ICSV13 - Vienna
The Thirteenth International Congress on Sound and Vibration
Vienna, Austria, July 2-6, 2006

BINAURAL PSYCHOACOUSTIC MODEL APPLICABLE TO SOUND QUALITY ESTIMATION

Zbynek Bures*[1], Frantisek Rund[1]

[1]Dept. of Radioelectronics, FEE, Czech Technical University in Prague
Technicka 2, 16627 Praha 6, Czech Republic
buresz@fel.cvut.cz

Abstract

In relation to lossy audio compression algorithms, the need for assessment of perceived sound quality of the compressed signal arises. Objective sound quality evaluation methods require appropriate models of human hearing system. In the paper a binaural psychoacoustic model is presented. The model comprises a monaural and a binaural processing stage, the former being based on an active nonlinear cochlear model including the effect of IHC, the latter consisting of a two-dimensional comparison of the right-ear and left-ear signal. Approximation of the influence of outer and middle ear employs the transfer properties of acoustic waveguides.

INTRODUCTION

Sound Quality Estimation

Assessment of the perceived sound quality is a prevailing problem. 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, that would provide results sufficiently correlated with the subjective tests [4]. 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.

Models of the Hearing System

Perception of a sound stimulus (both in terms of physiological sensation and conscious evaluation) is influenced by many factors: propagation of the sound wave in the external ear canal, impedance adjustment in the middle ear, complicated vibrations taking place in the cochlea and the transformation of basilar membrane motion to the sequences of neural spikes. Higher stages of neural processing influence the subjective properties of the perceived sound substantially.

More generally, a system for estimation of any subjective attribute of a given stimulus must comprise a psychoacoustic model along with subsequent cognitive model which by means of extraction and evaluation of proper parameters provides estimate of the examined attribute.

This paper introduces a psychoacoustic model incorporating the models of peripheral ear, outer and inner hair cells and binaural processing stage.

PROPAGATION OF AUDIO INFORMATION IN THE AUDITORY PATHWAY

Peripheral Ear

The outer ear consist of the pinna and the external ear canal terminated by the tympanic membrane. The adjacent middle ear system serves as an impedance transformer, allowing effective transmission of vibrations into the fluid of the cochlea. The transfer properties of outer and middle ear, especially the resonances in the region from 3 to 4kHz, affect the sensitivity of the hearing system to different frequencies. For our purposes, the outer and middle ears are considered linear.

The fluid-filled cochlea is a crucial subsystem of the auditory pathway. It contains the basilar membrane (BM), and hair cells, responsible for the transduction of the sound pressure wave into neural signals. Vibrations of the liquid result in motion of the basilar membrane which is detected by the outer and inner hair cells. The cochlea provides a certain kind of frequency-to-place transformation – each sound frequency leads to the excitation of specific region of the basilar membrane. This dependence is not linear and is often expressed in terms of critical band (Bark) rate. Tonotopical structure – the bijective assignment of frequency and place – is preserved in the entire auditory system, starting at cochlea.

Inner and Outer 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.

The inner hair cells (IHC), on the other hand, 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, 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.

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 just noticeable interaural delay is 10μs for human subject.

The first neural nuclei that gather information from both auditory channels are lateral superior olive (LSO) and 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. MSO is responsible for sound localization in the low frequency band, making use of interaural time difference, ITD. The detected binaural parameters, ITD and ILD, are coded by means of variable spike rate.

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.

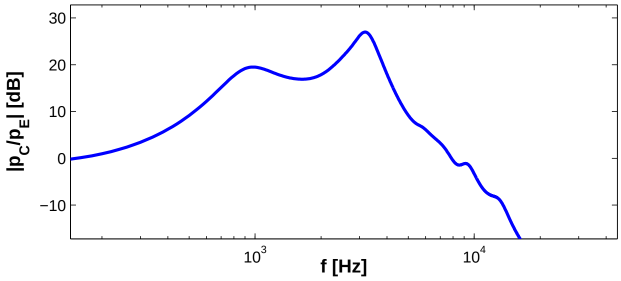

Figure 1: Transfer of free field sound pressure to pressure in cochlea. Magnitude response.

MODELS OF AUDITORY PATHWAY

Outer and Middle Ear

The analytical model describing transmission of sound from the sound source to the fluid in cochlea is based on electro-acoustic analogy. It extends the work published in [3], deriving a more precise model of the external ear canal [7]. Consecutive parts of the model are described by their cascade matrices, allowing smooth connection of the sections and effective calculations. In order to retain reasonable complexity, a constraint of only one dimension was accepted, thus the model is fully valid only for wave lengths larger than the cross-section of considered objects. Moreover, the active processes in the middle ear, i.e. the middle ear reflex, are not taken into account.

The original underlying theory, as published in [7], neglects damping in the ear canal. For purposes of this work, the damping was implemented using complex wave number

$$k^* = k - j\delta, \tag{1}$$

With respect to experimentally observed data, the damping factor δ was approximated as

$$\delta = \alpha f^2, \tag{2}$$

where $\alpha = 2.342710^{-7}$ and f is sound frequency. According to the derived analytical model, a digital filter with adequate magnitude and phase response was designed that reflects the important properties. The magnitude frequency response of the filter is depicted in Fig. 1.

Cochlea

Masking properties, frequency selectivity and sensitivity of the human hearing system are determined by the processes in cochlea, therefore a correct representation of cochlea and its subsystems is an important part of any psychoacoustic model. The basilar membrane motion caused by the sound wave propagating in the liquid is adaptively amplified by the outer hair cells which change the resonant properties of the BM according to the sound intensity [6]. Nonlinear behavior, assymetry of masking patterns and dependence of the response on the level of stimulus are difficult to represent by Fourier transform or a filterbank. The presented model is based on electro-mechanical analogy, with a simplified presumption of basilar membrane acting as a one-dimensional system of resonators [8], [9].

The entire length of cochlea is divided into electrically coupled sections. Current implementation of the model uses ten sections per critical band (Bark); that means the frequency resolution in the low frequency band is ca. 10Hz.

The hydromechanics and motion of the basilar membrane (BM) are simulated by a cascade of resonators, tuned to the particular sound frequencies. Such an arrangement corresponds to the tonotopical (frequency-to-place) transformation. The active amplification of the outer hair cells (OHC) is taken into consideration by adding the lateral electrical networks connected to individual resonators [1], [5]. These circuits provide the desired amplification of low level signals as well as the observed non-linearity (see Fig. 2).

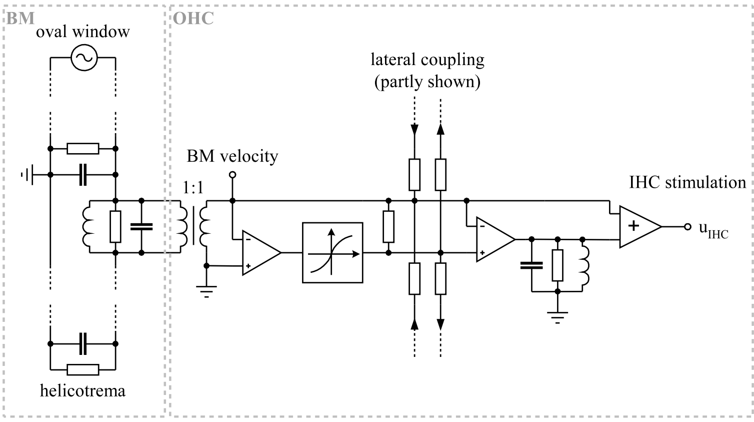

Figure 2: Cochlear model. Basilar membrane (BM) and outer hair cell (OHC) outlined.

In the continuous vibrating system of cochlea, even a sinusoidal signal results in excitation of a specific *region* of basilar membrane. Moreover, the neighbouring hair cells are not electrically isolated and influence each other to a certain extent. This fact is reflected in the model by means of lateral electrical coupling of the neighbouring OHC sections.

The software implementation of the cochlear model is based on wave digital filter (WDF) technology [5].

Inner Hair Cells

The inner hair cells translate the vibrations in cochlea into the series of neural impulses. Detailed models of individual cells are available; for our purposes, however, their straightforward application would result in unbearable computing complexity. The information coding in the auditory nerve employs huge parallelism in combination with rather stochastic character of the spike trains. For this reason, a more abstract model was sought.

The two elementary attributes of the stimulus that are to be transmitted to higher neural levels, frequency and intensity, are directly available at the output of the cochlear model. In our approach, the assumption of lossless information coding of these parameters is used, thus the troublesome translation to the spike trains is not necessary. Our model tries to capture only the fundamental properties, combining effects of several subsystems.

Besides frequency and intensity, the former being represented predominantly by the place of excitation on the basilar membrane, the latter by amplitude of basilar membrane vibrations, the IHC-coded spike sequence contains also some information about changes of the signal envelope: the individual hair cells exhibit adaptive behavior, their response during a steady tone is reduced in comparison with the sound onset. Therefore, the simultaneous masking is influenced by the fluctuations of the signal envelope. Furthermore, with respect to the consequent binaural processing, the information about the signal phase has to be preserved at low frequencies: the IHC fire mostly during the positive halfwaves of the signal while the negative halfwaves lead to inhibition of the spontaneous neural activity [6].

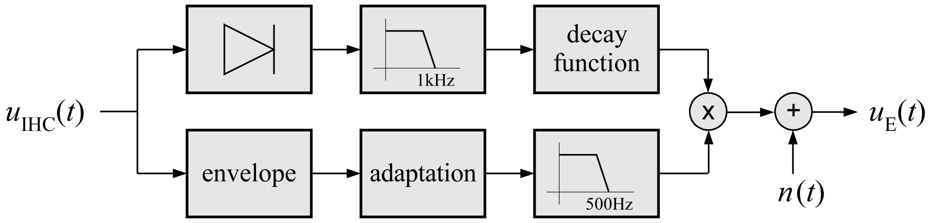

Figure 3: Principle of the IHC model.

Accordingly, the essential properties of the IHC model include:

- adaptation to a stationary signal and a more pronounced response to abrupt changes,
- higher importance of positive halfwaves, and
- well defined phase structure of a low frequency signal versus more prominent envelope in the high frequency band.

The principle of the IHC model is shown in Fig. 3. The cochlear model output is led to the IHC model input, the output of each model section is processed independently. The envelope of the signal is computed, the signal is then half-wave rectified and lowpass filtered. The decay function lays foundations for non-simultaneous masking and is also based on a low pass filter. The original envelope is adapted to meet the above stated requirements and applied to the modified signal to obtain the final result. The physiological systems participating in the information coding and transmission inevitably inject some amount of noise to the signal. This phenomenon is reflected by means of adding white noise to the output of the inner hair cell model.

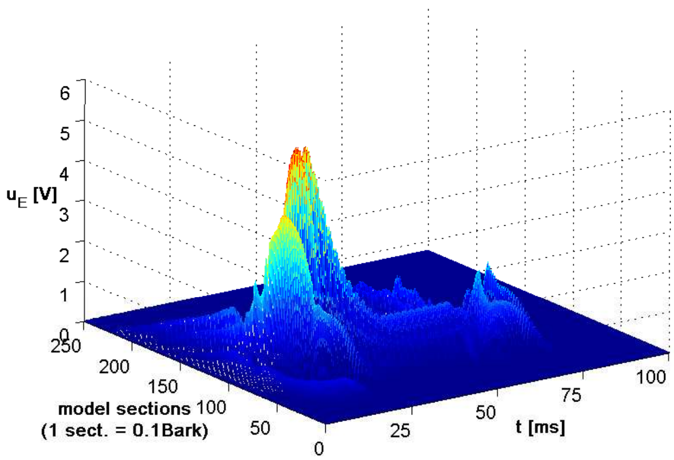

Figure 4: Output of IHC model for a tone burst.

An example of the IHC model output is depicted in Fig. 4, a 500Hz, 60dB sinusoidal tone burst was used as the stimulus. The x-axis shows time in ms, the y-axis shows model sections, the z-axis shows the IHC output without the additional noise.

Binaural Processing Stage

In humans, two specialized neural circuits at each side of the brainstem are dedicated to evaluation of basic binaural parameters – interaural time difference (ITD) and interaural level, or intensity, difference (ILD). Encoding of these parameters at the output employs variable spike rate. The tonotopical structure is preserved, the ILD and ITD are thus evaluated in parallel over the relevant frequency range.

In agreement with human physiology, evaluation of the binaural parameters is divided in two independent parts in the algorithm. For binaural hearing, the relevant frequency range is approximately 100Hz to 10000Hz. The calculation of the ITD is meaningful for frequencies below ca. 2kHz, whereas the ILD can be used in frequency range above ca. 500Hz. To reflect the limited time resolution of human hearing system, calculation of the parameters is performed in consecutive time windows. The result is low pass filtered to eliminate possible sharp peaks in the response.

The interaural time difference assessment is based on cross correlation of the signal, the interaural level difference detection relies upon the energy estimation in the two channels. With human subjects, a gap in the precision of sound localization is observed, sometimes called sensitivity notch: in the vicinity of ca. 1500Hz, the localization of a the sound source is less accurate. This corresponds to the frequency range where the two physical mechanisms overlap. The effect of sensitivity notch can be modeled by addition of noise with energy dependent on the critical band.

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, as shown in Figure 1. 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 phase information is disappearing towards high frequencies, the envelope transformation reflects the fact that signal fluctuations have strong impact on simultaneous masking and also perception of the signal onset is emphasized in comparison with the stationary part. The binaural processing stage reflects for example the binaural masking level difference (BMLD) – masking threshold decreases in case of different localization of the probe and the masker.

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 CR) no. 102/05/2054 ”Qualitative Aspects of Audiovisual Information Processing in Multimedia Systems” and research project MSM 6840770014 ”Research in the Area of Prospective Information and Communication Technologies”.

References

[1] F. Baumgarte, ”A physiological ear model for auditory masking applicable to perceptual coding”. Preprint 4511, 103rd AES Convention (1997)

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

[3] H. Hudde, A. Engel, ”Measuring and Modeling Basic Properties of the Human Middle Ear and Ear Canal”, ACUSTICA - acta acustica, **84**, 720-738, 894-913, 1091-1109 (1998)

[4] L. Husnik, M. Herrera Martinez, ”Comparison of two methods used for subjective evaluation of compressed sound signals”, Proceedings of Forum Acusticum [CD-ROM], Paper Nr. 333-0, 1787-1791. (Budapest, 29. Aug.-2.Sep. 2005)

[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] F. Rund, *Transmission of Acoustic Pressure through the Human External Ear Canal*. PhD Thesis in Czech. (CTU, Faculty of Electrical Engineering, Prague, 2004)

[8] E. Zwicker, ”A hardware cochlear nonlinear preprocessing model with active feedback”, J. Acoust. Soc. Am., **80**, 146-153 (1986)

[9] E. Zwicker, H. Fastl, *Psychoacoustics.Facts and Models*. (Springer, Berlin, 1990)