On the Perceptually Acceptable Noise Level in Binaural Room Impulse Responses

Wiebke Hahne, Vera Erbes, Sascha Spors

Institute of Communications Engineering, University of Rostock, Germany

Email: {wiebke.hahne, vera.erbes, sascha.spors}@uni-rostock.de

Introduction

A binaural room impulse response (BRIR) describes the acoustic properties of a room from a sound source to the left and the right ear of a listener. A captured BRIR convolved with a sound signal presented by a headphone gives a listener the impression of being in the recording room with the sound source location at the recorded position.

For the measurement of BRIRs an excitation signal $x[k]$ is played by a loudspeaker. With microphones in the ears of a dummy head the signal $y[k]$ is captured including the acoustic response of the outer ear and the acoustic characteristics of the room. In every measurement noise of the environment or of the measurement equipment is recorded. Therefore every measured BRIR is contaminated with a noise floor. The resulting question is at which level of the noise floor a listener cannot detect the noise in a stimulus convolved with the noisy BRIR.

As a measure of the noise level in the BRIR the peak-to-noise ratio (PNR) is evaluated. The PNR is used since the maximum value h_{max} of the impulse response has a dominant influence on the hearing experience of the listener. Therefore the PNR is defined by h_{max} and the power of the noise with the variance σ^2_{noise} as

$$PNR = 10 \cdot \lg \left(\frac{h^2_{max}}{\sigma^2_{noise}} \right). \qquad (1)$$

For the detection of the threshold of noise a listening test with an adaptive method is advisable due to its efficiency. Depending on the selected test method a threshold on the psychometric function is estimated. In adaptive listening tests stimuli with defined noise levels are needed. It would be possible to generate the noise levels while measuring the BRIRs, for instance by playing additional background noise with variable noise level. For every noise level one measurement would have to be done.

A second possibility is to impair the BRIR artificially through additive noise. At the DAGA 2018 in Munich (Germany) Häußler [1] presented the detection threshold of noise in HRTFs. They added the noise with different levels to the calculated impulse response. For efficiency and the difficulties in generating evenly distributed noise fields the second possibility is chosen. The corresponding system model is shown in Figure 1 with $x[k]$ as the excitation signal. Typically an exponential sine sweep is used. The dummy head captures the signal $y[k]$ with microphones. In contrast to the study of [1] the white gaussian noise $n_0[k]$ is added to the recorded signal before the deconvolution. It models the technical equipment noise as microphone noise and superposes therefore the recorded signal resulting in $y_n[k]$. The level of the added noise is chosen in accordance to the desired PNR-levels.

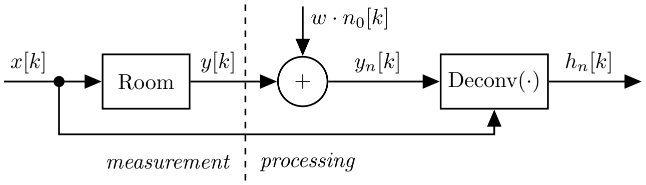

Figure 1: Schematic description of the system model which estimates the noisy impulse response.

BRIR Generation

Generating a BRIR with a high PNR-level the PNR-level can be lowered in a controlled way by additive white noise. To determine the noise threshold, the PNR of the measured BRIR needs to be much higher than the expected threshold. In the following the BRIR generation is described.

BRIR Measurement

A BRIR is a two channel signal which consists of the impulse response from a source to the left and the right ear. For ease of illustration h is used for both channels. The room response $y[k]$ is calculated by

$$y[k] = x[k] * h[k] \qquad (2)$$

in the time domain. Here k is the sample index and $*$ represents the convolution. With the help of the Discrete Time Fourier Transform $\mathcal{F}_*\{\cdot\}$ it is possible to dissolve the equation (2)

$$Y(e^{j\Omega}) = X(e^{j\Omega}) \cdot H(e^{j\Omega}). \qquad (3)$$

This results in

$$H(e^{j\Omega}) = \frac{Y(e^{j\Omega})}{X(e^{j\Omega})} \qquad (4)$$

which is known as deconvolution. As excitation signal $x[k]$ an exponential sine sweep [2] is used with $f_{\text{start}} = 20\,\text{Hz}$ and $f_{\text{stop}} = 20\,\text{kHz}$ as start and stop frequency of the sine sweep. These two frequencies are chosen in dependence of the human hearing capabilities. The exponential sine sweep has been proven to be a convenient measuring signal due to its crest factor and the absence of zeros in the spectral domain. The total sweep duration $T \approx 47.6\,\text{s}$ is chosen due to efficient calculations with the power of two at a sampling frequency of $f_s = 44100\,\text{Hz}$.

The measurement of the BRIR took place in the audio laboratory of the University of Rostock. The room with a size of $5.75\,\text{m} \times 5.0\,\text{m} \times 3.0\,\text{m}$ is acoustically treated and has a low reverberation time of $T_{60} \approx 0.3\,\text{s}$. The signal is captured by a KEMAR dummy head which is positioned in 2 m distance facing the loudspeaker.

To achieve a high PNR the mean BRIR $h_M[k]$

$$h_M[k] = \frac{1}{I} \cdot \sum_{i=0}^{I} h_i[k] \tag{5}$$

over I measurements is calculated. With the normalized system distance [3] defined as

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

a measure of the quality depending on the preceding impulse response $i-1$ is determined to detect outliers. Here is K the length of the impuls response. If the system distance D between these two measurements is greater than a self-defined threshold of 0.017, the measurement i is neglected. Then the system distance is calculated again with the next following measurement $i+1$. After measuring BRIRs a whole night, from the 600 measurements $I = 576$ are selected to calculate the mean BRIR h_M. This results in a PNR-level of 101.9 dB.

BRIR Processing

For the realization of the different PNR levels additional white noise is added to the recorded signal $y[k]$. This can be written as

$$\begin{aligned} H(e^{j\Omega}) &= \frac{Y_N(e^{j\Omega})}{X(e^{j\Omega})} \\ &= \frac{Y(e^{j\Omega}) + w \cdot N_0(e^{j\Omega})}{X(e^{j\Omega})} \\ &= \underbrace{\frac{Y(e^{j\Omega})}{X(e^{j\Omega})}}_{H_M(e^{j\Omega})} + \underbrace{\frac{w \cdot N_0(e^{j\Omega})}{X(e^{j\Omega})}}_{H_N(e^{j\Omega})} \end{aligned} \tag{7}$$

in the frequency domain. Hereby $N_0(e^{j\Omega})$ is the additive white gaussian noise with unity variance and $H_M(e^{j\Omega})$ the measured mean BRIR. The deconvolution of the white noise with the exponential sweep $H_N(e^{j\Omega})$ results in blue noise where the power density increases with 3 dB per octave.
To get different PNR levels the scaling factor w is introduced. For realizing the desired PNR steps a mathematical relationship between the weighting and the resulting PNR is needed. Connecting equation (1) with (4) yields an exponential relation of the two parameters:

$$w = 10^{\frac{1}{20} \cdot (PNR_0 - PNR)}. \tag{8}$$

The constant PNR_0 is calculated by equation (1) with the use of the noisy impulse response h_n with the factor $w = 1$. For the calculation of all BRIRs the same noise sequence is used in order to avoid differences in colouration as perceptual cue.

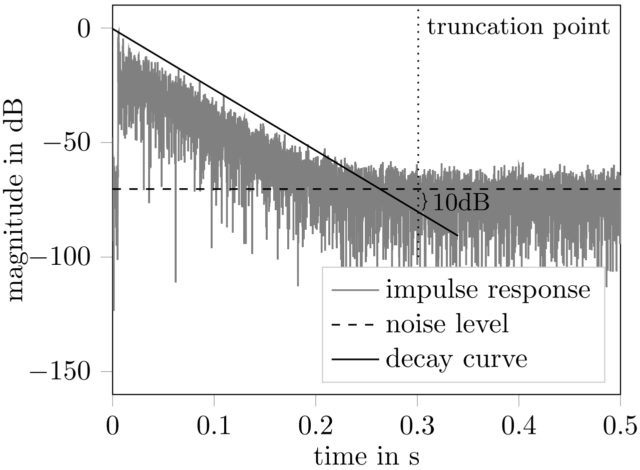

Figure 2: Truncation of the impulse response 10 dB below the intersection point of the decay slope and the noise level.

BRIR Truncation

The measured BRIR with the added noise is truncated due to the fact that no further information are contained in the noise tail. Additionally preliminary experiments showed that a long noise tail is audible in speech pauses between sentences what might serve as a cue for detection of noise in BRIRs. The truncation time is determined 10 dB below the intersection point of the decay curve to the noise floor in a logarithmic scale as shown in Figure 2 for an impulse response with a PNR = 70 dB. This point is chosen for the reason that information in the impulse response may still be audible through the noise. To identify this point the linear decay curve of the impulse response and the noise level are estimated. The PNR steps leads to different truncation points at every generated impulse response. Therefore an automatic algorithm is used for truncation.
Oversampling of the impulse response did not change the maximum value more than 0.5 dB and was therefore omitted for estimating the peak level h_{max}^2.

To estimate the noise power the last 10 % of the 1 s long impulse response were used. This way, an influence of measurement artefacts at the end of the deconvolved impulse response, which had the same length as the captured signal, is avoided.
A good approach for approximating the decay curve of the impulse response is described by Karjalainen et al. [4]. First the impulse response envelope is smoothed by the Schroeder Integration [5] which starts theoretically at infinity [6]. Through the contamination with noise it results in a bias at the late part of the decay slope [4]. To avoid this, the Schroeder integration is calculated in the interval $[k_{max}, k_I]$. k_I is the point at which the decay curve meets the noise level [6] and k_{max} is the point at which the impulse response has its maximum value. $L[k]$ is the discrete time version of the Schroeder

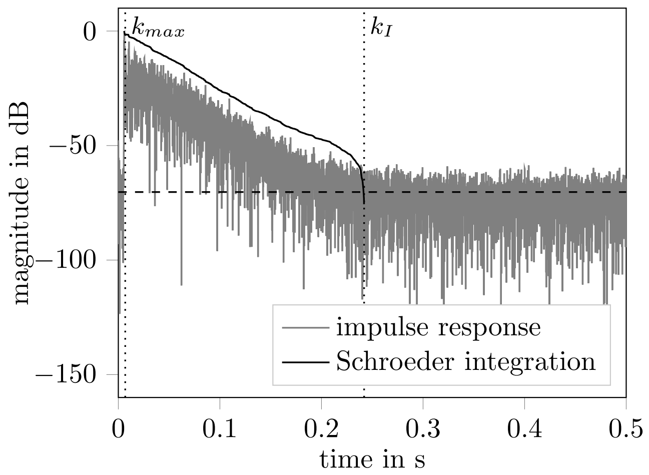

Figure 3: The backward integration of the impulse response with the interval boundary $[k_{max}, k_I]$.

integration [4]:

$$L[k] = 10 \cdot \lg \left(\frac{\sum_{i=k}^{k_I} h_i^2}{\sum_{j=k_{max}}^{k_I} h_j^2} \right). \quad (9)$$

With the iterative Lundeby algorithm [6] the interval boundary k_I as intersection point of the noise level and the decay curve is determined.

In Figure 3 the resulting smoothed decay curve of the Schroeder integration is shown exemplarily at an impulse response with PNR = 70 dB. From the smoothed envelope of the Schroeder integration the parameters of the linear decay curve are estimated with the least squares algorithm.

Stimuli Generation

The modified and truncated BRIR is convolved with a speech signal. The headphone transfer function of the used headphone type AKG K601 is equalized. A fade-in and fade-out of 0.01 s of the stimuli is realized by a Blackman-window. Finally the stimuli are normalized by the maximum value of all stimuli for the left and the right ear.

Design of the Listening Test

Using a listening test, a specific target value on the psychometric function is determined [7]. Due to the precise and efficient threshold detection, adaptive listening tests are preferred for this task [8]. Particularly the time minimization for the subject is important. In an adaptive method the stimulus level is dependent on the preceding stimulus level and the response of the subject [9]. With the transformed up-down staircase method no specific information about the psychometric function is needed [8]. Combined with the 3AFC (alternative forced choice) method which is the most efficient method [10], the listening test is designed as shown in Figure 4. The estimated point at the psychometric function results in 56.1% through the choice of the 3AFC paradigm with the 2Up-1Down rule.

As reference a stimulus with a PNR = 90 dB is used.

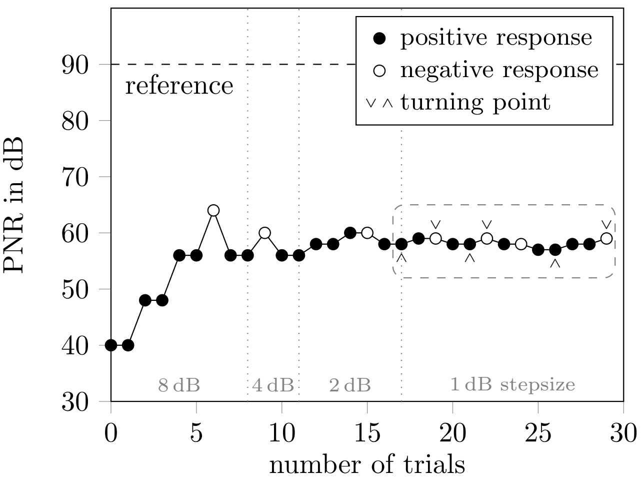

Figure 4: Exemplary procedure of the adaptive 3AFC listening test.

At this PNR-level no noise artefacts were audible in pre-tests. The listening test starts with a stimulus with a PNR = 40 dB. If the threshold and the psychometric function are not known it is recommended to use a larger step size in the beginning and reduce the step size during the experiment [9]. Therefore the step size is halved after the second, fourth and sixth turning point. The estimate of the threshold is done by averaging the turning points. It is recommended to test until six or eight turning points are reached [9] with the smallest step size.

The smallest step size is given by 1 dB. Keeping the concentration time for the subjects as short as possible the threshold is estimated after six turning points by averaging these and the listening test is finished. If the answers of a subject do not converge towards a threshold, a stop criterion is defined by 60 trials.

To execute the listening test, the Matlab based software package 'WhisPER' by the TU Berlin is used [11]. The listening test starts with a training phase where the subject gets familiar with the task. A staircase procedure with 10 trials is used for this purpose. Here, the stimuli are castanets. In order to make the subjects understand the test principle a signal is used which is sensitive.

Results of the Listening Test

The listening tests were performed in the audio laboratory of the University of Rostock. The subjects were sitting in the well-damped and quiet room. Altogether 20 subjects at the age from 24 to 61 years participated. Six subjects had no experience in the field of audio, 11 subjects through their hobbies and three through their work. One of the subjects announced hearing impairment after the test. Therefore the result of this subject has been removed afterwards. From two subjects no threshold could be estimated because they reached the stop criterion. After the first subject the sound level was raised at the recommendation of the subject. So the

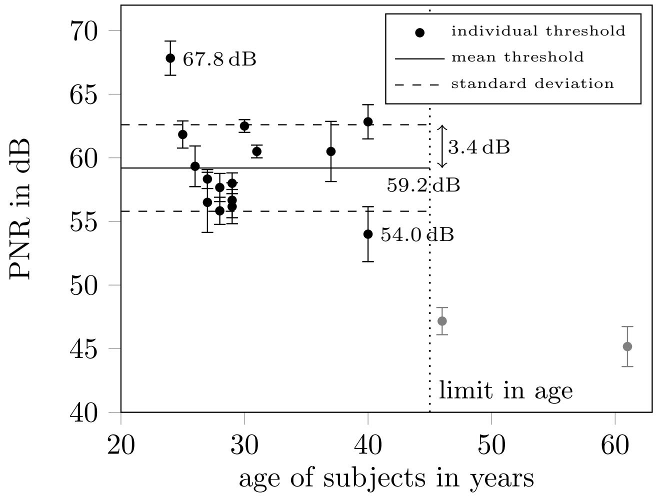

Figure 5: Results of the listening test in dependence of the age of the subjects.

subject did the listening test twice and the first result has not been evaluated. The sound level for all subjects was 58.4 dB$(A)_{eq}$, measured by the sound pressure level meter NTi XL2 used with the KEMAR dummy head. The time the subjects needed was between 33 min and 60 min. The number of trials is between 28 and 52 except for the ones who reached the stop criteria of 60 trials. Most subjects found a key word in the sentence where they could identify the different stimuli the easiest. They reported to hear differences in vocals or consonants or as noise in short breaks.

In Figure 5 the results of the listening test are presented in dependence of the age of the subjects. Two subjects reached results which are under 1.5 times the interquartile range and can be seen as outliers. Additionally these were the oldest subjects. Therefore only the subjects under the age of 45 years are evaluated in the following. An additional evaluation depending on the experience in the field of audio of the subjects is not presented because the results showed no dependence. Additionally to the individual thresholds of the subjects the standard deviation of the six turning points are shown in Figure 5. There seems to be no obvious relationship between the mean and the standard deviation of the subjects.
The results of 15 subjects are evaluated to estimate the threshold of noise detection in a BRIR. The mean threshold of the subjects is 59.2 dB with a standard deviation of ±3.4 dB. The wide range of 13.4 dB between the minimum threshold of 54.0 dB and the maximum threshold of 67.3 dB could result from the hearing abilities of the subjects. Hence the standard deviation of ±3.4 dB appears acceptable.
The confidence interval is calculated with a confidence level of 95 % to ±1.9 dB. The resulting threshold is valid for the tested sound level 58.4 dB$(A)_{eq}$. It is to be expected that the threshold will be lower with a lower sound level and higher with a higher sound level.

Conclusion

The threshold of detection of noise in a BRIR is determined with an adaptive 3AFC listening test. Through the 2Up-1Down rule the 56.1% point on the psychometric function is estimated by calculating the mean of the last six turning points. The used stimuli for the listening test are generated with manipulated BRIRs. White noise with defined level is added to an averaged BRIR with a high PNR in order to generate BRIRs with different PNR-levels. The BRIRs are truncated in dependence of the noise floor. With the results of 15 subjects, the mean threshold was 59.2 dB with a standard deviation of ±3.4 dB. This is valid for the given sound level at 58.4 dB$(A)_{eq}$. In future work the dependence between the PNR threshold and the sound level is analysed.
The mean BRIR, the BRIRs with different PNR-levels and the listening test results of each subject can be found under DOI: 10.18453/rosdok_id00002434.

Acknowledgement

This research was supported by a grant of the Deutsche Forschungsgemeinschaft (DFG) SP 1295/6-1 and 1295/10-1.

References

[1] Andreas Häußler, Henning Kuewen, Joachim Thiemann, and Steven van de Par. Detection threshold of uncorrelated (measurement) noise in hrtf. DAGA, 2018.

[2] Angelo Farina. Advancements in impulse response measurements by sine sweeps. In *Audio Engineering Society Convention 122*, May 2007.

[3] Nara Hahn and Sascha Spors. Comparison of continuous measurement techniques for spatial room impulse responses. In *Signal Processing Conference (EUSIPCO), 2016 24th European*, pages 1638–1642. IEEE, 2016.

[4] Poju Antsalo, Aki Makivirta, Vesa Valimaki, Timo Peltonen, and Matti Karjalainen. Estimation of modal decay parameters from noisy response measurements. In *Audio Engineering Society Convention 110*. Audio Engineering Society, May 2001.

[5] Manfred R Schroeder. New method of measuring reverberation time. *The Journal of the Acoustical Society of America*, 37(3):409–412, 1965.

[6] Anders Lundeby, Tor Erik Vigran, Heinrich Bietz, and Michael Vorländer. Uncertainties of measurements in room acoustics. *Acta Acustica united with Acustica*, 81(4):344–355, 1995.

[7] Bernhard Treutwein. Adaptive psychophysical procedures. *Vision research*, 35(17):2503–2522, 1995.

[8] Stefanie Otto and Stefan Weinzierl. Comparative simulations of adaptive psychometric procedures. *Jahrestagung der Deutschen Gesellschaft für Akustik*, pages 1276–1279, 2009.

[9] H. C. C. H. Levitt. Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical society of America*, 49(2B):467–477, 1971.

[10] Birger Kollmeier, Robert H Gilkey, and Ulrich K Sieben. Adaptive staircase techniques in psychoacoustics: A comparison of human data and a mathematical model. *The Journal of the Acoustical Society of America*, 83(5):1852–1862, 1988.

[11] TU Berlin. Whisper. DOI: 10.14279/depositonce-31.3, Version 1.9.1.