

Detection and localization of selected acoustic events in acoustic field for smart surveillance applications

Jozef Kotus · Kuba Lopatka · Andrzej Czyzewski

Published online: 31 July 2012

© The Author(s) 2012. This article is published with open access at Springerlink.com

Abstract A method for automatic determination of position of chosen sound events such as speech signals and impulse sounds in 3-dimensional space is presented. The events are localized in the presence of sound reflections employing acoustic vector sensors. Human voice and impulsive sounds are detected using adaptive detectors based on modified peak-valley difference (PVD) parameter and sound pressure level. Localization based on signals from the multichannel acoustic vector probe is performed upon the detection. The described algorithms can be employed in surveillance systems to monitor behavior of public events participants. The results can be used to detect sound source position in real time or to calculate the spatial distribution of sound energy in the environment. Moreover, the spatial filtration can be performed to separate sounds arriving from a chosen direction.

Keywords Acoustic event detection · Sound localization · Audio surveillance

1 Introduction

The paper addresses the problem of detecting and localizing some selected acoustic events in 3-dimensional acoustic field. The known solutions for localization of acoustic events in most cases use a microphone (pressure sensor) array and are limited to the calculation of acoustic wave direction of arrival (DoA) [14, 17]. In this work a novel approach is introduced employing the acoustic vector sensor (AVS) which makes possible to calculate not only

J. Kotus (✉) · K. Lopatka · A. Czyzewski
Multimedia Systems Department, Gdansk University of Technology,
Narutowicza 11-12, 80-233 Gdansk, Poland
e-mail: joseph@sound.eti.pg.gda.pl

K. Lopatka
e-mail: klopotka@sound.eti.pg.gda.pl

A. Czyzewski
e-mail: andcz@sound.eti.pg.gda.pl

the direction of arrival, but also the exact position of the sound source in 3D space [4]. The AVS comprises a pressure sensor and 3 orthogonally placed air particle velocity sensors (v_x , v_y , v_z) [3, 9]. The multichannel output of the probe allows the calculation of direction of acoustic wave front arrival with only one sensor without the use of any microphone array [1]. The proposed technology is a part of the developed automatic surveillance system [2]. Detecting and localizing acoustic events is particularly useful in monitoring of public events such as sports events or conventions in order to detect potential security threats. The described setup of the demonstration system was installed in a lecture hall at the Gdansk University of Technology.

The block diagram of the algorithm is presented in Fig. 1. The detected events are speech sounds and other sounds, which are detected by speakers or impulse detection modules, respectively. The detectors use only the pressure channel; of the AVS sensor. If the acoustic event is detected, the sound source can be localized in the audience using sound localization algorithms, operating on four channels of the AVS sensor (pressure p , particle velocity v_x , v_y , v_z) [1]. The results of the algorithm operation can be used to monitor the space for potential threats or to control the PTZ (*Pan-Tilt-Zoom*) camera to point the camera towards the direction of the event [1]. Spatial filtration of the sound field can also be employed to discern the acoustic wave coming from the particular direction.

In order to evaluate the system's ability to detect potentially hazardous events a series of measurements was conducted. Two setups of the applied measurement system are presented. The preliminary setup was used for evaluating the detection and the localization of individual sound sources in some selected regions of the audience. Next, some more precise measurements were carried out to evaluate the spatial resolution of the sound localization algorithm. Finally, the errors in the position calculation were analyzed and adequate correcting functions were established to enhance the spatial accuracy.

The remaining part of the paper is organized as follows. In Section 2 and in Section 3 the employed signal processing methods are described. Section 2 is devoted to detection algorithms, whereas Section 3 describes the operation of calculating the position of the sound source. In Section 4 measurement results are presented, which were conducted during the preliminary tests and for the fixed setup. Further, in Section 4 the accuracy analysis is performed and the methods for reducing the error rate are presented. Finally, the usability and further developments of the system are described in Section 5.

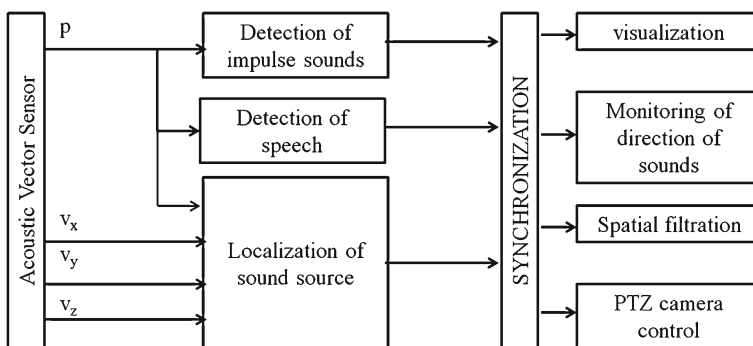


Fig. 1 Concept diagram of the system

2 Detection of acoustic events

The goal of the proposed system is to localize some selected acoustic events, i.e. speech signals and impulsive sounds in 3D acoustical space. Therefore, the algorithm for detection of such events had to be developed. The engineered system employs separate algorithms for detection of the two types of signals. The speech sounds are detected using an adaptive threshold algorithm based on the modified peak-valley difference (PVD) parameter. The detection of impulse sounds also utilizes adaptive threshold, however the parameter used for detection is simpler, namely it represents the equivalent level of sound pressure.

2.1 Detection of speech signals

Vocal activity is connected with appearance of harmonic components in the spectrum of the signal. Therefore, the key to detect speech sounds is to find the repetitive peaks in the power spectrum. The parameter known as peak-valley difference parameter (PVD) is used for voice activity detectors (VAD), which are part of speech processing or recognition systems [16]. The parameter is based on the difference between spectral peaks and valleys of vowels spectral representation,. In the proposed method the commonly used parameter was modified due to following reasons:

- in speech processing the sampling rate is assumed to be equal to 22050 samples per second [S/s], since it covers the significant bandwidth for speech analysis. In the present application it is sampled at 48000 S/s, which is a standard sampling rate in measurements of acoustic pressure from environmental microphones, which are employed in the smart surveillance system
- the localization frame covers 4096 samples. In most speech processing applications, shorter frames are used
- the 4096 Discrete Fourier Transform (DFT) representation of the signal is used to find the spectral peaks of sound
- the distribution of peaks and valleys in the spectrum of speech signals is dependent on the fundamental frequency of speech
- in classic PVD detection the model of peak distribution in vowels needs to be established before calculating the parameter.

To calculate the modified PVD first, the magnitude spectrum of the signal is estimated, using 4096 point DFT. Next, it is assumed that the fundamental frequency of speech (f_0) is located in the range of 80–300 Hz (for speakers of both genders). The fundamental frequency is expressed in the domain of DFT and denoted as k_0 . Consequently, the expected difference between spectral peaks equals k_0 . Thus the distribution of peaks for the assumed fundamental frequency can be resolved without the need for establishing a model of vowel spectral peak distribution. The PVD parameter is then calculated according to Eq. (1):

$$PVD = \frac{\sum_{k=1}^{N/2} X(k) \cdot P(k)}{\sum_{k=1}^{N/2} P(k)} - \frac{\sum_{k=1}^{N/2} X(k) \cdot (1 - P(k))}{\sum_{k=1}^{N/2} (1 - P(k))} \quad (1)$$

where: $X(k)$ is the magnitude spectrum, $N=4096$ is the length of the packet used for Fourier Transform computing and $P(k)$ is the function, which equals 1 if k is the position of spectral peak, 0 otherwise.

The PVD parameter is extracted iteratively for every value of k_0 from the range corresponding to the assumed range of fundamental frequencies. The maximum value is achieved when the current k_0 matches the actual fundamental frequency of the present speech signal and is assigned as a result of PVD calculation. For non-periodic signals the PVD is bound to achieve smaller values than for periodic signals, due to smaller difference between the two components of Eq. (1). The presented parameter is sensitive to signals which are rich in harmonic components (like speech signals). Such signals have a comb-like shaped spectrum and yield a significant difference between the level of peaks and valleys in the spectrum. Signals such as random noise or impulses have a flat magnitude spectrum, thus yielding small values of PVD. The results of PVD calculation from 16 frames are buffered and the mean value of \overline{PVD} is calculated in order to automatically determine the detection threshold. The instantaneous threshold T_i equals $m \cdot \overline{PVD}$ where m is the threshold multiplication factor. For example, $m=3$ means that in order to trigger speech detection, the PVD should exceed 3 times the average value from 16 last frames. The parameter m can be adjusted to change the sensitivity of the detector. Finally, the adaptation of the threshold T in the frame number i is calculated using exponential averaging with the constant $\alpha=0.001$ according to Eq. (2):

$$T = T_{old} \cdot (1 - \alpha) + T_{new} \cdot \alpha \quad (2)$$

This routine allows the detector to change acoustic conditions, i.e. the threshold is updated to follow the profile of the acoustic background. The constant α relates to the adaptation time, i.e. the time after which the former values of threshold are no longer meaningful for Eq. (2). The relation between α and adaptation time T_{adapt} is defined in Eq. (3):

$$T_{adapt}[s] = \frac{N}{F_s \cdot \alpha} \quad (3)$$

where N is the number of samples in the frame and F_s is the sampling rate (4096 samples and 48 kS/s were used).

2.2 Detection of impulsive sounds

The impulsive sounds are detected based on the energy of the signal. The level L of the current frame is calculated according to Eq. (4):

$$L = 20 \cdot \log \left(\sqrt{\frac{1}{N} \sum_{n=1}^N (x(n) \cdot L_{norm})^2} \right) \quad (4)$$

where $N=4096$ is the number of samples in the frame, L_{norm} is the normalization level corresponding to the maximum sample value. The signal level is expressed in dBFS (Full Scale dB). It was verified during measurements that the full scale of the signal corresponds to 120 dB SPL. The current threshold of the impulse detector equals $T=L+m$, where m denotes the margin, which is a sensitivity parameter of the detector (in our

application the margin equals 10 dB). The threshold is automatically updated according to Eq. (5):

$$\begin{aligned} T(0) &= L(0) + m \\ T(i) &= (1 - \alpha) \cdot T(i - 1) + \alpha \cdot (L(i) + m) \quad \text{for } i > 0 \end{aligned} \quad (5)$$

where $T(i)$ denotes the value of detection threshold in the frame number i .

3 Localization of sound sources

After the detection of acoustic events the additional information, i.e. the location of the sound source is extracted from the signal. The localization is based on the processing of signals received from the multichannel acoustic vector sensor. This sensor can provide sufficient data to calculate the acoustic direction of arrival, yet it is not enough to determine the position of the sound source in 3-dimensional space exactly. It is because φ and θ . polar coordinates are known, but the radius r is missing. The key feature of the proposed method for detecting acoustic events in 3D space is to use the information about the geometry of the room. Assuming that the source is located near the floor of the room, the distance between the sound source and the intensity probe can be estimated, thus the exact location of the detected event can be determined. Therefore, the algorithm of localization of the sound source comprises two operations. The first operation is the calculation of components of the intensity vector I of the acoustic field, using the signals from the multichannel acoustic vector sensor, as is defined in Eq. (6):

$$\vec{I} = I_x \vec{e}_x + I_y \vec{e}_y + I_z \vec{e}_z \quad (6)$$

The components of I are calculated by multiplying the signals of acoustic pressure and air particle velocity provided by the AVS, according to the physical dependence expressed in Eq. 7 [7]:

$$\vec{I} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T p(t) \vec{v}(t) dt \quad (7)$$

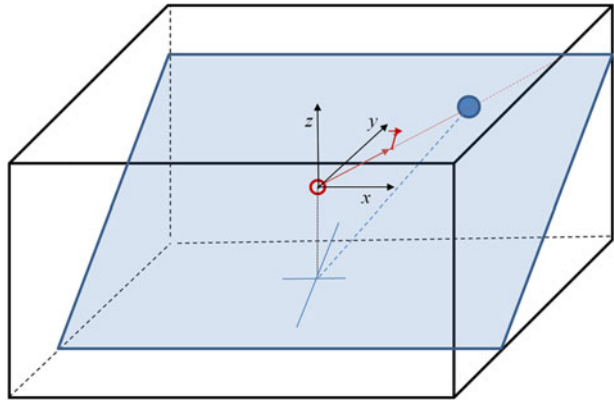
Instead of components of vector I , the polar coordinates are used, denoting the azimuth angle (φ) and the elevation angle (θ). They can be obtained from the components of the intensity vector according to Eq. (8):

$$\begin{aligned} \varphi &= \arctan\left(\frac{I_x}{I_y}\right) \\ \theta &= \arcsin\left(\frac{I_z}{\sqrt{I_x^2 + I_y^2 + I_z^2}}\right) \end{aligned} \quad (8)$$

Once the direction of arrival is computed, the position of the sound source can be determined. To calculate the precise location of the sound source, the information about the position of the AVS and the shape of the room is necessary. The employed method for detecting the sound source inside the room using the acoustic vector sensor is presented in Fig. 2.

The large cuboid represents the shape of the interior, whereas rhomboid models represent the floor plane. The AVS is placed in the room above the floor in the spot marked by the red empty dot. The black dotted line corresponds to the height of the AVS placement. The two intersecting blue lines indicate the point of the perpendicular line to the plane of the floor

Fig. 2 Illustration of employed method for detection of sound source inside a room using acoustic vector sensor (AVS)



directed to the location of the AVS. The full dot marks the position of the sound source inside the interior. The vector I of the intensity of the acoustic field calculated in the xyz space, has the direction of the arrow. The coordinate system, starting in the location of the AVS, is drawn. The intersection of the direction of the intensity vector with the floor plane indicates the position of the sound source. The location of the sound source is expressed by a set of coordinates (x, y, z) .

4 Measurements

The conducted measurements comprise following stages. First, a preliminary measurement system was set up in the lecture hall in order to evaluate the proposed method of calculation of sound source position. Measuring system consist of: USP regular probe (multichannel Acoustic Vector Sensor), conditioning module type MFSC-4, connection cables, multichannel sound card ESI Maya44 USB and laptop ASUS type B50A [3]. It involved 6 positions of a selected sound source in the audience. Next, full measurements were conducted, covering the whole area of the lecture hall. Finally, the error of calculation of sound source position for every seat was calculated and the correction procedure to increase the accuracy was introduced.

4.1 Preliminary measurement setup

The preliminary measurement system was composed of a fixed camera covering the audience, an acoustic vector sensor, the AVS conditioning module and a computer used for data acquisition. Signals used for evaluation of algorithms were recorded employing this demonstration system. In Fig. 3 the setup of the preliminary measurement system is presented. The placement of the acoustic vector sensor and the positions of sound sources (denoted as 1–6) are presented on top of the layout of the lecture hall.

The emitted signals included speech sounds (male's voice counting from 1 to 10) and impulsive sounds (shots from a signal pistol). Signals registered during the experiment with the described measurement system were analyzed using aforementioned audio signal processing algorithms. Some example results of detection and localization of impulse sound sources are presented in subsequent sections. The output of the speech detection algorithm is shown. Basing on the results of the detection of speech sounds, an effort was made to localize the position of the speaker.

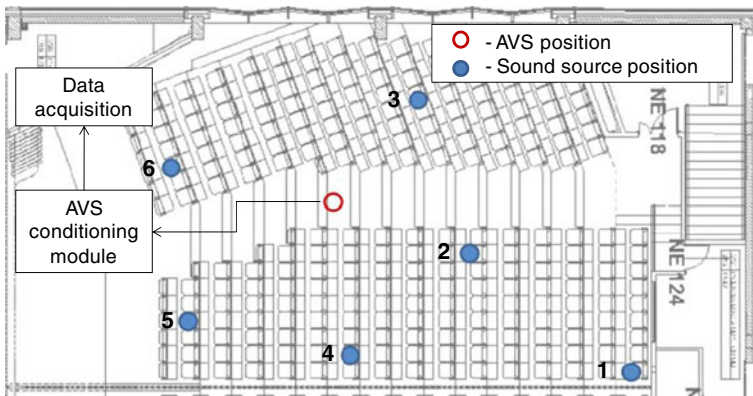


Fig. 3 Setup of measurement system

4.2 Preliminary detection results

A fragment of the measured signal of acoustic pressure was chosen in order to assess the ability of the algorithm for detecting speech signals. It contains words uttered by two speakers located at opposite areas in the auditory room (sources 2 and 3 in Fig. 3). The results of speech detection are presented in Fig. 4. The solid line on the top chart represents the plot of the PVD parameter. The dashed line shows the adaptive threshold of detection. The bottom plot illustrates the decision of the detector. The detection results of impulsive sounds are presented using a fragment of the test signal containing sound of the spire of the signal pistol (shot without bulls).

The spire produces a short-time click, causing a short burst of acoustic energy, whose instantaneous level (measured in 10 ms frames) exceeded the sound pressure level of acoustic background by 20 dB approximately. The results are presented in Fig. 5.

A number of 18 shots was emitted, namely 3 shots from the position of every sound source 1–6. The dashed line on the top plot indicates the 75 dB threshold of detection of impulsive sounds. No additional noise was emitted, so the signals can be considered quiet. The detection of impulsive sounds in this short sample was 100 % correct. The detection of speech was not assessed, since some overlap was present in the words uttered by the speakers. The evaluation of detection error, however, is not the scope of this work, in which detection serves as a preliminary operation before the calculation of sound source position.

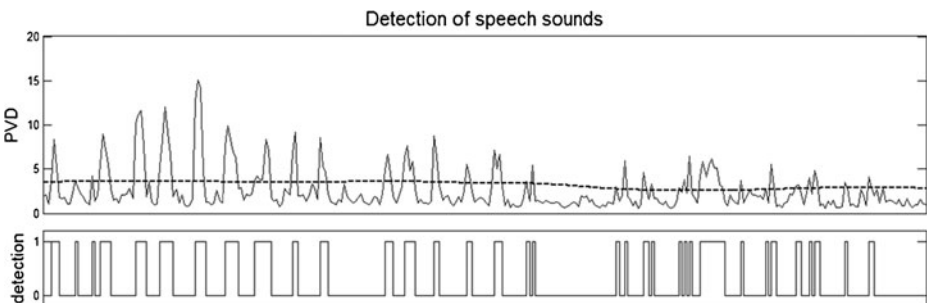


Fig. 4 Detection of speech sounds

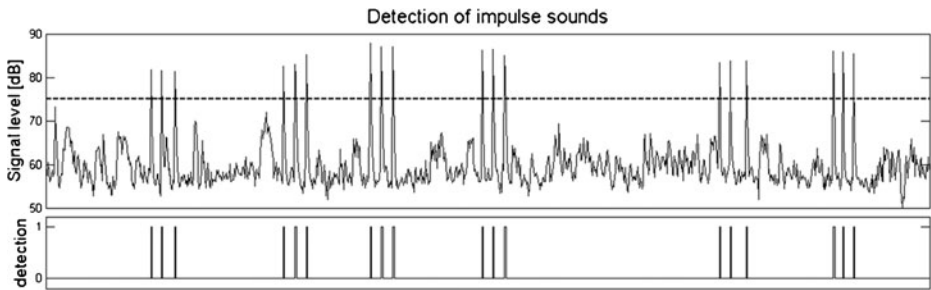


Fig. 5 Results of detecting impulsive sounds

4.3 Preliminary localization results

A proper sound source localization can take place when the AVS captures the front wave related to an acoustic event. Subsequent fragment of sound can include reverberant components produced by reflections from the walls or by objects present inside the room. For that reason, the sound event detection algorithm was modified to determine the sound attack phase properly. The impulsive sounds were analyzed employing frames of 1024 samples with a sampling rate of 48 kS/s. For speech sound events 4096 sample frame length was used. To improve the localization efficiency the band-pass filtration from 300 Hz to 3 kHz was used [13]. This frequency range was suitable not only to speech signals but also for impulsive ones. In this way many reflections, especially at higher frequencies were eliminated. In Fig. 6 computed results of the sound source localization in two dimensions were presented. The left figure presents results for broadband impulsive sounds, the right plot was created for sounds processed with the pass-band filtration. In Fig. 7 computed results of the speech sound source localization in two dimensions were presented. The left figure presents results for broadband sounds, whereas the right plot was obtained using pass-band filtration. The greater colored circles determine the position of the sound source. In Figs. 8 and 9 visualizations of 3D localization results were shown. The lines represent the direction of the computed sound intensity vector. For impulsive sound we did not observe the crosscut between the intensity vector and the plane of the floor. It is because the impulsive sound events were produced using the signal pistol emitted above the volunteers head - see Fig. 8 (right). In Fig. 9 obtained results

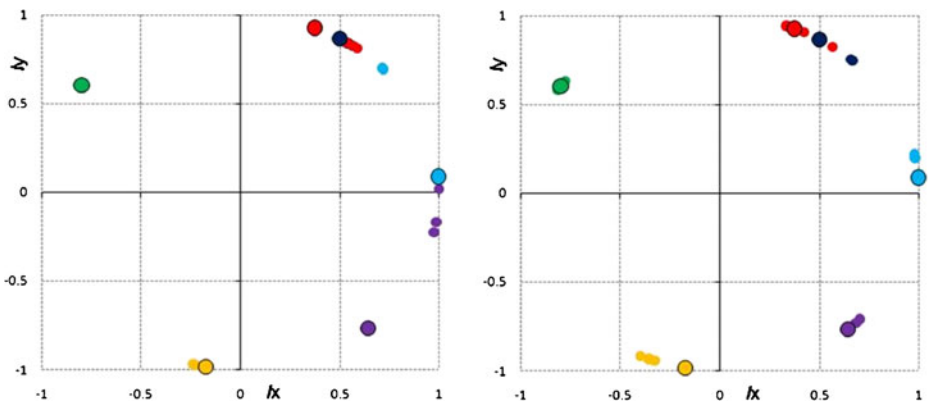


Fig. 6 2D localization results for impulsive sound events

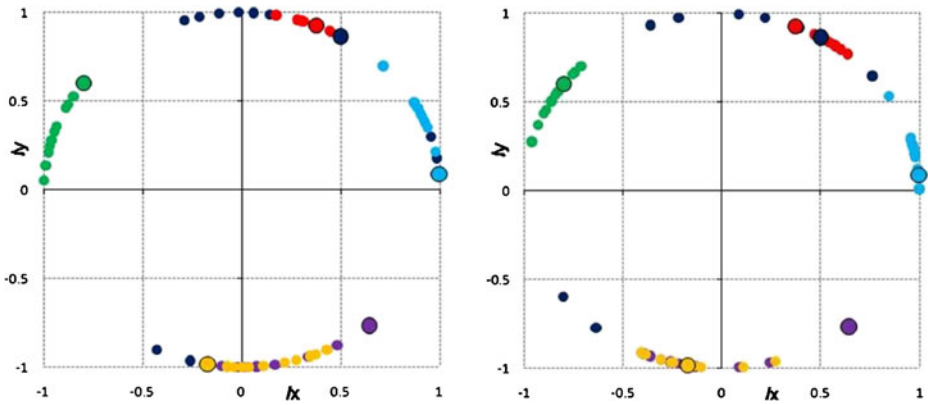


Fig. 7 Localization results for speech sounds events in 2D

for speaker localization were shown. The left part of Fig. 9 presents the proper sound source localization (the intensity vector crosscuts the plane of the floor). In the right part of Fig. 9 the intensity components did not crosscut the plane of the floor. Such situation happened because the sensor was located behind the speaker (see the Fig. 3). In fact the sound, which was localized by the USP sensor represented the reflection from the wall.

The root mean squared angle error (RMSE) indicator was used for evaluating the presented algorithm [11]. The computed values of RMSE for impulsive sounds were equal to 8.0° with filtration and 24.4° without filtration. For speech signals these values were equal to 39.1° and 42.5° , respectively. More information about sensitivity and accuracy can be found in our previous papers [2, 10]. Only dominant sound source was localized in the same time for typical acoustic background. When more than one sound source produced the acoustic energy simultaneously, determination their positions was very difficult because the acoustic energy produced by particular sound sources affects the final intensity vector. This is the limitation of the described sound source localization algorithm [8]. Quite different approach to multiple sound sources localization in real time using the acoustic vector sensor was presented in this study [5, 6]. Direction of arrival (DOA) for considered source was determined based on sound intensity method supported by Fourier analysis. Obtained spectrum components for considered signal allowed to determine the DOA value for the particular frequency independently. Such approach can be applied to localization of multiple different sound sources.

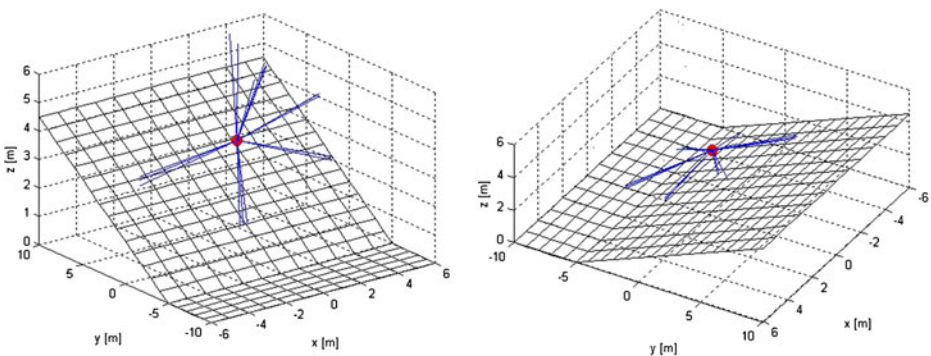


Fig. 8 Impulsive sound source localization results for all positions (in different orientations)

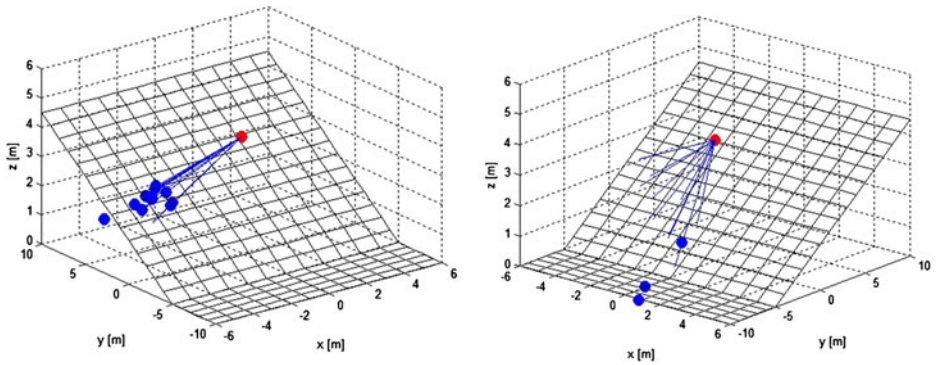


Fig. 9 Speaker localization, example results for speaker 3 and 6 (confront Fig. 3 for details)

4.4 Fixed measurement setup

Based on the preliminary results described in the previous section, the fixed installation in the considered auditory room was made. The location of the USP probe was 8.35 m above the center of the coordinate system. Measurements were conducted to evaluate the accuracy of the described algorithm of localization of sound events in the audience of the lecture hall. The probe was installed under the ceiling of the lecture hall. The signals from the sensor were connected to the dedicated USP signal conditioning module. The analysis was done on a computer with the ESI MAYA 44 USB sound card [15]. The signals were analyzed by a dedicated application and the results were stored. The setup of this measurement system is presented in Fig. 10. The system of (x, y, z) coordinates is indicated.

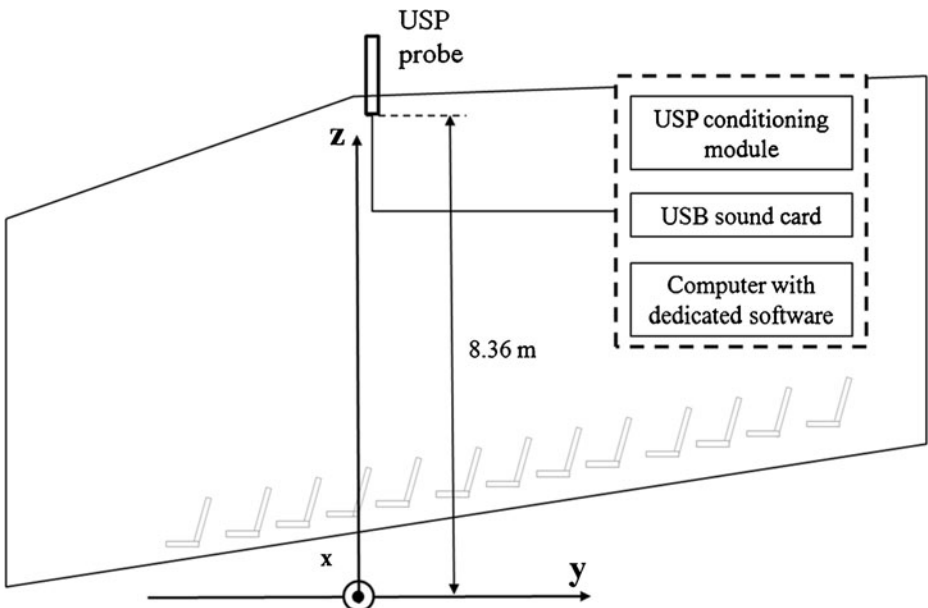


Fig. 10 Measurement system setup

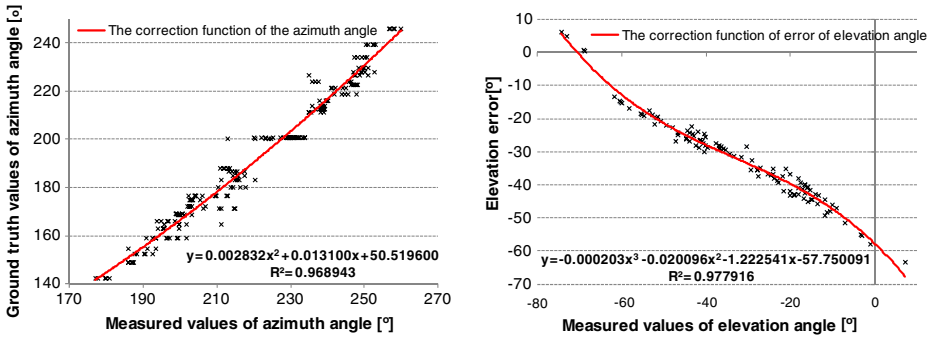


Fig. 11 Correction functions for azimuth and for elevation angle, R^2 – squared correlation coefficient, angle values are expressed in degrees

From each seat in the lecture hall 5 bursts of acoustic energy were emitted (sound of spire of a signal pistol). The result of the sound source localization operation is the direction of incoming sound (azimuth and elevation) and the position of the sound source in the audience (x, y, z coordinates). The results of measurements of error occurring during the calculation of the position of the sound source were presented in related work [11]. The error of calculation of x and y coordinates, as well as azimuth and elevation angle were depicted on the layout of the room. It was shown that the system is prone to errors which might occur due to various reasons related to sound

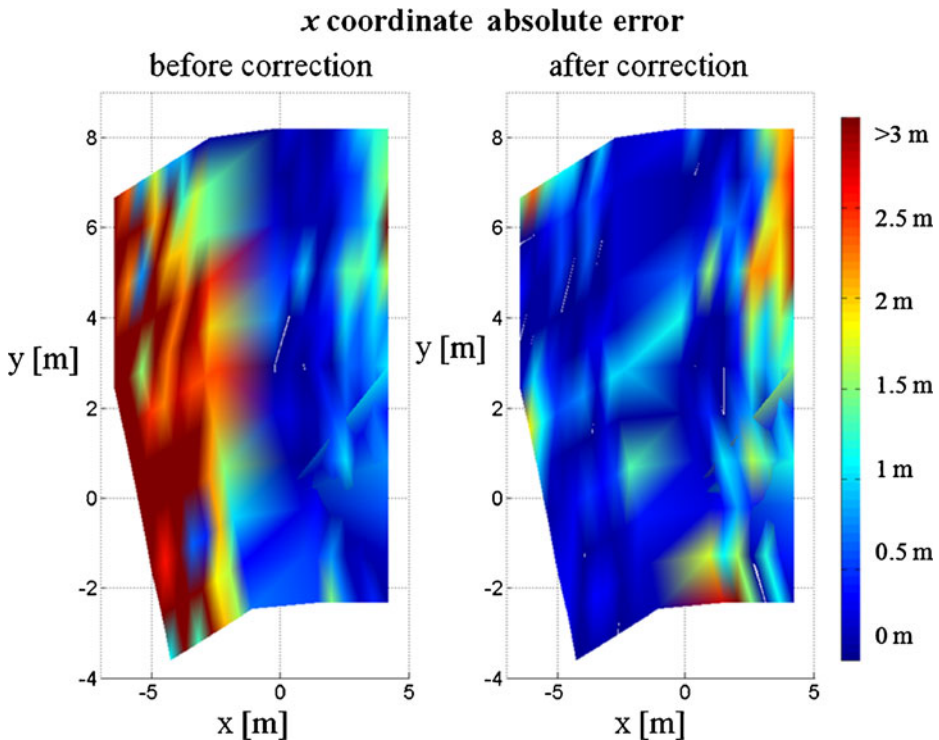


Fig. 12 Absolute error versus x coordinate

propagation, imperfections of the model and the shape of the room. Thus a calibration procedure was introduced to correct the errors of this algorithm. In the next section the proposed correction functions are described.

4.5 Correction functions

A comparison of results obtained in the experiment described in Section 4.4 with the Ground Truth values derived from the architectural plans of the building led to forming calibration functions to correct the computed acoustic wave direction of arrival.

On the basis of real acoustic calibration data and Ground Truth values, the detailed evaluation of the localization accuracy was done. The several kinds of errors were taken into account: absolute error versus x coordinate, absolute error versus y coordinate, absolute error versus azimuth angle, absolute error versus elevation angle [12]. Its distributions were presented in Figs: from 12 to 15 respectively. At the beginning no correction was applied. The error value were high, especially for left part of the room. The obtained error results were analyzed to find the relation between position of the sound source inside the room and the localization error for that position.

The two step correction procedure was proposed. In the first step, the correction of the azimuth angle was done. Next, for the corrected values of the azimuth angle, the

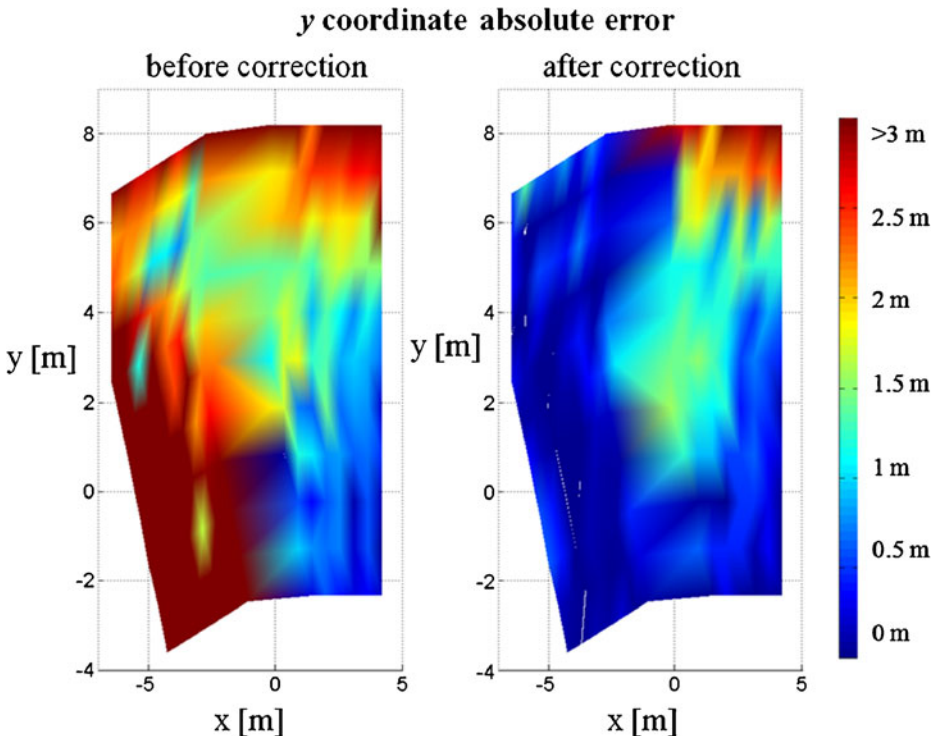


Fig. 13 Absolute error versus y coordinate

prediction of the elevation error (the difference between the ground truth and measured value) is calculated. Finally the predicted error is used to compute the corrected value of the elevation angle (the originally measured elevation angle is reduced by the predicted error). The x and y coordinates are then calculated, accordingly (expressed in meters) on the basis of azimuth and elevation angle values. The correction functions for azimuth and for elevation angles are presented in the Fig. 11.

In the following figures the comparison of system's accuracy before and after employing the correction is presented. Figs. 12, 13, 14 and 15 show the spatial distribution of error versus respective coordinates (x , y , azimuth, elevation). The shape in the figures corresponds to the vertical projection of the room. The coordinates comply with the system presented in Fig. 10. The z coordinate is omitted. It is visible that the employed correction procedure leads to an improvement of calculation accuracy of the position of the sound source in the lecture hall. Such a calibration should be performed after the system is installed in a room. The errors related to x , y coordinates which are smaller than 1 m can be interpreted as a good accuracy, since it yields resolution of 1–2 seats in the audience, which is usually satisfactory for the application of monitoring public events. As it can be seen in Figs. 12 and 13 such a resolution is achieved for most regions of the hall.

Practical experiments with application of proposed correction methodology indicated greater accuracy of the sound source localization. The number of people present inside the room can produce greater level of background noise and can influence on

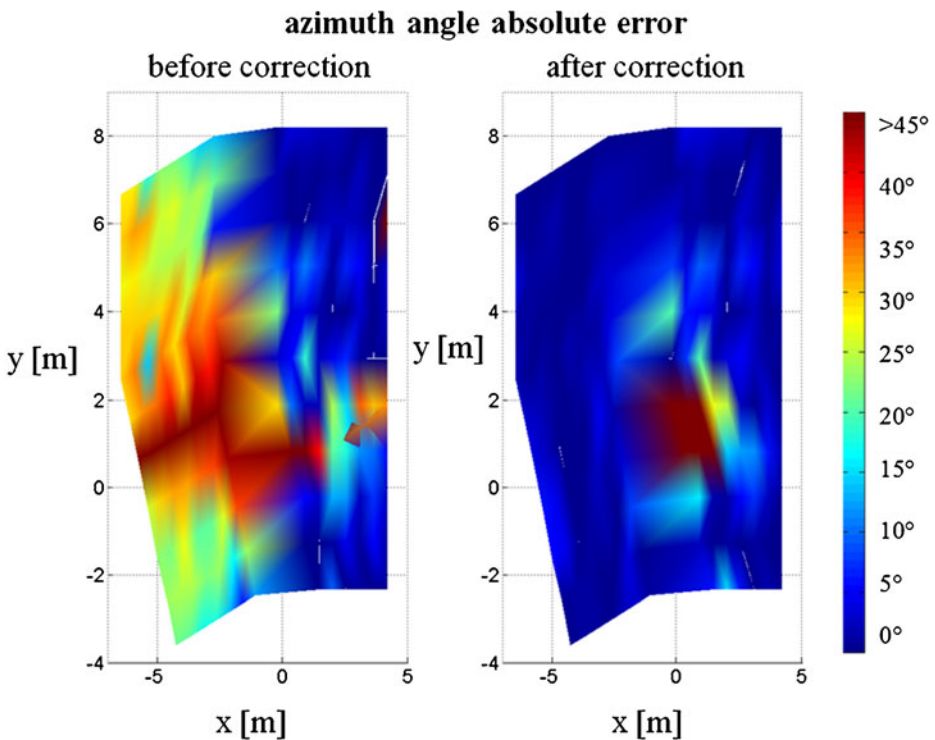


Fig. 14 Absolute error versus azimuth angle

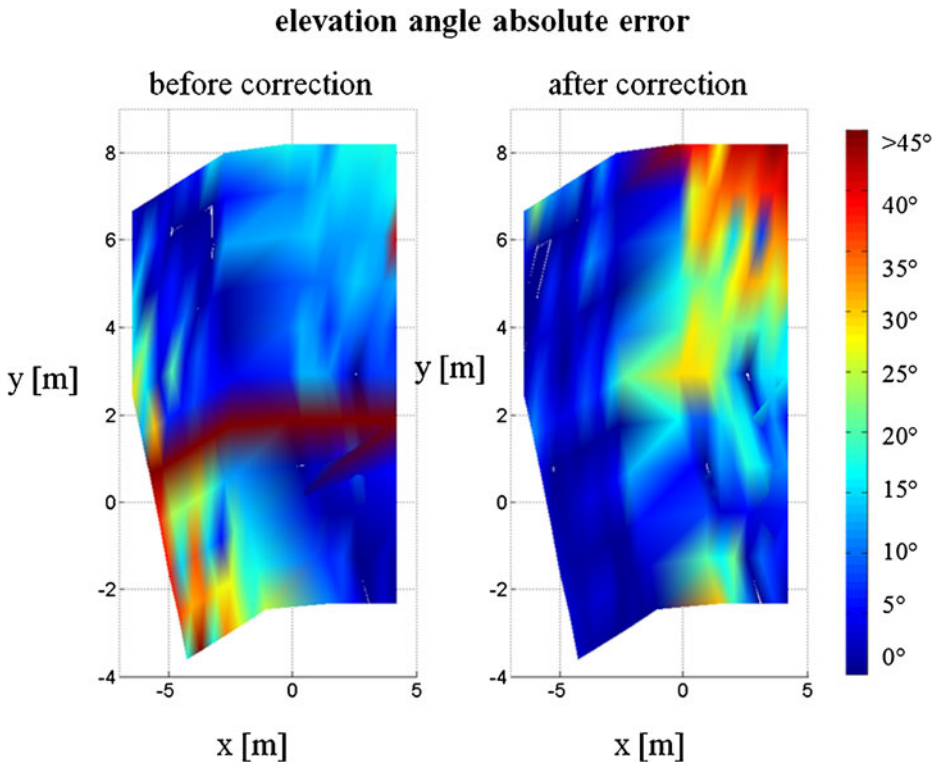


Fig. 15 Absolute error versus elevation angle

acoustic condition inside the room. The scattering of sound should increase, the number of reflections should decrease, it means that the difference between the direct and reflected sound should increase. For that reason, the accuracy of the sound source localization could be the same or greater than for empty room (if difference between the energy of the acoustic event and background noise will be greater than 10 dB, it means that the energy of the background noise can be neglected).

5 Conclusions

The presented method for detection and localization of acoustic events in 3D space was found to be adequate for identifying sound sources inside auditory halls. More generally, the results show that the spatial resolution is sufficient for the monitoring of public events. Obviously, the proposed correction procedure is crucial for achieving this accuracy. A proper sound source localization in the presence of reverberations in real time was possible, not only for impulsive sounds, but also for speech-related sound events. The sound source position was determined using a single 3D acoustic vector sensor. It provides a novel solution while compared to traditional microphone arrays. The information about the sound source position is present in the rising edge of the sound wave. Therefore, a proper detection of the wave front of the acoustic event is crucial for the sound source localization accuracy.

The pass-band filtration significantly improved the localization of the sound source. The presented algorithm is based on sound intensity computation in time domain. The broadband

signal analysis can be disturbed by sound coming from other rooms (especially the low frequency components) and by numerous reflections of high frequency components. The pass-band filtration reduce the level of low and high frequency components and increase the difference between the direct and reflected sound. For that reason the application of filtration improves the accuracy of the sound source localization.

The experimental results are promising, as far as the functionality of acoustical monitoring of activity of people is concerned. The described solutions can be useful for surveillance systems monitoring the behavior of participants of public events. In the future some more complex algorithms for localizing sound sources can be employed, e. g. ray tracing can be utilized to reduce errors related to acoustic wave reflections arriving from the walls of the interior.

Acknowledgements The presented research is subsidized by the European Commission within FP7 project “INDECT” (Grant Agreement No. 218086).

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

1. Basten T, de Bree H-E, Tijs E (2007) Localization and tracking of aircraft with ground based 3D sound probes, ERF33. Kazan, Russia
2. Czyżewski A, Kotus J (2010) Automatic localization and continuous tracking of mobile sound source using passive acoustic radar. In: Military University of Technology Press, pp. 441–453, Warsaw
3. de Bree H-E (2009) The Microflown E-book. <http://www.microflown.com/library/books/the-microflown-e-book.htm>
4. de Bree HE, Leussink P, Korthorst T, Jansen H, Lammerink T, Elwenspoek M (1996) The Microflown: a novel device measuring acoustical flows. *Sensors and Actuators A-Physical* 54:552–557
5. de Bree HE, Wind J, de Theije P (2011) Detection, localization and tracking of aircraft using acoustic vector sensors. In: Inter Noise 2011 Proceedings, Osaka
6. de Bree HE, Wind J, Sadasivan S (2010) Broad banded acoustic vector sensors for outdoor monitoring propeller driven aircraft. In: Proceedings of DAGA 2010, Berlin
7. Fahy FJ (1995) *Sound Intensity*, 2nd edn. E&FN Spon, London
8. Hawkes M, Nehorai A (2003) Wideband source localization using a distributed acoustic vector sensor array. *IEEE T Signal Proces* 51(1479):1491
9. Kotus J (2010) Application of passive acoustic radar to automatic localization, tracking and classification of sound sources. In: Proceedings of the 2nd International Conference on Information Technology, ICIT 2010, pp. 67–70, 28–30 June 2010, Gdansk, Poland
10. Kotus J, Łopatka K, Kopaczewski K, Czyżewski A (2010) Automatic Audio-Visual Threat Detection. In: Proceedings of the IEEE International Conference on Multimedia Communications, Services and Security MCSS 2010, pp. 140–144, Kraków, 6.5.2010–7.5.2010
11. Lehmann EL, Casella G (1998) *Theory of Point Estimation*, 2nd edn. New York, Springer
12. Łopatka K, Kotus J, Czyżewski A (2011) Application of vector sensors to acoustic surveillance of a public interior space. *Arch Acoust* 36:851–860
13. Smith SW (1997) *The Scientist and Engineer’s Guide to Digital Signal Processing*. California Technical Publishing, San Diego
14. Valenzise G, Gerosa L, Tagliasacchi M, Antonacci F, Sarti A (2007) Scream and Gunshot Detection and Localization for Audio-Surveillance Systems. In: Proceedings of the IEEE Conf. on Advanced Video and Signal Based Surveillance, pp. 21–26, London, 5–7.08.2007
15. www.esi-audio.com/products/maya44usb/
16. Yoo I, Yook D (2009) Robust voice activity detection using the spectral peaks of vowel sounds. *Journal of the Electronics and Telecommunication Research Institute* 31:451–453
17. Zhuang X, Zhou X, Hasegawa-Johnson M, Huang T (2010) Real-world acoustic event detection. *Pattern Recog Let* 31:1543–1551



Dr. Jozef Kotus graduated from the Faculty of Electronics Telecommunications and Informatics, Gdansk University of Technology in 2001. In 2008 he completed his Ph.D. under the supervision of prof. Bożena Kostek. His Ph.D. work concerned issues connected with application of information technology to the noise monitoring and prevention of the noise-induced hearing loss. Solutions developed during his work were presented at various domestic and international trade fairs and exhibitions of inventions (received 6 prizes, 3 medals, including gold with mention). Until now he is an author and co-author more than 40 scientific publications, including 6 articles from the ISI Master Journal List and 32 articles in reviewed papers. Also 1 chapter of book published by Springer was issued.



Mr. Kuba Lopatka is a Ph.D. student at the Multimedia Systems Department, Gdansk University of Technology. He graduated from the faculty of Electronics, Telecommunication and Informatics in 2009, majoring in sound and vision engineering and enrolled on the program of doctorate studies at the Multimedia Systems Department. The subject of his Ph.D. thesis is automatic recognition of sounds related to danger. So far he has published, as author or co-author, 14 publications including international journal conference papers and chapters in books published by Springer.



Prof. Andrzej Czyzewski Head of the Multimedia Systems Department is author of more than 400 scientific papers in international journals and conference proceedings. He has led more than 30 R&D projects funded by the Polish Government and participated in 5 European projects. He is also author of 8 Polish patents and 4 international patents. He has extensive experience in soft computing algorithms and sound & image processing for applications among others in surveillance.