On the perceptual difference of correlated and uncorrelated measurement noise in binaural room impulse responses

Wiebke Hahne and Sascha Spors
Institut für Nachrichtentechnik, Universität Rostock, Deutschland
Email: {wiebke.hahne, sascha.spors}@uni-rostock.de

Introduction

In practical measurements of binaural room impulse responses (BRIRs) the recorded response is superimposed by noise. This noise floor is caused either by background noise or the technical equipment. Both noise types result in a different interaural coherence. Background noise resulting from a single noise source is correlated, whereas the noise of technical equipment is considered to be uncorrelated. Noise sequences with different interaural sequences is localized at different positions in the head of a subject [1]. Correlated noise is located towards the center of the head, whereas uncorrelated noise is located more towards the ears. The noise floor in a BRIR might get perceptible when using it for auralization purposes. In this study we analyze whether a difference can be perceived between correlated and uncorrelated auralized noise floors. For this we artificially deteriorated first a BRIR with additive noise in the measurement process. Hereby, the interaural coherence and the peak to noise ratio (PNR) of the BRIR was varied. The resulting BRIRs were then used for auralization in an ABX listening test.

Signal Model

A continuous BRIR $h_i(t)$ is defined as the acoustic room impulse response from a source to a defined position in the left and right ear canal. In the measurement process the BRIRs are impaired by noise. In Figure 1 a common signal model is introduced which illustrates the measurement process. In a recording room an excitation signal $x[k]$ is played back by a loudspeaker. The continuous BRIR of the room is assumed to be weakly time-variant with temporal variations for each measurement i. It is superimposed by background and equipment noise which is modelled as an additive random signal $n_i(t)$. The microphones capture the recorded response $y_i[k]$ in the recording room with the discrete-time index k. We aim at deriving the discrete time model $h_i[k]$ correspond to the continuous BRIR $h_i(t)$. It may be calculated by the deconvolution of the recorded response $y_i[k]$ with the excitation signal $x[k]$. In spectral domain this yields in

$$h_i[k] = \mathcal{F}_*^{-1} \underbrace{\left\{ \frac{Y_i\left(\mathrm{e}^{\mathrm{j}\Omega}\right)}{X\left(\mathrm{e}^{\mathrm{j}\Omega}\right)} \right\}}_{H_i(\mathrm{e}^{\mathrm{j}\Omega})}, \tag{1}$$

with $Y_i\left(\mathrm{e}^{\mathrm{j}\Omega}\right) = \mathcal{F}_*\{y_i[k]\}$ and $X\left(\mathrm{e}^{\mathrm{j}\Omega}\right) = \mathcal{F}_*\{x[k]\}$ denoting Discrete-Time Fourier Transforms (DTFTs) of the captured and the excitation signal, and $\Omega = 2\pi\frac{f}{f_\mathrm{s}}$ the normalized angular frequency with the sampling frequency f_s. The imaginary unit is denoted by j.

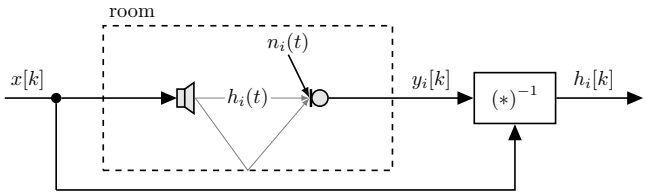

Figure 1: Model of BRIR-Measurement with natural noise impairment.

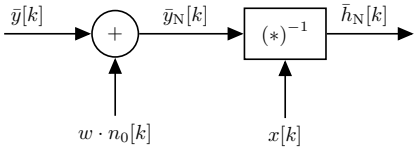

Figure 2: Signal flow diagram of captured response impairment.

$H_i\left(\mathrm{e}^{\mathrm{j}\Omega}\right) = \mathcal{F}_*\{h_i[k]\}$ is termed as binaural room transfer function for the i-th measurement.

BRIRs with defined noise level and noise coherence are required to address the research question. They are generated by firstly averaging over a series of repeated measurements $y_i[k]$ in order to derive a captured signal $\bar{y}[k]$ with very low (perceptually irrelevant measurement) noise floor [2]:

$$\bar{y}[k] = \frac{1}{I}\sum_{i=0}^{I-1} y_i[k]. \tag{2}$$

I denotes the total number of individual measurements. Secondly, this averaged signal is artificially impaired by controllable additive (design) noise with defined coherence to yield $\bar{y}_\mathrm{N}[k]$. White gaussian correlated and uncorrelated noise $n_0[k]$ is used to simulate on one hand ambient noise and on the other hand technical equipment noise. At the end, the averaged, noisy signal $\bar{y}_\mathrm{N}[k]$ is deconvolved by the excitation signal $x[k]$, see Figure 2, resulting in

$$\bar{H}_\mathrm{N}(\mathrm{e}^{\mathrm{j}\Omega}) = \frac{\bar{Y}_\mathrm{N}(\mathrm{e}^{\mathrm{j}\Omega})}{X(\mathrm{e}^{\mathrm{j}\Omega})} = \frac{\bar{Y}(\mathrm{e}^{\mathrm{j}\Omega})}{X(\mathrm{e}^{\mathrm{j}\Omega})} + \frac{w \cdot N_0(\mathrm{e}^{\mathrm{j}\Omega})}{X(\mathrm{e}^{\mathrm{j}\Omega})}. \tag{3}$$

If in the model no design noise is added ($w = 0$) the deconvolution of the mean recorded response $\bar{y}[k]$ by the excitation signal $x[k]$ yields the mean impulse response $\bar{h}[k]$ with very low (measurement) noise floor.

As a measure of the noise level in a BRIR, the peak-to-noise ratio (PNR) is used [3] here. It is defined as

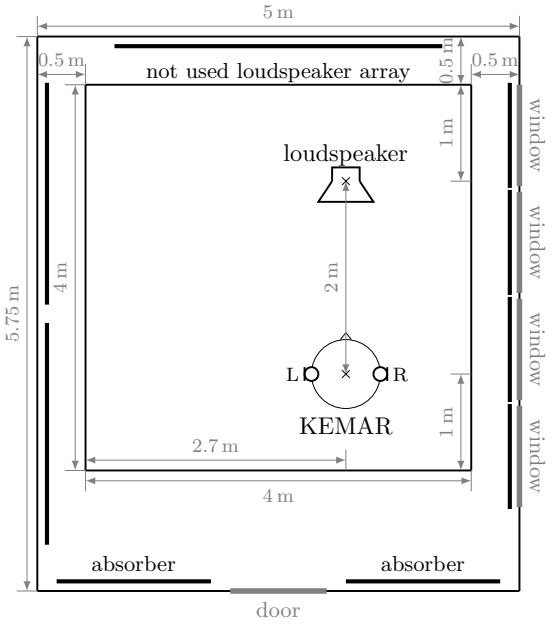

Figure 3: The BRIR measurement setup.

maximum value of the squared impulse response to its noise variance σ_{N}^2

$$\mathrm{PNR} = 10\log_{10}\left(\frac{\max\{h[k]^2\}}{\sigma_{\mathrm{N}}^2}\right) \tag{4}$$

in dB. The desired PNR in the noisy impulse response $\bar{h}_{\mathrm{N}}[k]$ is controlled by the scaling factor

$$w = 10^{\frac{1}{20}\cdot(\mathrm{PNR}_0 - \mathrm{PNR})}. \tag{5}$$

The reference PNR_0 is calculated with (4) choosing $w = 1$ in the design process of the impulse response $\bar{h}_{\mathrm{N}}[k]$.

BRIR Acquisition

BRIR Measurement: The measurement of the BRIRs took place in the audio laboratory of the University of Rostock with the measurement setup illustrated in Figure 3. The room with dimensions $5.75\,\mathrm{m} \times 5.0\,\mathrm{m} \times 3.0\,\mathrm{m}$ has an average reverberation time of $T_{60} \approx 0.3\,\mathrm{s}$. An exponential sine sweep [4] with the length of $K = 2^{21}$ samples at a sampling frequency of $f_{\mathrm{s}} = 44.1\,\mathrm{kHz}$ is used as excitation signal $x[k]$. The lower and upper corner frequency of the sweep signal are $f_{\mathrm{start}} = 20\,\mathrm{Hz}$ and $f_{\mathrm{stop}} = 20\,\mathrm{kHz}$. The excitation signal $x[k]$ is emitted by the dodecahedral omnidirectional power loudspeaker Brüel & Kjær type 4292-L. The room response $y_i[k]$ is recorded by the G.R.A.S. type 45BA KEMAR head and torso simulator (HATS) using the large pinnae type KB0065/KB0066 with pressure microphones type 40AO. The KEMAR is positioned in 2 m distance facing the loudspeaker. In total 600 measurements of the room response $y_i[k]$ were taken over one night. The PNR of one of these individual BRIRs is $\approx 68\,\mathrm{dB}$.

Outlier Detection: For outlier detection after the measurement session, the normalized system distance between consecutive BRIR measurements [5] is used. It is defined as

$$D_i = \sqrt{\frac{\sum_{k=0}^{K-1}(h_i[k] - h_{i-1}[k])^2}{\sum_{k=0}^{K-1}(h_{i-1}[k])^2}} \tag{6}$$

with K denoting the total length of the impulse response. If the system distance D_i between measurement i and $i-1$ exceeds a pre-defined threshold of $D_i > 0.017$, the measurement i is neglected for the computation of the average in (2). The system distance calculation proceeds then with the next measurement. The first 15 measurements were skipped in order to allow the measurement procedure to settle. From the 600 measurements $I = 544$ were selected to calculate the low noise recorded response of equation (2). The resulting low noise BRIR exhibits a PNR of 101.9 dB.

Truncation: For the experiment BRIRs with defined PNRs are needed. The corresponding weighting factors w are calculated with (5) and equation (3) is used to calculate the BRIRs. Pre-tests have indicated that the noise floor in the BRIR may become audible in pauses of the audio material it has been convolved with. To discard this perceptual cue from the experiment the BRIRs are further truncated. The noise floor shown in Fig. 4 is a result of the additive (design) noise and thus does not contain any information on the room itself. The truncation is applied at the time index k_{tr} where the BRIR level is 10 dB less than the noise floor level. In order to estimate the intersection point, it is assumed that the magnitude of the BRIR decays linear in dB without measurement noise. As a consequence, the truncation index k_{tr} depends on the linear decay parameters of the BRIR and the average level of the noise floor. Both have to be estimated from the individual BRIRs. As recommended in [6] the average power σ_N^2 of the noise floor is estimated from the last 100 ms of the 1 s long impulse responses. The estimation of the decay curve follows the method described by Karjalainen et al. [6]. First, a smoothed envelope of the BRIR is computed by Schroeder integration [7] in the interval $[k_{\mathrm{start}}, k_{\mathrm{stop}}]$. Hereby, k_{start} denotes the sample index at which the BRIR reaches its maximum value and k_{stop} the index at which the decay curve meets the level of the noise floor. The discrete-time version of the Schroeder integration is given as cf. [6, eq. (8)]

$$L[k] = 10\log_{10}\left(\frac{\sum_{j=k}^{k_{\mathrm{stop}}} h[j]^2}{\sum_{j=k_{\mathrm{start}}}^{k_{\mathrm{stop}}} h[j]^2}\right), \tag{7}$$

where $k \geq k_{\mathrm{start}}$. In order to estimate the upper limit k_{stop} of the summations, the iterative Lundeby algorithm [8] is applied. For details on its implementation please refer to the accompanying data publication [9]. In a second step, the bias and slope of the decay curve are estimated from the smoothed envelope $L[k]$ by applying

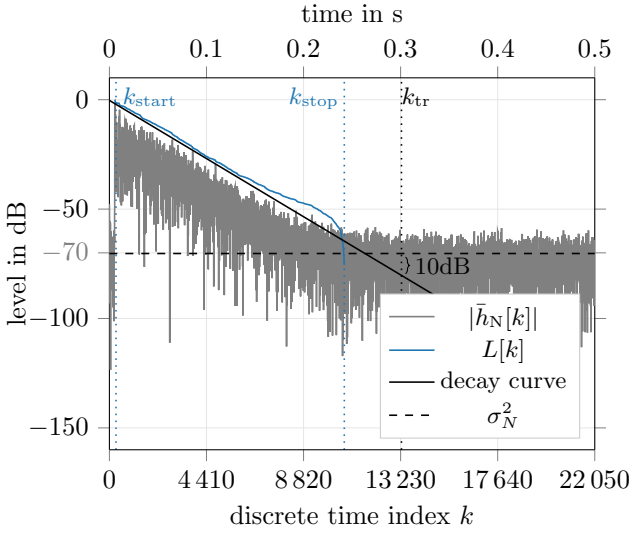

Figure 4: Magnitude $|\bar{h}_N[k]|$ in dB of the BRIR for a PNR of 70 dB. The time index k_{tr} for truncation of the BRIR is given by the index where its decay curve has dropped to 10 dB below the noise level σ_{N}^2.

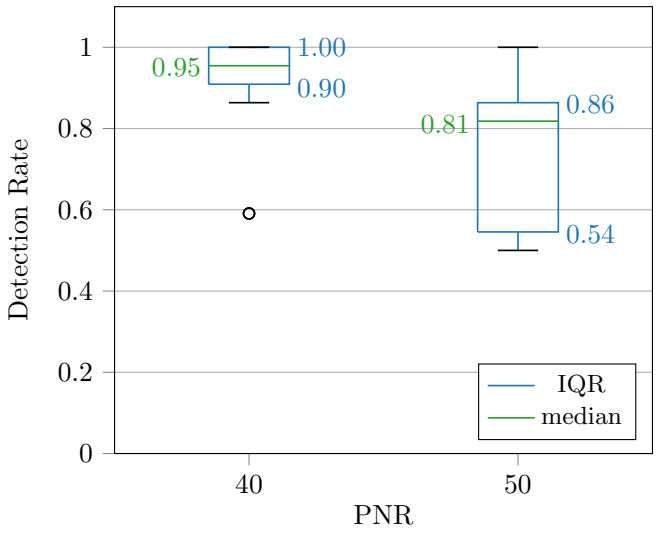

Figure 5: Boxplot diagram of listening test results.

a linear least squares regression. The time index k_{tr} is then determined from the decay curve and the BRIRs are truncated accordingly using a rectangular window. For each PNR, the overall procedure is applied separately to the left and the right BRIR which might result in slightly different lengths. As final step, the BRIR channel with shorter duration is zero-padded to the length of the longer one.

Listening Test

Design: As listening test an ABX-design is used. The subject has the choice between three stimuli "A", "B" and "X". Hereby, "X" equals randomly either "A" or "B". After listening to all signals the subject decides whether "X"="A" or "X"="B". If the subject can't detect the equality it will guess. As audio signal a dry castanet signal of 7 s length is used. It is considered as a critical signal since it is highly transient and impulse-like. As stimulus the castanet signal is convolved with BRIRs which differ in their additional noise coherence. For the "A" and "B" signal the additional noise in the BRIR is at one hand correlated and at the other hand uncorrelated white Gaussian noise. With the ABX listening test two treatments concerning the noise level (PNR = 40 dB, PNR = 50 dB) in the BRIR are tested. Both noise levels are perceptible by the subjects in the castanet stimulus as tested before. Each treatment is tested with 22 trials in a random order. Since every subject tests for both PNR levels it results in a repeated measurement design. The listening test is divided into first a training phase and second the actual listening test. As training stimuli correlated and uncorrelated white Gaussian noise with a length of 7 s is used.

Realization: The listening test is realized in the web-based application "webMUSHRA" Version 1.4.0 [10]. It took place in the audio laboratory of the University of Rostock. The subject was sitting alone in the laboratory and controls the experiment using a mouse. A screen is placed in front of it. The connected computer and sound card was placed outside the room. 13 subjects (11 male, 2 female) participated in the age between 27 and 54 years. All reported no known hearing impairments. 6 of them had no experience in listening test tasks and 7 had some experiences. The playback sound pressure level (SPL) is 66.3 $\text{dB}_{\text{LA eq,T=30s}}$. It is measured using an ear simulator type RA0039 (G.R.A.S.) and the SPL meter Brüel&Kjær type 2270 in accordance to IEC 60318-1 [11]. The stimulus with a PNR of 40 dB with correlated noise was used. It has a digital level of -16.6 LUFS [12] which is measured using the virtual studio technology (VST) plugin TBProAudio dpMeter3 (V1.3.1) and Audacity as VST host. After all subjects had taken place in the listening test the SPL was remeasured. It arose that the SPL had changed over the time of the listening test between the subjects. The playback (SPL) was at 59.2 $\text{dB}_{\text{LA eq,T=30s}}$. This undesired change is probably attributed to a Windows update.

Results

Descriptive Statistic: In Figure 5 the results of the listening test are visualized as a boxplot diagram in dependence of the PNR. The detection rate is defined in percent as how often the subject identified "X" correct as stimulus "A" or "B". Depending on the PNR the results are differently distributed. With a higher noise level (PNR = 40 dB) the median is at 0.95. Conspicuously, the upper inter quartile range (IQR) is at 1.0 and no whisker exists. It shows that 50% of the subjects detected the difference between correlated and uncorrelated noise in a BRIR with a reliability between 95% and 100%. The lower IQR is at 0.9. Additionally, there are 2 outliers which are lower located then the whisker at 0.59. At the lower noise level (PNR = 50 dB) the distribution of the results is wider. The median is lower with 81% as detection rate. 25% of the results are very close in the upper IQR between 81% and 86% detection rate. Obviously, one subject detected "X" to 100% correct at the lower noise level. The lower IQR is with 0.54 in a wider range

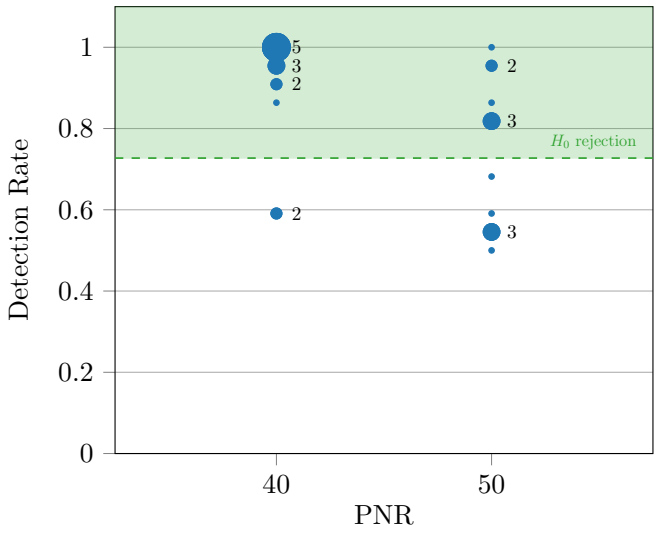

Figure 6: Hypothesis test on Binomial distribution of the results. The rejection rate of the H_0-Hypothesis is given with 0.72% and indicated with the green shaded area. A small dot without number visualizes the result of one subject. If more than one subject reached one result the dot is increased and has the number of subjects aside.

than the upper one.

Hypothesis Testing: To test the statistical significance a Binomial Distribution is used [13]. The results are visualized in a scatter plot in Figure 6. With the shaded area the results are marked at which guessing is improbable with a confidence level of 95%. The H_0-Hypothesis will be rejected if there are greater or equal than 16 correct answers which result in a rejection rate of 72%. 11 subjects detected the difference of correlated and uncorrelated noise in BRIR at the higher noise level (PNR = 40 dB). Two subjects could not detect a difference between these treatments. Whereas at a lower noise level (PNR = 50 dB) only 7 subjects could detect significantly the difference and 6 subjects have results below the rejection rate.

Discussion: The results show that at both noise levels the difference between correlated and uncorrelated noise in a BRIR can be detected. Decreasing the noise level (higher PNR) the detection rate lowers as well. At the lower noise level (PNR = 50 dB) there are more subjects who could not detect the difference. Additionally, the most subjects reported after the test that they could not hear a difference in the low noise stimulus. Even the subjects with a high detection rate reported that. So the difference detection at the low noise BRIR seems to be based on unconscious cues. A statistical evaluation of the central tendency of the two noise levels is not possible due to the fact of the changing playback SPL between the subjects. Evaluating the tendency of each subject, except for the one who detected the effect at both noise levels with 100%, the detection rate lowers with decreasing noise level. This leads to the conclusion, that lowering the noise level has the effect that the difference in correlated and uncorrelated noise in a BRIR has less influence.

Summary

In this study the perceptual difference in auralisation using BRIRs imparied by noise with different interaural coherences is analysed. As extreme cases in terms of interaural coherence correlated and uncorrelated noise is used. The effect is tested at two different PNR levels using an ABX listening test. All together, the results of 13 subjects are evaluated. It can be stated that a difference between correlated and uncorrelated noise in a BRIR is significantly for both noise levels detectable. Lowering the noise level leads to a decrease in the detection rate. Additionally, more subjects could not detect the difference significantly compared to the higher noise level.

References

[1] Blauert, J.: Spatial hearing: The psychophysics of human sound localization. MIT press (1997), p.241

[2] Müller, S. and Massarani, P.: Transfer-Function Measurement with Sweeps. Journal of the Audio Engineering Society 49.6 (2001), 443-471

[3] Guski, M. and Vorländer, M.: Comparison of noise compensation methods for room acoustic impulse response evaluations. Acta Acustica united with Acustica 100.2 (2014), 320-327

[4] Farina, A.: Advancements in Impulse Response Measurements by Sine Sweeps. Audio Engineering Society Convention 122 (May 2007)

[5] Hahn, N. and Spors, S.: Comparison of continuous measurement techniques for spatial room impulse responses. European Signal Processing Conference 24 (2016), 1638-1642

[6] Karjalainen, M. et al.: Estimation of Modal Decay Parameters from Noisy Response Measurements. Journal of the Audio Engineering Society 50.11 (2002), 867-878

[7] Schroeder, M.: New Method of Measuring Reverberation Time. The Journal of the Acoustical Society of America 37.3 (1965), 409-412

[8] Lundeby, A. et al.: Uncertainties of measurements in room acoustics. Acta Acustica united with Acustica 81.4 (1995), 344-355

[9] Hahne, W. et al.: On the Perceptually Acceptable Noise Level in Binaural Room Impulse Responses - Accompanying Dataset. `https://doi.org/10.18453/rosdok_id00002434` (2019)

[10] Schoeffler, M. et al.: webMUSHRA — A Comprehensive Framework for Web-based Listening Tests. Journal of Open Research Software. 6.1 (2018), p.8

[11] Standard IEC 60318-3: Electroacoustics – Simulators of human head and ear – Part 3: Acoustic coupler for the calibration of supra-aural earphones used in audiometry. International Electrotechnical Commission (2014)

[12] Standard EBU R 128: Loudness normalisation and permitted maximum level of audio signals. European Broadcasting Union, Geneva (2014)

[13] Boley, J. and Michael L.: Statistical analysis of ABX results using signal detection theory. Audio Engineering Society Convention 127, Audio Engineering Society (2009)