# Towards a binaural modelling toolbox

Peter L. Søndergaard, John F. Culling, Torsten Dau, Nicolas Le Goff,
Morten L. Jepsen, Piotr Majdak and Hagen Wierstorf.

## Abstract

The Auditory Modelling Toolbox (AMToolbox) is a new Matlab / Octave toolbox for developing and applying auditory perceptual models and in particular binaural models. The philosophy behind the project is that the models should be implemented in a consistent manner, well documented and user-friendly in order to allow students and researchers to actively work with current models and further develop existing ones. In addition to providing the models, it is a goal of the project to collect published human data and definitions of model experiments. This will simplify the verification of models by running the model experiments and comparing the predictions to human data. The software is released under the GNU Public License (GPL) version 3, and can be downloaded from `http://amtoolbox.sourceforge.net`.

## 1 Introduction

An auditory model is a mathematical algorithm that mimics a part of the human auditory system. There are at least two main reasons why auditory processing models are constructed: to represent the results from a variety of experiments within one framework and to explain the functioning of the system. Specifically, processing models help generate hypotheses that can be explicitly stated and quantitatively tested for complex systems. The models can also help evaluate how a deficit in one or more components affects the overall operation of the system. Some of the models can be useful for technical and clinical applications, such as the improved human-machine communication by employing auditory modelling based processing techniques, or new processing strategies in digital hearing aids and cochlear implants.

The development of auditory models has been hampered by the complexity of the individual auditory processing stages and their interactions. This resulted in a multiplicity of auditory models described in the literature. Models of auditory processing may be roughly classified into biophysical, physiological, mathematical (or statistical) and perceptual models depending on which aspects of processing are considered.

The major goals of the AMToolbox project is to simplify the development of new auditory models, and to make it easier for students and researchers to enter the field. This is possible because of the following three virtues of the toolbox:

1. Accessibility: AMToolbox can be obtained under a free software license, the GNU Public License version 3. Is is free to download and use by anyone.

2. Consistency: All functions are written in the same style, using the same name for key concepts and conventions for conversion of physical units to numbers in Matlab.

3. Reproducibility: AMToolbox contains test scripts and data functions to reproduce results (figures and tables) from selected papers. This provides validation of existing models and makes it easier to develop new models.

These virtues of research and software development are gaining traction (see for instance [1] about reproducible research in signal processing).

AMToolbox is build on top of the Linear Time-Frequency Analysis toolbox (LTFAT) [2], which provides a modern and stable foundation for the signal processing done in the models. Much of the cooperation on the AMToolbox takes place within the framework of the "Aural Assessment by means of Binaural Algorithms" (AABBA) project, [3].

In Section 2, the currently implemented stages of auditory signal processing are summarised, in-

cluding descriptions of the signal processing stages in the periphery, and decision stages after the assumed preprocessing. In Section 3, we present full models build from the model stages. These models have all been previously published and verified against human data.

In Section 4, we present the validation framework of the toolbox. This consists of human data from published studies and scripts to reproduce the data.

In this paper, text appearing in typewriter-style denotes names of functions in the toolbox, i.e. `gammatone`.

# 2 Model stages

## 2.1 Auditory scales

Several phenomena of the human auditory system show a linear frequency-dependence at low frequencies, and a logarithmic dependence at higher frequencies. These include the just-noticeable difference in frequency, giving rise to the mel scale [4] and a variant reported by Fant [5], the notion of critical bands by Zwicker [6] giving rise to the Bark scale and the equivalent rectangular bandwidth of the auditory filters giving rise to the ERB scale [7], which was revised in [8]. These scales (including their revisions) are available in the toolbox, and may be used for perceptually-related visualisation purposes and filtering.

## 2.2 Basilar membrane models

A classical model of the basilar membrane (BM) processing is the gammatone filterbank, of which there exist many variations. An overview is presented in [9]. In the toolbox, the original IIR approximation from [10] and the all-pole approximation proposed by Lyon [11] have been implemented for both real- and complex-valued filters. To build a complete filterbank covering the audible frequency range, the centre frequencies of the gammatone filters are typically chosen to be equidistantly spaced on an auditory scale, using bandwidths that are proportional to the distance between neighbouring centre frequencies. The toolbox implements the auditory-filter-bandwidth function from [8] which is consistent with the ERB-scale. The linear approximation to the basilar membrane processing is done by the function `auditoryfilterbank`, which in turn obtains the filters from the `gammatone` function.

The dual-resonance nonlinear (DRNL; [12, 13]) stage introduces the modelling of the nonlinearities in the peripheral processing. The most striking feature is a compressive input-output function, and consequently level-dependent tuning. The DRNL is computed by the `drnl` function. Parameters sets from [13, 14] are supported.

## 2.3 Inner hair cell envelope extraction

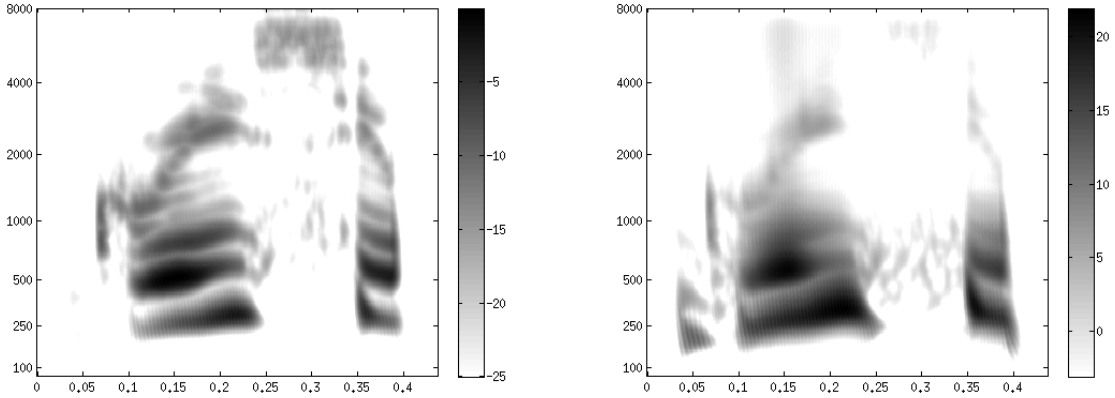
The envelope extraction process performed by the inner hair cell (IHC) is typically modelled by a half-wave rectification followed by a lowpass filtering. However, there are many variations to this scheme, where each auditory model uses a variation on the type of filter and cutoff frequency. Binaural models typically use a lower cutoff frequency for the lowpass filtering than monaural models. In the toolbox, the IHC models used by Bernstein ([15], 425 Hz cutoff), Breebaart ([16], 770 Hz cutoff), Dau ([17], 1000 Hz cutoff) and Lindemann ([18], 800 Hz cutoff) have been implemented. The IHC models are collected in the `ihcenvelope` function.

## 2.4 Adaption

Adaptation loops is a method to model the adaptive properties of the auditory periphery by using a chain of (typically 5) feedback loops in series. Each loop has a different time constant. The toolbox contains adaptation loops using the linearly spaced constants originally found by Püschel [19] and later used by Breebaart [16], and the constants found in [17] to better approximate forward masking data. In [20] it was found that the original definition from [19] behaves erratically if the input changes from complete silence, so the original definition was modified to include a minimum level (to avoid the transition from complete silence) and an overshoot limitation. The adaptation loop processing is done by the `adaptloop` function.

## 2.5 Modulation processing

The modulation filter bank is a processing stage that accounts for amplitude modulation (AM) de-




(a) Speech signal 'greasy' at 50 dB SPL.

(b) Speech signal 'greasy' at 90 dB SPL.

Figure 2.1: Output of the DRNL and inner hair cell envelope extraction of an input signal presented at two different levels. The dynamic range of the plots has been limited to 25 dB, to increase visibility. The widening of the DRNL filters at higher levels can be seen by comparing the plots. This is part of the output of the `demo_drnl` script.

tection and AM masking, [21, 20]. The input to the modulation filter bank is lowpass filtered using a first order Butterworth filter with a cutoff frequency at 150 Hz. This filter simulates a decreasing sensitivity to sinusoidal modulation as a function of modulation frequency.

By default, the modulation filters have centre frequencies of 0, 5, 10, 16.6, 27.77, ... where each next centre frequency is $5/3$ times the previous one. For modulation frequencies below (and including) 10 Hz, the real value of the filters are returned, and for higher modulation centre frequencies, the absolute value (the envelope) is returned. Modulation filter bank processing is done by the `modfilterbank` function.

### 2.6 Optimal detector

The optimal detector is a signal detection method based on signal detection theory [22]. The method works by deriving a "template" of the signal to be detected. This template is calculated as the normalised difference between the internal representation of the masker plus a suprathreshold signal representation and that of the masker alone. During the simulation procedure, the internal representation of the masker alone is calculated and subtracted from the internal representation in each interval of a given trial. Thus, in the signal interval, the difference contains the signal, embedded in internal noise, while the reference intervals contain internal noise only. A decision is made on the basis of the cross-correlation values obtained in the different intervals. The interval that produces the largest value is assumed to be the signal interval.

In the Dau version of the optimal detector [17], the template is derived from a supra-threshold value of the stimulus, and the template is fixed during the experiment. This corresponds to the subject having a prior knowledge of the signal. In the Breebaart version [16], the template is updated in the course of the experiment. This means that the model "learns" from the experiment by updating the template.

## 3 Models

### 3.1 The Dau et al. models

In 1996, Dau et. al. [17] proposed a model of human auditory perception. The model included stages of linear BM filtering, inner-haircell transduction, nonlinear neural adaptation, a modulation lowpass filter and an optimal detector as the deci-

sion device. The model was shown to quantitatively account for a variety of psychoacoustical data associated with simultaneous and non-simultaneous masking, [23]. In subsequent studies [20, 24], the modulation lowpass filter was replaced by a modulation filterbank, which enabled the model to account for AM detection and AM masking.

Jepsen et al. [14] further improved the predictive power of the model by replacing the linear BM stage by the DRNL, [13]. It was shown that the model's ability to account for the data of the previous studies was preserved, and the updated model could further account for psychoacoustical data associated with nonlinear and level dependent auditory processing. The internal representation described by these models can be computed by the `dau1996preproc`, `dau1997preproc` and `jepsen2008preproc` functions. To use the models in an AFC setting, they must be combined with the optimal detector step, which is done in the `dau1996`, `dau1997` and `jepsen2008` functions.

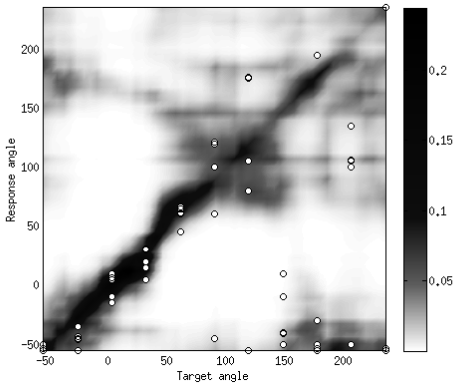



Figure 3.1: Output of the `demo_langendijk` script: Response-target plot for the median plane. Darker colours represent larger probability for the response calculated by the model. Circles represent the actual responses from a sound localization experiment.

## 3.2 The Breebaart model

In 2001, Breebaart et al. [16, 25, 26] proposed a binaural model of human auditory perception. The model is essentially an extension of the monaural model proposed by Dau et al. in 1996 [17], from which it uses the peripheral stages: linear basilar membrane filtering, inner-haircell transduction and non-linear neural adaptation. The peripheral internal representations for the left and right ear are then combined in an equalisation-cancellation (EC-type) binaural processor consisting of excitation-inhibitions (EI) elements which produces binaural internal representations fed to an optimal detector used as a decision device. The model has been shown to predict a large range of binaural detection tasks. It has also been shown to account for basic lateralisation tasks and was successively adapted by Park [27] to evaluate sound localisation performance for stereophonic systems. The internal representation of the Breebaart model is calculated by `breebaart2001preproc`, whereas `breebaart2001` is the complete model to be used in an AFC framework.

## 3.3 The Zakarauskas and Langendijk models

Zakarauskas et al. [28] introduced a model to predict vertical-plane sound-localisation performance based on the analysis of the incoming monaural sound spectrum. The model relies on head-related transfer functions (HRTFs) which describe the filtering of the incoming sound by the torso, head, and pinna. The model uses a peripherally-processed set of HRTFs to mimic the representation of the localisation cues in the auditory system. The decision process is simulated by minimising the spectral difference (in terms of mean or variance) between the peripherally-processed incoming sound spectrum and HRTFs from the set. The estimated position is given by the HRTF with the smallest difference. Langendijk et al. [29] proposed an improved model by considering only the directional information of the HRTFs and incorporating Bayes' statistics into the decision process. Both models, considering the monaural spectral information only, are able to predict perceived positions of the stationary wide-band sounds within the median plane.

### 3.4 The Culling model

Lavandier and Culling [30] developed a model of spatial unmasking for speech in noise and reverberation and validated it against human speech reception thresholds (SRTs). The underlying structure of the model has since been improved [31]. The method now operates directly upon binaural room impulse responses (BRIRs). It has two components, better-ear listening and binaural unmasking, which are assumed to be additive. The BRIRs are filtered into different frequency channels using an auditory filterbank [10]. The better-ear listening component assumes that the listener can select sound from either ear at each frequency according to which one has the better signal-to-noise ratio (SNR). The better-ear SNRs are then weighted and summed across frequency according to Table I of the Speech Intelligibility Index [32]. The binaural unmasking component calculates the binaural masking level difference within each frequency channel from equalisation-cancellation theory [33, 34]. These values are similarly weighted and summed across frequency. The summed output is the effective binaural SNR, which can be used to predict differences in SRT across different listening situations. The model has been validated against a number of different sets of SRTs both from the literature and from the authors own measurements [31]. The output of the model can be used to predict the effects of noise and reverberation on speech communication for both normally-hearing listeners and users of auditory prostheses and to predict the benefit of optimal head orientation.

### 3.5 The Lindemann and Gaik models

Lindemann [18] introduced a binaural model for predicting the lateralisation of a sound. His model extended the delay line principle introduced by Jeffress [35] by contralateral inhibition and monaural processors. The Lindemann model relies on a running interaural cross correlation process to calculate the interaural time difference (ITD). The contralateral inhibition sharpens the cross correlation peak and integrates interaural level differences (ILD) into the model by shifting the peak. This is an inherent part of the inhibition process due to its dependence on the amplitude. The whole

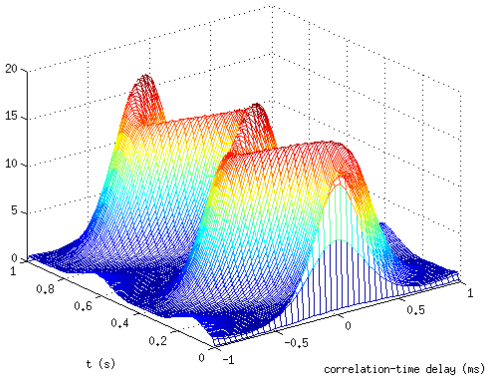



Figure 3.2: Output of the `demo_lindemann` script: Binaural activity map for one frequency channel. Input is a 500 Hz sinusoid with a binaural modulation rate of 2 Hz.

model consists of linear BM filtering, inner-haircell transduction by simple half-wave rectification and low pass filtering at a cutoff frequency of 800 Hz and finally the cross correlation and inhibition step. The output of the model is the interaural cross-correlation time in each frequency band of the BM filterbank. The Lindemann model can handle stimuli with a combination of ITD and ILD and can predict split images for unnatural combinations of the two. This has been optimised by Gaik [36] by extracting natural combinations of ITDs and ILDs from HRTFs and adding an additional weighting step to the Lindemann model. The Gaik model can be used with any suitable HRTF set.

## 4 Human data and experiments

The goal the human data from psychoacoustic experiments and psychophysical measurements included in AMToolbox is twofold:

1. To support auditory research and development of auditory models by providing a quick access to already existing data.

2. To provide a 'target' for the evaluation of models. This makes it easy to evaluate models against a large set of existing data.

The data is provided either by the nature of the data, i.e. data about the absolute thresholds of hearing recorded using various reproduction mechanism collected in one function (i.e. `absolutethreshold`), or by the figure/table of the underlying publication (i.e. `data_lindemann1986a`). The last method provides a very intuitive access of the data to the user, as the documentation for the data is provided in the referenced paper.

The toolbox includes data for the absolute threshold of hearing in a free field as defined in [37]. These data can be converted to provide information about the minimal audible pressure (MAP) at the eardrum using the method from [38]. MAP data for the ER-3A insert earphones [39], the ER-2A insert earphone [40] and the Sennheiser HDA-200 [41] is provided. Absolute thresholds for the ER2A and HDA-200 from [42] are provided up to 16 kHz.

The toolbox includes a small alternative forced choice (AFC) framework. The models in the toolbox that support it (`dau1996`, `dau1997`, `jepsen2008` and `breebaart2001`) can then be used with the framework to make predictions in place of a human listener. This way, the same experiment definitions can be used both for experiments with humans or models.

# 5 Conclusion

Work has just begun on the auditory modelling toolbox. The authors hope that the project will continue to flourish in the coming years and that it will grow to encompass more models, and become the platform for future state-of-the-art auditory models.

## Acknowledgements

The authors would like to thank all the original developers of the models, Jens Blaurt and Armin Kohlrausch for organising the AABBA project and Robert Baumgartner, Marton Marschall and Katarina Egger for contributing to the project.

# References

[1] P. Vandewalle, J. Kovacevic, and M. Vetterli. Reproducible research in signal processing - what, why, and how. *IEEE Signal Processing Magazine*, 26(3):37–47, May 2009.

[2] P. L. Søndergaard, B. Torrésani, and P. Balazs. The Linear Time Frequency Analysis Toolbox. *International Journal of Wavelets, Multiresolution Analysis and Information Processing*, submitted, 2011.

[3] A. Kohlrausch, J. Blauert, J. Braasch, H. Colburn, J. Culling, T. Dau, V. Hohmann, U. Jekosch, J. Mourjopoulos, V. Pulkki, A. Raake, P. Søndergaard, and D. Kolossa. Introducing the Topic 'Aural Assessment by means of Binaural Algorithms'. In *Proceedings of the Forum Acousticum*, 2011.

[4] S. Stevens, J. Volkmann, and E. Newman. A scale for the measurement of the psychological magnitude pitch. *J. Acoust. Soc. Am.*, 8:185, 1937.

[5] G. Fant. Analysis and synthesis of speech processes. In B. Malmberg, editor, *Manual of phonetics*. North-Holland, 1968.

[6] E. Zwicker. Subdivision of the audible frequency range into critical bands (frequenzgruppen). *J. Acoust. Soc. Am.*, 33(2):248–248, 1961.

[7] B. Moore and B. Glasberg. Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *J. Acoust. Soc. Am.*, 74:750, 1983.

[8] B. R. Glasberg and B. Moore. Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47(1-2):103, 1990.

[9] R. Lyon, A. Katsiamis, and E. Drakakis. History and future of auditory filter models. In *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*, pages 3809–3812. IEEE, 2010.

[10] R. Patterson, I. Nimmo-Smith, J. Holdsworth, and P. Rice. An efficient auditory filterbank based on the gammatone function. *APU report*, 2341, 1988.

[11] R. Lyon. All pole models of auditory filtering. *Diversity in auditory mechanics, World Scientific Publishing, Singapore*, 1997.

[12] R. Meddis, L. O'Mard, and E. Lopez-Poveda. A computational algorithm for computing nonlinear auditory frequency selectivity. *J. Acoust. Soc. Am.*, 109:2852, 2001.

[13] E. Lopez-Poveda and R. Meddis. A human nonlinear cochlear filterbank. *J. Acoust. Soc. Am.*, 110:3107, 2001.

[14] M. Jepsen, S. Ewert, and T. Dau. A computational model of human auditory signal processing and perception. *J. Acoust. Soc. Am.*, 124(1):422–438, 2008.

[15] L. Bernstein, S. van de Par, and C. Trahiotis. The normalized interaural correlation: Accounting for NoS$\pi$ thresholds obtained with Gaussian and l̈ow-noiseẗmasking noise. *J. Acoust. Soc. Am.*, 106:870, 1999.

[16] J. Breebaart, S. van de Par, and A. Kohlrausch. Binaural processing model based on contralateral inhibition. I. Model structure. *J. Acoust. Soc. Am.*, 110:1074–1088, August 2001.

[17] T. Dau, D. Püschel, and A. Kohlrausch. A quantitative model of the effective signal processing in the auditory system. I. Model structure. *J. Acoust. Soc. Am.*, 99(6):3615–3622, 1996a.

[18] W. Lindemann. Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals. *J. Acoust. Soc. Am.*, 80:1608, 1986.

[19] D. Püschel. *Prinzipien der zeitlichen Analyse beim Hören*. PhD thesis, Universität Göttingen, 1988.

[20] T. Dau, B. Kollmeier, and A. Kohlrausch. Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. *J. Acoust. Soc. Am.*, 102:2892, 1997a.

[21] R. Fassel and D. Püschel. Modulation detection and masking using deterministic and random maskers. *Contributions to Psychological Acoustics, edited by A. Schick (Universitätsgesellschaft Oldenburg, Oldenburg)*, pages 419–429, 1993.

[22] D. Green and J. Swets. *Signal detection theory and psychophysics*. Wiley New York, 1966.

[23] T. Dau, D. Püschel, and A. Kohlrausch. A quantitative model of the "effective" signal processing in the auditory system. II. Simulations and measurements. *J. Acoust. Soc. Am.*, 99:3623, 1996b.

[24] T. Dau, B. Kollmeier, and A. Kohlrausch. Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration. *J. Acoust. Soc. Am.*, 102:2906, 1997b.

[25] J. Breebaart, S. van de Par, and A. Kohlrausch. Binaural processing model based on contralateral inhibition. II. Dependence on spectral parameters. *J. Acoust. Soc. Am.*, 110:1089–1104, August 2001.

[26] J. Breebaart, S. van de Par, and A. Kohlrausch. Binaural processing model based on contralateral inhibition. III. Dependence on temporal parameters. *J. Acoust. Soc. Am.*, 110:1105–1117, August 2001.

[27] M. Park, P. A. Nelson, and K. Kang. A model of sound localisation applied to the evaluation of systems for stereophony. *Acta Acustica united Acoustica*, 94:825–839, 2008.

[28] P. Zakarauskas and M. Cynader. A computational theory of spectral cue localization. *J. Acoust. Soc. Am.*, 94:1323, 1993.

[29] E. Langendijk and A. Bronkhorst. Contribution of spectral cues to human sound localization. *J. Acoust. Soc. Am.*, 112:1583, 2002.

[30] M. Lavandier and J. Culling. Prediction of binaural speech intelligibility against noise in rooms. *J. Acoust. Soc. Am.*, 127:387, 2010.

[31] S. Jelfs, J. Culling, and M. Lavandier. Revision and validation of a binaural model for speech intelligibility in noise. *Hearing Research*, 2010.

[32] American National Standards Institute, New York. *Methods for calculation of the speech intelligibility index*, ANSI S3.5-1997 edition, 1997.

[33] N. I. Durlach. Binaural signal detection: equalization and cancellation theory. In J. V. Tobias, editor, *Foundations of Modern Auditory Theory. Vol. II*, pages 369–462. Academic, New York, 1972.

[34] J. Culling. Evidence specifically favoring the equalization-cancellation theory of binaural unmasking. *J. Acoust. Soc. Am.*, 122:2803, 2007.

[35] L. Jeffress. A place theory of sound localization. *Journal of comparative and physiological psychology*, 41(1):35–39, 1948.

[36] W. Gaik. Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling. *J. Acoust. Soc. Am.*, 94:98, 1993.

[37] ISO 226:2003. *Acoustics – Normal equal-loudness-level contours*. International Organization for Standardization, Geneva, Switzerland, 2003.

[38] R. A. Bentler and C. V. Pavlovic. Transfer Functions and Correction Factors used in Hearing Aid Evaluation and Research. *Ear and Hearing*, 10:58–63, 1989.

[39] ISO 389-2:1994(E). *Acoustics – Reference zero for the calibration of audiometric equipment – Part 2: Reference equivalent threshold sound pressure levels for pure tones and insert earphones*. International Organization for Standardization, Geneva, Switzerland, 1994.

[40] L. Han and T. Poulsen. Equivalent threshold sound pressure levels for Sennheiser HDA 200 earphone and Etymotic Research ER-2 insert earphone in the frequency range 125 Hz to 16 kHz. *Scandinavian Audiology*, 27(2):105–112, 1998.

[41] ISO 389-8:2004. *Acoustics – Reference zero for the calibration of audiometric equipment – Part 8: Reference equivalent threshold sound pressure levels for pure tones and circumaural earphones*. International Organization for Standardization, Geneva, Switzerland, 2004.

[42] ISO 389-5:2006. *Acoustics – Reference zero for the calibration of audiometric equipment – Part 5: Reference equivalent threshold sound pressure levels for pure tones in the frequency range 8 kHz to 16 kHz*. International Organization for Standardization, Geneva, Switzerland, 2006.