DOI: 10.61782/fa.2023.0575

forumacusticum 2023

TOWARDS A COMPUTATIONALLY EFFICIENT MODEL FOR COMBINED ASSESSMENT OF MONAURAL AND BINAURAL AUDIO QUALITY

Bernhard Eurich* **Thomas Biberger** **Stephan D. Ewert** **Mathias Dietz**

Department for Medical Physics and Acoustics, Universität Oldenburg, Germany

ABSTRACT

Audio signal processing is a core element in hearing devices, allowing for adaptation of the signal properties to the listener's needs in a specific listening situation. Besides the desired signal manipulations, e.g., spectral equalization and binaural noise reduction, signal processing may also introduce monaural or binaural distortions. Auditory models can be applied to predict the perceptual relevance of such distortions and for their minimization. This, however, requires a balance between prediction accuracy and computational efficiency of the model. In this work, the simplistic binaural processing model of Eurich et al. (2022 JASA, 151(6), pp. 3927–3936), based on the hemispheric two-channel code and consistent with binaural psychophysics, was combined with a modified version of the monaural generalized power spectrum model of quality (Biberger et al. 2018, JAES., vol. 66, no. 7/8, pp. 578–593) to cover binaural and monaural audio quality aspects. The suggested model was evaluated with several databases including music and speech signals processed by loudspeakers and algorithms typically applied in modern hearing devices. The presented model performed similar to previously employed computationally more complex models, which makes it applicable to hearing devices and algorithms.

Keywords: *audio quality, binaural hearing, monaural hearing, auditory modeling*

1. INTRODUCTION

Cutting-edge hearing technology benefits from real-time assessment of audio quality, as it allows for quick adjustment of running algorithms to better suit the listener's needs. To achieve this, computationally efficient models for combined monaural and binaural audio quality assessments are essential. While several binaural audio quality models have been developed in the past, e.g., [1–4], they are often too computationally expensive for real-time evaluations on devices. Recently, Fleßner et al. [5] combined the outputs of the monaural generalized power spectrum model for quality (GPSM^q; [6]) and the binaural auditory-model-based quality prediction (BAM-Q; [1]) to predict overall audio quality. However, such approach is not very efficient as stimuli are required to be processed by each of the two quality models. Here, we propose a lean and psychoacoustically validated model for combined monaural and binaural audio quality assessment, which is a first step towards real-time applications in the field of hearing device technology.

2. MODEL

The block diagram of the suggested computationally efficient model for combined assessment of monaural and binaural audio quality is shown in Figure 2.

2.1 Binaural front end

The binaural model is a modified version of the model proposed by Eurich et al. 2022 [7]. After basilar mem-

Corresponding author: bernhard.eurich@uol.de.

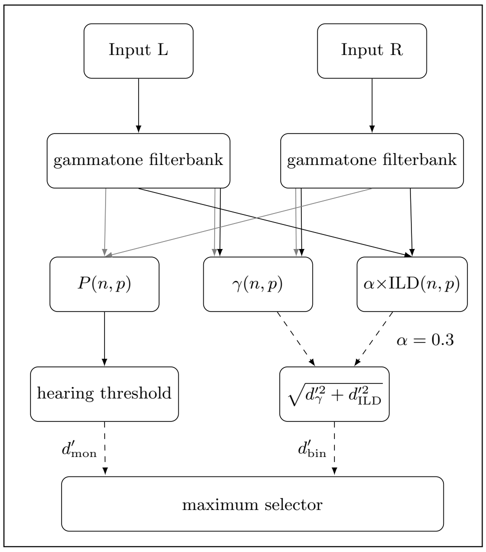

Figure 1. Block diagram of the proposed model. Gray lines denote low-pass filtered Hilbert envelopes, dashed lines indicate differences between reference and test signals.

brane processing using a linear fourth order gammatone filterbank [8], two binaural features are extracted for each frequency band in consecutive time frames of 400 ms:

1. the complex correlation coefficient γ, which is a mathematical formulation of the two-hemispheric channel code [7, 9] represents both the interaural phase difference and the magnitude of its temporal fluctuations. For frequency bands with center frequencies below 1400 Hz, this operates on the temporal fine structure of the bandpass signals, while above of 1400 Hz it operates on their Hilbert envelopes. A first-order lowpass filter with a 150 Hz cutoff frequency is applied to the envelope [10]. To avoid infinite sensitivity, the coherence (i.e. $|\gamma|$) is multiplied by 0.9. As in Eurich et al. [7], Fisher's z transform is applied to the coherence to account for the higher sensitivity to coherence deviations from unity [7].

2. Additionally, interaural level differences (ILDs), i.e. the logarithmic power ratio between left and right signals, were extracted.

2.2 Monaural front end

The monaural front end was adapted from the GPSM^q [6]. For each time-frequency segment, the local DC power was extracted. Time-frequency segment with a local DC power below the hearing threshold in quiet [11] were set to that threshold. Based on the local DC power derived from the test and reference signals, local increment and decrement SNRs were calculated and their dynamic range was limited to 13 dB.

2.3 Back end

The model's sensitivity to distortions was obtained as the difference between the front end outputs of reference and test signals. In the binaural features, information was optimally combined across time frames n and frequency bands p:

$$d' = \sqrt{\sum_{n}^{N} \sum_{p}^{P} [d'(n,p)]^2} \tag{1}$$

The relative weighting ILD and γ features as well as the lowest and highest predictable quality was calibrated using the database from experiment 1 in Fleßner et al. [1].The optimal combination of the two binaural feature sensitivity indices gives the output of the binaural model:

$$d'_{\text{bin}} = \sqrt{d'^2_{\gamma} + d'^2_{\text{ILD}}}. \tag{2}$$

The monaural sensitivity indices were averaged across the time frames and optimally combined across frequency bands.

The monaural and binaural sensitivity indices obtained for all items of a database were normalized to the range [0; 1]. While the sensitivity indices of the model, d', represent the perceptual distance between reference and test signals, the predicted audio quality was obtained as 1-d'. As the overall audio quality has been shown to be dominated by the lower quality aspect [5], the monaural or binaural aspect of the model that predicted the lower quality was selected.

3. EVALUATION

The database of [1] was used for calibration of the binaural path of the suggested model, while three further databases covering a broad variety of monaural, binaural and combined monaural and binaural distortions as they typically occur in loudspeakers and hearables were used to evaluate the "calibrated" model.

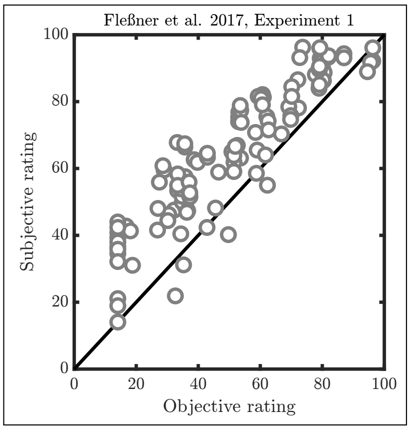

Figure 2. Model predictions for the database provided by Fleßner et al. [1], Experiment 1. This database was used to calibrate the weight of the ILD feature relative to the γ feature of the binaural model pathway.

The binaural calibration database [1] has 114 items, consisting of speech, music, and pink noise signals with a duration of 10 s. The reference signals were diotic and thus perceived in the middle of the head as a narrow spatial image. The test signals were manipulated in ILDs and ITDs to change the perceived apparent source width, listening envelopment and the direction of arrival of the sound source. The listeners rated the perceived difference between a reference and various test signals on a numerical rating scale ranging from 100 (”no difference”) to 0 (”very strong difference”) by using a procedure similar to the MUSHRA (Multiple Stimulus with Hidden Reference and Anchor) method. The following three databases were used for model evaluation. The loudspeaker database, taken from [6], consists of 336 items, based on the ratings of 10 well-trained NH listeners (”expert listeners”) for the perceived overall sound quality difference between a high-quality three-way reference loudspeaker and 59 low-to-mid quality three-way and two-way test speaker systems playing 15 music excerpts (20-30 s). All loudspeakers were digitally equalized in order to evaluated quality differences between test loudspeakers with digitally compensated frequency response and a high-quality three-way reference loudspeaker. The played-back music signals were recorded by a dummy head (Neutric Cortex MK2). The perceived sound quality differences between reference and test signals were rated by using a quasi-continuous rating scale ranging from 0 (imperceptible differences) to 4 (significant differences).

The binaural magnification database, including 8 items, was taken from [1] and comprises binaural hearing aid algorithms ([12]), that magnifies binaural ILD- and ITD-cues to improve the spatial separation between sound sources. The algorithm was applied to one speaker in a conversation scenario who talks with another (unprocessed) speaker. Such processing shifts the perceived location of the processed speaker, while the spatial position of the other speaker does not change. In the unprocessed reference signal both speakers were perceived in front of the receiver. Different degrees of magnifications were tested and 10 NH listeners rated the overall difference between the reference signal and the test signals by using a procedure similar to MUSHRA.

The acoustic transparency database was taken from a study of Schepker et al. [13] and consists of 140 speech and music items. They evaluated the audio quality of a real-time hearing device prototype, aiming at an acoustically transparent sound reproduction, by applying feedback suppression based on a null-steering beamformer and individualized equalization of the sound pressure at the eardrum. A dummy head with inserted hearing devices was used for recordings. The dummy head open-ear recordings served as the reference signals for acoustical transparency. Fifteen NH listeners rated the perceived overall sound quality of each stimulus relative to the (open-ear) reference by using a MUSHRA-like procedure.

The subjective quality ratings for all mentioned databases were measured in headphone experiments with NH subjects in sound-isolated booths.

Table 1. Prediction performance of the suggested model in terms of Pearson linear and Spearman rank correlation coefficient between subjective and predicted quality assessments.

Database	r_{Pearson}	r_{rank}
Binaural calibration [1]	0.91	0.92
Binaural magnification [1, 12]	0.91	1.00
Loudspeaker [6]	0.92	0.88
Acoustic transparency [13]	0.88	0.87

4. RESULTS AND DISCUSSION

For each database, the prediction performance of the suggested model was quantified by the Pearson linear correlation coefficient ($r_{Pearson}$) and the Spearman rank correlation coefficient (r_{rank}) between measured and predicted data.

Table 1 shows the prediction performance for the calibration database (”Binaural calibration”) and the three evaluation databases. For the considered databases, including purely monaural, binaural and combined monaural and binaural distortions, the proposed audio quality model gave accurate predictions, which is indicated by $r_{Pearson} \geq 0.88$ and $r_{rank} \geq 0.87$, respectively.

Despite such good prediction performance, further model evaluations with other types of distortions related to hearables and smart headphones are required to draw a more conclusive picture about its predictive power and limitations. Since the overarching goal is to have a computationally efficient model, one of the next steps is to systematically assess how far the complexity of this model can be further reduced, e.g., reducing the density of auditory filters, while preserving its predictive power.

5. ACKNOWLEDGMENTS

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project-ID 352015383–SFB 1330 A2 and B4, as well as from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (Grant No. 716800).

6. REFERENCES

[1] J.-H. Flessner, R. Huber, and S. Ewert, “Assessment and Prediction of Binaural Aspects of Audio Quality,” *Journal of the Audio Engineering Society*, vol. 65, pp. 929–942, Nov. 2017.

[2] M. Schäfer, M. Bahram, and P. Vary, “An extension of the PEAQ measure by a binaural hearing model,” in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 8164–8168, May 2013.

[3] P. Manocha, A. Kumar, B. Xu, A. Menon, I. D. Gebru, V. K. Ithapu, and P. Calamia, “SAQAM: Spatial Audio Quality Assessment Metric,” in *Interspeech 2022*, pp. 649–653, ISCA, Sept. 2022.

[4] A. Raake, H. Wierstorf, and J. Blauert, “A case for TWO!EARS in audio quality assessment,” *Proceedings of Forum Acusticum, Krakow, 2014*, Sept. 2014.

[5] J.-H. Flesner, T. Biberger, and S. D. Ewert, “Subjective and Objective Assessment of Monaural and Binaural Aspects of Audio Quality,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, pp. 1112–1125, July 2019.

[6] T. Biberger, J.-H. Fleßner, R. Huber, and S. Ewert, “An Objective Audio Quality Measure Based on Power and Envelope Power Cues,” *Journal of the Audio Engineering Society*, vol. 66, pp. 578–593, Aug. 2018.

[7] B. Eurich, J. Encke, S. D. Ewert, and M. Dietz, “Lower interaural coherence in off-signal bands impairs binaural detection,” *The Journal of the Acoustical Society of America*, vol. 151, pp. 3927–3936, June 2022.

[8] V. Hohmann, “Frequency analysis and synthesis using a Gammatone filterbank,” *Acta Acustica united with Acustica*, vol. 88, pp. 433–442, May 2002.

[9] J. Encke and M. Dietz, “A hemispheric two-channel code accounts for binaural unmasking in humans,” *Communications Biology*, vol. 5, p. 1122, Oct. 2022.

[10] A. Kohlrausch, R. Fassel, and T. Dau, “The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers,” *The Journal of the Acoustical Society of America*, vol. 108, pp. 723–734, Aug. 2000.

[11] “ISO (2005). 389-7, Acoustics-Reference Zero for the Calibration of Audiometric Equipment. Part 7: Reference Threshold of Hearing Under Free-Field and Diffuse-Field Listening Conditions (International Organization for Standardization, Geneva, Switzerland),” 2005.

[12] B. Kollmeier and J. Peissig, “Speech Intelligibility Enhancement by Interaural Magnification,” *Acta Oto-Laryngologica*, vol. 109, pp. 215–223, Jan. 1990.

[13] H. Schepker, F. Denk, B. Kollmeier, and S. Doclo, “Subjective Sound Quality Evaluation of an Acoustically Transparent Hearing Device,” in *Audio Engineering Society Conference: 2019 AES INTERNATIONAL CONFERENCE ON HEADPHONE TECHNOLOGY*, Audio Engineering Society, Aug. 2019.