HRTF ANALYSIS BY MEANS OF AN AUDITORY MODEL

F. Rund, V. Vencovský

Department of Radioelectronics, Faculty of Electrical Engineering, Czech Technical University in Prague, Czech Republic

Abstract

Pinna, head and body causes distortion of the perceived sound and it, in turn, help to sound source localization in medial plane, where the interaural time and level differences between the signals coming to both ears are the same. Auditory model is used in this paper to analyse which spectral parts of transient sound signal are mostly affected by the pinna, head and body in dependence on the sound source position in space. The presented results show that mainly influenced bands are between 5 and 12 kHz, as was already shown in literature, and additionally to these data, some other bands located in lower frequencies (between 1 and 4 kHz). Subjective measurements need to be done in order to evaluate these results that lower frequencies are also important for the sound source localization in the medial plane.

1 Introduction

Sound signal entering the ear is distorted by the head, pinna and body. This, together with differences between sound signals entering both ears, allows humans to localize the sound source. If the sound coming from the medial plane, sound entering the both ears should be theoretically identical. Distortion of the sound by pinna, head and body is thus the only cue which can human use to localize the sound signal. It was that sound signal has to be broadband and contain some spectral components above 7 kHz to be localize in medial plane [1], [6] and [10].

Auditory model is used in this paper to analyse the sound signals coming from the different places in the medial plane. It was done in order to assess which parts in the sound signal spectrum are mainly affected by the pinna, head and body transfer function.

2 Head Related Transfer Function

Head Related Transfer Function (HRTF) captures the spectral changes of signal that occurs when a sound wave propagates from sound source to the listener's ear. The HRTF is by means of (inverse) Fourier transform connected with the HRIR (Head Related Impulse Response) The HRIR can be take out by means of measurement, or can be also modelled (e.g. series of filters simulating effect of torso, shoulders, head, pinna ...). The set of HRIR's used in the model is taken from [2]. As is known from theory of signal processing, an output signal of a linear system can be computed as convolution of an input signal and the impulse response of the system. Thus by the convolution of a signal with HRIR for corresponding direction we can get signal entering the ear.

Implementation in Matlab: Implementation of the convolution is done by using the `conv` function in Matlab. Parameters of the function are proper input signal and the appropriate HRIR taken taken from [2].

3 Peripheral Ear Model

Auditory model used in the paper is as well as the human auditory system composed of different stages. The first stage is the model of auditory canal and the last stage models sound signal processing in the neural parts of brain. Different parts of model was taken from literature and implemented in Matlab environment.

3.1 Outer and Middle Ear Model

The outer and middle ear model was taken from the ITU norm [7]. It is reverse function to the absolute threshold of hearing.

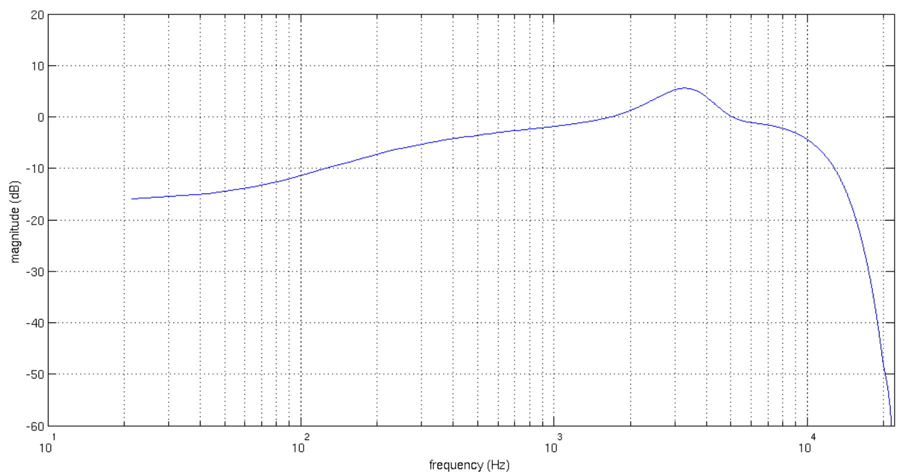

Figure 1: Response of the FIR filter designed according to the ITU norm [7].

Implementation in Matlab: Implementation of the model is done by using the `filter` function in Matlab. Parameters of the 512th-order, FIR filter where calculated by means of the `fir2` function. Sample function used for the data processing was 44.1 kHz.

3.2 Inner Ear Model

The second stage of the model is a cochlear frequency selectivity model. It was designed by [9]. Diagram of one model section (dual resonance nonlinear filter) is shown in Fig. 2. The overall model is composed of dual resonance nonlinear filterbank with 321 DRNL filters. The output of each DRNL filter represents the velocity of vibrations of the certain point along the basilar membrane. Values of the DRNL filter center frequencies were distributed between 235 Hz to 14.5 kHz. the separation of center frequencies is one equivalent rectangular bandwidth (ERB) [5].

The DRNL filter response is nonlinear for signal frequencies close to the filter's center frequency. It is in accordance with physiological observation of the basilar membrane response in mammals [9].

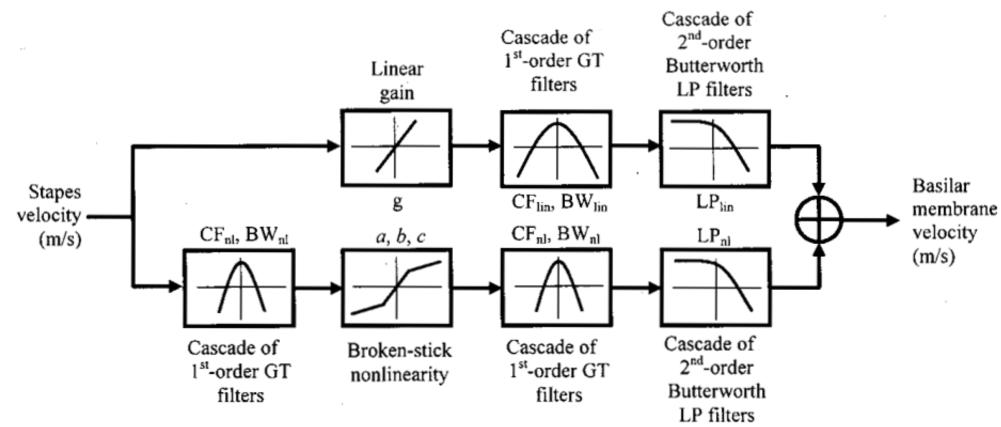

Figure 2: Diagram of the dual resonance nonlinear filter. The filter has two parts. The first one is linear and the second one nonlinear. Nonlinear stage is dominant for the signal frequencies close to the filter's center frequency and for the low level signals [9].

Implementation in Matlab: Matlab implementation of the DRNL filterbank is one function `[out cfArray] = drnl(fs,CFmin,CFmax,NrDRNL,otherPar,x)`, where `out` is output matrix. Each column of matrix contain sound samples in one channel. `cfArray` is a vector with the DRNL filter center frequencies. Input parameters to the function are sample frequency `fs`, minimal and maximal value of center frequency and number of channels in the filterbank `CFmin,CFmax` and `NrDRNL`, respectively. Other parameters of the DRNL filterbank are in the structural variable `otherPar`. Vector of the input signal to the DRNL filterbank is in the variable `x`.

The DRNL filterbank input signal is then sequentially processed by the different parts of the DRNL filter. Processing is done separately for each channel by using the `for` function. Filtration of the signal is done by means of Matlab `filter` function.

4 Central Ear Models

IHC model, function of auditory nerves and models of higher neural parts of the auditory system are included in a central ear model.

4.1 Inner Hair Cell Model

BM vibrations causes displacement of cilias on the top of the inner hair cells. The cilia displacement opens ion channels and it induces other chemical processes inside the IHC [3]. Consequence of this processes is the changing of the potential between inside of the IHC and outer space. Change of the membrane potential then causes generation of electrical action potentials which are transmitted by the auditory nerves into the brain.

The probabilistic model of IHC was designed by Dau [3]. The BM vibration signal is firstly half-wave rectified and then filtered by the first order, butterworth low-pass filter with cutoff frequency 1 kHz. This process roughly simulates transformation of the BM vibration velocity to the membrane potential inside the IHC. The signal is then compared with the threshold value in order to model the absolute threshold of hearing and then processed by five non-linear feedback loops with different time constants. The model response is nonlinear for stationary input signals and it shows enhancement for rapid signal fluctuations (faster then time constants in the feedback loops) and for the signal onset.

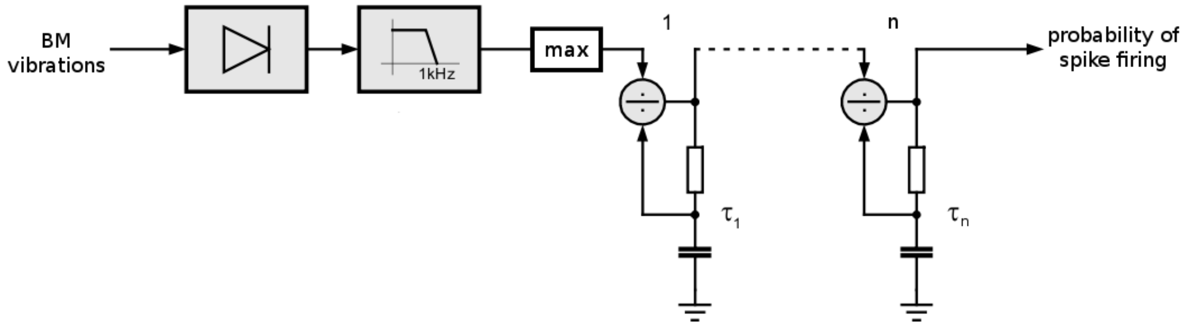

Figure 3: Diagram of the IHC model with feedback loops.

This model has a high degree of compression for stationary sound signals ($\sqrt[2^n]{x}$, where n is the number of feedback loops and x is the input signal. Five feedback loops where used according to [3] with time constants ranging from 5 to 500 ms.

Implementation in Matlab: Inner hair cell model is implemented in the function `out = IHC_Dau(fs,parIHC,in)`, where `out` is the output matrix. Timedomain signal for each channel is in one column. Sample frequency and model parameters are in the variables `fs` and `parIHC`, respectively. Input signal to the model `in` is the output matrix of the DRNL filterbank.

BM velocity signal is firstly half-wave rectified. All signal samples whose value is less than 0 are set equal to 0 (`in(in<0) = 0`). Signal is then filtered by means of the `filter` function. Filter coefficients are calculated by means of the `butter` function. Sound signal is then powered by 2 and compared with the threshold, it is implemented similarly like the aforementioned half-wave rectification.

Five stages feedback loop is implemented in the for cycle. It is direct implementation of

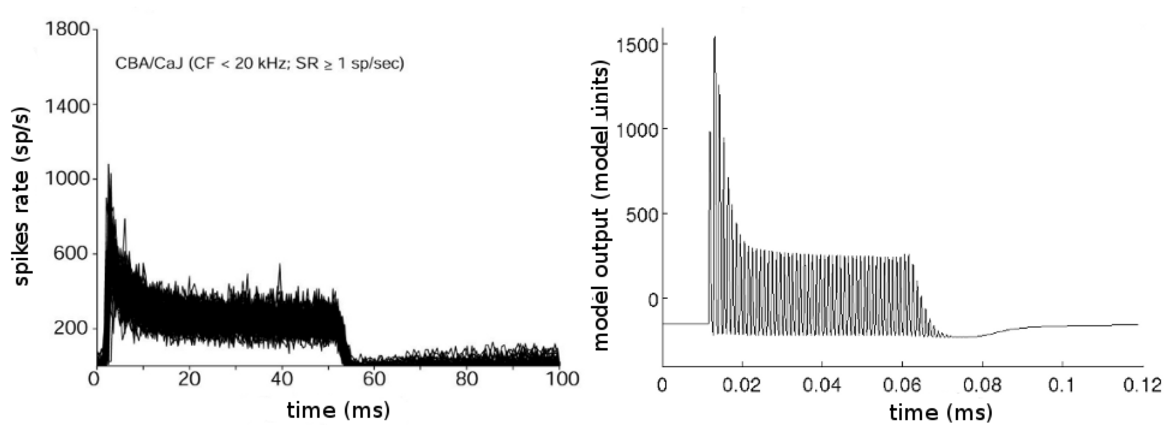

Figure 4: On the left: Histogram of the electrical action potential rate in one auditory nerve of mouse in response to pure tone stimulation [11]. On the right: Output of the IHC model in response to 1 kHz 60 dB pure tone. CF of the model is 1 kHz too.

differenc equation. It is done by means of the `for` cycle.

4.2 Modulation filterbank

Human auditory system can detect amplitude modulation (AM) [3]. The model simulating this phenomenon was taken from Jepsen et al. [8]. The input signal representing the probability of the spike generation (the output signal of the IHC model presented above) is processed by a first-order lowpass filter with cutoff frequency at 150 Hz. The filter simulates decreasing sensitivity to amplitude modulation as a function of the modulation frequency. Since it was observed that thresholds measured with narrow-band noise carriers show low-pass and also high-pass characteristics depending on the bandwidth of noise, the signal is then processed by a modulation filterbank [8].

The lowest modulation filter is lowpass filter with cutoff frequency at 2.5 Hz. Other modulation filters are bandpass filters. Bandwidth of the modulation filter tuned to 5 and 10 Hz is 5 Hz. Center frequencies of modulation filters above 10 Hz are logarithmically scaled and the filters have constant quality value of 2. The maximum center frequency of the modulation bandpass filter is set to quarter of the CF value of the each model section. The upper limit is 1 kHz and thus the maximum number of modulation filters is 12. Magnitude transfer functions of the filters are overlapped at their -3 dB points. The filters are designed as complex frequency-shifted first-order lowpass filters and thus the output is complex value. For the filters centered at and below 10 Hz, the real value of the output is considered and for the filters centered above 10 Hz the absolute value is considered and moreover, the output of the filters is attenuated by a factor of $\sqrt{2}$. Attenuation ensures that RMS values of the all channels are same in response to the sinusoidal AM input signal of the same modulation depth [8].

Implementation in Matlab: Modulation filterbank is implemented in the function `MFB`. The function declaration is `[out] = MFB(fs,cfBM,par,in)`. Sample frequency and vector of DRNL filterbank center frequencies is in the variables `fs` and `cfBM`, respectively. Parameters of the filterbank are in the structural variable `par`. Matrix of the input signal which is the output of the adaptation feedback loop is in the variable `in`. The output matrix `out` is three-dimensional.

Filtration is done by the Matlab function `filter`. Two `for` cycles are used in the `MFB` function. One of them for processing different channels of the input signal and one of them for filtration by different modulation filters.

5 Evaluation of the Model Output Signals

Sound signals coming to the auditory canal from different angles in the medial plane are independently processed by the auditory model. The cross-correlation coefficients are then calculated in 10 ms long frames of the corresponding model output signals. It is schematically shown in Fig. 5 . The equation for calculating the cross-correlation coefficients is

$$r_{fm} = \frac{\sum_{t=1}^{N}(x_{tfm} - \overline{x_{fm}})(y_{tfm} - \overline{y_{fm}})}{\sqrt{\sum_{t=1}^{N}(x_{tfm} - \overline{x_{fm}})^2 \sum_{t=1}^{N}(y_{tfm} - \overline{y_{fm}})^2}}, \tag{1}$$

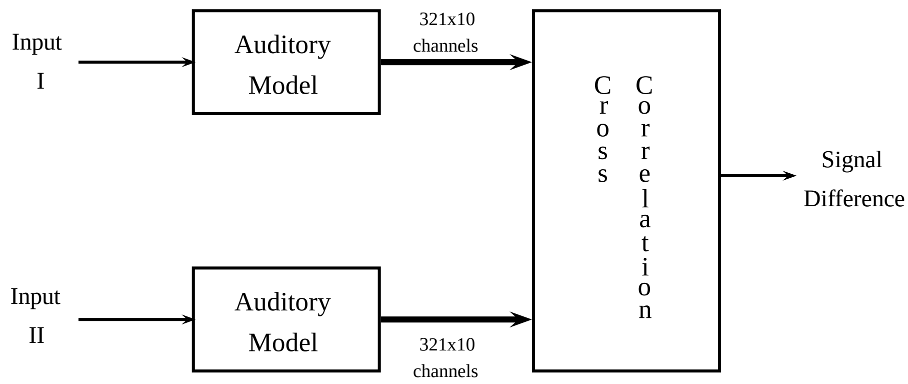

Figure 5: Diagram of the evaluation of the input sound signals coming to the ear from medial plane

where x_t and y_t are the model output signals in one frequency and modulation frequency channel and $\overline{x}$ and $\overline{y}$ are the mean values calculated over time (for 10 millisecond frames). It means that the output signal r_{fm} is a two-dimensional matrix. Mean value of the r_{fm} signal is then calculated across modulation frequencies by this equation

$$r_f = \frac{\sum_{m=1}^{M} r_{fm}}{M}, \tag{2}$$

where M is the number of modulation filters. The overall output signal can be called correlation matrix or difference pattern.

Implementation in Matlab: Implementation of the evaluation stage in the Matlab is done simply just by implementation of the aforementioned equations. It is done in the function called `kross` whose declaration is `r_m = kross(fs,x_in,y_in,t_frame,aux)`, where `r_m` is the output correlation matrix, `fs`, `x_in`, `y_in` is sample frequency, the first input matrix and the second input matrix (outputs of the auditory models), respectively. The last two parameters `t_frame` and `aux` is the duration of the frame in which is the crosscorrelation calculated and auxiliary input giving information which modulation channels are without signal. It is because not all modulation channels are used for the lowest center frequencies of the cochlear model.

One `for` cycle is used in the `kross` function. It is because signal is processed sequentially in the frames.

6 Results

6.1 Model output signal

Since the modulation filterbank is used as the last stage of the auditory model, the output signal is a three-dimensional matrix. The output signal obtained by means of the model in response to the transient input sound signal (castanets) is in Fig. 6. The output signals where summed across the modulation channels in order to be visualized.

The model outputs obtained in response to two sound signals coming from different angles in medial plane were compared. Obtained results tell us how high are the audible differences

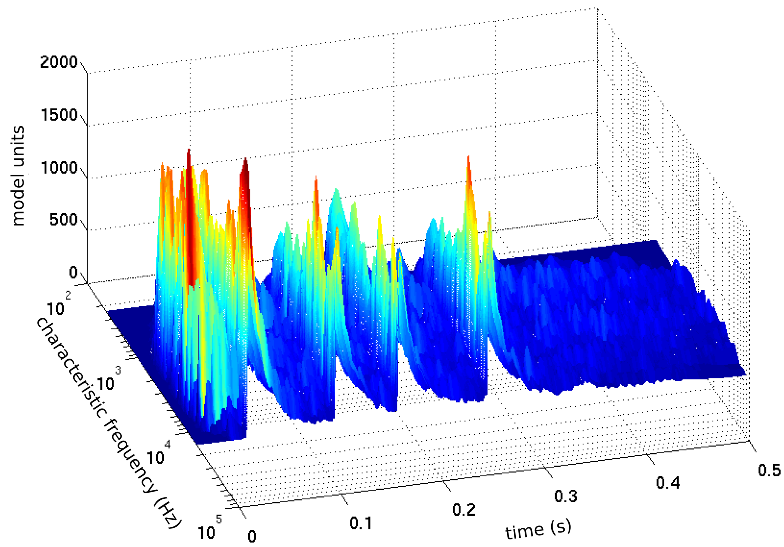

Figure 6: The auditory model output signal in response to the castanets

between two signals. It can tell us in which band is the localization of the sound signal mostly influence by pinna, head and body distortion.

6.2 Analysis of the HRTF in the medial plane

Analysis was done for the sound signal composed of castanets. The elevation was changed from -45 degrees to 85 degrees.

Responses on sound signals filtered by the HRTF function for different elevations ranging from -45 degrees to 85 degrees were compared with responses in case of not filtered sound signal (reference sound signal). Crosscorrelation matrix obtained by comparing the reference signal with the sound signal filtered by HRTF function for azimuth 0 degrees and elevation -45 degrees is shown in Fig. 7. It can be concluded from the obtained data that the pinna and head affects the signal for the lowest elevations mainly in the frequency band around 6.5 kHz and 10 kHz. The change in the obtained results for the highest elevations is mainly at the higher frequencies, between 10 kHz and 15 kHz. Some accordance with the presented HRTF data from the same subject and analysis of the minimum and maximum of the HRTF function can be seen between the obtained results and the analysis of HRTF function for the same subject presented in the paper [4].

The difference patterns obtained by means of the auditory model in response to the sound signals filtered by the HRTF function for neighbouring elevations. The analysis was done with the elevation step approximately 6 degrees.

Here are the frequencies where the largest changes in the model output data were observed.

- -45 and -39 degrees – 1 kHz, 2.5 kHz, 7 kHz and 10 kHz.
- -22 and -16 degrees – 2 kHz, 5.5 kHz, 7 kHz and 10 kHz.
- 0 and 6 degrees – 1.1 kHz 3 kHz, 5.3 kHz and 8.5 kHz

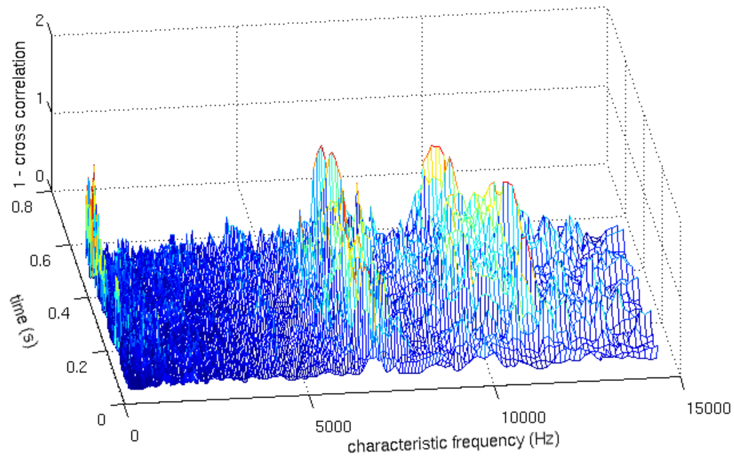

Figure 7: Crosscorrelation matrix obtained by comparison the model outputs in response to the sound signal filtered by the HRTF function for -45 degree elevation and signal which was not filtered by the HRTF.

- 18 and 22 degrees – 1 kHz, 2.3 kHz, 3.1 kHz, 9.01 kHz and around 14 kHz
- 39 and 45 degrees – 2 kHz, 3 kHz and 10 kHz
- 78 and 84 degrees – 2.7 kHz and between 10 kHz and 13 kHz

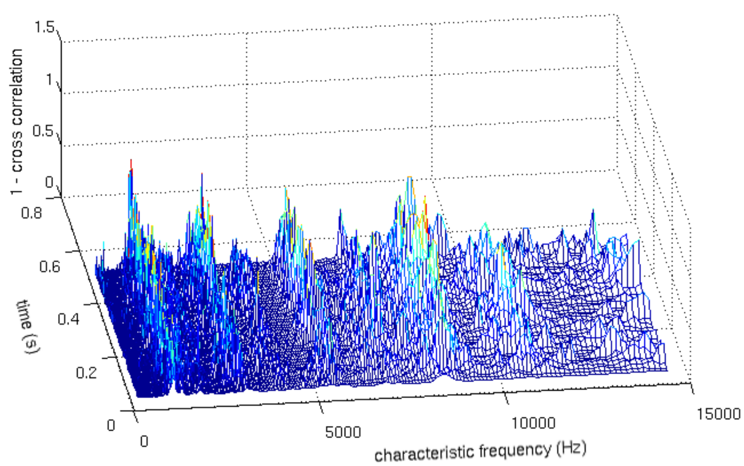

Figure 8: Crosscorrelation matrix obtained by comparison the model outputs in response to the sound signal filtered by the HRTF function for 0 degree elevation and 6 degree elevation.

The fact that HRTF function affects sound signal mainly at high frequencies is in accordance with publications [1] and [10].

7 Conclusion

Sound signals coming from the different directions in the medial plane were analysed by means of an auditory model. Transient sound signal (castanets) was used. It was filtered by HRTF filters [2] to model different directions of incoming sound and then processed by the auditory model. Crosscorrelation of the model outputs was done in order to obtain similarity patterns (crosscorrelation matrix) for neighbouring directions and also for the signal coming from different directions and reference (not filtered) sound signal. Matlab environment was used for the implementation of the auditory model, HRTF filtration of the sound signals and other needed sound signal processing in the presented simulations.

It can be concluded from the obtained data, that people are able to localize sound signals in medial plane mainly because HRTF filtration is different in the frequency region higher than 5 kHz. The same results were obtained in the literature (see [1], [6] and [10]). It means that proposed method for analysis of HRTF function gave relevant data. It should be also mentioned, that the results in comparison with the current literature, indicates that important spectral components for the sound signal localization are also in the spectral region between 1 and 4 kHz. This results should be connected with distortion of the sound signal by human body. Psychophysical measurements need to be done in order to verify these results.

Acknowledgments

This work was supported by the research program MSM 6840770014 "Research in the Area of Prospective Information and Communication Technologies" and by by the Grant Agency of the Czech Technical University in Prague, grant No. SGS10/082/OHK3/1T/13. Presented algorithms use data from CIPIC HRTF database [2].

References

[1] BLAUERT, J. *Spatial Hearing*

[2] ALGAZI, V.R., DUDA, R.O., THOMPSON, D.M., AVEDANO, C. CIPIC HRTF Database, *Proc. IEEE WASPAA-01*, New Platz, NY, 2001, http://interface.cipic.ucdavis.edu/.

[3] DAU, T. *Modeling auditory processing of amplitude modulation.* Ph.D. thesis, Fachbereich Physik der Universität Oldenburg, 1996

[4] GERANZZO, M., SPAGNOL, S., AVANZINI, F. Estimation and Modelling of Pinna-related Transfer Functions, *Proc. of the 13th Int. Conference on Digital Audio Effects (DAFx-10)*, Graz, Austria , September 6-10, 2010

[5] GLASBERG, B. R., and MOORE, B. C. J. Derivation of auditory filter shapes from notched-noise data, *Hear. Res.* 47, p. 103 – 138.

[6] IIDA, K., MOTOKUNI, I., ITAGAKI, A., MORIMOTO, M. Median plane localization using a parametric model of the head-related transfer function based on spectral cues *Applied Acoustics* 68 (2007) 835850

[7] International Telecommunication Union (ITU). *RECOMMENDATION ITU-R BS.1387: Method for Objective Measurement of Perceived Audio Quality.* 1998.

[8] JEPSEN, M. L., EWERT, S. D., DAU, T. A computational model of human auditory signal processing and perception *J. Acoust. Soc. Am.*, July 2008, vol. 124, no. 1, p.422–438.

[9] MEDDIS, R., O'MARD, L.P. A computational algorithm for computing nonlinear auditory frequency selectivity *J. Acoust. Soc. Am.*, vol. 109, no. 6, June 2001

[10] ROFFLER, S. K., BUTLER, R. A., Factors That Influencethe Localization of Sound in the Vertical Plane *J. Acoust. Soc. Am.*, 1968, vol. 43, no. 6, p.1255–1259.

[11] TABERNER, A. M., LIBERMAN, M. C. Response Properties of Single Auditory Nerve Fibers in the Mouse *J. Neurophysiol.* 2005, vol. 93, p. 557–569

F. Rund
Department of Radioelectronics, FEE, CTU in Prague, Technická 2, 166 27 Praha 6, Czech Republic, email:xrund@fel.cvut.cz

V. Vencovský
Department of Radioelectronics, FEE, CTU in Prague, Technická 2, 166 27 Praha 6, Czech Republic, email: vencovac@fel.cvut.cz