ARCHIVES OF ACOUSTICS
36, 4, 851–860 (2011)

DOI: 10.2478/v10168-011-0056-2

Application of Vector Sensors to Acoustic Surveillance of a Public Interior Space

Kuba ŁOPATKA, Józef KOTUS, Andrzej CZYŻEWSKI

Gdańsk University of Technology
Faculty of Electronics, Telecommunication and Informatics
Multimedia Systems Department
Narutowicza 11/12, 80-233 Gdańsk, Poland
e-mail: {klopatka, joseph, andcz}@sound.eti.pg.gda.pl

(*received June 20, 2011; accepted August 22, 2011*)

A method for precise sound sources detection and localization in interiors is presented. Acoustic vector sensors, which provide multichannel output signals of acoustic pressure and particle velocity were employed. Methods for detecting acoustic events are introduced. The algorithm for localizing sound events in the audience is presented. The system set up in a lecture hall, which serves as a demonstrator of the proposed technology, is described. The accurracy of the proposed method is evaluated by the described measurement results. The analysis of the results is followed by conclusions pertaining the usability of the proposed system. The concept of the multimodal audio-visual detection of events in the audience is also introduced.

Keywords: sound localization, acoustic vector sensors, acoustic events.

1. Introduction

In audio surveillance system sound is often used, along image, to detect alarming situations. Sound recognition algorithms can be employed to recognize audio events which are symptoms of hazardous situations (Grigoras, 2009). Apart from recognition of audio events, techniques for sound localization can be used to determine the place where the alarming event took place. Most solutions for localization of sound sources for security applications employ microphone arrays. Such works has been widely described in the literature (Valenzise *et al.*, 2007; Julian *et al.*, 2004; Li *et al.*, 2002). The proposed technology is unique, because it employs only one sensor to determine the direction of coming sound. It is a multichannel acoustic vector sensor (AVS) – a USP probe manufactured by Microflown. The accuracy of the probe was evaluated and it was proved that the

precision of this device is sufficient to localize acoustic events (Wind *et al.*, 2010; Czyżewski, Kotus, 2010).

The application of sound localization, using an acoustic vector sensor, which is proposed in this paper, is related to the monitoring of public events (concerts, sports events etc.). The aim of this application is to detect and localize alarming sound events in the audience. Our work on this subject resulted in setting up a demonstration system located in a lecture hall in Gdańsk University of Technology. The demonstration system comprises of an acoustic vector sensor used to determine the localization of the sound source and two cameras. The fixed high resolution camera sees the entire audience, while the moveable PTZ (pan-tilt-zoom) camera can zoom in and point in the direction of a detected acoustic event, as described in previous papers on the subject (Kotus *et al.*, 2010). Apart from the demonstration of technology which can be used in monitoring of public events, the system has another purpose, which is to monitor the activity of the audience during lectures or conference sessions. In this case of use speech signals are detected and the speaker is localized. The camera can then point automatically to the speaker, who is asking the lecturer a question or taking part in the discussion. The image of the speaking person can be automatically displayed on the main screen. In this case the systems enables automatic, intelligent realization and mixing of the video image.

In the paper the method of localization the sound source in the room is presented. The algorithms of detection of acoustic events are described. The detection of events is focused on detecting impulse sounds (usage case of safety monitoring) and speech sounds (in case of monitoring of audience activity). In the following sections the formulas for calculating the position of the sound source in 3-dimensional space are introduced. The realization of controlling the PTZ camera is then briefly described. The paper is concluded with the discussion of the system accuracy based on the conducted measurements. As a conclusion the plans for future work on improving the accuracy and usability are explained.

2. Detection of sound events

In an acoustic surveillance system the sound event detection algorithm separates the foreground events from the acoustic background. Several methods for detection of environmental, natural and hazardous sounds are described in the literature (Valenzise *et al.*, 2007; Li *et al.*, 2002, Abouchacra *et al.*, 2007; Zhuang *et al.*, 2010). Concerning the potential usage of the presented technology, two types of sound events are detected: impulse sounds and speech sounds (Kotus *et al.*, 2011). Two adaptive detectors of the respective sounds are employed. The block diagram of the impulse sounds detector is presented in Fig. 1.

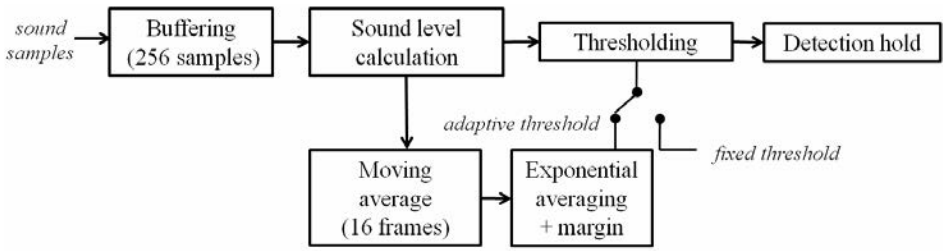

Fig. 1. Block diagram of the detector of impulse sounds.

The frame employed for detection of impulse sounds is of 256 samples in length (5.3 milliseconds at 48000 samples per second). Sound level of each frame is calculated. The decision is based on comparing the current sound level with the threshold value. The detector can work in two modes: with fixed threshold and with adaptive threshold. While operating in adaptive mode, the sound level of the background is averaged first using moving average from 16 frames, then using exponential averaging. A margin is added to the averaged sound level (typically 10–20 dB). Finally, the detection is held for an amount of time (0.5 second) to cover the whole duration of the acoustic event. The operation of the impulse sounds detector is presented in Fig. 2.

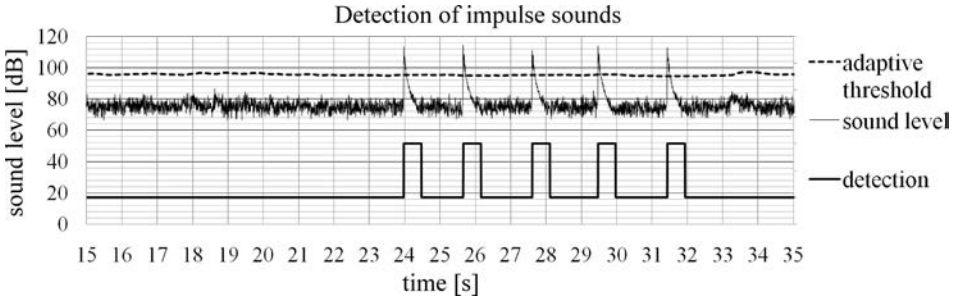

Fig. 2. Example of detection of impulse sounds.

Detection of speech sounds is performed similarly to the detection of impulse sounds. Instead of the sound level, the modified peak-valley difference (PVD) feature is employed. The details of this detection method were described in previous work (Kotus *et al.*, 2011).

3. Localization of the sound source

The employed acoustic vector sensor is sensitive to the velocity of oscillation of air particles (Czyżewski, Kotus, 2010). The output of the probe is 4-channels: p – acoustic pressure signal from built-in microphone and 3 channels of particle velocity signals in orthogonal directions – v_x, v_y, v_z. These signals can

be used to calculate the components of the sound intensity vector **I** according to Eq. (1).

$$\mathbf{I} = \begin{bmatrix} I_x \\ I_y \\ I_z \end{bmatrix} = \int_T p(t) \cdot \mathbf{v}(t)\, \mathrm{d}t. \tag{1}$$

The time constant T is related to the frame, in which the direction of coming sound is computed. In this work the length of the frame was equal to 1024 samples, which at 48000 samples per second corresponds to 21.3 milliseconds. The sound intensity vector can also be presented in the polar coordinate system, according to Eq. (2).

$$\begin{aligned} I &= \sqrt{I_x^2 + I_y^2 + I_z^2}, \\ \varphi &= \arctan\left(\frac{I_x}{I_y}\right), \\ \theta &= \arcsin\left(\frac{I_z}{I}\right), \end{aligned} \tag{2}$$

where φ is the azimuth angle, and θ is the elevation angle. The coordinates of the sound intensity vector provide the information about the direction of coming sound, not about the exact localization of the sound source. To determine the localization of the sound source, the geometry of the room has to be taken into account.

In Fig. 3 the room, in which the sound is localized is presented. The center of the coordinate system (x, y, z) is placed under the surface of the seats, directly below the sound probe. The height of the sound probe in this system equals H. The surface of the seats is sloped in respect to the floor of the lecture hall and is modeled by the plane p (3).

$$p : Ax + By + Cz + D = 0. \tag{3}$$

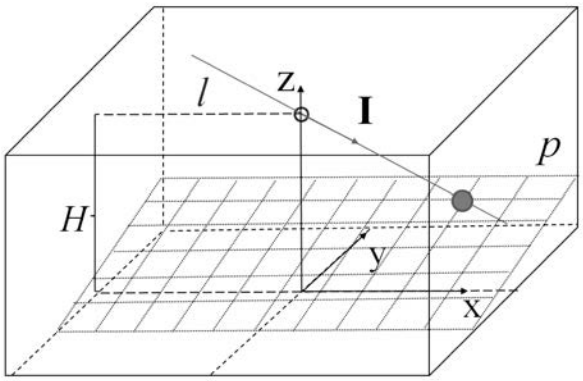

Fig. 3. Illustration of the method of localizing the sound source inside the lecture hall.

The formula for the plane p is calculated basing on the measurement of three points lying on this plane. The line l marks the direction of the sound, indicated by the vector **I**. The formula for l is given by (4).

$$\begin{aligned} y &= ax + b, \qquad a = \frac{I_y}{I_x}, \qquad b = 0, \\ z &= ex + d, \qquad e = \frac{I_z}{I_x}, \qquad d = H. \end{aligned} \tag{4}$$

The sound source lies on the line l and on the plane p, thus the coordinates of the sound source (s_x, s_y, s_z) is the solution of the system of Eqs. (3) and (4). When the coordinates of the sound source are obtained, the number of row and seat in the audience is determined, based on the known layout of the seats in the room. The coordinates (x, y) of every seat and every row are extracted from the architectural plans of the room and tabularized. The seat resulted from the analysis is the one, the distance of which from the calculated point is minimal.

4. PTZ camera control

The result of localization of the sound source is the number of the row and the seat, in which the sound is localized. To steer the PTZ camera these coordinates need to be transformed to the coordinates which can be processed by the camera. The employed camera was manufactured by AXIS and the program communicates with it using the Vapix standard. In this standard the view of the camera is described in a polar-like system of coordinates: p – pan, t – tilt, z – zoom. The placement of the camera is different than the placement of the AVS, though. The calibration procedure and the transformation of (*seat, row*) coordinates into (p, t, z) coordinates is needed. The problem is depicted in Fig. 4.

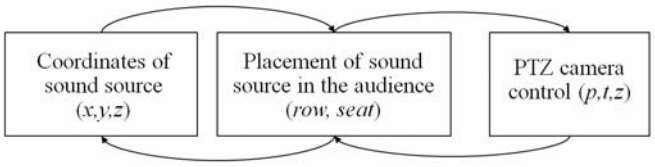

Fig. 4. Transformation of the systems of coordinates.

5. Evaluation

Measurements were conducted to evaluate the accuracy of the described algorithm of localization of sound events in the audience of a lecture hall. The measurement setup, presented in Fig. 5, was composed of the USP probe installed under the ceiling of the lecture hall, the dedicated USP signal conditioning module and a computer equipped with a MAYA 44 USB sound card and a dedicated

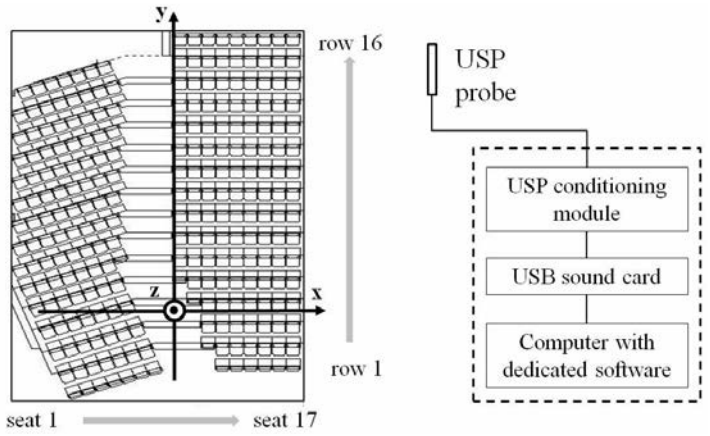

Fig. 5. Measurement system setup.

software for analyzing and recording signals. The system of (x, y, z) coordinates is indicated. The location of the USP probe is 8.35 meters above the center of the coordinate system.

The test signals contained sounds of the needle of a noise gun (without bullets). Such sound has a short duration and contains a significant burst of energy. A number of 5 signals were emitted from each of the places in the auditorium. This yields a total number of 1040 acoustic events. The signals were processed online and the results were saved for further analysis. The dedicated computer program computed the direction of coming sound (azimuth and elevation) and the position of the sound source in the audience (x, y, z coordinates) according to the method described in Sec. 3. The computed values were compared with the *ground truth* values. Ground truth position of the sound source (x_{GT}, y_{GT}, z_{GT}) was extracted from the architectural plans of the building. The assumed system of coordinates was the same as was described in Sec. 3 and presented in Fig. 3. Ground truth azimuth and elevation angle $(\varphi_{GT}, \theta_{GT})$ were calculated according to Eqs. (5) and (6).

$$\varphi_{GT} = \arctan\left(\frac{y_{GT}}{x_{GT}}\right), \tag{5}$$

$$\theta_{GT} = \arcsin\left(\frac{H - z_{GT}}{\sqrt{x_{GT}^2 + y_{GT}^2 + z_{GT}^2}}\right), \tag{6}$$

where $H = 8.35$ m equals the height of the USP probe. The absolute errors of measurements of the position of the sound source and the azimuth and elevation angles were calculated according to the formula $\Delta x = |x - x_{GT}|$, where x is the measured value. The results of such measurement for an example seat in the audience are presented in Table 1.

Table 1. Results of measurement from an example point in the audience.

ground truth						measurement				error			
row	seat	x	y	φ_{GT}	θ_{GT}	x	y	φ	θ	Δx	Δy	$\Delta\varphi$	$\Delta\theta$
7	9	−0.22	3.98	93.15	−59.52	0.73	1.78	67.70	−75.30	0.95	2.20	25.45	15.78
						−0.95	1.99	115.60	−73.20	0.74	2.00	22.45	13.68
						−1.02	2.06	116.20	−72.40	0.80	1.92	23.05	12.88
						−1.27	2.49	116.90	−68.70	1.05	1.49	23.75	9.18
						−0.97	1.74	119.10	−74.80	0.75	2.24	25.95	15.28

From the 5 computed values of error for every seat the minimum value (denoted $\Delta x_{\min}$, $\Delta y_{\min}$ etc.) is selected for further analysis. The spatial distribution of error on the (x, y) plane is presented in Figs. 6–9. Figure 6 presents the spatial distribution of the error of calculation of the x coordinate $\Delta x_{\min}$, and Fig. 7 presents the error of y coordinate $\Delta y_{\min}$. In Figs. 8 and 9 errors of azimuth and elevation angle ($\Delta\varphi_{\min}$, $\Delta\theta_{\min}$ respectively) are plotted.

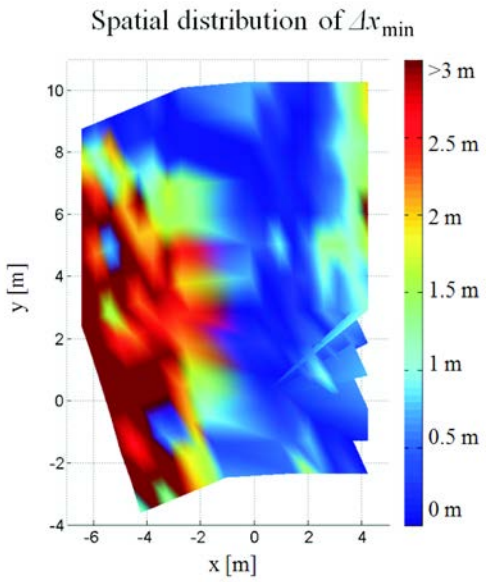

Fig. 6. Distribution of x coordinate error.

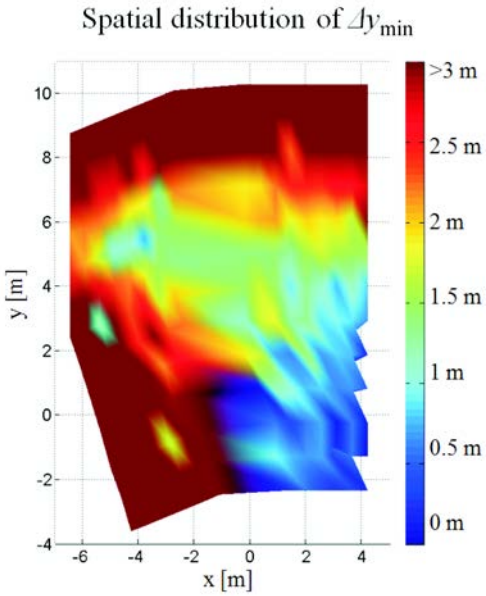

Fig. 7. Distribution of y coordinate error.

It is visible that in some regions of the audience the localization is more precise than in others. The best region is the center seats of the right side of the hall. In the left side of the audience a greater error is observed. It is also apparent that the error of y coordinate is larger than the error of x coordinate. The experiment proves that the sound source can be localized with the precision of ±2 seats/rows in the right side of the audience. It is often impossible to localize the acoustic event in most of the left part of the audience, though.

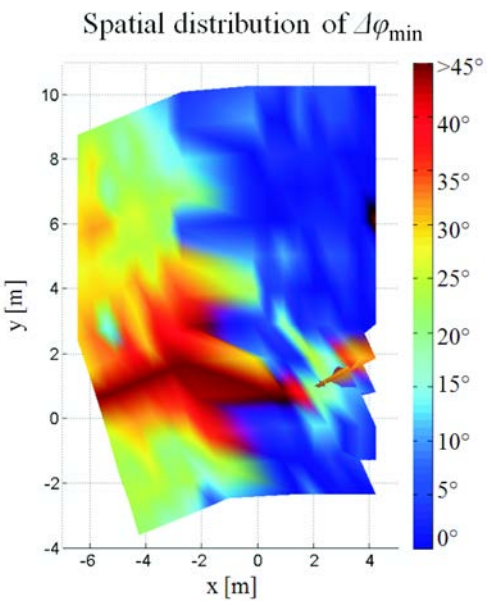

Fig. 8. Distribution of azimuth angle error.

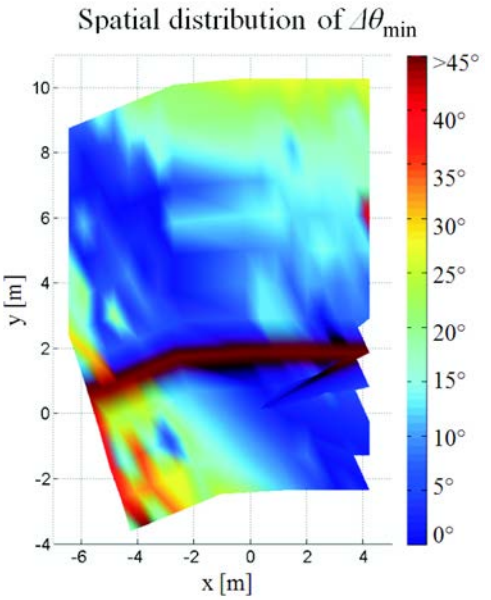

Fig. 9. Distribution of elevation angle error.

As far as the accuracy of angle calculation is concerned, in more than half of the lecture hall the azimuth and elevation error is smaller than 10°. There are however regions, where the angle is calculated wrongfully. A significant error of elevation is observed in row 5. The azimuth angle is inaccurate in the left side of the hall. The possible causes and ways for dealing with this inaccuracy will be discussed in the following section.

6. Conclusions

The method for localization of acoustic events in the audience of a public event was introduced. The demonstration system established in the Gdansk University of Technology was presented. The architecture of the system and the employed sound processing algorithms were explained. The accuracy of the determination of the position of the sound source was measured and discussed. In future work we plan to improve this accuracy by dealing with the following problems:

1. In the assumed model the surface of the seats is a regularly sloped plane. In fact the left side of the auditorium is sloped at a different angle than the right side. A more precise model should be employed to improve the accuracy of the system. Useful techniques for modeling sound propagation in rooms were presented in (Meissner, 2009).

2. The air in the room is heated with heaters placed near the left wall. In such a room the air temperature is not constant in relation to height. Therefore, sound waves do not propagate along straight lines, which leads to change in the direction of coming sound. A correction curve should be employed to diminish the significance of this phenomenon.
3. The probe is located on the level of the sound directing panels hung under the ceiling. It is possible that the panels reflect the direct sound from some localization in such a way, that the reflections from the walls reach the probe first. Thus, the angle of coming sound is calculated with error. There can also be reflections from the ceiling present above the panels (possible cause of the elevation error in row 5). The ray-tracing algorithm could also be applied to localize the sound sources basing on the direction of arrival of reflected sound (Alpkocak, Sis, 2010).
4. The error of calculating the y coordinate is probably caused by the inaccurate assumption of the seat surface plane. The characteristic points in the hall, used to find the formula of the plane (Eq. (3)), were measured with error. More precise measurements should be performed. Moreover, the distance of the sound source from the ground was assumed to be equal to 1.5 m. In the experiment the height of the sound source possibly changed.
5. The error of azimuth and elevation angle can also be caused by the possible tilt of the probe. Precise measurements should be conducted to determine if the probe is in correct position.

The overall accuracy of the system can also be improved by determining the dependence between the elevation and azimuth angle error and applying a correction to the calculation of the parameters.

The key innovation which can lead to improvement of the usability and accuracy of the system is to integrate audio and video analysis. Thanks to such fusion of data, events can be detected multimodally and with greater certainty (Kotus *et al.*, 2010). The integration of video processing techniques with the audio processing algorithms described in this paper in a public events monitoring application will be the subject of future work. It will also be attempted to localize speech sounds, which is more challenging than localization of impulse sounds. Methods for improving quality of speech signal, which can be employed here, were presented in the literature (Drgas *et al.*, 2008; Latos, Pawełczyk, 2010).

Acknowledgments

Research funded within the project No. POIG.02.03.03-00-008/08, entitled "MAYDAY EURO 2012 – the supercomputer platform of context-depended analysis of multimedia data streams for identifying specified objects or safety threads". The project is subsidized by the European regional development fund and by the Polish State budget.

References

1. Abouchacra K., Łętowski T., Gothie J. (2007), *Detection and Recognition of Natural Sounds*, Archives of Acoustics, **32**, 3, 603–616.
2. Alpkocak A., Sis M.K. (2010), *Computing impulse response of room acoustics using the ray-tracing method in time domain*, Archives of Acoustics, **35**, 4, 505–519.
3. Czyżewski A., Kotus J. (2010), *Automatic localization and continuous tracking of mobile sound source using passive acoustic radar*, Military University of Technology, 441–453.
4. Drgas S., Kociński J., Sęk A.P. (2008), *Logatom Articulation Index Evaluation of Speech Enhanced by Blind Source Separation and Single-Channel Noise Reduction*, Archives of Acoustics, **33**, 4, 455–474.
5. Grigoras C. (2009), *Applications of ENF Analysis in Forensic Authentication of Digital and Video Recordings*, Journal of Audio Engineering Society, **57**, 9, 643–661.
6. Julián P., Andreou A.G., Riddle L., Shamma S., Goldberg D.H., Cauwenberghs G. (2004), *A comparative study of sound localization algorithms for energy aware sensor network nodes*, IEEE Transactions on Circuits and Systems, **51**, 4, 640–648.
7. Kotus J., Łopatka K., Czyżewski A. (2011), *Detection and localization of selected acoustic events in 3D acoustic field for smart surveillance applications*, 4th International conference on Multimedia Communications, Services and Security, Kraków.
8. Kotus J., Łopatka K., Kopaczewski K., Czyżewski A. (2010), *Automatic audio-visual threat detection*, IEEE International Conference on Multimedia, Communications, Services and Security, 140–144, Kraków.
9. Latos M., Pawełczyk M. (2010), *Adaptive algorithms for enhancement of speech subject to a high-level noise*, Archives of Acoustics, **35**, 2, 203–212.
10. Li D., Wong K.D., Hu Y., Sayeed A. (2002), *Detection, classification and tracking of targets*, IEEE Signal Processing Magazine, **2**, 17–29.
11. Meissner M. (2009), *Computer Modelling of Coupled Spaces: Variations of Eigenmodes Frequency Due to a Change in Coupling Area*, Archives of Acoustics, **34**, 2, 157–168.
12. Valenzise G., Gerosa L., Tagliasacchi M., Antonacci F. , Sarti A. (2004), *Scream and gunshot detection and localization for audio-surveillance systems*, IEEE Conference on Advanced Video and Signal Based Surveillance, 21–26.
13. Wind J., De Bree H., Xu B. (2010), *3D sound source localization and sound mapping using a PU sensor array*, 16th AIAA/CEAS Aeroacoustics Conference, Stockholm.
14. Zhuang X., Zhou X., Hasegawa–Johnson M., Huang T. (2010), *Real-world acoustic event detection*, Pattern Recognition Letters, **31**, 1543–1551.