# Estimation of Direction of Arrival of Multiple Sound Sources in 3D Space Using B-Format

Hasan Khaddour, Jiří Schimmel, and Michal Trzos

***Abstract*—This contribution deals with sound source direction estimation in the three-dimensional space. An energetic analysis method based on B-format signals processing is presented in this paper. This method is able to estimate the direction of arrival for multiple sound sources in the three dimensional space. A single SoundField microphone can be used to pick-up B-format signals indirectly. The method has been simulated in Matlab and tested in a real environment. Experimental results demonstrate the validity of this method.**

***Keywords*—Sound source localization, B-format signals, energetic analysis method.**

## I. Introduction

In the last years, several sound source localization methods have been invented to localize targets. They can be mainly divided into active and passive systems. Active systems send a sound pulse and receive the echo coming back after reaching a target, and then calculate the distance between the target and the main station. This method is used in active SONAR (sound navigation and ranging) [1]. The passive systems listen to the sound coming from the targets to locate them. Such method is used in passive SONAR. The passive systems can be divided into groups depending on the physical principle they use to localize the sound sources. The most physical principles used to localize the sound sources are the time delay estimation [2] and the phase difference [3]. Physical principle of the phase difference and time delay is essentially the same but the methods differ in approach to the estimation. Two or more microphones are used to pick-up the sound coming from the sound sources and then some methods are used to calculate the time delay. The time delay can be calculated as the time which gives the maximum correlation between the sound signals that picked up by the microphones. In case where the method is used to localize several sound sources, more microphones are needed. The phase difference depends on the frequency of the sound signal and on the propagation path difference. The phase difference should be calculated in the frequency domain after using short time Fourier transform with Hanning window for instance. The corresponding outputs for each signal are then multiplied to achieve the cross spectrum. The cross spectrum is then overlapped and averaged to get the phase difference spectrum [4].

Many sound source localization methods have been proposed in the last decade. They differ in the number of sound sources they can localize and the ability of localization in the three dimensional space. The new methods try to reduce the number of used microphones. A method proposed in [5] uses three microphones to localize the sound sources in three dimensional space. However, that method needs special reflector and source counting, and it is used to localize a dominant sound source. Other methods can be used to localize multiple sound sources, whereas they use more microphones. For instance, in [6] an array of eight microphones is used for sound source localization and tracking. However, the previous method is able to estimate the distance of the sound source too.

This paper presents an approach referred to as sound source direction estimation using energetic analysis, which aims at estimating the direction of arrival for multiple sound sources in three dimensional space depending on energetic analysis of B-format signals, i.e., the direction of the sound sources. Three B-format signals are needed to estimate the direction of the sound sources in the horizontal plane only, while four B-format signals are needed to estimate the direction of the sound sources in three dimensional space.

The paper is organized as follows: B-format signals are described in Section 2. The energetic analysis method is introduced in Section 3. Section 4 presents the simulation results. Experimental results in both horizontal and vertical planes are presented in Section 5 and conclusion can be found in Section 6.

## II. B-format Signals

### A. B-format Principle

B-format signals consist of four signals namely $x(t)$, $y(t)$, $z(t)$ and $w(t)$, which carry the information about the acoustic field near to the microphone [7]. The signals $x(t)$ and $y(t)$ carry information about horizontal plane, $z(t)$ carries information about vertical plane and $w(t)$ is an omnidirectional signal, see Fig. 1.

The encoding equations for B-format signals are [7]

$$\begin{aligned} x(t) &= \cos\alpha \cos\beta \, s(t), \\ y(t) &= \sin\alpha \cos\beta \, s(t), \\ z(t) &= \sin\beta \, s(t), \end{aligned} \tag{1}$$

Manuscript received November 10, 2012. The described research was performed in laboratories supported by the SIX project; the registration number CZ.1.05/2.1.00/03.0072, the operational program Research and Development for Innovation.

H. Khaddour is with the Department of Telecommunication FEEC, Brno University of Technology, Brno, Czech Republic (phone: +420-541-149-210; fax: +420-541-149-192; e-mail: xkhadd00@stud.feec.vutbr.cz).

J. Schimmel is with the Department of Telecommunication FEEC, Brno University of Technology, Brno, Czech Republic (phone: +420-541-149-210; fax: +420-541-149-167; e-mail: schimmel@feec.vutbr.cz).

M. Trzos is with the Department of Telecommunication FEEC, Brno University of Technology, Brno, Czech Republic (phone: +420-541-149-195; fax: +420-541-149-192; e-mail: xtrzos00@stud.feec.vutbr.cz).

doi: 10.11601/ijates.v2i2.35

$$w(t) = \frac{1}{\sqrt{2}} s(t)$$

where $\alpha$ represents the azimuth angle of the source, $\beta$ represents the elevation angle of the source and $s$ represents the sound signal.

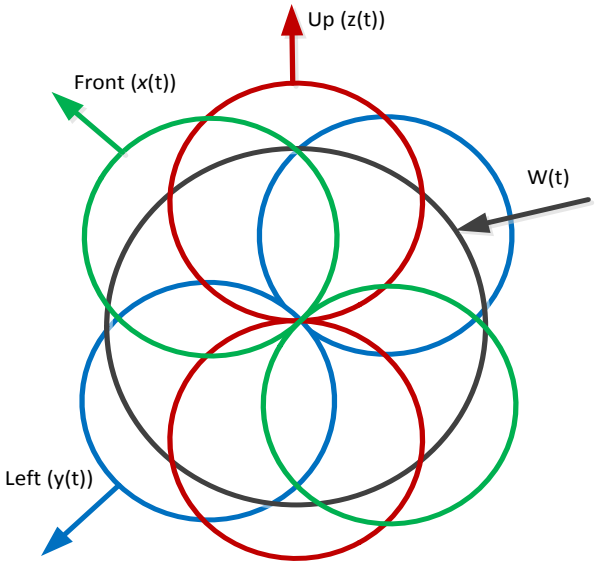


Fig.1. Polar patterns of B-format components.

In order to record B-format signals directly, a combination of coincident conventional microphones is needed, whereas three figure-of-eight microphones are used to pick-up the signals $x(t)$, $y(t)$, and $z(t)$, an omnidirectional microphone is used to pick up the $w(t)$ signal.

### B. A-format Signals

B-format signals can be derived from A-format signals. A single SoundField microphone can be used to pick-up A-format signals [8]. As can be seen in Fig.2, the microphone consists of four capsules to pick up the sound in the directions left-front, right-front, left-back and right-back.

Fig.2. SPS200 SoundField microphone used to record A-format signals.

After recording A-format signals, B-format signals can be derived by the equations [9]

$$\begin{aligned} x(t) &= 0.5((LF - LB) + (RF - RB)), \\ y(t) &= 0.5\big((LF - RB) - (RF - LB)\big), \\ z(t) &= 0.5\big((LF - LB) + (RF - RB)\big), \\ w(t) &= 0.5\big((LF + LB) + (RF + RB)\big) \end{aligned} \quad (2)$$

where $x(t)$, $y(t)$, $z(t)$ and $w(t)$ are B-format signals, and $LF, RF, LB$ and $RB$ correspond to the signals recorded by the capsules left-front, right-front, left-back and right-back respectively.

## III. Energetic Analysis Method

The principle of energetic analysis method is that the sound source direction is the opposite direction of the intensity vector of the sound. This principle is used also in directional audio coding (DirAC) [10].

In time domain, the instantaneous acoustic intensity can be written as [11]

$$\vec{I}(t) = p(t)\vec{v}(t) \quad (3)$$

where $p(t)$ is the acoustic pressure and $\vec{v}(t)$ represents the particle velocity vector.

In energetic analysis method, the sound signals are first divided in time and then in frequency using short Fourier transform method (STFT). For each time frame, the intensity vectors are computed in frequency domain. The instantaneous intensity vector can be derived from the B-format signals, it can be written as [12]

$$\boldsymbol{I}(t,f) = \left[I_x(t,f), I_y(t,f), I_z(t,f)\right]^T \quad (4)$$

where its component can be derived from the equations

$$\begin{aligned} I_x(t,f) &= \frac{1}{\sqrt{2}Z_0}\,\mathrm{Re}\{W^*(t,f).X(t,f)\}, \\ I_y(t,f) &= \frac{1}{\sqrt{2}Z_0}\,\mathrm{Re}\{W^*(t,f).Y(t,f)\}, \\ I_z(t,f) &= \frac{1}{\sqrt{2}Z_0}\,\mathrm{Re}\{W^*(t,f).Z(t,f)\} \end{aligned} \quad (5)$$

where $Z_0$ is the acoustic impedance of the air, $t$ is time, $f$ is frequency, * denotes complex conjugate, $X(t,f), Y(t,f)$, $Z(t,f)$ and $W(t,f)$ are the Fourier transform for the B-format signals $x(t)$, $y(t)$, $z(t)$ and $w(t)$ respectively.

After calculating the intensity vector for each time frame, the direction of sound can be calculated using these equations for the azimuth [11]

$$\alpha(t,f) = \begin{cases} tan^{-1}\left[\dfrac{-I_y(t,f)}{-I_x(t,f)}\right] & \text{for} \quad I_y(t,f) \geq 0, \\ tan^{-1}\left[\dfrac{-I_y(t,f)}{-I_x(t,f)}\right] - 180° & \text{for } I_y(t,f) < 0, \end{cases} \quad (6)$$

and this equation is used to estimate the elevation

$$\beta(t,f) = tan^{-1}\left[\frac{-I_z(t,f)}{\sqrt{I_x(t,f)^2+I_y(t,f)^2}}\right]. \quad (7)$$

As it can be seen from the previous equations, the azimuth and the elevation is calculated for each frequency bin in each time frame, and then the azimuth and the elevation can be determined, see Fig.3.

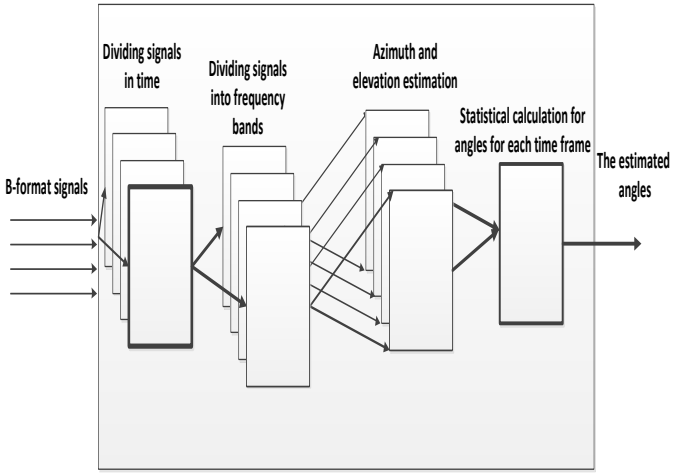


Fig.3. Energetic analysis method's diagram.

During a single time frame, each frequency bin carries information about direction of one sound source with dominant intensity in given frequency bin. We assume that only one single sound source is dominant in this case. This assumption can be hold since the sound signals differ from each other, and they have different spectral intensity in each time frame. After calculating the azimuth and elevation, a statistical process should be done to choose the most likelihood angles, from which the sound comes from as follows: assuming we have only one sound source, the estimation of direction or arrival of sound is determined as the angle that maximizes the summation of function $(\alpha(t,f))$ on the whole frequency interval for each time frame, and it could be written as

$$N(\alpha) = \sum_{k=0}^{K} p(\alpha(t,f_k) \mid \alpha),$$
$$\alpha_{estimted} = \arg max\, N(\alpha), \quad (8)$$

and the elevation as

$$N(\beta) = \sum_{k=0}^{K} p(\beta(t,f_k) \mid \beta),$$
$$\beta_{estimted} = \arg max\, N(\beta), \quad (9)$$

where $\alpha_{estimted}$ , $\beta_{estimted}$ are the estimated sound source angles (azimuth and elevation respectively), $K$ is the number of the frequency bins for $\alpha \in (-\pi,\pi)$, and $\beta \in (0,\pi)$, $\alpha(t,f_k)$ is the vector of azimuths, $t$ denotes the time frame index, $k$ is the frequency bin, and $(\alpha(t,f_k) \mid \alpha)$ is the probability that this signal comes from the direction $\alpha$ which is estimated from each frequency bin according to (6).

## IV. Simulation Results

Simulation results show the ability of this method to estimate direction of arrival of sound sources in both vertical and horizontal planes. Assuming we have three sound sources around the microphone, B-format signals can be generated from these signals according to (1). In the first simulation scenario, three sound sources were assumed to be around the microphone, with absence of noise. As can be seen in Fig. 4, the method was able to estimate the sound sources directions correctly, where the peaks denote the three estimated angles.

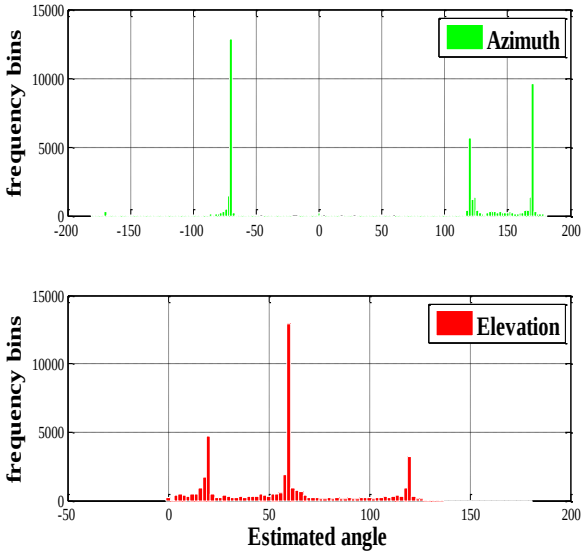


Fig.4. Simulation results in absence of noise.

In this simulation part, two different noise signals were added to each B-format signal. The first noise signal is a fan's noise signal, the spectral density distribution of this signal is shown in Fig.5. The second noise signal is pseudo-random noise with a normal distribution with mean zero and standard deviation of one which is generated by Matlab.

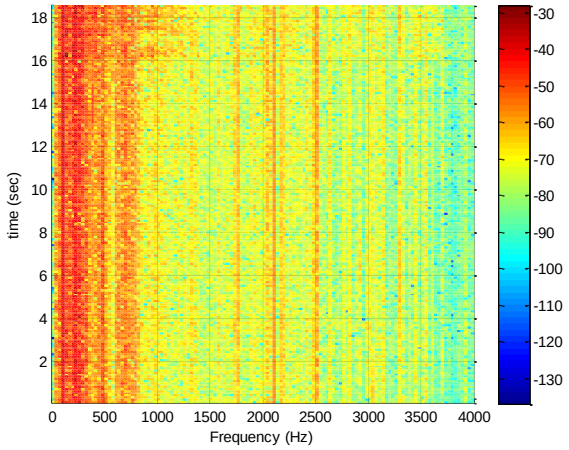


Fig.5. Spectral density distribution for a fan noise sound signal.

The two noise signals were assumed to surround the

microphones in both horizontal and vertical planes. The signals were assumed to be equidistantly separated (i.e. 4 degrees from each other in the horizontal plane and 5 degrees from each other in the vertical plane).

Simulation results are shown in Fig.6. As can be seen, the method is able to determine the direction of the sound sources in both vertical and horizontal plane, where the peaks denote the sound sources direction of arrival. The present of the noise signals affected the accuracy of the method, where some frequency bins denote to the direction of the noise signal sources. The SNR between $w(t)$ and the noise signal in our simulation is about $-26$ dB.

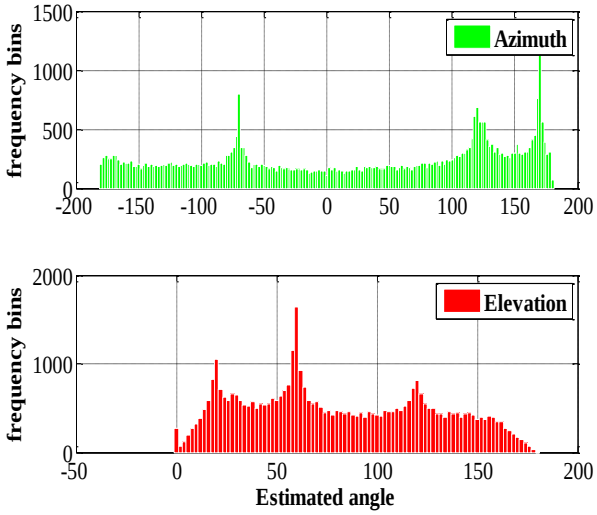


Fig.6. Simulation result with the present of pseudo-random noise signal and a fan noise signal.

## V. Experimental Results

The measurements were carried out in the acoustic laboratory at Department of Telecommunications FEEC, Brno University of Technology, where the conditions of the experiment were same as in sound control rooms, listening rooms, or in living rooms with high quality listening environment; the laboratory provides semi-diffuse field with reverberation time RT60 < 0.3 s in all octave bands.

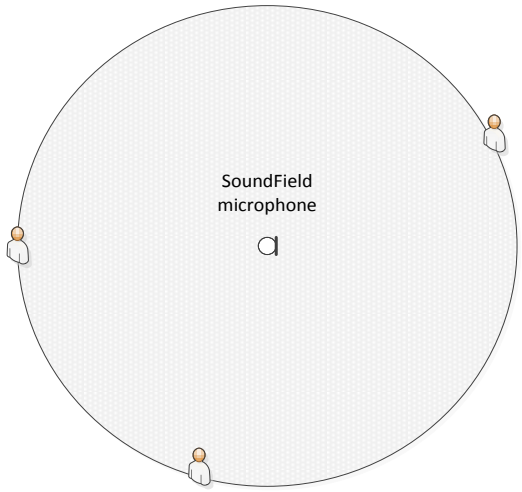


Fig.7. Recording the sound using soundfield microphone.

The measurements were carried out in both horizontal and vertical plane. The recording was made for three speakers (three men), who stood around the microphone in different arbitrary positions, see Fig.7.

Soundfield microphone was used to pick-up the sound, after recording the A-format signals, the B-format signals were derived according to (1).

In the first part of our experiment, three men were talking simultaneously in three arbitrary positions around the microphones, see Fig.7; the measurements were repeated forty times. The results for those forty measurements in the horizontal plane are shown in Fig.8. The results are shown using box plots. The boxes have lines at lower quartile, median, and upper quartile values. The whiskers show the extent of the rest of the data. The outliers are presented by red cross outside of the whiskers. As can be seen in Fig.8, the median error for the speakers was about 5 degrees for the first speaker, and 4 degrees for the second and the third speaker.

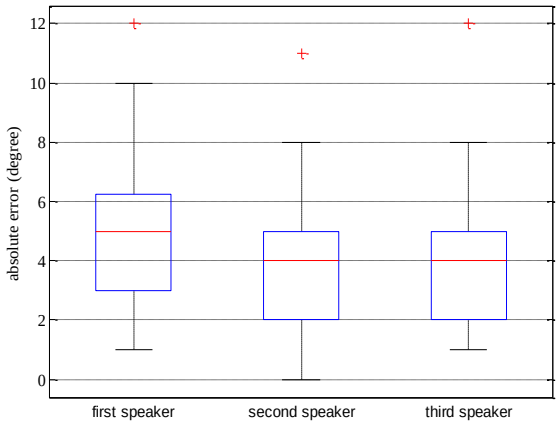


Fig.8. Average absolute angle error for the three speakers in the horizontal plane.

In the second part of the experiment, the same three men, as in the first part, were talking simultaneously in vertical plane; the measurement was repeated twenty times. The absolute angle error in the vertical plane is shown in Fig.9, it can be seen that the median error in this case was about 5 degrees for the first and second speakers and 4 degrees for the third speaker.

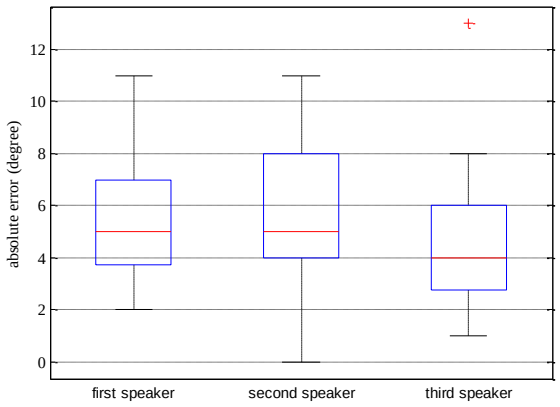


Fig.9. Average absolute angle error for the three speakers in the vertical plane.

The error that happens when this method is used comes mostly from the reverberation in the room and from the noise signals.

As can be seen in Fig.8 and Fig. 9 the method is able to estimate the direction of arrival for multiple sound sources in both horizontal and vertical plane, the median error was about 4 degrees.

Compared to our method, eight microphones are used in a method presented in [6] for three dimensional localization and tracking of sound sources, whereas our method is able to estimate the direction of the sound sources in three dimensional space using four signals. However, the absolute angle error is bigger in our method. The angle absolute error in our method is about 4 degrees whereas the angular accuracy was better than one degree for a stationary source at 1.5 meter distance in the method presented in [6]. The simulation results for the method presented in [5] showed that the method was able to localize a dominant sound source using three microphones. The angle of arrival absolute error for this method differs depending on the kind of added noise and the SNR. The simulation results for this method showed that the angle error in absence of white Gaussian noise was about 3% when SNR was about -20 dB, and the angle error was 100% in absence of pink noise and SNR less than 0 dB. However, our method is able to localize multiple sound source using only three signals in the horizontal plane and four signals in the three dimensional space with absence of mixed fan's noise and pseudorandom noise and SNR about -26 dB.

## VI. Conclusion

A method for three dimensional sound sources direction estimation was presented. This method is able to estimate the direction of multiple sound sources in both horizontal and vertical plane. Simulation results showed the affectivity of this method in both absence and presence of the noise signals. Experimental results showed that this method was able to estimate the direction of sound sources in three dimensional space.

## References


[1] de Moura, N.N.; Seixas, J.M.; Filho, W.S.; Greco, A.V.; , "Independent Component Analysis for Optimal Passive Sonar Signal Detection," *Intelligent Systems Design and Applications, 2007. ISDA 2007. Seventh International Conference on* , vol., no., pp.671-678, 20-24 Oct. 2007.

[2] Carter, G.C.; , "Coherence and time delay estimation," *Proceedings of the IEEE* , vol.75, no.2, pp. 236- 255, Feb. 1987.

[3] Schmidt, R.; , "Multiple emitter location and signal parameter estimation," *Antennas and Propagation, IEEE Transactions on* , vol.34, no.3, pp. 276- 280, Mar 1986.

[4] Shimoyama, R.; Yamazaki, K..; , "Acoustical source localization using phase difference spectrum images", Acoust. Sci. & Tech., 24 pp.161-171 February 2003.

[5] Pourmohammad, A.; Ahadi, S.M.; , "TDE-ILD-HRTF-Based 3D entire-space sound source localization using only three microphones and source counting," *Electrical Engineering and Informatics(ICEEI), 2011 International Conference on* , vol., no., pp.1-6, 17-19 July 2011.

[6] Valin, J.-M.; Michaud, F.; Rouat, J.; , "Robust 3D Localization and Tracking of Sound Sources Using Beamforming and Particle Filtering," *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on* , vol.4, no., pp.IV, 14-19 May 2006.

[7] Benjamin, E.; Lee, R.; Heller, A. ;, "Localization in horizontal-only ambisonic systems, " in *Proc*. 121st *Convention of the Audio Engineering Society*, San Francisco, pp.12, 2006.

[8] *Sound field technology, how was it work*. [Online]. [Citied 25.10.2012]. Accessible from <http://www.soundfield.com/soundfield/soundfield.php>.

[9] Rusemy, F., McCormick, T. *Sound and Recording*. Linacre House, Jordan Hill, Oxford OX2 8DP, UK. 2009.

[10] Pulkki, V. ; "Spatial Sound Reproduction with Directional audio coding" *J. Audio Eng.Soc*.,vol.55,pp.503-516,Jun 2007.

[11] Williams, E.; *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* , Academic Press, USA. 1999.

[12] E. Ahonen, J.; Pulkki, V.; Kuech, F.; Kallinger, M.; Schultz-Amling, R.; "Directional analysis of sound field with linear microphone array and applications in sound reproduction*"*. In *Proc. AES 124th Convention*, Amsterdam, The Netherlands, May 2008.



**Hasan Khaddour** received his Eng. title from the Department of Telecommunications and Electronics, Faculty of Mechanical and Electrical Engineering, Tishreen University, Syria, in 2007. Since 2009, he is a Ph.D. candidate at the Department of Telecommunications, Faculty of Electrical Engineering, Brno University of Technology (BUT), Czech Republic. His current research is focused on sound source localization, acoustical zooming, and sound rendering methods.

**Jiří Schimmel** received his M.Sc. and Ph.D. degrees in Electronics and Communications in 1999 and in Teleinformatics in 2006. He is currently an assistant professor at the Department of Telecommunications of the Faculty of Electrical Engineering and Communication of Brno University of Technology, Czech Republic. His research is focused on acoustics, multichannel digital audio signal processing, and software and hardware development for real-time audio signal processing systems. He is a member of the AES and IEEE.

**Michal Trzos** received B.Sc. degree in teleinformatics and M.Sc. degree in telecommunications and informatics from the Faculty of Electrical Engineering, Brno University of Technology (BUT), CZE, in 2007 and 2009respectively. Since 2009, he is a Ph.D. candidate at the Department of Telecommunications. His main task is to explore new methods of time and frequency warping of audio signals. His research interests include: general-purpose computing on graphics processing units, audio algorithm parallelization, time-frequency transformations, and speech processing.