Just noticeable notch smoothing of head-related transfer functions

Michael Kohnen, Ramona Bomhardt, Janina Fels, Michael Vorländer

Institute of Technical Acoustics, 52074 Aachen, Germany, Email: mko@akustik.rwth-aachen.de

Abstract

To understand the influence of notch depth mismatches in head-related transfer functions (HRTFs) on the subjective perception, an expert listening test was conducted to find the just noticeable difference (JND) of notch smoothing. The findings are also intended to be used for binaural reproduction over loudspeakers to avoid high peaks due to HRTF inversion (e.g. in crosstalk cancellation systems (CTC)). For the determination of the subjective threshold a QUEST procedure was implemented. Six positions on the right hemisphere were tested using individual HRTFs and resulted that notches can be smoothed at least with a factor of 0.5 towards their one octave moving average window smoothed equivalent.

Introduction

3-D visual recording and projection systems allow us nowadays to fully recreate a captured or synthesized visual scene. Yet, common acoustic representation is static and impeding the full immersion into that scene. To overcome this issue, recent audio capturing technology has been advanced with Ambisonics recording. The binaural down mix of this flexible broadcasting format is available for everyone owning a simple headphone device. Furthermore, synthesizing full 3-D acoustic scenes can be achieved using binaural synthesis and reproduction over headphones or loudspeakers (CTC). In both cases a carefully chosen or pre-processed HRTF is vital for spatial externalization of the perceived acoustic scene and localization of its containing sound sources. This ongoing development from plane stereo to binaural reproduction evokes a wider interest into HRTF measuring techniques, either consumer orientated home-use technology, parametric adaption methods or scientific high precision measurements. However, all techniques differ in their 3-D sound reproduction quality. Nowadays, the subjective effect of these differences is not extensively investigated and the quality of a HRTF is hard to evaluate. In general, the quality is related to the match of the HRTF used in comparison to the individual one. Nevertheless, the comparison is time- and frequency-dependent. While time differences are more important at low frequencies, especially high frequent notches help to localize sound sources. In case of a non-matching HRTF with deviating notches, the sound localization can be disturbed and the sound colorized. Especially, in case that the inverse HRTF is needed for reproduction, like in CTC systems, undesired effects occur due to high, narrow band amplification caused by deep notches in the HRTF. To avoid delocalization or colorization, the subjective effect of notch depth in an measured HRTF will be investigated in this paper.

Setup of the listening experiment

To determine the JND of smoothed notches in HRTFs, a listening test was performed. In total, 15 subjects participated in the listening experiment. They were 29±6 years old (mean ± standard deviation), one was female. All of the subjects reported normal hearing and had previous experience with spatial audio.

Head-related transfer functions

The participants individual HRTFs were measured previously for the ITA HRTF database [1]. The coordinate systems used has its origin in the center of the subject's head. The representation of the elevation and azimuth angle $(0°, 0°)$ defines the view direction. $(0°, 90°)$ is located at the left ear and $(0°, 90°)$ located above the head. Due to the right ear advantage [2], the directions used for the listening experiment were on the right side of the subject. To avoid fatigue effects, only six sound source directions were tested in the listening experiment. Four directions were located in the horizontal plane at $(0°, 0°)$, $(0°, 240°)$, $(0°, 300°)$, and $(0°, 340°)$. Additionally, two elevated directions at $(30°, 340°)$ and $(60°, 340°)$ were chosen.

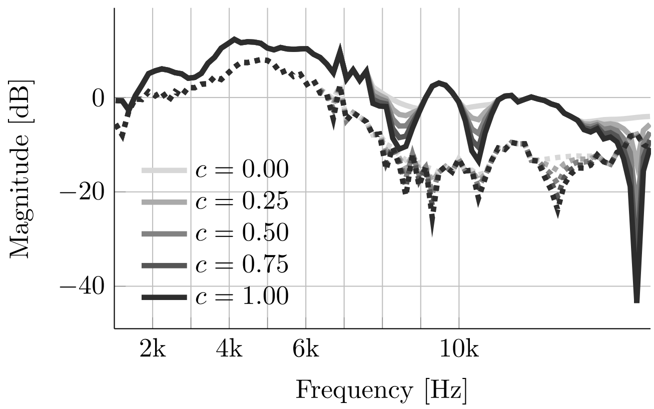

Figure 1: Smoothed HRTFs at $(0°, 340°)$ for notch conservation factors $c = 0$ to 1 in 0.25 steps (bright to dark). HRTFs of the right ear are plotted by solid lines and the left ones by dotted lines. A notch conservation factor of 0 leads to one octave band smoothed notches.

Headphone transfer functions

The HpTFs were measured according to Masiero and Fels [4]: The subject had to reposition the headphones (Sennheiser HD650) eight times on the head for the measurement of the transfer functions to the microphones (Sennheiser KE3) at the ear canal entrances. To obtain a smooth and robust spectrum, the individual averaged HpTF was calculated using the inverted mean plus twice the standard deviation of these measurements. Finally, a minimum phase from the magnitude spectrum of the HpTF was calculated and added [6, pp. 788-789].

Smoothing algorithm

To smooth the notches, a compare-and-squeeze function [5] based on a moving average filter was used. The original frequency response is compared to its duplicate smoothed with an one octave moving average filter. Once the difference between these two curves exceeds a threshold of 3 dB, the notch will be smoothed. If the difference is less than the threshold, no smoothing was applied. The degree of smoothing is influenced by a notch conservation factor which is defined between $0 \leq c \leq 1$. While $c = 0$ smooths the notches as an one octave moving average filter, $c = 1$ results in the original data (see Fig. 1).

Adaptive psychometric procedure

The JND of smoothed notches in HRTFs was determined using the adaptive psychometric procedure QUEST [7]. This approach adapted an initial conservation factor $c = 0.5$ dependent on the subject's response. For this purpose, the HpTF and each original HRTF pair, right and left side of one measured direction, were multiplied in the frequency domain. Subsequently, the Fourier transform of the product was convolved with a pulsed white noise stimulus. Each of the three pulses had a length of $250\,\mu$s followed by a pause of $100\,\mu$s. This stimulus is the so-called *reference*. In addition to this, a smoothed HRTF pair was used in the same manner for comparison. Both stimuli were played back. Afterwards, the subjects had to answer the question "Did you hear a difference?" by choosing one of the two options *Yes* or *No*. Based on the subject's decision, the QUEST algorithm [3] [1] determines the new threshold c which is used for the following trail. If the answer was *Yes*, the conservation factor c increases and otherwise it decreases. After 20 trials, the conservation factor c converges against the JND c_{JND} and is assumed as converged (see Fig. 2).

The JNDs were determined in six blocks in the listening experiment, one direction in one block respectively. These blocks were randomized and separated by a break of one minute. The result for one tested position can be seen in Fig. 2. The x-axis is the trial number while the y-axis shows the notch conservation factor. For each trial the step size is decreased to converge to the JND. The dashed line shows the mean value over all subjects whereas the gray area indicates the standard deviation between the subjects. The solid horizontal line is the estimated mean value for all subjects for the JND threshold. For trial number zero all subjects had the same starting value and no difference was heart which lead to the same testing condition for trial one. From there on the answers of the subjects and therefore the conditions changed between subjects. The estimated just noticeable threshold of the notch conservation factor is 0.18 as shown in the legend.

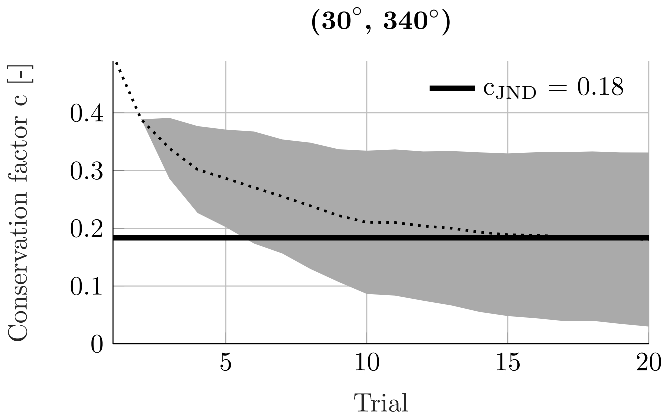

Figure 2: Exemplary results at (30°, 340°). c denotes the notch conservation factor where $c = 0$ is maximum smoothing and $c = 1$ is is the original HRTF. The dotted line indicates the mean value, the gray area the standard deviation. The solid line is the calculated threshold value.

Results

The experiments JND with respect to the direction tested, are shown in Fig. 3. For each position the mean value of the JND conservation factor and the standard deviation is plotted. For the positions tested no standard deviation exceeds a conservation factor of 0.5. Neither the mean values nor the standard deviations indicate a difference between horizontal and elevated directions or a difference between frontal or lateral directions. No significant differences could be found between the positions.

The notch depth was defined as the difference between the mean value between 100 Hz and 1000 Hz and the value of the deepest notch[2]. This notch difference is plotted in Fig. 4 for the left ear and in Fig. 5 for the right ear with respect to the original notch depth. Consequently, each circle marks one the notch of one direction per subject. The different notch depths were then clusterd in 10 dB ranges and the mean value with its standard deviation plotted. For the right, i.e. the ipsilateral, ear the gradient of the mean values is higher than for the left one, which indicates that the audible smoothed notch is on this side.

Conclusion

The results show that notches can be smoothed at least half their depth compared to their depth when filtered with a one octave band moving average window. No differences could be found between the positions tested. Further investigations will follow that investigate the perceptive performance for CTC systems with this filters. Additionally, it was shown that the critical notch was located on the ipsilateral ear side which leads to the as-

[1]The initial distribution had a mean of $c = 0.5$ and a standard deviation of 0.15. The psychometric function had a false positive rate $\gamma = 0$ due to the Yes-No design, a slope of $\beta = 0.5$, a false negative rate of $\delta = 0.1$ which were determined in pretests. Since the conservation factor decreases for false answers, the psychometric function was flipped.

[2]This can lead to negative notch depths

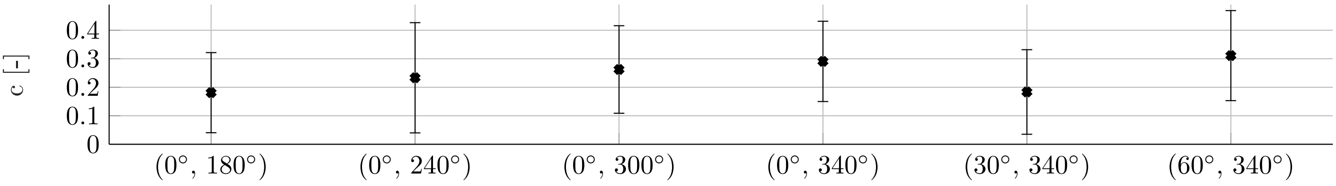

Figure 3: The figure shows the mean values of the estimated threshold for the notch conservation factor c for different positions and the standard deviation.

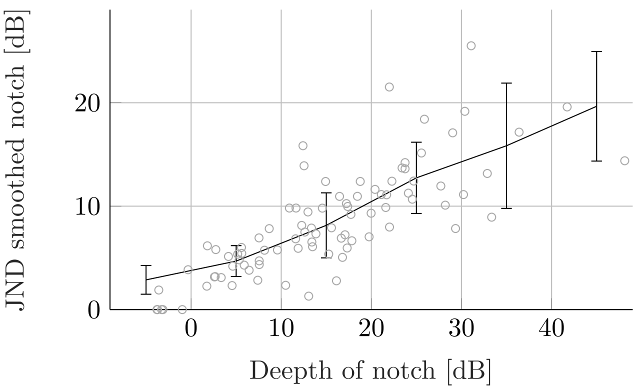

Figure 4: The applicable smoothing is shown in relation to the depth of the deepest notch for the left (i.e. the contralateral) ear. Dots are single values whereas the straight line indicates the mean values together with standard deviation whiskers.

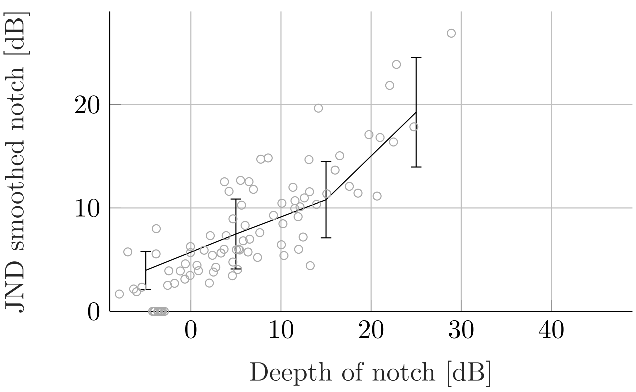

Figure 5: The applicable smoothing is shown in relation to the depth of the deepest notch for the right (i.e. the ipsilateral) ear. Dots are single values whereas the straight line indicates the mean value together with standard deviation whiskers.

sumption that the contralateral ear is less sensitive to smoothing.

Acknowledgement

Special thanks go to Lukas Vollmer who implemented and performed the listening experiment. Additionally, we thank the persons who participated in the experiment.

This research was financed by the Head Genuit Foundation under the project ID P-16/4-W.

References

[1] R. Bomhardt, M. de la Fuente Klein, and J. Fels. A high-resolution head-related transfer function and three-dimensional ear model database. In Acoustical Society of America, editor, *5th Joint Meeting of the Acoustical Society of America and Acoustical Society of Japan*, page 050002, 2016.

[2] D. S. Emmerich, J. Harris, W. S. Brown, and S. P. Springer. The relationship between auditory sensitivity and ear asymmetry on a dichotic listening task. *Neuropsychologia*, 26(1):133–143, 1988.

[3] P. E. King-Smith, S. S. Grigsby, A. J. Vingrys, S. C. Benes, and A. Supowit. Efficient and unbiased modifications of the QUEST threshold method: theory, simulations, experimental evaluation and practical implementation. *Vision Res*, 34(7):885–912, 1994.

[4] B. Masiero and J. Fels. Perceptually robust headphone equalization for binaural reproduction. In Audio Engineering Society, editor, *Audio Engineering Society 130th Convention*, pages 1–7, 2011.

[5] S. Müller. *Digitale Signalverarbeitung für Lautsprecher.* PhD thesis, RWTH Aachen University, 1999.

[6] A. V. Oppenheim and R. W. Schafer. *Discrete-Time Signal Processing.* Prentice Hall signal processing series. Prentice Hall, Upper Saddle River, N.J, 2nd edition, 2014.

[7] A. B. Watson. QUEST+: A general multidimensional Bayesian adaptive psychometric method. *Journal of Vision*, 17(3):10, 2017.