BINAURAL PSYCHOACOUSTIC MODEL: IMPLEMENTATION AND EVALUATION

Z. Bureš

Czech Technical University in Prague, FEE, Dept. of Radioelectronics

Abstract

Objective sound quality evaluation methods and audio compression tools require appropriate models of human hearing system, that would predict the audibility or annoyance of a given signal component. In this paper, the implementation and evaluation of a psychoacoustic model is described. The proposed model comprises all the fundamental stages of human auditory pathway: outer/middle ear model, active nonlinear cochlear model, model of the inner hair cell and auditory nerve complex and also the binaural processing stage.

Introduction

Due to the development of digital sound processing, massive expansion of communication technologies etc., new phenomena that affect the overall sound quality perceived by the listener appear. In order to avoid the demanding subjective listening tests, algorithmic methods of quality estimation are sought. A basic feedback is desirable in many real-time applications – automated sound processing systems need adequate automated sound quality control. The requirement of a relevant model of human auditory pathway is apparent. A way how to approximate satisfactorily the capabilities of a real listener is to model more precisely the processes taking place in the human hearing system, including the effects of binaural information processing.

Perception of a sound stimulus is affected by the processes taking place at the individual stages of the auditory pathway: propagation of the sound wave in the ear canal, impedance adjustment in the middle ear, cochlear hydromechanics and transformation of the vibrations to the neural spike trains. Higher neural processing influences the subjective properties of the stimulus substantially. The presented model comprises all the respective parts: outer/middle ear model, active nonlinear cochlear model, model of the inner hair cell and auditory nerve complex and also the binaural processing stage.

Physiology of Human Auditory Pathway

1.1 Outer and Middle Ear

The outer ear consist of the pinna and the external ear canal terminated by the tympanic membrane [6]. Its transfer properties affect the spectral image of the stimulus; the resonances of the ear canal result in pronouced peaks in the transfer function of the outer/middle ear complex, appearing in the vicinity of 3 and 11 kHz. The adjacent middle ear system serves as an impedance transformer, allowing effective transmission of vibrations into the fluids of cochlea (see Fig. 1).

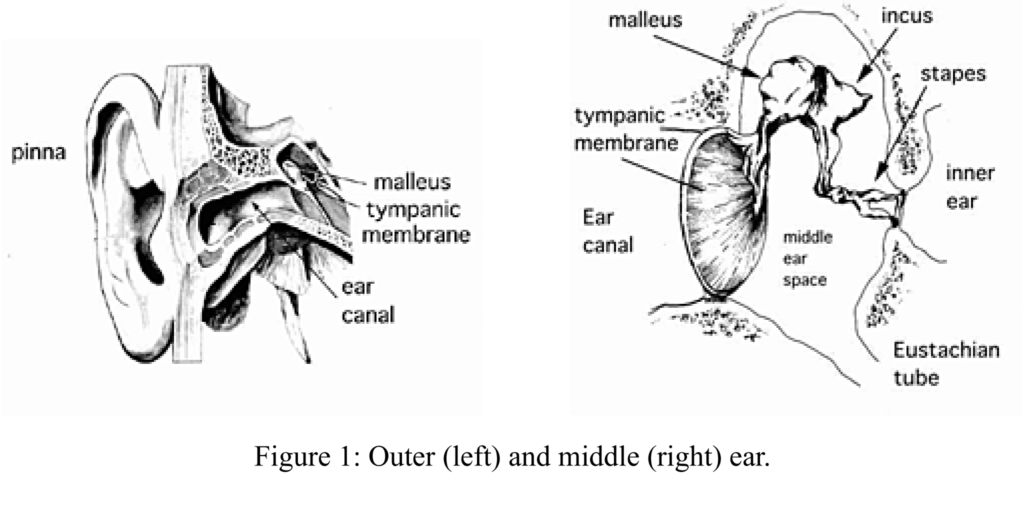

Figure 1: Outer (left) and middle (right) ear.

1.2 Cochlea

Cochlea is a fluid-filled tube divided longitudinally in three canals; *scala media* forms an elastic partition between *scala vestibuli* and *scala tympani* (Fig. 2). These two scalae are connected at the tip of the cochlea (*helicotrema*), to enable the flow of the fluids. At the cochlear base, scala vestibuli opens to the middle ear cavity with a membraneous *oval window* representing the input to the cochlea. The oval window is connected to the ear drum via the middle ear ossicles (malleus, incus & stapes, see Fig. 1). The auditory receptor itself, the *organ of Corti*, is placed inside scala media, along the *basilar membrane* (BM). It contains two sets of sensory cells: the *outer* and *inner hair cells* (see Fig. 3).

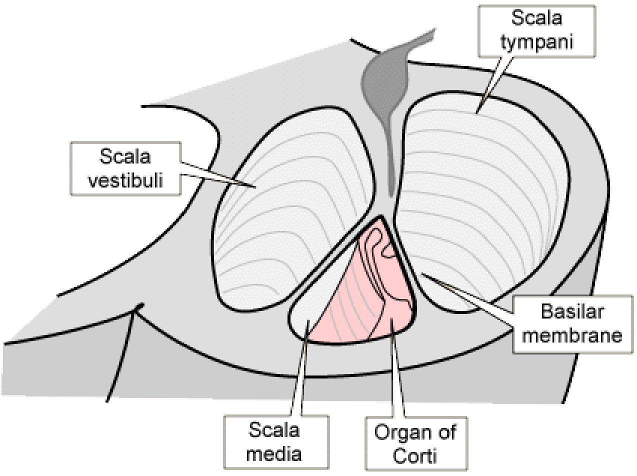

Figure 2: Cochlea, cross-section.

The vibrations of the stapes are transmitted to the cochlear liquids via the oval window. Pressure changes in the liquid result in basilar membrane motion registered by the hair cells. The outer hair cells (OHC) act as a saturating amplifier, adaptively amplifying vibrations of the BM. The inner hair cells (IHC) act as the very receptor, translating the vibrations to the series of neural impulses.

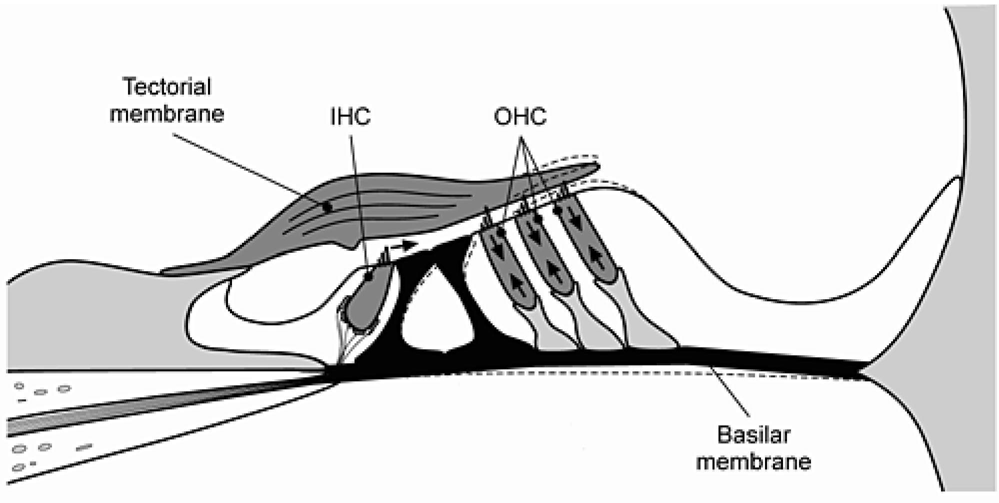

Figure 3: The organ of Corti, cross-section.

The cochlea provides a certain kind of frequency-to-place transformation: each sound frequency causes excitation of a different region of the BM. Higher frequencies are resolved at the base of the cochlea, low frequencies excite regions near the apex. Therefore, a unique *characteristic frequency* (CF) can be assigned to a given position along the BM. This principle evokes the structure of a filterbank. The positioning of the filters along the frequency axis is often expressed in terms of so called critical-band, or Bark, scale. Tonotopical structure – the bijective assignment of frequency and place – is preserved in the entire auditory system beyond cochlea.

1.3 Outer And Inner Hair Cells

The outer and inner hair cells are distributed along the basilar membrane. The outer hair cells (OHC) act predominantly as a saturating amplifier; OHC significantly broaden the range of audible sound intensity levels, lowering the hearing threshold in quiet by as much as 50dB. The effect of OHC is particularly pronounced for low level stimuli and becomes less prominent with increasing sound intensity. As a consequence of the active non-linear amplification, phenomena like cubic difference tones or otoacoustic emissions occur. Another consequence is the dependence of sharpness of the cochlear filter on the stimulus level. For low level sounds, the BM vibrations are sharply tuned, displaying a high peak at the characteristic frequency. As the stimulus level increases, the vibrations become more broadly tuned and the peak amplitude shifts towards higher CF (towards the cochlear base) [7].

The inner hair cells (IHC) serve as a source of majority of information advancing through the auditory nerve. IHC translate the motion of basilar membrane into the series of neural impulses (spikes), employing three basic features of the auditory system [6]:

- the tonotopical organisation of the auditory pathway is used for frequency coding,
- the spike rate (number of spikes over a time period) is used for intensity coding, and
- the phase lock, ie. synchronization of spikes with the sound period, is used independently for frequency coding in the low frequency band.

The response of the inner hair cells is adaptive, that means that the average spike rate elicited by a steady sound is reduced in comparison with the sound onset. An important feature of translation of the vibrations to the series of binary events is the envelope coding; the signal is effectively half way rectified (IHC fire mostly during the positive half-wave), which in combination with low-pass filtering gives an estimate of the envelope.

Humans are capable of perceiving sounds with intensities varying in the range of more than 120dB. However, such a wide range of amplitudes would require an immense sensitivity of the receptors. For this reason, our hearing system exhibits compressive behavior: the amplitudes of BM vibration fall into the range of only 60dB [7]. Moreover, weak stimuli presented simultaneously with a loud sound of similar frequency are not perceived (this phenomenon is called *masking*) [8]. The time resolution of human hearing system is also limited, though the olivary complex exhibits an extraordinary precision (see Sect. 2.4).

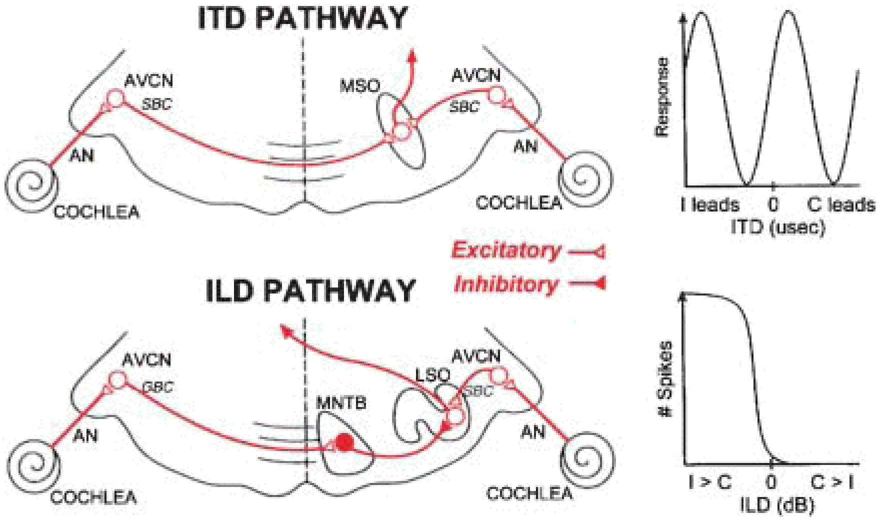

Figure 4: Structure of the olivary complex (left), LSO and MSO response (right).

1.4 Binaural hearing

One of the main purposes of binaural hearing is to allow the subject to localize the source of the incoming sound. Employing the differences in the times of arrival (*interaural time difference* – ITD) and the amplitudes (*interaural level difference* – ILD) of the stimuli at the ears, location of the sound source in the horizontal plane can be determined. The order of magnitude of just noticeable interaural delay is approximately 10µs for human subject.

The first neural nuclei (see Fig. 4) that gather information directly from both auditory nerves are the *lateral superior olive* (LSO) and the *medial superior olive* (MSO) [6]. LSO is responsible for evaluation of ILD, interaural level difference, which is the prevailing mechanism of sound localization in the high frequency band (above ca. 500 Hz). MSO is responsible for sound localization in the low frequency band (below ca. 2 kHz), making use of interaural time difference, ITD. The detected binaural parameters, ITD and ILD, are coded by means of variable spike rate. Additionally, the interaural delay of the signal envelopes is evaluated in the high frequency band. The tonotopical structure is preserved, both the ILD and ITD are thus evaluated in parallel over their relevant frequency range.

The above mentioned branching is in agreement with physical reality. In the high frequency band, evaluation of the phase shift between the two channels leads to ambiguous results. At the same time, however, acoustic shadow caused by the head and torso of the subject results in the distinguishable intensity difference [2]. The implications for the neural spike trains are obvious: in the low frequency band, the translation in the inner hair cells has to preserve the time shift whereas for the higher frequencies, the intensity has to be encoded properly.

Model Structure

The proposed model consist of several processing stages, reflecting the above described structure of human auditory pathway (see Fig. 5) [4].

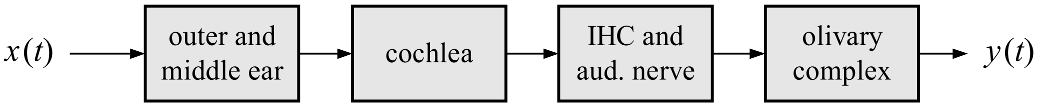

Figure 5: Block diagram of the binaral psychoacoustic model.

1.5 Outer and Middle Ear

For our purposes, the outer and middle ear complex is considered linear, although the transfer properties of middle ear may be affected by the activity of middle ear muscles in case of high level sounds. Under this assumption, it is possible to model the outer and middle ear complex utilizing only a digital filter with appropriate frequency response (see Fig. 6).

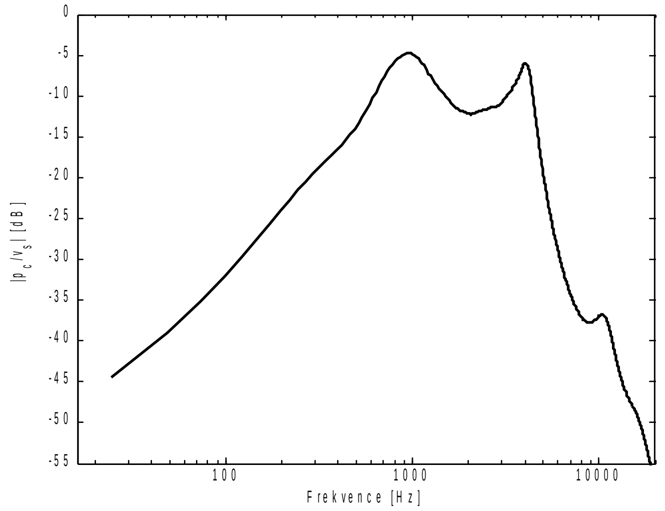

Figure 6: Transfer function of the outer and middle ear complex.

1.6 Cochlea And Outer Hair Cells

The cochlear model [1, 5] is based on electro-mechanical analogy. In the model, the entire length of cochlea is divided into electrically coupled sections with the resolution of ten sections per critical band (Bark). The hydromechanics and motion of the BM are simulated by a cascade of resonators, tuned to the particular characteristic frequencies. Such an arrangement corresponds to the frequency-to-place transformation (see Fig. 7).

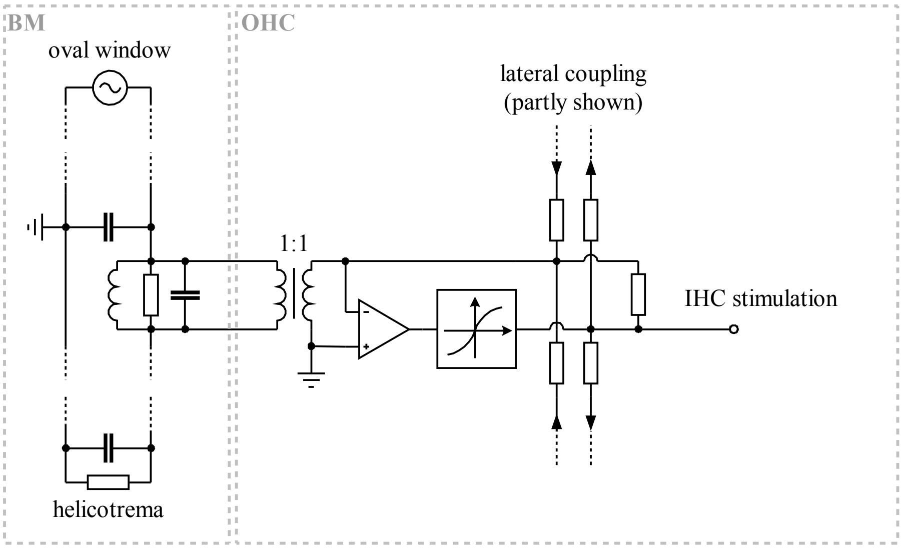

Figure 7: One section of the cochlear model. Basilar membrane and OHC circuits outlined.

The active amplification of the OHC is taken into consideration by adding the lateral electrical networks connected to individual resonators. These circuits comprise an ideal amplifier plus a nonlinear element, providing the desired amplification of low level signals as well as the observed nonlinearity (see Fig. 7). The neighbouring hair cells are not totally electrically isolated and influence each other to a certain extent, this fact is reflected in the model by means of lateral coupling of the neighbouring OHC sections.

1.7 Inner Hair Cells

The IHC model is designed to capture the most important properties of the transformation of the BM vibrations to the neural spike trains [4] (see also Section 2.3). The IHC model structure is shown in Fig. 8.

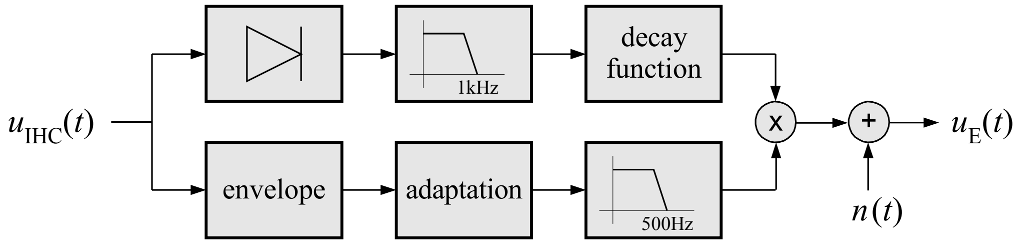

Figure 8: Structure of the IHC model.

1.8 Binaural Processing

Two alternative models of binaural hearing were implemented and tested:

1) a model designed by Breebaart [3], which introduces a two-dimensional matrix of excitation-inhibition elements; the binaural parameters (ITD and ILD) correspond to the place of minimum activity in the matrix, and

2) our own analytic model [4]. In agreement with human physiology, evaluation of the binaural parameters is divided in two independent parts in the algorithm. The ITD assessment is based on cross correlation of the signal, the ILD detection relies upon the energy estimation in the two channels.

Implementation of the Model

All the respective processing stages of the auditory model are implemented using standard MATLAB code. The input files use 100 kHz sample rate and 32-bit resolution. Each processing stage output is written to a file to allow multiple usage of pre-processed signals. The output of the cochlear model can be downsampled by a specified factor to reduce the data rate.

1.9 Outer and Middle Ear

The transfer properties of the outer/middle ear complex are modelled by a linear 512-tap FIR filter. The filter coefficients were obtained by fitting the response to the desired amplitude and phase characteristics using `invfreqz` function [4]. The input is read from a single-channel 100 kHz/32-bit raw PCM file, it is filtered using `filter` function and the result is written to a file with the same data format.

1.10 Cochlea

Current implementation of the cochlear model consists of 251 sections, each comprising a resonator plus an OHC circuit (see also Fig. 7). The model resolution is thus 0.1 Bark. Using Wave Digital Filter (WDF) technology, the above described electrical structure is transformed to WDF domain, employing the correspondence of basic electrical and WDF elements. Such an approach allows to avoid the troublesome differential equations and to maintain reasonable computation speed. According to electrical elements values, the coefficients for calculating the forward and backward propagating waves [1, 5] are computed.

Many of the parameters of the model can be changed easily from a parameter file. These include the sample rate, downsample factor, number of model sections, resolution in critical-bands or number of laterally coupled sections. Moreover, it is possible to switch between linear/non-linear and active/passive processing modes.

The OHC non-linearity follows Equation 1:

$$u_{out} = 100\frac{u_{in}}{\sqrt{1+20|u_{in}|+10u_{in}^{2}}} \qquad (1)$$

The cochlear model output forms a (*N_Sections*, *N_Samples*) matrix, where *N_Sections* is total number of mode sections and *N_Samples* is total number of data samples. The matrix is written to a binary file with `float32` precision.

1.11 Inner Hair Cells

The output of each section of the cochlear model is lead to its own IHC model instance, the individual model sections are processed independently.

The envelope of the signal is computed using RMS assessment (`norm()/sqrt()` functions) in consecutive time windows, the window length is altered according to the characteristic frequency of currently processed model section. The result is lowpass filtered to obtain smooth response. The envelope is adapted to meet the above stated requirements, employing the first derivative assessment (`diff` function) and a special „peak-hold-and-decay“ loop.

Subsequently, the signal is half-wave rectified and lowpass filtered (see Fig. 8). The decay function lays foundations for non-simultaneous masking and is also based on a low pass filter. The adapted envelope is then applied to the modified signal. Finally, a certain amount of internal noise is added to model the spontaneous activity of the hair cells.

The output of the IHC model can be interpreted as the probability of spike occurrence at given time, the important properties of the spike trains are captured. If necessary, the IHC model output can be transformed to the sequences of spikes using the obtained probability and a random generator.

The output of the IHC model is written to a file with structure analogous to that of the cochlear model output file.

1.12 Binaural Processing

In the following paragraph, two implemented binaural models are described. The input to both of them consists of two IHC model outputs, the output comprises the estimation of ILD and ITD at each sample time. Preserving the tonotopical organisation of human auditory pathway, the ILD and ITD are assessed independently in each model section. From these binaural parameters, a spatial image of the incoming sound may be constructed; however, this function is not incorporated into the current model.

One of the two implemented binaural models was originally developed by Breebaart [3]. This model introduces a matrix of excitation-inhibition (EI) elements, see Fig. 9. Each EI element is excited by the left channel signal and inhibited by the right channel signal. Therefore, equal signals cancel each other and the EI element response is close to zero. The two binaural parameters to evaluate, ITD and ILD, are projected to the respective dimensions of the matrix. The EI matrix works as a 2D delay line: the signals from the two ears are both simultaneously delayed and amplified by an elementary step (Δτ and Δα, respectively). An EI element is placed at every cross-section of the delay lines, comparing the signals. If the response from all the elements is plotted as a 2D function of ITD and ILD, the ILD and ITD sought is defined by the place of minimum activity in the (*ITD*, *ILD*) plane, as the signals cancel each other at the place corresponding to their mutual time and amplitude shift.

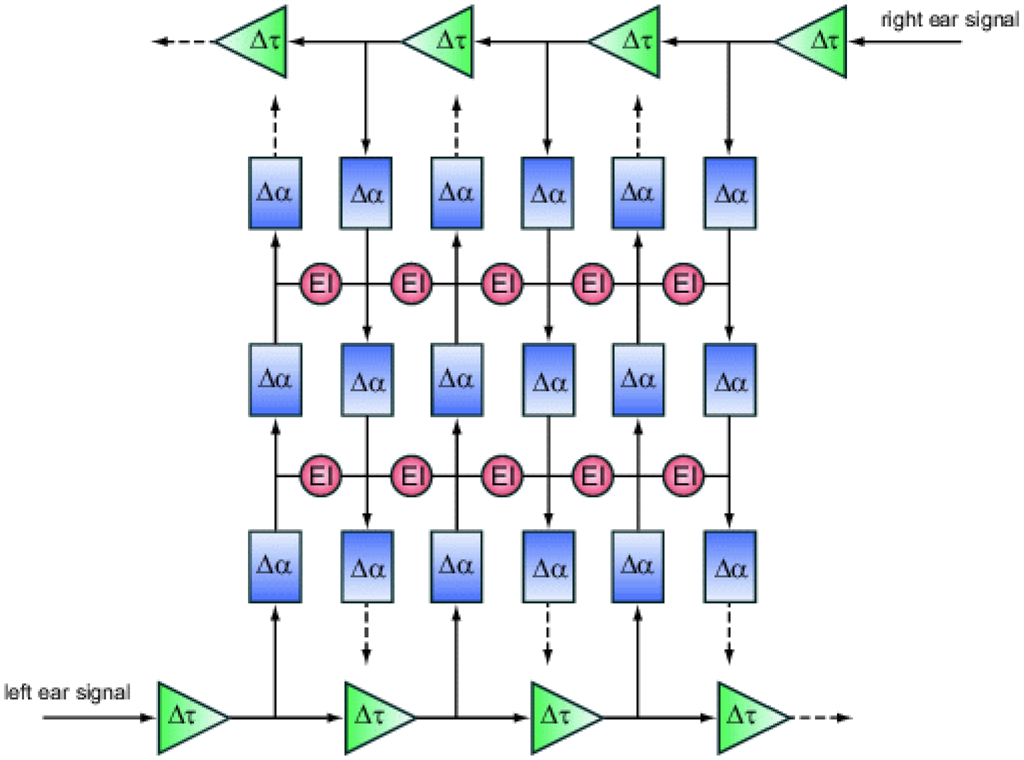

Figure 9: Principle of the binaural model by Breebaart.

The implementation of the 2D matrix is based on computing the response of each EI element according to the following code:

```
Ei(t) = ((10^alfa)*dataL(t+tau) - (10^(-alfa))*dataR(t-tau))^2;
```

where `alfa` and `tau` are the elementary delay/amplitude steps, `dataL` and `dataR` are the left and right channel signals and `t` is time. The EI response has to be computed for all relevant ITDs and ILDs with a physiologically relevant resolution. The matrix is computed for every sample in each model section. Subsequently, each EI output is integrated over a short time period. To assess the ITD and ILD, minimum response in the (*ITD*, *ILD*) plane in each channel has to be found.

Obviously, this procedure is computationally burdensome: for implementation of the model with 100 kHz sample rate, 251 sections and at least 2500 EI elements in each matrix, 62,750,000 EI calculations have to be performed in 1 ms period! The evaluation of the model was therefore limited to only one model section, also the sample rate was reduced to 20 kHz.

The second binaural model is purely analytic [4], it is based on cross-correlation and energy estimation. The evaluation of the binaural parameters is performed in two independent loops in the code. The calculation of the ITD is meaningful for frequencies below ca. 2 kHz, whereas the ILD can be used in frequency range above ca. 500 Hz, therefore the frequency ranges of the loops overlap.

The ITD assessment employs cross correlation of the left ear and right ear signals, performed in partially overlapping time windows. The ILD detection relies upon the energy estimation in the two channels; the RMS of the signals is computed in partially overlapping time windows and the results are subtracted. The output of both ITD and ILD estimation is lowpass filtered to reduce the effect of windowing.

Evaluation of the Model

1.13 Signal Generator

To evaluate the model, it is necessary to relate its outputs to known psychoacoustic and physiological facts. Despite their purely analytic nature, the psychophysical measurements commonly utilize harmonic (sinusoidal) signals and their combinations. A signal generator was designed to create the measuring signals in the desired data format.

Currently implemented version of the generator allows to create mixtures of sinusoidal signals with defined frequencies, phases and amplitudes. The length of the signal, the leading and trailing silence may be specified, as well as the sample rate. The specified signal length can be automatically modified to comprise integer multiples of the sound period. The implemented signal generator can automatically create sets of measuring signals with defined parameters according to given input vectors.

In future versions, the ramping of the signal will be implemented. Additionaly, more signal types (such as bursts of noise, clicks, amplitude modulated sounds etc.) will be supported.

The output of the generator is written to a RAW binary file accepted by the model (given sample rate – usually 100 kHz, and bit depth – either 16 or 32 bits).

1.14 Visualisation Tools

To check the performance of the peripheral ear model (outer/middle ear + cochlea + IHC models), a simple tool for visualisation of the outputs was implemented. The tool plots the model outputs depending on time (in samples) and model section. It is possible to choose either a `mesh` plot or an `imagesc` plot. The tool also allows to compare two model outputs relatively, to assess for example the relative change in basilar membrane excitation.

Two different tools for basic evaluation of the binaural processing stage were implemented, one was used for the model by Breebaart, the other one for our own model. The function of Breebaart's model had to be tested by observing the output of an EI matrix. For this purpose, the consecutive EI matrices (for a given time interval) relating to one model section were plotted onscreen as a movie; just a keystroke was required to update the screen. This way it was possible to watch the effect of the short time integration and also the position of the minimum activity.

The analytical binaural model was tested by simulating the binaural level difference phenomenon (see the text below), utilizing the dependence on time of ITD and ILD in all the model sections.

1.15 Evaluation of the Peripheral Ear Model

One of the ways how to evaluate the performance of the peripheral ear model is to relate its outputs to known masking data, assuming that simultaneous masking reflects directly the excitation of the basilar membrane. For this purpose, a set of 1 kHz sinusoidal signals with varying intensity was generated and processed by the model. The signal amplitudes ranged from 0 dB to 100 dB with a 20 dB step. The excitation of the basilar membrane was estimated according to the envelope of the vibrations at the model output (see Fig. 10).

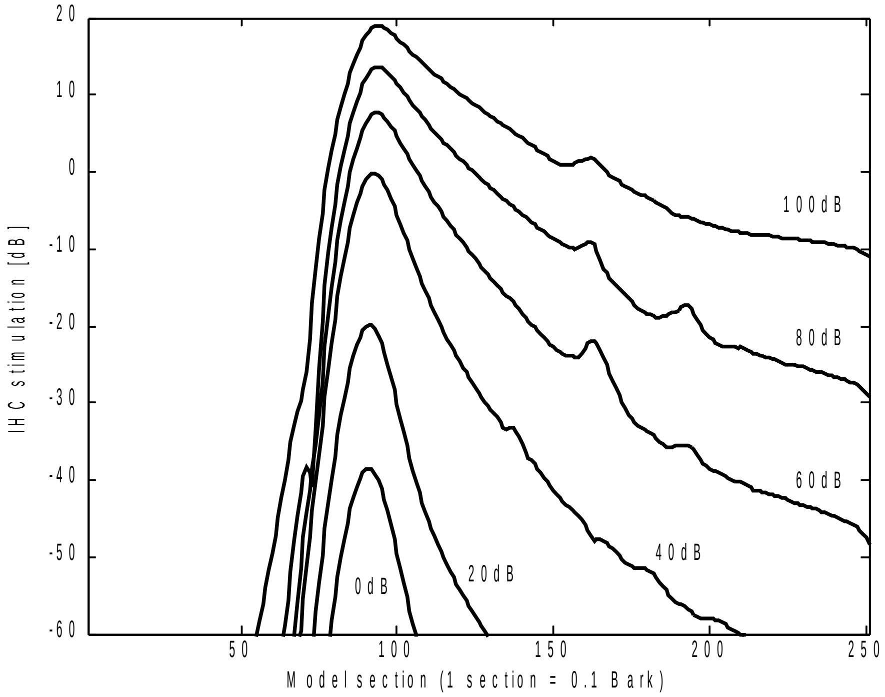

Figure 10: Excitation of BM by a 1 kHz sinusoidal stimulus of varying intensity.

The important facts to note are the level dependent response that closely resembles human masking patterns, and the minor peaks at the right-hand skirt of the high level responses – a consequence of the model non-linearity. It is obvious that the BM response grows linearly at places far from the characteristic frequency and it exhibits compressive behavior at the place of CF. This model does not account for the amplitude dependent shifts of the maximum BM response [7].

The output of the IHC model was evaluated using the above mentioned tool for visualisation, the model response for a tone burst is plotted in Fig. 11. Note the adaptive behavior of the IHC response.

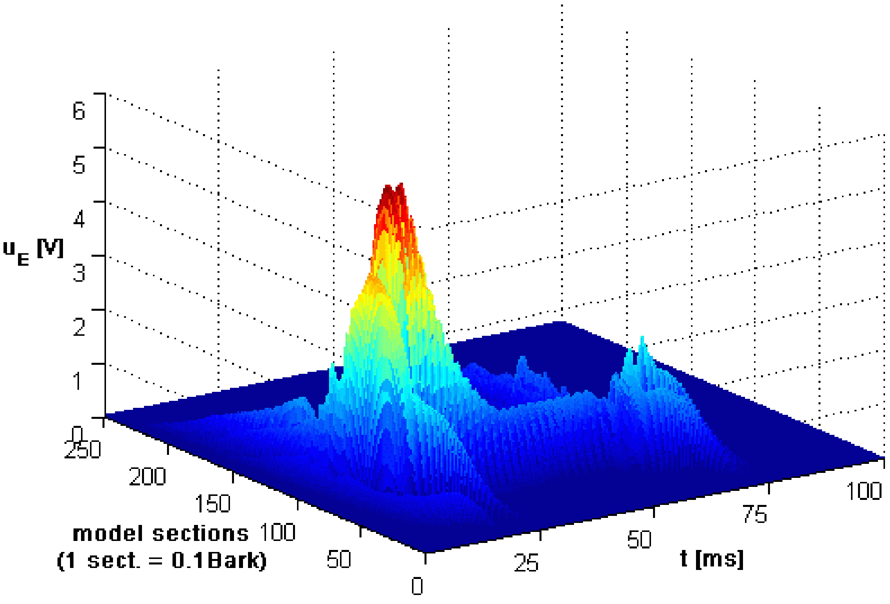

Figure 11: IHC model response for a tone burst.

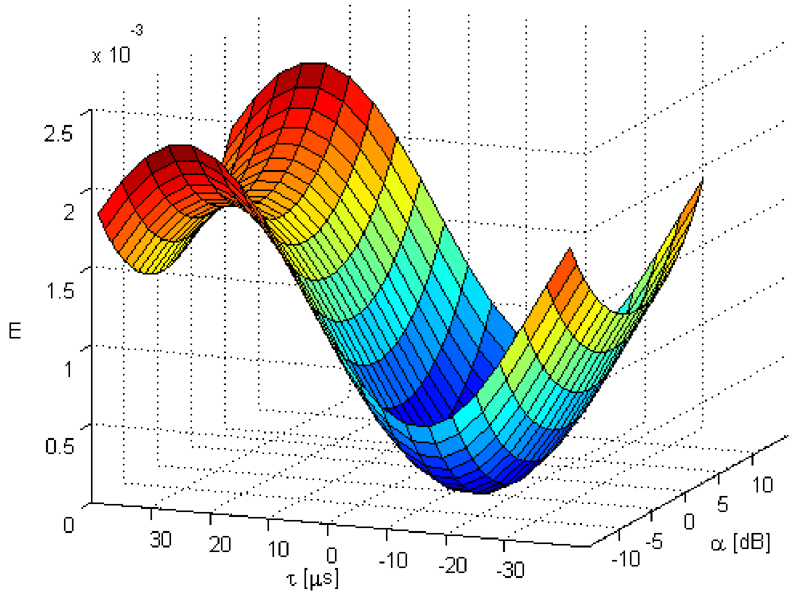

Figure 12: Output of Breebaart's binaural model. The position of minimum activity corresponds to the ILD and ITD sought.

1.16 Evaluation of the Binaural Processing Stage

The behavior of the binaural model by Breebaart was tested only for a simple harmonic stimulus with given time and amplitude shifts. For the reason of unbearable computing complexity, this model was abandoned and is currently not used as a part of our auditory model. As an example of its output, one of the EI matrices is plotted in Fig. 12.

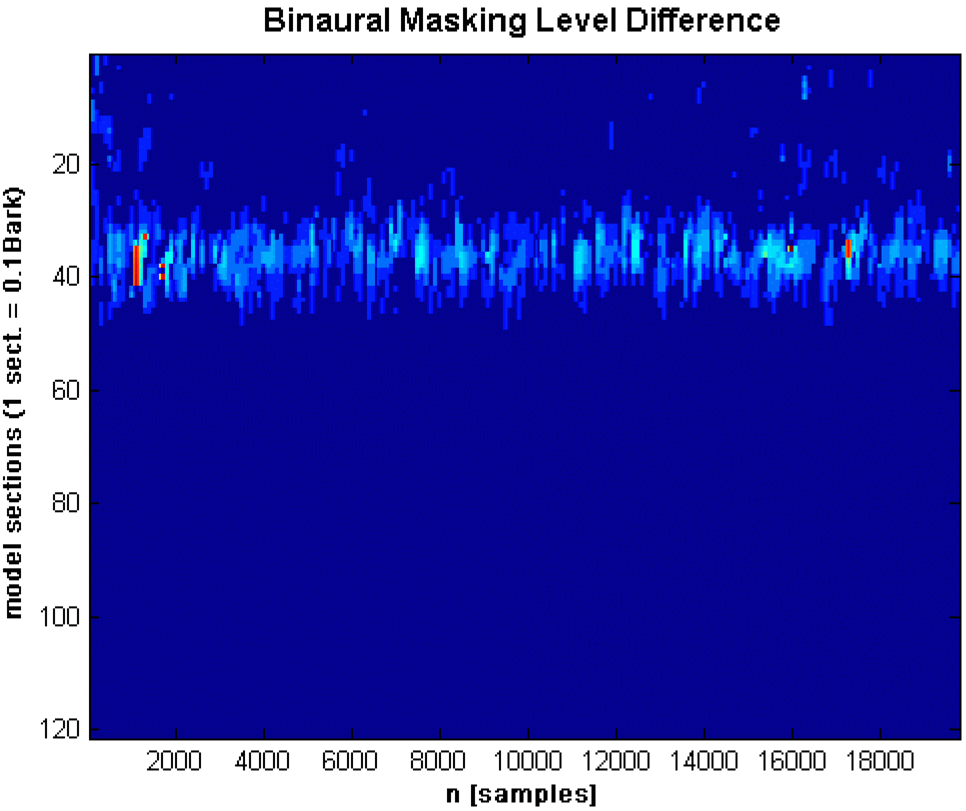

Figure 13: Effect of BMLD.

One of the phenomena that emerge as a consequence of binaural hearing is so called *binaural masking level difference* (BMLD): when the spatial location of the masker and test signal differs, themasking threshold is significantly lowered compared to the monaural perception. This phenomenon can be modelled by the proposed binaural model (see Fig. 13), a non-zero signal appears at those model sections where the binaural parameters differ. If these parameters are taken into account, the masking threshold may be shifted downwards to match the psychoacoustically observed results.

Fig. 13 shows the result of the basic BMLD test. Two signals were comprade, the first one consists of white noise in phase in both ears plus a 1 kHz tone in phase in both ears (N0S0 configuration), the second signal consists of white noise in phase in both ears and a tone burst out of phase (N0Sπ configuration). The result of comparison of these two configurations is plotted in Fig. 13. Although each of the signals, if presented monaurally, would result in the same masking threshold, binaural evaluation reveals a difference and pushes the masking thershold downwards.

Conclusion

The presented model reflects the essential phenomena that are known to occur in human auditory pathway, including the effect of the first neural nuclei for evaluation of binaural information.

The transfer properties of the external ear canal and middle ear system is modeled using a filter with appropriate response. Fundamentals of simultaneous masking, non-linear distortion and active amplification are provided by the cochlear model, which is directly followed by the IHC model. The basic effects of the IHC transformation are captured. The binaural processing stage reflects for example the binaural masking level difference (BMLD).

The details of implementation and evaluation of the model were given, including the description of signal generator and tools for visualisation. Several examples of outputs of the individual processing stages was shown.

In order to be able to use the model for evaluation of any subjective attribute of the presented sound, a subsequent cognitive stage is necessary which would transform the information at the model output into respective quantity. For example, the information about ILD and ITD from the individual model sections has to be analyzed together to assess the azimuth of the sound source correctly.

The model will be subject to further testing to verify the expected results. Psychoacoustic experiment will be organized to correlate the model output with the response of human listerens. Currently, a physiological alternative of the binaural model is being developed, employing the properties of the neural spike trains.

Acknowledgment

This work has been supported by the grant of Czech Science Foundation (GA ČR) no. 102/05/2054 ”Qualitative Aspects of Audiovisual Information Processing in Multimedia Systems”.

References

[1] F. BAUMGARTE. *Ein psychophysiologisches Gehörmodell zur Nachbildung von Wahrnehmungsschwellen für die Audiocodierung*, PhD Thesis in German, University of Hannover. 2000.

[2] J. BLAUERT. *Spatial hearing. The Psychophysics of Human Sound Localization.* Revised Edition, The MIT Press, London, 1996

[3] J. BREEBAART, S. VAN DE PAR, A. KOHLRAUSH. Binaural processing model based on contralateral inhibition. I. Model structure. *J. Acoust. Soc. Am*, vol. 110, pp. 1074 – 1089. 2001.

[4] Z. BUREŠ, F. RUND. Binaural Psychoacoustic Model Applicable To Sound Quality Estimation. In *Proceedings of 13th Intl. Congress on Sound and Vibration* [CD-ROM], Wien, Technische Universität, ISBN 3-9501554-5-7, 2006.

[5] W. PEISL. *Beschreibung aktiver nichtlinearer Effekte der peripheren Schallverarbeitung des Gehoesdurchein Rechnermodell*. PhD Thesis in German. Technical University of Munich. 1990.

[6] J.O. PICKLES. *Introduction To The Physiology Of Hearing*. Academic Press, London, 1982.

[7] L. ROBLES, M.A. RUGGERO. Mechanics of the Mammalian Cochlea. *Physiological Reviews*. vol. 81, no. 3. pp. 1305 – 1352. July 2001.

[8] E. ZWICKER, H. FASTL. *Psychoacoustics. Facts and Models*. Springer, Berlin, 1990.

Ing. Zbyněk Bureš
Dept. of Radioelectronics, FEE
Czech Technical University in Prague
Technická 2
166 27 Praha 6
tel.: 2 2435 2108
email: buresz@fel.cvut.cz