Audio-Visual Surveillance System for Application in Bank Operating Room

Józef Kotus[1], Kuba Lopatka[1], Andrzej Czyżewski[1], and Georgis Bogdanis[2]

[1] Gdańsk University of Technology, Faculty of Electronics, Telecommunications and Informatics, Multimedia Systems Department, Gdańsk, Poland
[2] Informatic Systems Designing and Applications Agency Microsystem, Sopot, Poland
{joseph,klopatka,andcz}@multimed.org

Abstract. An audio-visual surveillance system able to detect, classify and to localize acoustic events in a bank operating room is presented. Algorithms for detection and classification of abnormal acoustic events, such as screams or gunshots are introduced. Two types of detectors are employed to detect impulsive sounds and vocal activity. A Support Vector Machine (SVM) classifier is used to discern between the different classes of acoustic events. The methods for calculating the direction of coming sound employing an acoustic vector sensor are presented. The localization is achieved by calculating the DOA (Direction of Arrival) histogram. The evaluation of the system based on experiments conducted in a real bank operating room is given. Results of sound event detection, classification and localization are given and discussed. The system proves efficient for the task of automatic surveillance of the bank operating room.

Keywords: sound detection, sound source localization, audio surveillance.

1 Introduction

An audio-visual surveillance system for application in bank operating room is presented. The system implements the acoustic event detection, classification and localization algorithms for recognizing safety threats. This work provides a continuation of the previous research in which the algorithms were developed [1, 2]. Here their practical application and evaluation in real conditions are presented. According to the concept diagram, presented in Fig. 1, the input signals originate from the multichannel acoustic vector sensor (AVS). Next, the signal processing is performed. Initially, the detection of acoustic events is carried out. Subsequently, the detected events are classified to determine if the event poses any threat or not. Finally, the acoustic direction of arrival (DOA) is determined using the multichannel signals from the AVS. The techniques for detection and classification, as well as the localization algorithm, are described in the next sections.

The algorithms operate in realistic conditions, with a significant level of disturbing noise and room reflections presence. In the following sections an attempt to assess the performance of the employed signal processing techniques in such difficult conditions is made. The results gathered from the analysis of live audio data and recorded situations including arranged threats are presented. It is shown that the performance of the system is sufficient to detect alarming situations.

A. Dziech and A. Czyżewski (Eds.): MCSS 2013, CCIS 368, pp. 107–120, 2013.

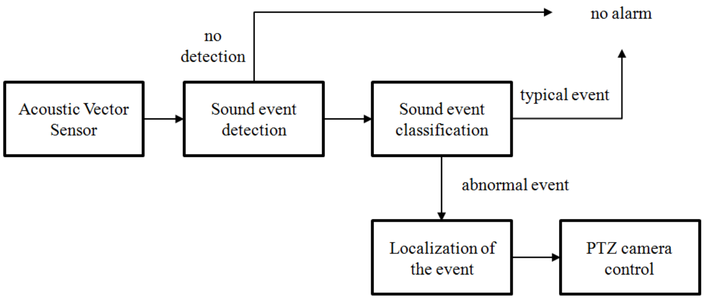

Fig. 1. Concept diagram of the audio-visual bank operating hall surveillance system

2 Acoustic Event Detection and Classification

In the state-of-the art research on acoustic event recognition the most popular approach is to employ generative (e.g. Gaussian Mixture Models or Hidden Markov Models) or discriminative (e.g. Support Vector Machines, Artificial Neural Networks) pattern recognition algorithms to discern between predefined types of sounds basing on the extraction or spectral, temporal or perceptual features [3-4]. Some works follow the so-called *detection by classification* approach, in which the classifier operates online and the decision concerning foreground acoustic event presence is made based on results of classification [4]. Another approach, denoted *detection and classification* separates the process into two operations – detection is responsible for separating the foreground event from the acoustic background; classification – for recognizing the type of detected sound [5]. In our work we follow the second approach. In consequtive subsections the algorithms for detecting and classifying acoustic events will be briefly introduced.

The bank operating hall is an indoor acoustic environment in which most sounds are generated by people and comprise:

- background cocktail-party noise,
- foreground voice activity,
- stamping,
- other sounds (e.g. chairs moving with a squeak, safes beeping, people steps, objects being put down on desks etc.).

These sounds are considered typical elements of the acoustic environment. We define the following classes of acoustic events in order to discern between them: *speech, scream, gunshot, stamp, chair, beep* and *other*. To detect and to recognize the events we use two types of detectors and one classification algorithm, which will be introduced in the following subsections.

2.1 Detection

Two types of detectors are employed. *Impulse detector* is designed to detect short impulsive sounds (e.g. stamps) [2]. *Speech detector* is intended to detect tonal sounds, and voice activity in particular [1]. The impulse detector algorithm is based on comparing the instantaneous equivalent sound level L with the threshold t. The sound level is calculated according to the formula as in Eq. (1):

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

where $x[n]$ represents the discrete time signal and N = 512 samples (at 48000 samples per second) equals to the length of the analysis frame. The parameter L_{norm} (normalization level) assures that the result is expressed in dB SPL (relative to 20 μPa). The speech detector is based on the parameter peak-valley difference (PVD) defined in Eq. (2) [6]:

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

The PVD parameter is calculated from the signal power spectrum $X(k)$ calculated in 4096 sample frames. The vector P contains locations of the spectral peaks. In order to calculate the parameter value, locations of the spectral peaks must be known. Herewith, we employed a simple search algorithm capable of finding the spectral peaks. The detection threshold t is obtained by adding a 10 dB margin to the average sound level in case of impulse detection or multiplying the median PVD value by 2 in case of speech detector. The threshold is smoothed using exponential averaging according to Eq. (3):

$$t = t_{old} \cdot (1 - \alpha) + t_{new} \cdot \alpha \tag{3}$$

The exponential averaging enables an adaptation of the detection algorithm. The new threshold value t_{new} utilized in Eq. 3 is introduced in order to consider the changes in the acoustic background. The constant α is related to the detector adaptation time and is obtained from the formula in Eq. (4), which in this case provides an assumed value of 10 minutes, yielding α equal to $1.8 \cdot 10^{-5}$.

$$T_{adapt}[s] = \frac{N}{SR \cdot \alpha} \tag{4}$$

where SR denotes the sampling rate – in this case 48000 S/s and N denotes the length of the detector frame – 512 samples for the impulse detector and 4096 samples for the speech detector.

2.2 Classification

A Support Vector Machine classifier (SVM) was used to classify the events. The seven classes of events (*speech, scream, gunshot, stamp, chair, beep, other*) are recognised. Therefore, a multiclass SVM with 1-vs-all approach was employed. A set of 55 features was used for the classification. Further details concerning classification and calculation of the features can be found in related publications [7-11].

3 Acoustic Events Localization

Acoustic vector sensors were first applied to acoustic source localization in the air by Raangs et al. in 2002, who used measured sound intensity vector to localize a single monopole source [12]. A more recent development is the application of acoustic vector sensors to the problem of localizing multiple sources in the far field. In 2009, Basten et al. applied the MUSIC method to localize up to two sources using a single acoustic vector sensor [13]. In the same year Wind et al. applied the same method to localize up to four sources using two acoustic vector sensors [14, 15].

The authors' experiences with the sound source localization based on the sound intensity methods performed in the time domain or in the frequency domain were presented in the previous papers [16, 17]. The developed algorithm used for acoustic events localization works in the frequency domain. Its block diagram is depicted in Fig. 2. The multichannel acoustic vector sensor produces the following signals: sound pressure p and three orthogonal particle velocity components u_x, u_y, u_z. The essential functionality of the localization algorithm is its connection with the acoustic event detection module (described in details in section 2.1). The detection module (block 4 indicated by gray shading) operates on acoustic pressure signal only. The detection procedure is performed in parallel to the localization process. Signal set by the detection module affects the operation of the units 5, 7 and 10.

In the block 2 the acoustical signals are buffered and prepared for FFT (Fast Fourier Transform) calculation. The Hanning window was applied. Subsequently, the 4096 point FFT calculation for each signal is performed, with the sampling frequency equal to 48 kS/s (frequency resolution: 11.7 Hz). Such parameters provide a sufficient spectral resolution for sound source localization. The overlap degree was equal to 50%. The FFT calculation was performed for each acoustic component (p, u_x, u_y, u_z), separately. This operation yields transformed signals: $X_p(i)$, $X_{ux}(i)$, X_{uy}(i), $X_{uz}(i)$, where i (ranging from 0 to 4095) denotes the index of the spectral bin. The matrix $\boldsymbol{X}$ now contains information about the direction of arrival of every spectral component of the signal.

$$\mathbf{X} = \lfloor X_p(i) \quad X_{ux}(i) \quad X_{uy}(i) \quad X_{uy}(i) \rfloor = \lfloor \Im\{p(n)\} \quad \Im\{u_x(n)\} \quad \Im\{u_y(n)\} \quad \Im\{u_z(n)\} \rfloor \tag{5}$$

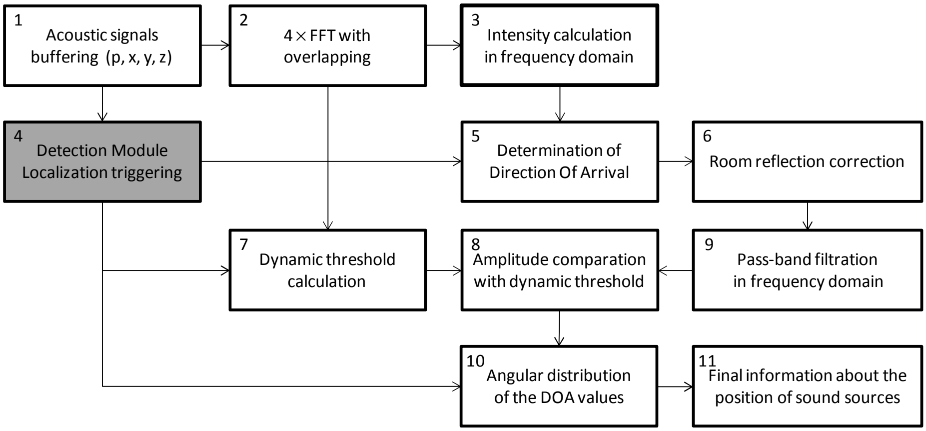

Fig. 2. The block diagram of the proposed algorithm

To extract the information about DOA from the signals in the frequency domain, the sound intensity in the frequency domain is computed (block 3 of Fig. 2). The sound intensity vector is defined and calculated according to Eq. (6).

$$\vec{I}(i) = \begin{bmatrix} I_x(i) \\ I_y(i) \\ I_z(i) \end{bmatrix} = \begin{bmatrix} X_p(i) \cdot \overline{X_{ux}(i)} \\ X_p(i) \cdot \overline{X_{uy}(i)} \\ X_p(i) \cdot \overline{X_{uz}(i)} \end{bmatrix} \tag{6}$$

where:
$I_x(i)$ – sound intensity component for x direction for i-th spectral components,
$X_p(i)$ – coefficients of complex spectrum for i-th spectral components for acoustic pressure signal,
$\overline{X_{ux}(i)}$, conjugated spectrum coefficients for particle velocity in x direction.

The calculation of sound intensity components in the frequency domain is the most important part of the localization algorithm. These data are used in the block 5. Before the calculation of the direction of arrival for particular frequencies, the signal from the detection module is included in the calculation.

The positive state on the detection module output initiates the following actions:

- the dynamic threshold calculation in the frequency domain is suspended (block 7),
- time distribution of direction of arrival (DOA) is prepared to be calculated (block 10),
- the calculation process of DOA for detected sound event begins (block 5).

In the block 5 the angular values for azimuth and elevation for i-th spectral components are computed according to Eq. (7):

$$\begin{cases} r(i) = \sqrt{I_x(i)^2 + I_y(i)^2 + I_z(i)^2} \\ \varphi(i) = \arctan\left(\dfrac{I_x(i)}{I_y(i)}\right) \\ \theta(i) = \arcsin\left(\dfrac{I_z(i)}{r(i)}\right) \end{cases} \tag{7}$$

where $\varphi(i)$ is the azimuth angle, $\theta(i)$ and is the elevation angle.

In the next step (block 6) the obtained angular values are corrected according to specular reflection [18]. This kind of correction is applied for azimuth angle values from the range: 0 to 180 degrees (reflections produced by the wall behind the sensor) and for elevation angle values from the range: 0 to -90 (reflections produced by the ceiling above the sensor) -see the Fig. 5 for details.

The most of the acoustic energy for the speech signal is contained in the frequency range from 200 to 4000 Hz. For better separation of the considered speech signal from the other disturbing components the additional filtration in the frequency domain was applied, calculated in the block 9. At the output of this block DOA values are limited to the frequency range from 200 to 4000 Hz. After that the next step of the localization algorithm can be executed, within the block 8. In such a case the data from blocks 7 and 9 are used. For a better understanding this step the explanation of block 7 is provided below. The computation of the dynamic threshold is performed for all frequency components in the FFT spectrum independently, according to Eq. (8):

$$\left|X_{p\_th}[n]\right| = \left|X_{p\_old}[n]\right| \cdot (1-\alpha) + \left|X_{p\_new}[n]\right| \cdot \alpha \tag{8}$$

where: $|X_{p\_th}[n]|$ - is the magnitude of the acoustic pressure which will be used to calculate the dynamic threshold, n – component number of the FFT spectrum, indices *old* and *new* mean the earlier and the present value of $|X_p[n]|$, respectively and the constant α is related to the spectrum adaptation time and in this case equal to 0.05.

The time averaging process is carried out when the detector module returns 0 (no event detected). As it was mentioned before, the averaging process is suspended when the event detection occurs. In that case the final average values are used as the dynamic threshold.

In the block 8 the values from the block 9 are compared with the dynamic threshold obtained from the block 7 according to the condition expressed by formula (9):

$$L_{p\_event}[n] \geq L_{p\_th}[n] + 10 \tag{9}$$

where: L_{p_event} – sound pressure level for considered acoustic event, L_{p_th} – sound pressure level of dynamic threshold, n – point number of the FFT spectrum.

If the condition expressed by Eq. (9) is true for a given FFT component, the DOA values computed in the block 9 (corrected and filtered azimuth and elevation values for particular FFT components) are used to compute the angular distribution of azimuth and elevation. This operation is performed in the block 10. It is important to emphasize that the azimuth and elevation distributions are prepared to be calculated every time when a new acoustic event is detected (because the algorithm is working in the real time, the data computed for the last event are erased and the new empty tables are prepared). Both angular distributions (both for azimuth and elevation) are computed using the sound intensity value for given angular values according to (10):

$$\begin{cases} F_A[\varphi] = \sum I_\varphi \\ F_E[\theta] = \sum I_\theta \end{cases} \tag{10}$$

where $F_A[\varphi]$ is the azimuth distribution, $F_E[\theta]$ is the elevation distribution, I_φ – the intensity value for given azimuth angle, I_θ – the intensity value for given elevation angle.

Finally, the values stored in the tables $F_A[\varphi]$ and $F_E[\theta]$ are smoothed by means of weighted moving average and Hanning window coefficients (window length was equal to 5), according to (11):

$$\begin{cases} F_{SA}[\varphi] = \sum_{n=-2}^{n=2} F_A[\varphi + n] \cdot h_n \\ F_{SE}[\theta] = \sum_{n=-2}^{n=2} F_E[\theta + n] \cdot h_n \end{cases} \tag{11}$$

where: $F_{SA}[\varphi]$ is the smoothed azimuth distribution, $F_{SE}[\theta]$ is the smoothed elevation distribution, n – index number, h_n – Hanning window coefficients (window length was equal to 5 frames).

The maximum values of the tables $F_{SA}[\varphi]$ and $F_{SE}[\theta]$ indicate the current position of the considered acoustic event.

4 Results

4.1 Detection and Classification Results

First, an experiment was conducted in order to assess the detection and the classification of typical acoustic events. A 13-hour-long signal registered in the bank operating hall was analysed (recorded from 7:00 a.m. to 8:00 p. m.). The ground truth data concerning the position of sound sources were unknown. The detailed results of sound source localization were presented in section 4.2. The results of event detection are presented in Tab. 1. In total 1026 acoustic events were detected. It is visible that stamping (547 detections) are the prevailing sounds. Other events are also frequently present (368 detections). Only 84 speech events were recognised, which can be explained by the fact that speech is often present in the background as cocktail-party noise and the detectors adapt to that background by adjusting the detection threshold according to Eqs. 3 and 4. The sound of moving chairs and money safe beeps were seldom detected (23 and 4 occurrences, respectively). No threatening sound events were detected during that period.

Table 1. Number of detected events

Type of event:	speech	other	stamp	chair	beep
number of detecteed events:	84	368	547	23	4
				total:	**1026**

In Fig. 3 the distribution of events per hour of operation is presented. It is visible that during the early hours – 7 to 9 a. m. – other sounds are more frequent than stamping. Stamping becomes more frequent during the operating hours, which are 9 a.m.to 6 p. m.. Information about the movement in the bank can also be derived from the event distribution. The peak hour appears to be on 1 p. m.. Speech events are evenly distributed, since they are caused by both: the clients and the staff of the bank, who are present during the whole period.

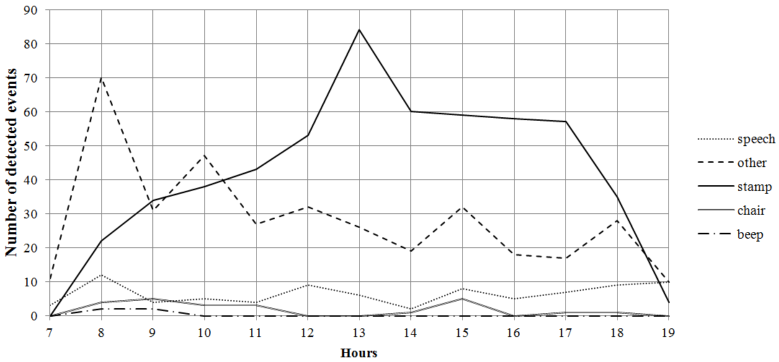

Fig. 3. Distribution of detected events per operating hour

Another important aspect is the level of the acoustic background. As it was stated in the previous sections, the threshold of acoustic event detection is adaptive, i. e. it follows the changes of sound level in the environment. The changes of the detection threshold of both detection algorithms employed are presented in Fig. 4. Again, it is visible that the peak hour is around 1 p.m. to 2 p. m.. The SPL level of the background and the median PVD reach a maximum. During the early and the late hours the threshold is lowered, which enables a detection of some more subtle sound events.

To assess the efficiency of threatening events detection, a hazardous situation was arranged. 17 people took part in an experiment organised in a real bank operating hall. The arranged situations included shouting, screams, typical conversation and robbery simulation with the use of a gun. A noise gun was used to generate gunshot sounds. A 31-minute-long signal was registered and then analysed. 74 events were detected. Sounds belonging to the following classes were emitted: *speech, scream, gunshot* and *other*. The analysis of the classification results is presented in Tab. 2. It can be observed that the threatening events are recognised with an acceptable accuracy. However, the speech sounds are often confused with other acoustic events.

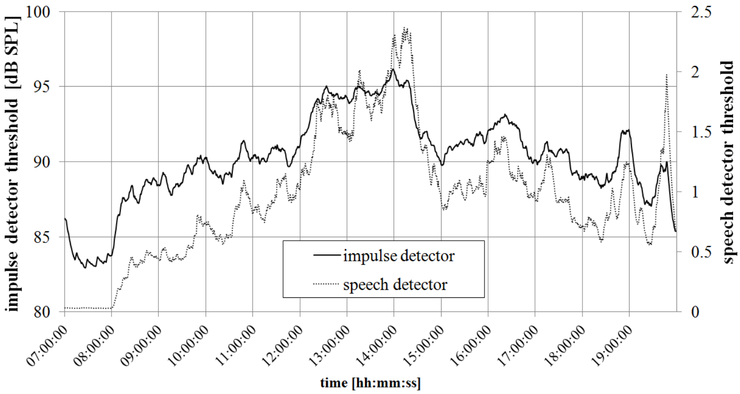

Fig. 4. Changes in detection threshold during operating hours

The analysis of the classification results is presented in Tab. 2. It can be observed that the threatening events are recognised with an acceptable accuracy. However, again speech sounds are often confused with other events. All gunshots were correctly classified, although only four shots were emitted. It is worth noting that the events in this experiment were recognised in fully realistic conditions, namely in the presence of noise (mostly cocktail-party noise).

Table 2. Results of recognition of threatening events

clasified as →	speech	scream	gunshot	other	precision	recall
Speech	**24**	7	0	10	**1**	**0.58**
Scream	0	**21**	0	4	**0.75**	**0.84**
Gunshot	0	0	**4**	0	**1**	**1**
Other	0	0	0	**4**	**0.22**	**1**
		overall correctly classified:			**[53/74]**	**71.6%**

4.2 Acoustic Events Localization Results

In the previous section the detection and classification results were presented. In this section the detection and localization functionality of the developed solution were shown. All experiments were conducted in the bank operating room The ground truth data of sound source position were available. First of all, the detection and localization tests for a single speaker were performed. The speaker was located successively in various customer service points (points 7 and 9 in Fig. 5) and the spoken sequence consisted of counting from 1 to 10. In Fig. 5 the layout of the bank operating room was shown. The position of the USP Sensor, direct sound and reflected sound paths were also depicted. The group of gray rectangles indicates the position of customers service places (tables and chairs). The size of the room: x: 20, y: 18, z: 3.5 m. 20 persons took part in the experiments.

The prepared algorithm detected the spoken sentence, then for this part of the signal both azimuth and elevation values were determined.

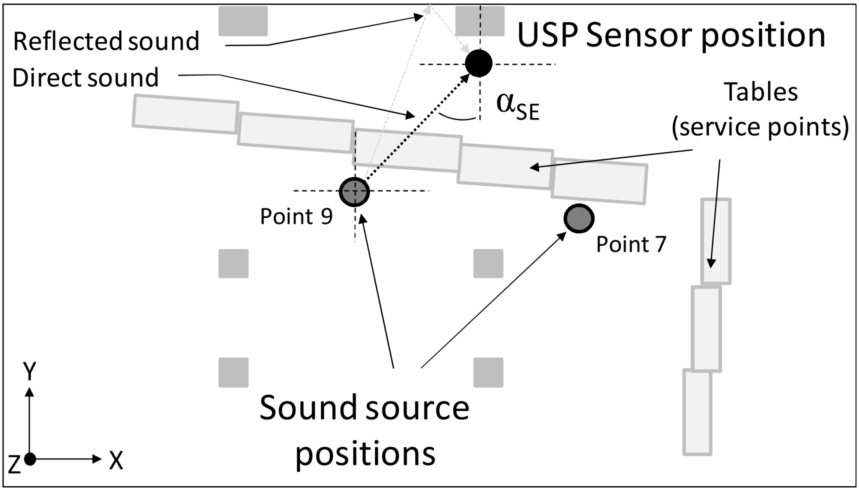

Fig. 5. Speaker positions during the sound source detection and localization experiments

The utterances were repeated three times for each particular point during this experiment. As was mentioned in section 3, only detected sound event is used for computing the azimuth and elevation values. It means that localization process is based only on the parts of the signal in which sound events were detected by the detection module (see the detection algorithm – section 2.1 and localization algorithm in section 3). In Fig. 6 the speaker detection results are depicted.

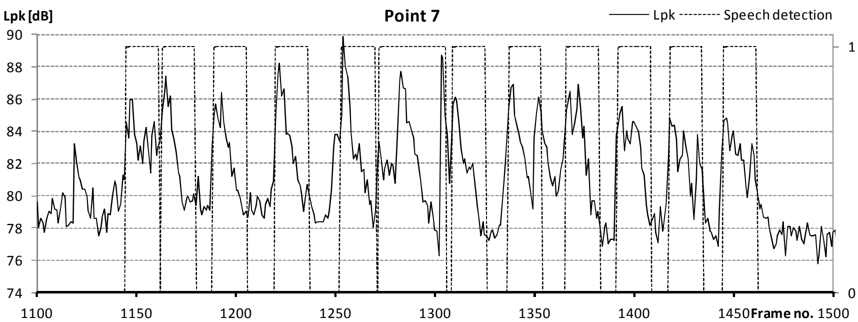

Fig. 6. Speaker detection results obtained in point 7. This signal triggered the sound source localization algorithm.

Two kinds of computation were done. First, the room reflection correction was disabled. In the second analysis the room reflections were considered. The sample results represented by the angular distributions (for azimuth and elevation) are presented in Fig. 7 (the reflection correction was disabled) and in Fig. 8 (the reflection correction was enabled). The ground truth data are also depicted. The gray dot indicates the point 7 and the black rhombus - point 9, respectively.

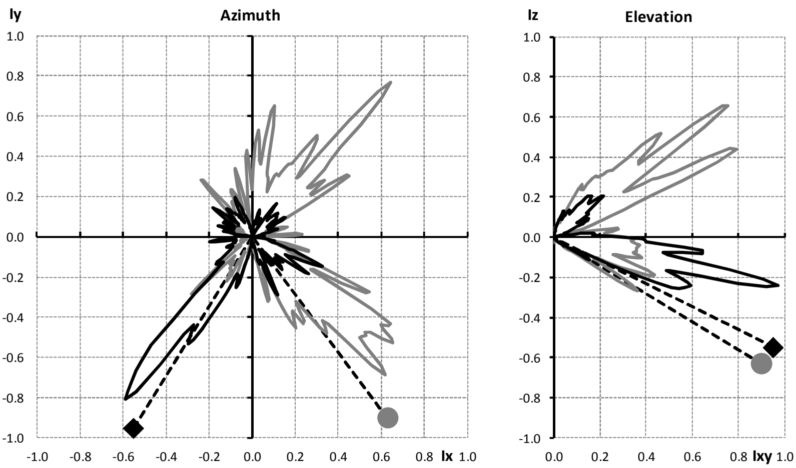

Fig. 7. Sample angular distributions for both: azimuth and elevation obtained for two speaker positions

Localization results for the same signals, however analyzed with the reflection correction were shown in Fig. 8. Imaginary sound sources were reduced, so therefore the proper position of the considered sound sources is indicated. It proves that the reflection correction is an essential part of the localization algorithm.

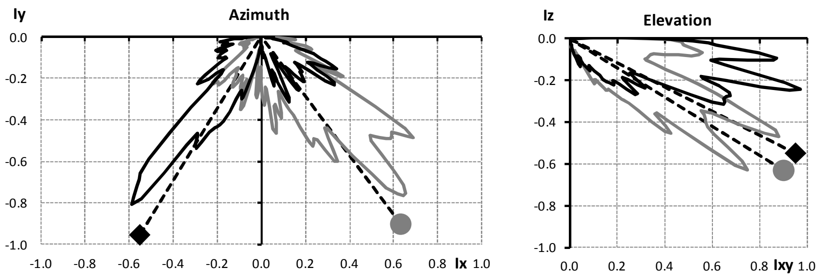

Fig. 8. Sample angular distributions both for azimuth and elevation obtained for two speaker positions

All localization results for every detected sentence are presented in Fig. 9. The average values of azimuth and elevation for considered speaker positions were calculated. They are presented in Fig. 9 .The grey empty circle corresponds to the position 7, the black empty rhombus indicates the position 9. It is important to emphasize that the obtained average angle values are very close to the reference points.

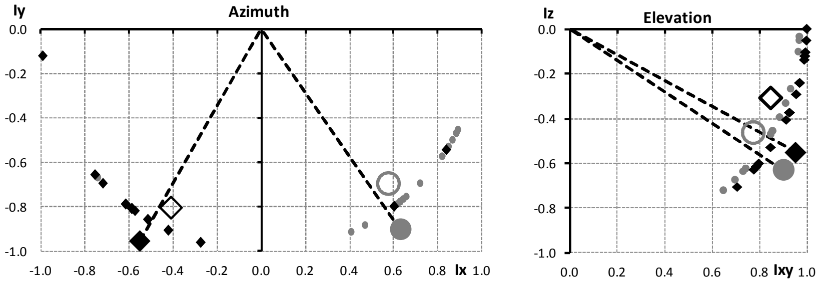

Fig. 9. Localization results expressed by the azimuth and elevation values for all detected sound events spoken in considered points

The average values of azimuth and elevation calculated for all detected sound events spoken in both customer service points were shown in Tab. 3. Reference angular values were also collected. All angular values were expressed in degrees. Tab. 3 contains also the estimated localization error (defined as a difference between the ground truth data and the calculated angles) for both settings: with and without reflection correction.

Table 3. Average results of sound source localization for particular points

Source	Ground Truth		Measurement results		Localization error	
Position	Azimuth	Elevation	Azimuth	Elevation	Azimuth	Elevation
P 7 r.c. on	300	35	309.5	30.8	9.5	4.2
P 9 r.c. on	245	30	243.0	19.9	2.0	10.1
P 7 r.c. off	300	35	203.9	-14.9	96.1	49.9
P 9 r.c. off	245	30	216.2	-0.8	28.8	30.8

5 Conclusions

On the basis of obtained results it was found that accurate acoustic event detection enables preparing essential acoustic data that were used by the sound source localization algorithm. The designed and implemented algorithm of sound source localization was practically tested in the real acoustic conditions. It was shown that the obtained average values of azimuth and elevation were close to the Ground Truth values. The error was less than 10 degrees for considered speaker positions.

It was shown that sound reflections inside operating room can seriously disturb the localization process. The proposed room acoustics influence correction procedure, which allows for obtaining better results, despite the presence of noticeable sound reflections.

The proposed methodology can significantly improve the functionality of the traditional surveillance monitoring systems. Application of acoustic vector sensor can be useful for localization of detected sound events. The described method can be applied to surveillance systems for monitoring and visualising the acoustic field of a specified area. The direction of arrival information can be used to control the Pan-Tilt-Zoom (PTZ) camera to automatically point it towards the direction of the detected sound source.

Acoustic modality offers many interesting functionalities in the automating detection and classification of hazardous situations. It is worth to emphasize that the proposed method can be a useful tool also during the offline forensic audio analysis.

Acknowledgements. Research is subsidized by the European Commission within FP7 project "INDECT" (Grant Agreement No. 218086). The presented work has been also co-financed by the European Regional Development Fund under the Innovative Economy Operational Programme, INSIGMA project no. POIG.01.01.02-00-062/09.

References

1. Łopatka, K., Kotus, J., Czyżewski, A.: Application of vector sensors to acoustic surveillance of a public interior space. Archives of Acoustics 36(4), 851–860 (2011)
2. Kotus, J., Łopatka, K., Czyżewski, A.: Detection and localization of selected acoustic events in 3d acoustic field for smart surveillance applications. Multimedia Tools and Applications (2012) (published online)
3. Cowling, M., Sitte, R.: Comparison of techniques for environmental sound recognition. Pattern Recognition Letters 24, 2895–2907 (2003)
4. Valenzise, G., Gerosa, L., Tagliasacchi, M., Antonacci, F., Sarti, A.: Scream and gunshot detection and localization for audio-surveillance systems. In: IEEE Conference on Advanced Video and Signal Based Surveillance, London, August 5–7, pp. 21–26 (2007)
5. Temko, A., Nadeu, C.: Acoustic event detection in meeting room environments. Pattern Recognition Letters 30, 1281–1288 (2009)
6. Yoo, I.-C., Yook, D.: Robust voice activity detection using the spectral peaks of vowel sounds. ETRI Journal 31(4), 451–453 (2009)
7. Zhuang, X., Zhou, X., Hasegawa-Johnson, M., Huang, T.: Real-world acoustic event detection. Pattern Recognition Letters 31, 1543–1551 (2010)
8. Rabaoui, A., et al.: Using robust features with multi-class SVMs to classify noisy sounds. In: International Symposium on Communications, Control and Signal Processing, Malta, March 12–14 (2008)
9. Łopatka, K., Zwan, P., Czyżewski, A.: Dangerous sound event recognition using support vector machine classifiers. In: Nguyen, N.T., Zgrzywa, A., Czyżewski, A. (eds.) Adv. in Multimed. and Netw. Inf. Syst. Technol. AISC, vol. 80, pp. 49–57. Springer, Heidelberg (2010)

10. Zwan, P., Czyzewski, A.: Verification of the parameterization methods in the context of automatic recognition of sounds related to danger. Journal of Digital Forensic Practice 3(1), 33–45 (2010)
11. Peeters, G.: A large set of audio features for sound description (similarity and classification) in the cuidado project. CUIDADO IST Project Report 54 (version 1.0), 1-25 (2004)
12. Raangs, R., Druyvesteyn, W.F.: Sound source localization using sound intensity measured by a three dimensional PU probe. In: AES Munich (2002)
13. Basten, T., de Bree, H.-E., Druyvesteyn, E.: Multiple incoherent sound source localization using a single vector sensor ICSV16, Krakow, Poland (2009)
14. Wind, J.W., Tijs, E., de Bree, H.-E.: Source localization using acoustic vector sensors, a MUSIC approach, NOVEM, Oxford (2009)
15. Wind, J.W.: Acoustic Source Localization, Exploring Theory and Practice. PhD Thesis, University of Twente, Enschede, The Netherlands (2009)
16. Kotus, J.: Application of passive acoustic radar to automatic localization, tracking and classification of sound sources. Information Technologies 18, 111–116 (2010)
17. Kotus, J.: Multiple sound sources localization in real time using acoustic vector sensor. In: Dziech, A., Czyżewski, A. (eds.) MCSS 2012. Communications in Computer and InfSCCISormation Science, vol. 287, pp. 168–179. Springer, Heidelberg (2012)
18. Weyna, S.: Identification of Reflection and Scattering Effects in Real Acoustic Flow Field. Archives of Acoustics 28(3), 191–203 (2003)