1

The Auditory Modeling Toolbox

P. L. Søndergaard and P. Majdak

Acoustics Research Institute, Austrian Academy of Sciences, Wien, Austria

Summary. The *Auditory Modeling Toolbox*, AMToolbox, is a Matlab/Octave toolbox for developing and applying auditory perceptual models with a particular focus on binaural models. The philosophy behind the AMToolbox is the consistent implementation of auditory models, good documentation, and user-friendly access in order to allow students and researchers to work with and to advance existing models. In addition to providing the model implementations, published human data and model demonstrations are provided. Further, model implementations can be evaluated by running so-called experiments aimed at reproducing results from the corresponding publications. AMToolbox includes many of the models described in this volume. It is freely available from http://amtoolbox.sourceforge.net

Cite this article and the AMToolbox as:

> Søndergaard, P. and Majdak, P. (2013). "The Auditory Modeling Toolbox,"
> in *The Technology of Binaural Listening*, edited by Blauert, J. (Springer,
> Berlin, Heidelberg), pp. 33-56.

1.1 Introduction

An auditory model is a mathematical algorithm that mimics part of the human auditory system. There are at least two main motivations for developing auditory processing models: First, to represent the results from a variety of experiments within one framework and to explain the functioning of the auditory system. In such cases, the models help to generate hypotheses that can be explicitly stated and quantitatively tested for complex systems. Second, models can help to evaluate how a deficit in one or more components affects the overall operation of the system. In those cases, some of the models can be useful for technical and clinical applications, such as the improvement of human-machine communication by employing auditory modeling based processing techniques, or the development of new processing strate-

gies in hearing-assist devices. The *auditory modeling toolbox*, AMToolbox, is a freely available collection of such auditory models.[1]

Often a new auditory model aims at improving an already existing one. Thus, auditory modeling begins with the process of comprehending and reproducing previously published models. Imagine a thesis adviser who wants to integrate a new feature Y into an existing model X. The student might spend months on the implementation of X, trying to reproduce the published results for X, before even being able to integrate the feature Y. While already the re-implementation of old models sounds like re-inventing the wheel, sometimes, it is even not possible to validate the new implementation of the old model because of lack of the original data as used in the original publication. This problem is not new, it has already been described in [9] as follows.

> "An article about computational science in a scientific publication is *not* the scholarship itself, it is merely *advertising* of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures"

In order to address this problem, the manuscript publication must go with the software publication, allowing to reproduce the published research, a strategy called reproducible research [10]. Reproducible research is becoming more and more popular - see for instance [76] - and the AMToolbox is an attempt to promote the reproducible research strategy within the hearing science by pursuing the following three virtues.

- *Reproducibility* in terms of
  - Valid reproduction of the published outcome like figures and tables from selected publications
  - Trust in the published models with no need for a repetition of the verification
  - Modular model implementation and documentation of each model stage with a clear description of the input and output data format
- *Accessibility*, namely, free and open source software, available to download, use, and contribute by anyone
- *Consistency*, achieved by all functions written in the same style, using the same names for key concepts and conventions for conversion of physical units to numbers

In the past, other toolboxes concerning auditory models have been published [71, 60, 63]. The *auditory toolbox* [71] was an early collection of implementations focused on auditory processing. It contains basic models of the human peripheral processing, but the development of that toolbox seems to have stopped. The *auditory-image-model toolbox*, AIM, [63] comprises a more up-to-date model of the neural responses at the level of the auditory nerve. It seems to be still actively developed. The *development system for auditory modeling*, DSAM, [60], includes various auditory nerve models including the AIM. Written in C, it provides a great basis for the development of computationally efficient applications. Note that while the source code of the DSAM is free, the documentation is only commercially available.

---

[1] Much of the cooperation on the AMToolbox takes place within the framework of the AABBA group, an open group of scientist dealing with *aural assessment by means of binaural algorithms*

In contrast to those toolboxes, the *auditory modeling toolbox*, AMToolbox, comprises a larger body of recent models, provides a rating system for the objective evaluation of the implementations, is freely available – both code and documentation – and offers high proficiency gain when it comes to understanding and further developing existing model implementations.

1.2 Structure and Implementation Conventions

The AMToolbox is published under the GNU general public license version 3, a free and open source license[2] that guarantees the freedom to share and modify it for all its users and all its future versions. The AMToolbox, including its source code, is available from SourceForge[3]. AMToolbox works not only in Matlab[4], versions 2009b and higher, but is in particular developed for Octave[5], version 3.6.0 and higher, in order to avoid the need for any commercial software. The development is open and transparent by keeping the source files in the software repository *Git*[6] allowing for independent contributions and developments by many people. AMToolbox has been tested on 64–bit Windows 7, on Mac OSX 10.7 *Lion*, and on several distributions of Linux. Note that for some models, a compiler for C or Fortran is required. While Octave is usually provided with a compiler, for Matlab the compiler must be installed separately. Therefore, binaries for major platforms are provided for Matlab.

 AMToolbox is build on top of the *large time-frequency-analysis toolbox*, LTFAT, [72]. LTFAT is a Matlab/Octave toolbox for time-frequency analysis and multichannel digital signal processing. LTFAT is free and open source. It provides a stable implementation of the signal processing stages used in the AMToolbox. LTFAT is intended to be used both as a computational tool and for teaching and didactic purposes. Its features are basic Fourier analysis and signal processing, stationary and non-stationary Gabor transforms, time-frequency bases like the modified discrete cosine transform, and filterbanks and systems with variable resolution over time and frequency. For all those transforms, inverse transforms are provided for a perfect reconstruction.

 Further, LTFAT provides general, not model-related auditory functions for the AMToolbox. Several phenomena of the human auditory system show a linear frequency dependence at low frequencies, and an approximately logarithmic dependence at higher frequencies. These include the just-noticeable difference in frequency, giving rise to the *mel scale* [74] and its variants [26]. The concept of critical bands giving rise to the *Bark scale* [82], and the equivalent rectangular bandwidth, ERB, of the auditory filters giving rise to the *ERB scale* [58] – later revised in [31]. All these scales, including their revisions, are available in the LTFAT toolbox as frequency-mapping functions.

---

[2] `http://www.gnu.org/licenses/gpl.html`, last viewed on 9.1.2013
[3] `http://sourceforge.net/projects/amtoolbox`, last viewed on 9.1.2013
[4] `http://www.mathworks.de/products/matlab/`, last viewed on 9.1.2013
[5] `http://www.gnu.org/software/octave/`, last viewed on 9.1.2013
[6] `http://git-scm.com/`, last viewed on 11.1.2013

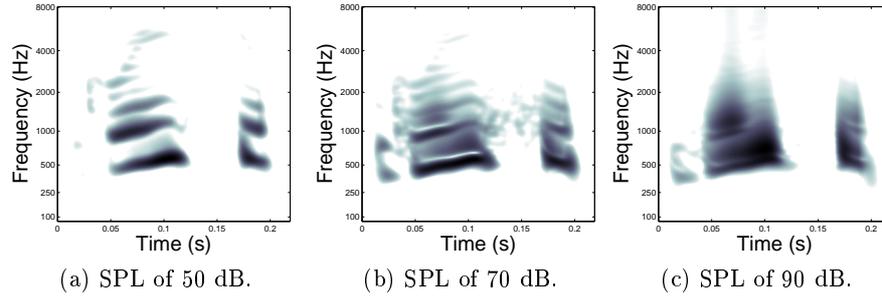(a) SPL of 50 dB.   (b) SPL of 70 dB.   (c) SPL of 90 dB.

Fig. 1.1: Example for a demonstration provided in the AMToolbox. The three panels show spectrograms of the DRNL filterbank and IHC envelope extraction of the speech signal [*greasy*] presented at different levels. The figure can be plotted by evaluating the code `demo_drnl`

1.2.1 Structure

AMToolbox consists of monaural and binaural auditory models, as described in the latter sections of this chapter, complemented with additional resources. The additional resources are

- *Data* from psychoacoustic experiments and acoustic measurements, used in and retrieved from selected publications
- *Experiments*, that is, applications of the models with the goal of simulating experimental runs from the corresponding publications
- *Demonstrations* of a simple kind, for getting started with a model or data

By providing both the data and the experiments, two types of verifications can be applied, namely,

- Verifications where the human data serve to reproduce figures from a given paper showing recorded human data
- Verification where *experiment functions* simulate experimental runs from a given paper and display the requested plots. Data collected from experiments with human can then be compared by visual inspection

Demonstrations are functions beginning with `demo_`. The aim of the demonstrations is to provide examples for the processing and output of a model in order to get quickly into the purpose and functionality of the model. Demonstrations do not require input parameters and provide a visual representation of a model output. Fig. 1.1 shows an example for a demonstration, `demo_drnl`, which plots the spectrograms of the dual-resonance nonlinear, DRNL, filterbank and inner-hair-cell, IHC, envelope extraction of the speech signal [*greasy*].

Data

The data provide a quick access to already existing data and a target for an easy evaluation of models against a large set of existing data. The data are provided

either by a collection of various measurement results in a single function, for example, `absolutethreshold`, where the absolute hearing thresholds as measured with various methods are provided – see Fig. 1.2 – or by refering to the corresponding publication, for example, `data_lindemann1986a`. The latter method provides a very intuitive access of the data to the user, as the documentation for the data is provided in the referenced publication. The corresponding functions begin with `data_`.

Currently, the AMToolbox provides the following publication-specific data.

- `data_zwicker1961`: Specification of critical bands [81]
- `data_lindemann1986a`: Perceived lateral position under various conditions [47]
- `data_neely1988`: Auditory brainstem responses (ABR) wave V latency as function of center frequency [59]
- `data_glasberg1990`: Notched-noise masking thresholds [31]
- `data_goode1994`: Stapes footplate displacement [32]
- `data_pralong1996`: Amplitudes of the headphone and outer ear frequency responses [64]
- `data_lopezpoveda2001`: Amplitudes of the outer and middle ear frequency responses [48]
- `data_langendijk2002`: Sound-localization performance in the median plane [45]
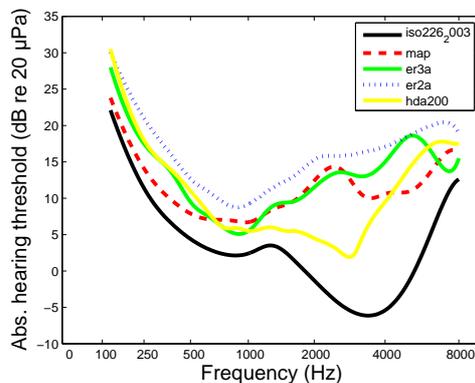

Fig. 1.2: Example for data provided in the AMToolbox. The figure shows the absolute hearing thresholds measured under various conditions and returned by the function `absolutethreshold` by evaluating the following code:

```
types = {'iso226_2003','map','er3a','er2a','hda200'};
symbols = {'k' ,'r--' ,'g' ,'b:','y' };
fc=125:125:8000; hold on; box on;
for ii=1:numel(types),
    opt={symbols{ii}, 'LineWidth', 3};
    semiaudplot(fc,absolutethreshold(fc,types{ii}),'opts',opt);
end;
legend(types); xlabel('Frequency (Hz)','FontSize',16);
ylabel('Abs. hearing threshold (dB re 20 Pa)','FontSize',16);
```

- `data_elberling2010`: ABR wave V data as function of level and sweeping rate [22]
- `data_roenne2012`: Unitary response reflecting the contributions from different cell populations within the auditory brainstem [68]
- `data_majdak2013`: Directional responses from a sound-localization experiment involving binaural listening with listener-specific HRTFs and matched and mismatched crosstalk-cancellation filters[52]
- `data_baumgartner2013`: Calibration and performance data for the sagittal-plane sound localization model [3]

The general data provided by the AMToolbox include data like the *speech intelligibility index* as a function of frequency [12] – `siiweightings`. Further, data for the *absolute threshold of hearing* in a free field [37] – `absolutethreshold`. The absolute thresholds are further provided as the *minimal audible pressures*, MAPs, at the eardrum [4] by using the flag `'map'`. The MAPs are provided for the insert earphones ER-3A (Etymotic) [38] and ER-2A (Etymotic) [33] as well as the circumaural headphone HDA-200 (Sennheiser) [40]. Absolute thresholds for the ER-2A and HDA-200 are provided for the frequency range up to 16 kHz [39].

Experiments

AMToolbox provides applications of the models that simulate experimental runs from the corresponding publications, and display the outcome in the form of numbers, figures or tables – see for instance Fig. 1.3. A model application is called `experiment`, the corresponding functions begin with `exp_`. Currently, the following experiments are provided.

- `exp_lindemann1986a`: Plots figures from [47] and can be used to visualize differences between the current implementation and the published results of the binaural cross-correlation model
- `exp_lopezpoveda2001`: Plots figures from [48] and can be used to verify the implementation of the DRNL filterbank
- `exp_langendijk2002`: Plots figures from [45] and can be used to verify the implementation of the median-plane localization model, see Fig. 1.3
- `exp_jelfs2011`: Plots figures from [42] and can be used to verify the implementation of the binaural model for speech intelligibility in noise
- `exp_roenne2012`: Plots figures from [68] and can be used to verify the implementation of the model of auditory evoked brainstem responses to transient stimuli
- `exp_baumgartner2013`: Plots figures from [3] and can be used to verify the implementation of the model of sagittal-plane sound localization performance

1.2.2 Documentation and Coding Conventions

In order to ensure traceability of each model and data, each implementation must be backed up by a publication in indexed articles, standards, or books. In the AMToolbox, the models are named after the first author and the year of the publication. This convention might appear unfair to the remaining contributing authors, yet, it establishes a straight-forward naming convention. Similarly, other files necessary for
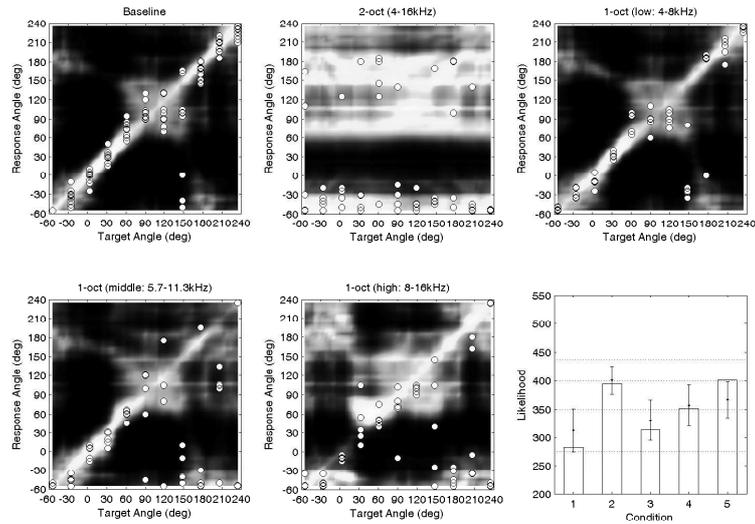
Fig. 1.3: Example for an experiment in the AMToolbox. The figure is plotted by evaluating the code `exp_langendijk2002('fig7')` and aims at reproducing Fig. 7 of the article describing the `langendijk2008` model [45].

a model are prefixed by the name of the model, that is, first author plus the year, to make it clear to which model they belong.

All function names are lowercase. This avoids a lot of confusion because some computer architectures respect upper/lower casing and others do not. Furthermore, in Matlab/Octave documentation, function names are traditionally converted to uppercase. It is also not allowed to use underscores in variable or function names because they are reserved for structural purposes, for example, as in `demo_gammatone` or `exp_lindemann1986a`. As much as possible, function names indicate what they do, rather than the algorithm they use, or the person who invented it. We do not allow to use global variables since they would make the code harder to debug and to parallelize. Variable names are allowed to be both lower and upper case.

Further details on the coding conventions used in the AMToolbox can be found at the website[7].

1.2.3 Level Conventions

Some auditory models are nonlinear and the numeric representation of physical quantities like pressure must be well-defined. The auditory models included in the AMToolbox have been developed with a variety of level conventions. Thus, the interpretation of the numeric unity, that is, the of value 1, varies. For of historical

---

[7] `http://amtoolbox.sourceforge.net/notes/amtnote003.pdf`, last viewed on 9.1.2013

reasons, per default, the unity represents the *sound-pressure level*, SPL, as the *root-mean-square* value, RMS, of 100 dB. The function `dbspl`, however, allows to globally change this representation in the AMToolbox. Currently, the following values for the interpretation of the unity are used by the models in the AMToolbox.

- SPL of 100 dB (default), used in the adaptation loops [13]. In this representation, the signals correspond to pressure in 0.5 Pa
- SPL of 93.98 dB, corresponding to the usual definition of the SPL in dB re 20 μPa. This representation corresponds to the international system of units, SI, namely, the signals are the direct representation of the pressure in Pa
- SPL of 30 dB, used in the inner-hair cell model [56]
- SPL of 0 dB, used in the DRNL filterbank [48] and in the model for binaural signal detection [6]

Note that when using linear models like the linear all-pole Gammatone filterbank, the level convention can be ignored.

1.3 Status of the Models

Ther description of a model implementation in the AMToolbox context can only be a snapshot of the development since the implementations in the toolbox are continuously developed, evaluated, and improved. In order to provide an overview of the development stage, a rating system is used in the AMToolbox. The rating status for the AMToolbox version 1.0 is provided in Table 1.1[8].

First, we rate the implementation of the model by considering its source code and documentation.

☆☆☆   *Submitted*   The model has been submitted to the AMToolbox, there is, however, no working code/documentation in the AMToolbox, or there are compilation errors, or some libraries are missing. The model neither appears on the website nor is available for download

★☆☆   *OK*   The code fits the AMToolbox conventions just enough for being available for download. The model and its documentation appear on the website, but major work is still required

★★☆   *Good*   The code/documentation follows our conventions, but there are open issues

★★★   *Perfect*   The code/documentation is fully up to our conventions, no open issues

Second, the implementation versus the corresponding publication is verified in experiments. In the best case, the experiments produce the same results as in the publication – up to some minor layout issues in the graphical representations. Verifications are rated at the following levels.

☆☆☆   *Unknown*   The AMToolbox can not run experiments for this model and can not produce results for the verification. This might be the case when the verification code has not been provided yet

---

[8] The current up-to-date status of the AMToolbox can be found under `http://amtoolbox.sourceforge.net/notes/amtnote006.pdf`, last viewed on 14.2.2013

★☆☆   *Untrusted*   The verification code is available but the experiments do not reproduce the relevant parts of the publication (yet). The current implementation can not be trusted as a basis for further developments

★★☆   *Qualified*   The experiments produce similar results as in the publication in terms of showing trends and explaining the effects, but not necessarily matching the numerical results. Explanation for the differences can be provided, for example, not all original data available, or publication affected by a known and documented bug

★★★   *Verified*   The experiments produce the same results as in the publication. Minor differences are allowed if randomness is involved in the model, for instance, noise as input signal, probabilistic modeling approaches, and a plausible explanation is provided

Table 1.1: Model status. Ch: chapter number with the model. The hyphen indicates a general model with no particular assignment to a specific chapter. D: Rating for the model documentation. C: Rating for the model source code. V: Rating for the model verification with experiments.

Model			Rating		
Name	**Function**	**Ch**	**D**	**C**	**V**
Peripheral models					
Continuous-azimuth HRTFs	enzner2008	3	★★☆	★★☆	★★☆
Directional time-of-arrival	ziegelwanger2013	-	★★★	★★★	★★★
Gammatone filterbank	gammatone	-	★★★	★★★	☆☆☆
Invertible Gammatone filterbank	hohmann2002	-	★☆☆	★☆☆	★★☆
Dual-resonance nonlin. filterbank	drnl	-	★★★	★★★	★★☆
Cochlear transmission-line model	verhulst2012	13	☆☆☆	★☆☆	☆☆☆
Auditory-nerve filterbank	zilany2007humanized	-	☆☆☆	★☆☆	☆☆☆
Inner hair cell	ihcenvelope	-	★★★	★★★	☆☆☆
Adaptation loops	adaptloop	-	★★★	★★★	☆☆☆
Modulation filterbank	modfilterbank	-	★★★	★★★	☆☆☆
Auditory brainstem responses	roenne2012	-	★★☆	★★☆	★★★
Signal detection models					
Monaural masking	dau1997preproc	-	★★★	★★★	☆☆☆
Binaural signal detection	breebaart2001preproc	5	★★★	★★★	☆☆☆
Spatial models					
Lateralization, cross-correlation	lindemann1986	12	★★★	★★★	★★☆
Concurrent-speakers lateral dir.	dietz2011	6, 10	★★★	★★★	★★★
Lateralization, supervised training	may2013	15	★☆☆	★☆☆	☆☆☆
Binaural activity map	takanen2013	13	★★★	★☆☆	☆☆☆
Median-plane localization	langendijk2002	4	★★☆	★★★	★★★
Sagittal-plane localization	baumgartner2013	4	★★★	★★★	★★★
Distance perception	georganti2013	7	★☆☆	★☆☆	☆☆☆
Speech perception models					
Speech intelligibility in noise	joergensen2011	-	★☆☆	★☆☆	★★★
Spatial unmasking for speech	jelfs2011	16	★★☆	★★★	★★☆

1.3.1 Peripheral Models

This section describes models of auditory processes involved in the periphery of the human auditory system like outer ear, middle ear, inner ear, and the auditory nerve.

Continuous-azimuth head-related transfer functions − enzner2008

Head-related transfer functions, HRTFs, describe the directional filtering of the incoming sound due to torso, head, and pinna. HRTFs are usually measured for discrete directions in a system-identification procedure aiming at fast acquisition and high spatial resolution of the HRTFs − compare [51]. The requirement of high-spatial-resolution HRTFs can also be addressed with a *continuous-azimuth* model of HRTFs [23, 24]. Based on this model, white noise is used as an excitation signal and normalized least-mean-square adaptive filters are employed to extract HRTFs from binaural recordings that are obtained during a continuous horizontal rotation of the listener. Recently, periodic perfect sweeps have been used as the excitation signal in order to increase the robustness against nonlinear distortions at the price of a potential time aliasing [2].

Within the AMToolbox, the excitation signal for the playback is generated and the binaurally recorded signal is processed. The excitation signal can be generated either by the means of Matlab/Octave internal functions [23] or with the function perfectsweep [2]. For both excitation signals, the processing of the binaural recordings [23] is implemented in enzner2008 that outputs HRTFs with arbitrary azimuthal resolution.

Directional time-of-arrival − ziegelwanger2013

The broadband delay between the incoming sound and the ear-canal entrance depends on the direction of the sound source. The delay, also called time-of-arrival, TOA, can be estimated from an HRTF. A *continuous-direction* TOA model, based on a geometric representation of the HRTF measurement setup has been proposed [78]. In the function ziegelwanger2013, TOAs, estimated from HRTFs separately for each direction, are used to fit the model parameters. Two model options are available, the on-axis model where the listener is assumed to be placed in the center of the measurement, and the off-axis model where a translation of the listener is considered. The corresponding functions, *ziegelwanger2013onaxis* and *ziegelwanger2013offaxis*, output the monaural directional delay of the incoming sound as a continuous function of the sound direction. It can be used to further analyze broadband-timing aspects of HRTFs, such as broadband interaural-time differences, ITDs, in the process of sound localization.

Gammatone filterbank − gammatone

A classical model of the human basilar membrane, BM, processing is the Gammatone filterbank, of which there exist many variations [50]. In the AMToolbox, the original IIR approximation [62] and the all-pole approximation [49] have been implemented for both real- and complex-valued filters in the function gammatone. To build a complete filterbank covering the audible frequency range, the center frequencies of the gammatone filters are typically chosen to be equidistantly spaced on an auditory frequency scale like the ERB scale [31], provided in the LTFAT.

Invertible Gammatone filterbank − hohmann2002

The classic version of the Gammatone filterbank does not provide for a method to reconstruct a signal from the output of the filters. A solution to this problem has been proposed [35] where the original signal can be reconstructed using a sampled all-pass filter and a delay line. The reconstruction is not perfect, but stays within 1 dB of error in magnitude between 1 and 7 kHz and, according to [35], the errors are barely audible. The filterbank has a total delay of 4 ms and uses 4th-order complex-valued all-pole Gammatone filters [49] – equidistantly scaled on the ERB scale.

Dual-resonance nonlinear filterbank − drnl

The DRNL filterbank introduces the modeling of the nonlinearities in peripheral processing [57, 48]. The most striking feature is a compressive input-output function, and, consequently, level-dependent tuning. The DRNL function drnl supports the parameter set for a human version of the nonlinear filterbank [48].

Auditory-nerve filterbank − zilany2007humanized

The auditory-nerve, AN, model implements the auditory periphery to predict the temporal response of AN fibers [79]. The implementation provides a "humanized" parameter set, which can be used to model the responses in human AN fibers [68]. In the AMToolbox, the function is called zilany2007humanized and outputs the temporal excitation of 500 AN fibers equally spaced on the BM.

Cochlear transmission-line model − verhulst2012

The model computes the BM velocity at a specified characteristic frequency by modeling the human cochlea as a nonlinear transmission-line and solving the corresponding ordinary differential equations in the time-domain. The model provides the user direct control over the poles of the BM admittance, and thus over the tuning and gain properties of the model along the cochlear partition. The passive structure of the model was designed [80] and a functional, rather than a micro-mechanical, approach for the nonlinearity design was followed with the purpose of realistically representing level-dependent BM impulse response behavior [67, 70]. The model simulates both forward and reverse traveling waves, which can be measured as the otoacoustic emissions, OAEs.

In the AMToolbox, the model is provided by the function verhulst2012. The model can be used to investigate time-dependent properties of cochlear mechanics and the generator mechanisms of OAEs. Furthermore, the model is a suitable pre-processor for human auditory perception models where realistic cochlear excitation patterns are required.

Inner hair cells − ihcenvelope

The functionality of the IHC is typically described as an envelope extractor. While the envelope extraction is usually modelled by a half-wave rectification followed

by a low-pass filtering, many variations to this scheme exist. For example, binaural models typically use a lower cutoff frequency for the low-pass filtering than monaural models.

In the AMToolbox, the IHC models are provided by the function `ihcenvelope` and selected by the corresponding flag. Models based on the low-pass filter with the following cutoff frequency are provided, namely, 425 Hz [5] – flag `'ihc_bernstein'`, 770 Hz [6] – flag `'ihc_breebaart'`, 800 Hz [47] – flag `'ihc_lindemann'`, and 1000 Hz [15] – flag `'ihc_dau'`. Further, the classical envelope extraction by the Hilbert transform is provided [28] – `'hilbert'`. Finally, a probabilistic approach for the synaptic mechanisms of the human inner hair cells [56] is provided – `'ihc_meddis'`.

Adaption loops – `adaptloop`

"Adaptation loops" is a simple method to model the temporal nonlinear properties of the human auditory periphery by using a chain of typically five feedback loops in series. Each loop has a different time constant. The AMToolbox implements the adaptation loops in the function `adaptloop` with the original, linearly spaced constants [66] – flag `'adt_puschel'`. In [6], the original definition was modified to include a minimum level to avoid the transition from complete silence and an overshoot limitation, `'adt_breebaart'`, because it behaved erratically if the input changes from complete silence [13]. Also, the constants from [15], `'adt_dau'`, are provided, which better approximate the forward masking data.

Modulation filterbank – `modfilterbank`

The modulation filterbank is a processing stage that accounts for amplitude modulation, AM, detection and AM masking in humans [27, 13]. In the AMToolbox, the modulation filterbank is provided in the function `modfilterbank`. The input to the modulation filterbank is low-pass 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 center frequencies of 0, 5, 10, 16.6, 27.77 ... Hz, where each next center frequency is 5/3 times the previous one. For modulation center frequencies below and including 10 Hz, the real values of the filters are returned and, for higher modulation frequencies, the absolute value, that is, the envelope, is returned.

Auditory brainstem responses – `roenne2012`

A quantitative model describing the formation of human auditory brainstem responses, ABRs, to tone pulses, clicks, and rising chirps as a function of stimulation level is provided in the function `roenne2012`. The model computes the convolution of the instantaneous discharge rates using the "humanized" nonlinear AN model [79] with an empirically derived unitary response function that is assumed to reflect contributions from different cell populations within the auditory brainstem, recorded at a given pair of electrodes on the scalp. The key stages in the model are (i) the nonlinear processing in the cochlea, including key properties such as compressive BM filtering, IHC transduction and IHC-AN synapse adaptation, and (ii) the linear transformation between the neural representation at the output of the AN and the recorded potential at the scalp.

1.4 Signal-detection Models

Signal detection models predict the ability to detect a signal or a signal change by human listeners. These models usually rely on a peripheral model and use a framework to simulate the decisions made by listeners. Note that functions with the prefix `preproc` are modeling the preprocessing part of the model only, while excluding the decision framework.

Preprocessing for modeling simultaneous and nonsimultaneous masking − dau1997preproc

A model of human auditory masking of a target stimulus by a noise stimulus has been proposed [15]. The model includes stages of linear BM filtering, IHC-transduction, adaptation loops, a modulation low-pass filter, and an optimal detector as the decision device. The model was shown to quantitatively account for a variety of psychoacoustical data associated with simultaneous and non-simultaneous masking [16]. In subsequent studies [13, 14], the cochlear processing was replaced by the GM filterbank and the modulation low-pass filter was replaced by a modulation filterbank, which enabled the model to account for AM detection and AM masking. The preprocessing part of this model consisting of the GM filterbank, the IHC stage, the adaptation loops, and the modulation filterbank is provided by the function `dau1997preproc`.

Preprocessing for modeling binaural signal detection based on contralateral inhibition − breebaart2001preproc

A model of human auditory perception in terms of the binaural signal detection has been proposed [6, 7, 8]. The model is essentially an extension of the monaural model [13, 14], from which it uses the peripheral stages, that is, linear BM filtering, IHC-transduction, and adaptation loops, and the optimal detector decision device. The peripheral internal representations for both ears are then fed to an equalization-cancellation binaural processor consisting of excitation-inhibition, EI, elements, resulting in a binaural internal representations that is finally fed into the decision device. Implemented in the function `breebaart2001preproc`, the preprocessing part of the model outputs the EI-matrix, which can be used to predict a large range of binaural detection tasks [8] or to evaluate sound localization performance for stereophonic systems [61].

1.5 Spatial Models

Spatial models consider the spatial position of a sound event in the modeling process. The model output can be the internal representation of the spatial event on a neural level. The output can also be a perceived quality of the event like the sound position, apparent source width, or the spatial distance, also in cases of multiple sources.
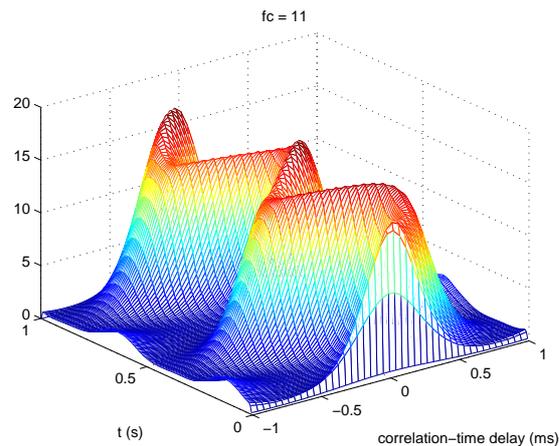
Fig. 1.4: Example for a model output in the AMToolbox. The figure shows the modeled binaural activity of the 500-Hz frequency channel in response to a 500-Hz sinusoid with a 2-Hz binaural modulation and a sampling frequency of 44.1 kHz as modeled by the cross-correlation model [47]. The figure can be plotted by evaluating the code below.

```
[cc,t] = lindemann1986(bmsin(500,2,44100),44100,'T_int',6);
plotlindemann(cc,t,'fc',500);
```

Modeling sound lateralization with cross-correlation
— `lindemann1986`

A binaural model for predicting the lateralization of a sound has been proposed [47]. This model extends the delay line principle [41] by contralateral inhibition and monaural processors. It relies on a running interaural cross-correlation process to calculate the dynamic ITD which are combined with the interaural level differences, ILDs. The peak of the cross-correlation is sharpened by contralateral inhibition and shifted by the ILD.
In the AMToolbox, the model is implemented in the function `lindemann1986` and consists of linear BM filtering, IHC-transduction, cross-correlation, and the inhibition step. The output of the model is the interaural cross-correlation in each characteristic frequency, see Fig. 1.4. The model can handle stimuli with a combination of ITD and ILD and predict split images for unnatural combinations of the two. An example is given in Fig. 1.4.

Modeling lateral-deviation estimation of concurrent speakers
— `dietz2011`

Most binaural models are based on the concept of place coding, namely, on coincidence neurons along counterdirected *delay lines* [41]. However, recent physiologic evidence from mammals, for example, [55], supports the concept of rate coding [77]

and argues against axonal delay lines. In [19], this idea is extended and the derivation of the interaural phase differences, IPDs, from both the temporal fine structure and the temporal envelope without employing mechanisms of delay compensation is proposed. This concept was further developed as a hemispheric rate comparison model in order to account for psychoacoustic data [17] and for auditory model based multi-talker lateralization [18]. The latter is the basis for further applications, such as multi-talker tracking and automatic speech recognition in multi-talker conditions [73]. The model, available in the AMToolbox in the function `dietz2011`, is functionally equivalent to both [18] and [73].

Modeling sound-source lateralization by supervised learning − `may2013`

A probabilistic model for sound-source lateralization based on the supervised learning of azimuth-dependent binaural cues in individual frequency channels is implemented in the function `may2013` [54]. The model jointly analyzes both ITDs and ILDs by approximating the two-dimensional feature distribution with a Gaussian-mixture model, GMM, classifier. In order to improve the robustness of the model, a multi-conditional training stage is employed to account for the uncertainty in ITDs and ILDs resulting from complex acoustic scenarios. The model is able to robustly estimate the position of multiple sound sources in the presence of reverberation [54]. The model can be used as a pre-processor for applications in computational auditory scene analysis such as missing data classification.

Modeling binaural activity − `takanen2013`

The decoding of the lateral direction of a sound event by the auditory system is modeled [75] and implemented in `takanen2013`. A binaural signal, processed by a peripheral model, is fed into functional count-comparison-based models of the medial and lateral superior olive [65] which decode the directional information from the binaural signal. In each frequency channel, both model outputs are combined and further processed to create the binaural activity map representing neural activity as a temporal function of lateral arrangement of the auditory scene.

Modeling median-plane localization − `langendijk2002`

Sound localization in the median planes relies on the analysis of the incoming monaural sound spectrum. The monaural directional spectral features arise due to the filtering of the incoming sound by the HRTFs. A model for the probability of listener's directional response to a sound in the median plane has been proposed [45]. The model uses a peripherally-processed set of HRTFs to mimic the representation of the localization cues in the auditory system. The decision process is simulated by minimizing the spectral distance between the peripherally-processed incoming sound spectrum and HRTFs from the set. Further, a probabilistic mapping is incorporated into the decision process.

In the AMToolbox, the model is provided by the function `langendijk2002`. The model considers the monaural spectral information only and outputs the prediction for the probability of responding at a vertical direction for stationary wideband sounds within the median-sagittal plane.

Modeling sagittal-plane-localization − baumgartner2013

The median-plane localization model [45] has been further improved with the focus to provide a good prediction of the localization *performance* for individual listeners in sagittal planes [3]. The model considers adaption to the actual bandwidth of the incoming sound and calibration to the listener-specific sensitivity. It considers a binaural signal and implements binaural weighting [34]. Thus, it allows for predicting target position in arbitrary sagittal planes, namely, parallel shifts of the median plane. The model further includes a stage to retrieve psychophysical performance parameters such as quadrant error rate, local polar RMS error, or polar bias from the probabilistic predictions, allowing to directly predict the localization performance for human listeners. For stationary, spectrally unmodulated sounds, this model incorporates the linear Gammatone filterbank [62], and has been evaluated under various conditions [3]. Optionally, the model can be extended by incorporating a variety of more physiology-related processing stages, for example, the DRNL filterbank [48] − flag `'drnl'` or the humanized AN model [79, 68] in order to model, for example, the level dependence of localization performance [53] − flag `'zilany2007humanized'`.

In the AMToolbox, the model is provided by the function `baumgartner2013`, which is the same implementation as that used in [3]. Further, a pool of listener-specific calibrations is provided, `data_baumgartner2013`, which can be used to assess the impact of arbitrary HRTF-based cues on the localization performance.

Modeling distance perception − georganti2013

A method for distance estimation in rooms based on binaural signals has been proposed [29, 30]. The method requires neither *a priori* knowledge of the room impulse response, nor the reverberation time, nor any other acoustical parameter. However, it requires training within the rooms under examination and relies on a set of features extracted from the reverberant binaural signals. The features are incorporated into a classification framework based on GMM classifier. For this method, a distance estimation feature has been introduced exploiting the standard deviation of the interaural spectral level differences in the binaural signals. This feature has been shown to be related to the statistics of the corresponding room transfer function [69, 43] and to be highly correlated with the distance between source and receiver. In the AMToolbox, the model is provided by the function `georganti2013`, which is the same implementation as that used in [30].

1.6 Speech-perception Models

Speech perception models incorporate the speech information into the modeling process and thus, they usually test the speech intelligibility under various conditions.

Modeling monaural speech intelligibility in noise − joergensen2011

A model for quantitative prediction of speech intelligibility based on the signal-to-noise envelope-power ratio, SNRenv, after modulation frequency selective processing has been proposed [44]. While the SNRenv-metric is inspired by the concept of the

signal-to-noise ratio in the modulation domain [20], the model framework is an extension of the envelope-power-spectrum model for modulation detection and masking [25], and is denoted as the speech-based envelope-power-spectrum model, sEPSM. Instead of comparing the modulation power of clean target speech with a noisy-speech mixture such as the modulation transfer function [36], the sEPSM compares an estimate of the modulation power of the clean speech within the mixture and the modulation power of the noise alone. This means that the sEPSM is sensitive to effects of nonlinear processing, such as spectral subtraction, which may increase the noise modulation power, where the classical speech models fail. In the AMToolbox, the model is provided by the function `joergensen2011`.

Modeling spatial unmasking for speech in noise and reverberation – `jelfs2011`

A model of spatial unmasking for speech in noise and reverberation has been proposed [46]. It has been validated it against human speech reception thresholds, SRTs. The underlying structure of the model has been further improved [42] and 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 [62]. 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 [1], see `siiweightings`. The binaural unmasking component calculates the binaural masking level difference within each frequency channel based on equalization-cancellation theory [21, 11]. 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. Implemented in the AMToolbox in the function `jelfs2011`, the model has been validated against a number of different sets of SRTs both from the literature and from [42]. The output of the model can be used to predict the effects of noise and reverberation on speech communication for both normal-hearing listeners and users of auditory prostheses and to predict the benefit of optimal head orientation.

1.7 Working with the AMToolbox

Assuming a working Matlab/Octave environment, the following steps are required for getting started.

1. Download the LTFAT from `http://ltfat.sourceforge.net`
2. Download the AMToolbox from `http://amtoolbox.sourceforge.net`
3. Start the LTFAT at the Matlab/Octave prompt: `ltfatstart`
4. Start the AMToolbox: `amtstart`

Further instructions on the setup for the AMToolbox can be found in the file INSTALL in the main directory of the AMToolbox. The further steps depend on the

particular tasks and application. As a general rule, a demonstration is a good start-
ing point, thus, a `demo_` function can be used to obtain a general impression of the
corresponding model, see for example Fig. 1.1. Then, an experiment, namely, an
`exp_` function can be used to see how the model output compares to the correspond-
ing publication, see for example Fig. 1.3. Editing the corresponding `exp_` function
will help to understand the particular experiment implementation, the call to the
model functions, and the parameters used in that experiment and generally available
for the model. By modifying the `exp_` function and saving as an own experiment, a
new application of the model can be easily created.

Note that some of the models require additional data not provided with the
AMToolbox because of size limitations. These data can be separately downloaded
with the corresponding link usually being provided by the particular model function.

1.8 Conclusion

AMToolbox is a continuously growing and developing collection of auditory models,
human data, and experiments. It is free as in "free beer", that is, *freeware*, and it
is free as in "free speech", in other words *liberty*[9] It is available for download[10],
and auditory researchers are welcome to contribute their models to the AMToolbox
in order to increase the pool of easily accessible and verified models and, thus, to
promote their models in the community.

Much effort has been put to the documentation. The documentation in the
software is directly linked with the documentation appearing at the web page[11],
providing a consistent documentation of the models. Finally, a rating system is
provided that which clearly shows the current stage of verification of each model
implementation. The ratings are continuously updated[12], and we hope that all the
implementations of the models from AMToolbox will reach the state of *Verified*
soon.

Acknowledgments

References

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

---

[9] `http://www.gnu.org/philosophy/free-sw.html`, last viewed on 9.1.2013
[10] from `http://amtoolbox.sourceforge.net`, last viewed on 9.1.2013
[11] see `http://amtoolbox.sourceforge.net/doc/`, last viewed on 9.1.2013
[12] see `http://amtoolbox.sourceforge.net/notes/amtnote006.pdf`, last viewed on
9.1.2013

2. C. Antweiler, A. Telle, P. Vary, and G. Enzner. Perfect-Sweep NLMS for Time-Variant Acoustic System Identification. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 517–529, Kyoto, Japan, 2012.

3. R. Baumgartner, P. Majdak, and B. Laback. Assessment of sagittal-plane sound-localization performance in spatial-audio applications. In J. Blauert, editor, *The technology of binaural listening*, chapter 4. Springer, Berlin–Heidelberg–New York NY, 2013.

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

5. L. Bernstein, S. van de Par, and C. Trahiotis. The normalized interaural correlation: Accounting for $NoS\pi$ thresholds obtained with Gaussian and "low-noise" masking noise. *J Acoust Soc Am*, 106:870–876, 1999.

6. 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, 2001.

7. 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, 2001.

8. 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, 2001.

9. J. Buckheit and D. Donoho. *Wavelab and Reproducible Research*, pages 55–81. Springer, New York NY, 1995.

10. J. Claerbout. Electronic documents give reproducible research a new meaning. *Expanded Abstracts, Soc Expl Geophys*, 92:601–604, 1992.

11. J. Culling. Evidence specifically favoring the equalization-cancellation theory of binaural unmasking. *J Acoust Soc Am*, 122:2803–2813, 2007.

12. J. Culling, S. Jelfs, and M. Lavandier. Mapping Speech Intelligibility in Noisy Rooms. In *Proceedings of the 128th convention of the Audio Engineering Society (AES)*, page Convention paper 8050, 2010.

13. 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*, 102:2892–2905, 1997.

14. T. Dau, B. Kollmeier, and A. Kohlrausch. Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration. *J Acoust Soc Am*, 102:2906–2919, 1997.

15. 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*, 99:3615–3622, 1996.

16. 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*, 99:3623–3631, 1996.

17. M. Dietz, S. D. Ewert, and V. Hohmann. Lateralization of stimuli with independent fine-structure and envelope-based temporal disparities. *J Acoust Soc Am*, 125:1622–1635, 2009.

18. M. Dietz, S. D. Ewert, and V. Hohmann. Auditory model based direction estimation of concurrent speakers from binaural signals. *Speech Comm*, 53:592–605, 2011.

19. M. Dietz, S. D. Ewert, V. Hohmann, and B. Kollmeier. Coding of temporally fluctuating interaural timing disparities in a binaural processing model based on phase differences. *Brain Res*, 1220:234–245, 2008.
20. F. Dubbelboer and T. Houtgast. The concept of signal-to-noise ratio in the modulation domain and speech intelligibility. *J Acoust Soc Am*, 124:3937–3946, 2008.
21. 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.
22. C. Elberling, J. Callø, and M. Don. Evaluating auditory brainstem responses to different chirp stimuli at three levels of stimulation. *J Acoust Soc Am*, 128:215–223, 2010.
23. G. Enzner. Analysis and optimal control of LMS-type adaptive filtering for continuous-azimuth acquisition of head related impulse responses. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 393 – 396, Las Vegas NV, 2008.
24. G. Enzner. 3D-continuous-azimuth acquisition of head-related impulse responses using multi-channel adaptive filtering. In *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WAS-PAA)*, pages 325 – 328, New Paltz NY, 2009.
25. S. Ewert and T. Dau. Characterizing frequency selectivity for envelope fluctuations. *J Acoust Soc Am*, 108:1181–1196, 2000.
26. G. Fant. Analysis and synthesis of speech processes. In B. Malmberg, editor, *Manual of phonetics*. North-Holland, Amsterdam, 1968.
27. R. Fassel and D. Püschel. *Modulation detection and masking using deterministic and random maskers*, pages 419–429. Universitätsgesellschaft, Oldenburg, 1993.
28. D. Gabor. Theory of communication. *J IEE*, 93:429–457, 1946.
29. E. Georganti, T. May, S. van de Par, and J. Mourjopoulos. Sound source distance estimation in rooms based on statistical properties of binaural signals. *IEEE Trans Audio Speech Lang Proc*, submitted.
30. E. Georganti, T. May, S. van de Par, and J. Mourjopoulos. Extracting sound-source-distance information from binaural signals. In J. Blauert, editor, *The technology of binaural listening*, chapter 7. Springer, Berlin–Heidelberg–New York NY, 2013.
31. B. R. Glasberg and B. Moore. Derivation of auditory filter shapes from notched-noise data. *Hear Res*, 47:103–138, 1990.
32. R. Goode, M. Killion, K. Nakamura, and S. Nishihara. New knowledge about the function of the human middle ear: development of an improved analog model. *Am J Otol*, 15:145–154, 1994.
33. 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*, 27:105–112, 1998.
34. M. Hofman and J. Van Opstal. Binaural weighting of pinna cues in human sound localization. *Exp Brain Res*, 148:458–70, 2003.
35. V. Hohmann. Frequency analysis and synthesis using a gammatone filterbank. *Acta Acust United Acust*, 88:433–442, 2002.

36. T. Houtgast, H. Steeneken, and R. Plomp. Predicting speech intelligibility in rooms from the modulation transfer function. i. general room acoustics. *Acustica*, 46:60–72, 1980.

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

38. 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.

39. 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.

40. 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.

41. L. Jeffress. A place theory of sound localization. *J Comp Physiol Psych*, 41:35–39, 1948.

42. S. Jelfs, J. Culling, and M. Lavandier. Revision and validation of a binaural model for speech intelligibility in noise. *Hear Res*, 2011.

43. J. Jetzt. Critical distance measurement of rooms from the sound energy spectral response. *J Acoust Soc Am*, 65:1204–1211, 1979.

44. S. Jørgensen and T. Dau. Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing. *J Acoust Soc Am*, 130:1475–1487, 2011.

45. E. Langendijk and A. Bronkhorst. Contribution of spectral cues to human sound localization. *J Acoust Soc Am*, 112:1583–1596, 2002.

46. M. Lavandier and J. Culling. Prediction of binaural speech intelligibility against noise in rooms. *J Acoust Soc Am*, 127:387–399, 2010.

47. W. Lindemann. Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals. *J Acoust Soc Am*, 80:1608–1622, 1986.

48. E. Lopez-Poveda and R. Meddis. A human nonlinear cochlear filterbank. *J Acoust Soc Am*, 110:3107–3118, 2001.

49. R. Lyon. All pole models of auditory filtering. In E. Lewis, G. Long, R. Lyon, P. Narins, C. Steele, and E. Hecht-Poinar, editors, *Diversity in Auditory Mechanics: Proceedings of the International Symposium, University of California, Berkeley*. World Scientific Publishing, 1996.

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

51. P. Majdak, P. Balazs, and B. Laback. Multiple exponential sweep method for fast measurement of head-related transfer functions. *J Audio Eng Soc*, 55:623–637, 2007.

52. P. Majdak, B. Masiero, and J. Fels. Sound localization in individualized and non-individualized crosstalk cancellation systems. *J Acoust Soc Am*, 133:2055–68, 2013.

53. P. Majdak, T. Necciari, B. Baumgartner, and B. Laback. Modeling sound-localization performance in vertical planes: level dependence. In *Poster at the 16th International Symposium on Hearing (ISH)*, Cambridge, UK, 2012.

54. T. May, S. van de Par, and A. Kohlrausch. Binaural localization and detection of speakers in complex acoustic scenes. In J. Blauert, editor, *The technology of binaural listening*, chapter 15. Springer, Berlin–Heidelberg–New York NY, 2013.

55. D. McAlpine and B. Grothe. Sound localization and delay lines-do mammals fit the model? *Trends in Neurosciences*, 26:347–350, 2003.

56. R. Meddis, M. J. Hewitt, and T. M. Shackleton. Implementation details of a computation model of the inner hair-cell auditory-nerve synapse. *J Acoust Soc Am*, 87:1813–1816, 1990.

57. R. Meddis, L. O'Mard, and E. Lopez-Poveda. A computational algorithm for computing nonlinear auditory frequency selectivity. *J Acoust Soc Am*, 109:2852–2861, 2001.

58. B. Moore and B. Glasberg. Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *J Acoust Soc Am*, 74:750–753, 1983.

59. S. Neely, S. Norton, M. Gorga, and J. W. Latency of auditory brain-stem responses and otoacoustic emissions using tone-burst stimuli. *J Acoust Soc Am*, 83:652–656, 1988.

60. P. O'Mard. Development system for auditory modelling. Technical report, Centre for the Neural Basis of Hearing, University of Essex, UK, 2004.

61. M. Park, P. A. Nelson, and K. Kang. A model of sound localisation applied to the evaluation of systems for stereophony. *Acta Acust United Acust*, 94:825–839, 2008.

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

63. R. D. Patterson, M. H. Allerhand, and C. Giguère. Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform. *J Acoust Soc Am*, 98:1890–1894, 1995.

64. D. Pralong and S. Carlile. The role of individualized headphone calibration for the generation of high fidelity virtual auditory space. *J Acoust Soc Am*, 100:3785–3793, 1996.

65. V. Pulkki and T. Hirvonen. Functional count-comparison model for binaural decoding. *Acta Acust united Acust*, 95:883–900, 2009.

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

67. A. Recio and W. Rhode. Basilar membrane responses to broadband stimuli. *J Acoust Soc Am*, 108:2281–2298, 2000.

68. F. Rønne, J. Harte, C. Elberling, and T. Dau. Modeling auditory evoked brain-stem responses to transient stimuli. *J Acoust Soc Am*, 131:3903–3913, 2012.

69. M. Schroeder. Die statistischen Parameter der Frequenzkurven von grossen Räumen. *Acustica*, 4:594–600, 1954.

70. C. Shera. Intensity-invariance of fine time structure in basilar-membrane click responses: Implications for cochlear mechanics. *J Acoust Soc Am*, 110:332–348, 2001.

71. M. Slaney. Auditory toolbox, 1994.

72. P. L. Søndergaard, B. Torrésani, and P. Balazs. The Linear Time Frequency Analysis Toolbox. *Int J Wavelets Multi*, 10:1250032 [27 pages], 2012.

73. C. Spille, B. Meyer, M. Dietz, and V. Hohmann. Binaural scene analysis with multi-dimensional statistical filters. In J. Blauert, editor, *The technology of binaural listening*, chapter 6. Springer, Berlin–Heidelberg–New York NY, 2013.
74. S. Stevens, J. Volkmann, and E. Newman. A scale for the measurement of the psychological magnitude pitch. *J Acoust Soc Am*, 8:185–190, 1937.
75. M. Takanen, O. Santala, and V. Pulkki. Binaural assessment of parametrically coded spatial audio signals. In J. Blauert, editor, *The technology of binaural listening*, chapter 13. Springer, Berlin–Heidelberg–New York NY, 2013.
76. P. Vandewalle, J. Kovacevic, and M. Vetterli. Reproducible research in signal processing - what, why, and how. *IEEE Signal Proc Mag*, 26:37–47, 2009.
77. G. von Békésy. Zur theorie des hörens; Über das Richtungshören bei einer Zeitdefferenz oder Lautstärkenungleichheit der beiderseitigen Schalleinwirkungen. *Phys Z*, 31:824–835, 1930.
78. P. Ziegelwanger, H Majdak. Continuous-direction model of the time-of-arrival in the head-related transfer functions. *J Acoust Soc Am*, submitted.
79. M. S. A. Zilany and I. C. Bruce. Representation of the vowel /ϵ/ in normal and impaired auditory nerve fibers: Model predictions of responses in cats. *J Acoust Soc Am*, 122:402–248, 2007.
80. G. Zweig. Finding the impedance of the organ of corti. *J Acoust Soc Am*, 89:1229–1254, 1991.
81. E. Zwicker. Subdivision of the audible frequency range into critical bands (frequenzgruppen). *J Acoust Soc Am*, 33:248–248, 1961.
82. E. Zwicker and H. Fastl. *Psychoacoustics: Facts and models*. Springer Berlin, 1999.