The Joint Database of Audio Events and Backgrounds for Monitoring of Urban Areas

PLEVA Matúš, VOZÁRIKOVÁ Eva, DOBOŠ Ľubomír, ČIŽMÁR Anton

Technical university of Košice, Slovakia
Department of Electronics and Multimedia Communications, FEI TU Košice
Park Komenského 13, 041 20 Košice, Slovak Republic
E-Mail: {Matus.Pleva, Eva.Vozarikova, Lubomir.Dobos, Anton.Cizmar}@tuke.sk

Abstract* – This paper describes the Joint Database of Audio Events (JDAE-TUKE), which was assembled for developing and evaluating audio events detection algorithms. These algorithms should be capable of recognizing dangerous situations from microphones installed in public urban areas. The database consist of different audio events recordings and it was extended with long recordings of different backgrounds to emulate different conditions during the evaluation of the audio events detection algorithms. After joining these resources the corpus was prepared for manual annotation. This database of approximately 800 recordings was collected for monitoring of the urban areas and indicating dangerous events as gunshots, explosions, car crashes etc. The definition, categorization and collection of appropriate corpus of sound (audio event) recordings, the annotation process, chosen file standards, used tools and strategies are provided. This database is available for INDECT project partners and it is planned to be extended after installing of the static outdoor noise monitoring station with continuous recording on the campus building.

Keywords: *audio events; audio monitoring; database*

I. INTRODUCTION

The urban areas security is assured mainly using PTZ (pan tilt zoom) surveillance cameras. The CCTV (Closed-circuit tv) operators are usually watching more cameras at once. This research is focused on utilization of noise monitoring stations in the urban areas, which are often used because of specific hygienic rules.

For example urban areas near factories, airports, big crossings and other noisy environments are monitored with static outdoor microphones and the noise levels are collected and controlled periodically. These control stations could be used to monitor also dangerous sound events and send an alert to the command & control center. For example the sound of shooting, screaming, explosions, etc.

These omnidirectional (or nondirectional) monitoring microphones used in monitoring stations are usually positioned in a way, that one is not able to recognize any spoken content from the received audio signal. The microphone is usually under-gained to prevent overdrive effect from noisy events as an aircraft take-off.

The first step of building a system, which generates alerts only if dangerous audio events occurred, is to build a database of different audio events and also comparably loudly regular events to prevent false alarms. It is possible to collect some real sounds from real environments, use clear studio recordings or use simulated recordings from TV shows or movies.

The second step is to emulate noisy environments using recordings of different backgrounds. These background recordings could also be used for testing the audio events recognition algorithms.

A. Outdoor event detection problem

Many of the current researches in the field of audio events detection are focused on indoor environments, especially smart room monitoring applications [1]. There are many possible approaches based mainly on microphone arrays [2].

Privacy protection of the citizens whose speech could be recorded on the monitoring storage databases is a problem of the outdoor sound monitoring.

There are research studies in the field of surveillance of hazardous situations [3], events detection for an audio-based surveillance system [4] or mixed audio-visual event recognition from surveillance cameras [5]. There are also research activities in the field of detecting events from real life environments [6] for indexing different recordings and to improve the searching strategies. There are no public databases available.

The new approach is in the method of building and the content of the database. JDAE-TUKE is focused on data with low gain, to prevent overloading from loud events, where one is not able to distinguish any spoken content and the detected events are only loud events which could be reported to security & rescue stuff in particular urban area.

B. Noise monitoring in urban areas

The noise monitoring stations in urban areas are common near industrial spaces, motorsport venues, wind

farms and especially airports [7] using specialized outdoor microphones (as you can see on the Fig. 1).

Also big cities usually work on environment monitoring to prevent complaints [8] of the community (including health problems, insurance, etc.), which includes exhaust, light noise, electromagnetic fields, noise monitors, etc. [9].

The noise monitoring stations installations are common in urban areas, so it is very useful to use the installed outdoor microphones and installed equipment for improving security or speed of public services response to dangerous situations.

Figure 1. Environmental outdoor microphone for noise monitoring applications

II. DATABASE PREPARATION

First of all it is required to choose the sound events which are interesting for our purposes. Then it is necessary to find out which sounds are similar to the interesting ones, and they often appear in monitoring environment audio recordings.

Then we need to prepare the recording procedure, the desired recording format, the equipment and also find out any other types of possible resources, in the case that during the recording procedure will not be recorded enough audio materials as it is needed for training and testing purposes.

We also need to decide what type of file formats and annotation techniques will be used after collecting the audio data.

A. Audio event categories

The selected audio events such as gunshots, breaking the glass, scream, car crash and explosion are interesting for the detection of abnormal sounds. Of course especially during outdoor background recordings there are a lot of loud audio events occur, which are important to annotate because of reducing the False Alarm Rate. We divided the events to these categories:

a) Speech-based audio events - This group of events consists of all events, which are produced by human beings in a form of speech or scream and relates to threats, violence, dangerous situations and any other loud vocal expressions (cheering, etc.).

b) Non-speech audio events - This group of events consists mainly of traffic sounds (including airplanes and helicopters), sounds accompanying threats (gunshots) and similar to them (fireworks), animal sounds (dogs, birds, etc.).

c) Ambient noises – The audio input of the monitoring system in outdoor environment contains also the ambient noise: music, sounds produced by the abnormal weather conditions (like strong rain, thunder storms, strong wind), etc.

B. Equipment used for recording the database

The recordings were mainly made using Olympus LS-10 PCM stereo recorder on the stand. They have uniform format: sampling frequency 48kHz, 16 bits per sample PCM encoded WAV files. We are planning to use professional outdoor microphones (Fig. 1) to extend the database in the future. Using this recording equipment especially gunshots from air-guns and breaking glass events was performed as you can see on the Fig. 2 and Fig. 3 below.

Figure 2. Glass breaking recording (bottles and windows)

Figure 3. Gunshots recording on snowy field

We are planning to make a stable outdoor noise monitoring stations which will produce recordings from noisy crossings from the top of the buildings, near or inside university campus. This noise monitoring stations will have stable outdoor microphones which are professionally used near airports for the noise level monitoring for environmental purposes (Fig. 1).

The monitoring stations will be connected to the external storage, and every loud event will be automatically included to the database for the annotation.

After the first stable release of the audio events detection system for outdoor noise environment will be completed, the events could be also automatically detected and online reported to operator if dangerous event occur.

C. Data collected from other sources

The corpus of recordings should to have a large number of each sound event realization, which is necessary for training of models for audio events recognition. That's the reason why the database also consists of recordings of real or actor played events collected from freely available sources, like recordings of crime TV series produced by the same television as it was broadcasted on (for easier approval procedure).

These recordings were made using Technisat AirStar 2 PCI internal DVB-T card, which is able to stream the broadcasted media on the hard drive in the PC. The quality of DVB-T recording is usually the same as DVB-S transport stream quality from satellite broadcast and cable operators DVB-C quality: 48kHz sample rate 128kbits CBR (constant bitrate) quality MPEG-1 Audio Layer 2 codec (depending on the TV station). The audio channel is usually not transcoded when passing the cable/terrestrial provider transcoder, because the biggest challenge is to save the video bitrate.

Second freely available source of real or actor/amateur played events are home recordings published on the websites, which could contain very rare events like car crashes, real emotions of scared people and so on.

Many professional car crashes (in the movies) are corrupted with additional explosions, music or other sounds which make these materials not usable for our purpose. The problem is also the quality of these recordings, because they are usually compressed using lossy algorithms which corrupt the original sound and the reconstructed one is sometimes very distorted and blurry.

D. Database structure and used audio tools

The database consists of audio recordings and metadata files. The audio recordings are divided to the directories named according to the recorded event. A special category is the background recordings, which consist of many events or no events (for example only background traffic noise).

Every audio recording has different metadata files with the same name as the audio recording but different extension.

First of all, the annotation file **.trs* (native Transcriber [10] tool XML format), **.stm* (the NIST Sclite [11] more simple text file format exported from Transcriber), **.txt* (any additional text notes from the annotator or the recording staff). In particular directories there could be any other documents with more detailed description of the content or the recording equipment and also specific notes from annotators about the content or problematic issues.

If there are some unwanted sounds as the operator voice or sound before switching off the recording equipment, we use the freely available Audacity [12] audio tool for editing or potentially down-sampling the recordings.

III. ANNOTATION PROCEDURE AND TOOLS

The annotation procedure of the recorded events and especially background recordings consist of these annotation levels:

a) the segmentation and labeling of the sound events from closed predefined item set – labeling the time intervals and if the rest of the time is quiet or some other event (music)

b) labeling the background noise (different time level)

c) specifying the recorder setup, place, time and recording equipment

The annotation is realized mainly using Transcriber tool [7], but in the future we plan to change the annotation tool with more flexible solution for this purpose: the ELAN tool developed by Technical group from Max Planck Institute for Psycholinguistics [13].

IV. DATABASE DESCRIPTION

The final database of joint audio event recordings and backgrounds for real-time audio event detection purposes consist of more than 790 recordings and more than 200 minutes of backgrounds from different places and environments as can be seen on the Table 1 below (busy street, quiet underground room, near public traffic stops, supermarkets, traffic roundabout, big crossing, near campus before lessons start in the morning, etc.).

The specified sound events recordings will produce alarms for the operator (*AP events* – Alarm Producing events): *explosions (23 recordings), broken glass (128), gunshots (162), screaming/calling for help (96), car crash (16).*

Other sound events located in the database for preventing false alarms are (*FA events* - False Alarm events): *fireworks (26 recordings), aircraft (10), helicopter (8), different alarms (45), church bells (6), air drill (4), thunder (22), power-saw (4), different siren sounds (20), accelerating car (4)*, etc.

TABLE 1. Database statistics.

Audio events	Number of files	Length [s]
gunshots	182	728
breaking the glass	167	1 002
other AP events	211	1 477
other FA events	209	2 010
backgrounds	23	12 568
Total	792	17 785

V. DATABASE UTILISATION

For evaluating of the database many tests were done, using MFCC features and from 1 to 64 Gaussian PDF mixtures HMM models [14] and also SVM classifiers [16] and MFCC & MPEG-7 feature vectors combination [17] (MPEG-7 features used: Audio Spectrum Centroid, Audio Spectrum Spread and Audio Spectrum Flatness). The accuracy of the detection was measured using Audio Event Detection Rate (DR) [%] defined by the formula [17]:

$$DR[\%] = \frac{Number\ of\ correct\ recognized\ models}{Number\ of\ all\ reference\ models} \times 100. \quad (1)$$

The results are promising for quiet environment recordings, where almost 100% detection rate was reached using HMM models and SVM classifiers [16].

Unfortunately when the background noise level is higher, the detection rate falls beyond 50% in case of breaking glass detection and using intensity of background environment sounds mixed to 5dB under detected event intensity level [14]. Gunshot models are more robust and the detection rate is usually around 90% also in noisy recordings [17]. See detailed results in references.

VI. CONCLUSIONS

We are planning to install our own static noise monitoring station with professional outdoor microphone, connected to new storage equipment. We are working on developing outdoor event detection system with a low false alarm rate and automatic report producing evaluating engine. This engine will send the reported event with the sound recording to the evaluator, and s/he will confirm or reject the event (and possibly categorize the event to correct item set).

In the scope of the INDECT project we are preparing also extension of the database with recorded crime scenes played by professional actors and police cadets (cooperation with PSNI - Police Service of Northern Ireland) and different gunshots recordings (cooperation with GHP – General Headquarters of Police Warsaw).

ACKNOWLEDGMENTS

The research presented in this paper was supported by the Slovak Research and Development Agency and Ministry of Education under research project APVV-0369-07, VMSP-P-0004-09 and VEGA-1/0065/10 and the EU ICT Project INDECT (FP7- 218086).

REFERENCES

[1] A. Temko, R. Malkin, C. Zieger, D. Macho, C. Nadeu, M. Omologo, “CLEAR Evaluation of Acoustic Event Detection and Classification systems”, In: Lecture Notes in Computer Science, 1st International Evaluation Workshop on Classification of Events, Activities and Relationships, CLEAR 2006, Southampton, April 6-7, 2006, LNCS 4122, pp. 311-322, ISSN 0302-9743, 2007.

[2] A. Temko, C. Nadeu, “Classification of meeting-room acoustic events with support vector machines and variable-feature-set clustering”, ICASSP 2005, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, Philadelphia, March 18-23, 2005, Volume V, article number 1416351, pp. V505-V508, ISBN: 0780388747, 2005.

[3] S. Ntalampiras, I. Potamitis, N. Fakotakis, “On acoustic surveillance of hazardous situations”, ICASSP 2009, IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 165-168, Taipei, Taiwan, April 19-24, ISBN: 978-1-4244-2353-8, 2009.

[4] C. Clavel, T. Ehrette, G. Richard, “Events detection for an audio-based surveillance system”, In proceedings of IEEE International Conference on Multimedia and Expo, ICME 2005, July 6-8, Article number 1521669, pp. 1306-1309, ISBN: 978-078039332-5, 2005.

[5] M. Cristiani, M. Bicego, V. Murino, “Audio-visual event recognition in surveillance video sequences”, In: IEEE Transactions on Multimedia, Volume 9, Number 2, pp. 257–266, February 2007, ISSN: 1520–9210, 2007.

[6] Mesaros, A., Heittola, T., Eronen, A., Virtanen, T., “Acoustic event detection in real life recordings”, In: 18th European Signal Processing Conference (EUSIPCO-2010), pp. 1267 —1271, Aalborg, Denmark, August 23-27, ISSN 2076-1465, 2010.

[7] Brüel & Kjær Environment Management Solutions (formerly Lochard) available online on www.lochard.com (update 11.3.2011)

[8] K. Hume, D. Terranova, C. Thomas, “Complaints and annoyance caused by aircraft operations - Temporal patterns and individual bias”, In: Noise and Health, Volume 4, Issue 15, April 2002, pp. 45–55, ISSN: 14631741, 2002.

[9] K. Hume, M. Gregg, C. Thomas, D. Terranova, “Complaints caused by aircraft operations: an assessment of annoyance by noise level and time of day”, In: Journal of Air Transport Management, Volume 9, Issue 3, May 2003, pp. 153-160, Elsevier, ISSN: 09696997, 2003.

[10] Transcriber – annotation tool: http://trans.sourceforge.net

[11] NIST SCLITE scoring toolkit: http://www.itl.nist.gov/iad/mig/tools/

[12] Audacity - Free Sound Editor and Recording Software: http://audacity.sourceforge.net/

[13] ELAN tool: http://www.lat-mpi.eu/tools/elan/

[14] E. Vozarikova, M. Pleva, J. Vavrek, S. Ondas, J. Juhar, A. Cizmar, “Detection and classification of audio events in noisy environment”, In JCSCS - Journal of Computer Science and Control Systems, Volume 3, Number 1, University of Oradea Publisher, ISSN: 1844-6043, 2010.

[15] M. Pleva, E. Vozarikova, S. Ondas, J. Juhar, A. Cizmar, ”Automatic detection of audio events indicating threats”, In: MCSS 2010, Multimedia Communications, Services and Security, IEEE international conference, Krakow, May 6-7, 2010, AGH, ISBN 978-83-88309-92-2, 2010.

[16] J. Vavrek, M. Pleva, J. Juhar, ”Acoustic events detection with support vector machines”, In: Electrical Engineering and Informatics, Proceeding of the Faculty of Electrical Engineering and Informatics of the Technical University of Košice, September, 2010, Kosice, pp. 796-801, ISBN 978-80-553-0460-1, 2010.

[17] E. Vozarikova, J. Juhar, A. Cizmar, “Acoustic events detection using MFCC and MPEG-7 descriptors”, In: MCSS 2011, Multimedia Communications, Services and Security, Krakow, June 2-3, 2011, AGH, Springer CCIS series in press, ISSN: 1865-0929, 2011.