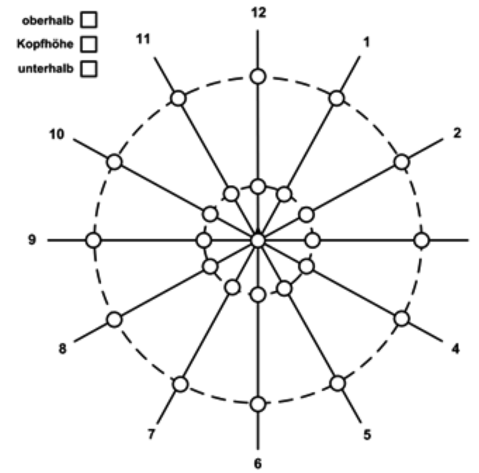


**Fig. 4:** Rating sheet for the listening test based upon Hammershøi (2009) and Møller *et al.* (1996).

## RESULTS

The ratings of the test persons for localization and externalization are counted as frequencies. The frequencies show low dependency from the used sound signal. Both signals are put together for analysis. All test persons found the reference signals (diotic presentations). The test on normal distribution is significant for some conditions (Shapiro-Wilk, $p<0.05$). Normal distribution cannot be assumed in general. Therefore, non-parametric statistics are used consequentially.

### Localization

The analysis of frequencies for the localization task shows strong differences between the individual and dummy head customization for the different source positions. Dummy head customization shows stronger sensitivity to front-back-confusions, within, and without cone-of-confusion errors. The correlation coefficients of the different room conditions and customization methods related to an ideal localization are shown in table 2.

| *Condition* | *R1R1* | *R1R2* | *R2R1* | *R2R2* |
|---|---|---|---|---|
| *Correlation Coefficient (individual)* | 0.8569 | 0.7112 | 0.8584 | 0.8584 |
| *Correlation Coefficient (dummy head)* | 0.7150 | 0.6175 | 0.6131 | 0.5671 |

**Table 2:** Correlation coefficients (*rho*) for conditions and customization methods related to ideal localization (correlation after Spearman; $p<0.05$)

A clear dependency between source positions and different room conditions is not apparent. Figure 5 shows the deviation in degrees of the ratings for the localization task for all four room conditions and all five source positions as a clockwise notation.

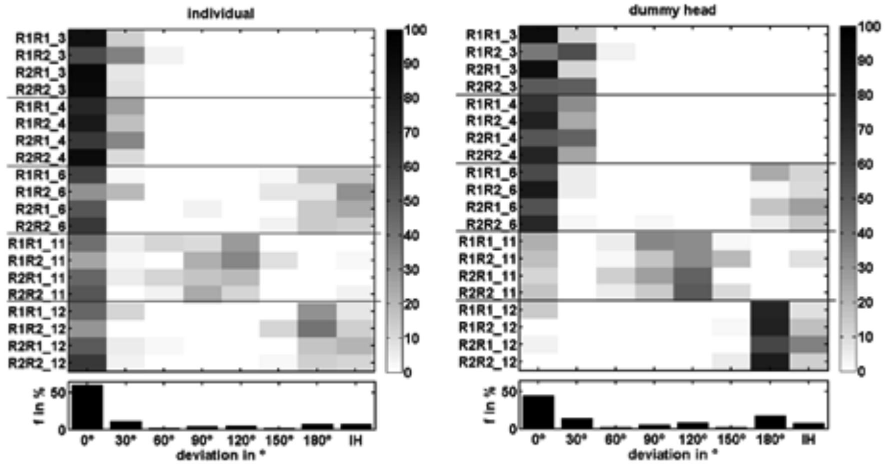


**Fig. 5:** Deviation of localization ratings in degree; top: different conditions and incidence angles indicated by numbers, bottom: summation over all conditions; frequency (*f*) in percent; left: individual customization, right: dummy head customization.

### Externalization

The analysis of frequencies for the externalization task shows no significant dependency from the kind of customization (Wilcox, $p<0.05$). Therefore, all ratings are put together for analysis of externalization. The externalization index is defined as ratio between the sum of ratings for total externalization (outer circle in figure 4) and the total number of ratings. Figure 6 shows the index for the four room conditions and the five source positions. Significant differences are found for condition R2R1 compared to all other room conditions and for all source positions (Wilcox, $p<0.05$). The hypothesis H1 can be accepted for playback of room 1. No consistent significant differences (Wilcox, $p>0.05$) can be found for conditions R1R1 and R1R2 (except source position four o´clock). Furthermore, no significant differences are found for R1R2 and R2R2. H1 cannot be accepted for playback of room 2. Statements from the probands confirm the results. They report consistently that it is more plausible to auralize the wet room (room 2) in the dry room (room 1) than the dry room in the wet room. The highest plausibility is reached if the recording room is the same as the playback room. The expected effect of a low externalization index, if the recording and playback rooms are different, can only be confirmed for the wet playback room.





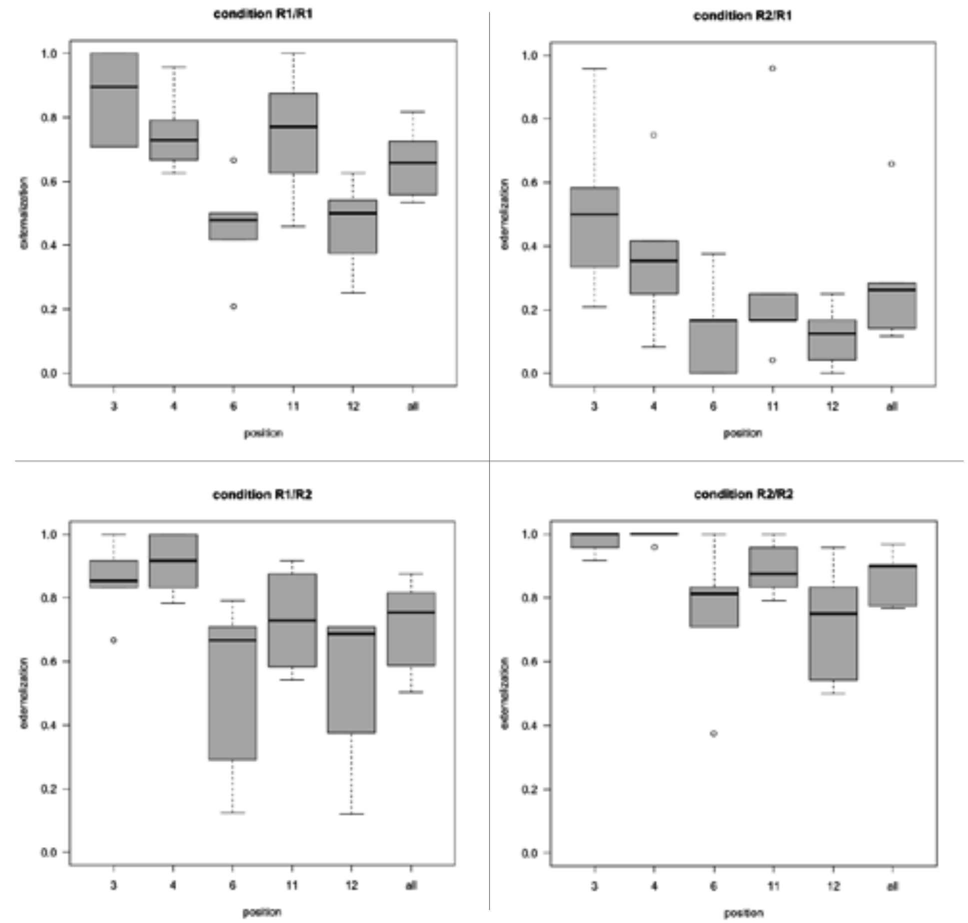


**Fig. 6:** Externalization index for room conditions R1/R1, R1/R2, R2/R1, R2/R2 and different positions; "all" indicate the mean rating over all positions; boxplots with medians, 25%, 75% quantile, and whiskers are shown

## CONCLUSIONS

A listening test with six test persons was accomplished to evaluate the localization accuracy and the perceived externalization of a binaural auralization via headphones. Five source positions, four combinations of playback and recording rooms, and two customization methods were investigated. The used rooms have distinct different room-acoustic parameters. A dependency of the used customization method of the binaural system can be shown for localization accuracy but not for the externalization task. A low externalization index was found for playback of test signals from the dry room in the wet room. A high externalization index was found for playback of test signals from the wet room in the same wet room. Lower indices were found for the other two conditions. Furthermore, a tendency becomes apparent that the externalization index is higher for playback of test signals from the reverberant room than from the dry room. The analysis indicates the assumed effect of room acoustics on binaural auralization. Further investigations with subtle graduation of room acoustic parameters, like the strength and ratio of early reflections related to the direct sound, are meaningful. Furthermore, the multimodal divergence of audio and visual context between recording and playback conditions are less investigated but seemed to play an important role in a plausible perception of a synthesized scene.

## REFERENCES

Klein, F. and Werner, S. (**2011**). "Individualization of head-related transfer functions", 37th annual convention for acoustics, DAGA, Düsseldorf.

Abou-Elleal, E. R. E. (**2003**). "Raumakustik – Interaktion visueller und auditiver Wahrnehmungen", dissertation, RWTH Aachen.

Sass, R., Werner, S., and Siegel, A. (**2010**). "Comparison of recording methods for measurements of head-related transfer functions". 26th, VDT Int. Convention, Tonmeistertagung, Leipzig.

Merimaa, J. and Hess, W. (**2004**). "Training of Listeners for Evaluation of Spatial Attributes of Sound". Proc. of 117th AES Conv., preprint 6237, San Francisco.

Hammershøi, D. (**2009**). "Human localization and performance measures". in Proceedings of ISAAR 2009: *Binaural Processing and Spatial Hearing*. 2nd International Symposium on Auditory and Audiological Research ISAAR, Denmark. Edited by J.M. Buchholz, T. Dau, J. Christensen-Dalsgaard, and T. Poulsen. ISBN: 87-990013-2-2. (The Danavox Jubilee Foundation, Copenhagen), pp. 103-111.

Møller, H., Sørensen, M. F., Jensen, C. B., and Hammershøi, D. (**1996**). "Binaural Technique: Do We Need Individual Recordings?". J. Audio Eng. Soc, 44(6), pp. 451-469.

Møller, H., F., Hammershøi, D., Jensen, C. B., and Sørensen, M. (**1999**). "Evaluation of artificial heads in listening tests". J. Audio Eng. Soc, 47, pp.83-100.

Hartmann, W. M. and Wittenberg, A. (**1996**). "On the externalization of sound images". J. Acoust. Soc. Am., 99(6), pp. 3678-3688.



# "HiST taleaudiometri" - A new Norwegian speech audiometry

JON ØYGARDEN

*Audiology, University College of Sør-Trøndelag, 7004 Trondheim, Norway*

"HiST taleaudiometri", a new material for speech audiometry measurements, has been made available from the Audiology Programme at Sør-Trøndelag University College, the institution that teaches the Norwegian audiologists. The material will replace the old Norwegian speech audiometry material made by Quist-Hanssen in the late 1950s. The test material is constructed from four different parts – five-word sentences, three-word utterances, monosyllabic words and digit triplets. Details about the development of the five-word sentences, simulations of speech audiometry measurements and special tests included in the set are given.

## INTRODUCTION

"HiST taleaudiometri" (Øygarden, 2009a and 2009b), consists of four types of speech material.

### Five-word sentences

This set is developed in the tradition of Hagerman (1982) and Wagener *et al.* (1999). The sentences are constructed so that each list of 10 sentences contains exactly the same 50 words. The same syntactical structure is used for all of the sentences: (name verb numeral adjective noun). The five-word sentences in "HiST taleaudiometri" are made with a different cut and splice point which may help us achieve even more natural-sounding sentences than earlier. The material is available for different types of speech in noise tests.

### Three-word utterances

These utterances are the three last words in the five-word sentences. The material has a steep slope of the performance intensity (PI) curve and is suited to measure speech recognition threshold.

### Monosyllabic words

This material is a revision of the words used in the old speech audiometry. Seldom used words and words found very easy, or very difficult to recognize in a listening test were excluded from the material. The selected 160 words were mixed and repeated to make nine lists of 50 words. A Matlab mixing procedure was developed to get both the nine lists and every 10 word-group as uniform as possible.

Proceedings of ISAAR 2011: Speech perception and auditory disorders. 3rd International Symposium on Auditory and Audiological Research. August 2011, Nyborg, Denmark. Edited by T. Dau, M. L. Jepsen, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN 87-990013-3-0. EAN 9788799001330. The Danavox Jubilee Foundation, 2012.