Florence, Italy 12-16 July 2015
ICSV22
The 22nd International Congress on Sound and Vibration

DEVELOPMENT OF AN ANOMALOUS NOISE EVENT DETECTION ALGORITHM FOR DYNAMIC ROAD TRAFFIC NOISE MAPPING

Joan Claudi Socoró, Gerard Ribera, Xavier Sevillano, Francesc Alías

GTM – Grup de Recerca en Tecnologies Mèdia, La Salle – Universitat Ramon Llull, C/Quatre Camins, 30. 08022 Barcelona, Spain, e-mail: jclaudi/si20753/falias/xavis@salleurl.edu

Dynamic road traffic noise maps should display, in real time, the noise levels generated by road infrastructures measured by the sensors located on the road. For this reason, any acoustic event produced by another source that could alter the measured noise levels (e.g. an aircraft flying over, nearby railways, church bells, crickets, etc.) should be detected and eliminated from the map computation to provide a reliable picture of the actual road noise impact. To that end, it becomes necessary to devise strategies to automatically identify anomalous noise events captured by the network of sensors. This work describes a first version of the anomalous noise event detection algorithm designed in the LIFE DYNAMAP project. The proposed algorithm follows a "detection-by-classification" approach based on a semi-supervised two-class classifier that does not require training with on-site collected "anomalous noise events" samples, thus being location-independent. Instead, it optimizes a decision threshold based on distance distributions with respect to the predominant "road traffic noise" class to maximize detection accuracy. The experimental results reveal that our proposal outperforms the baseline two-class supervised detector especially in scenarios in which anomalous events show higher noise levels and, thus, are more likely to alter the levels represented in dynamic road traffic noise maps.

1. Introduction

Traffic noise is one of the multiple sound sources that, especially in urban environments, cause a negative impact on the quality of life of the population [1]. In fact, environmental noise has been found to cause harmful health effects, for instance, being highly correlated with cardiovascular diseases such as myocardial infarction and hypertension [2]. In order to face this situation, the European Commission promoted the Environmental Noise Directive 2002/49/EC (END) with the aim of informing citizens about their exposure to noise and drawing up appropriate action plans to prevent the harmful effects derived from noise exposure [3]. The reporting of acoustic levels caused by different sources (among which traffic noise is a relevant one) by means of noise maps is one strategic action plan to promote policies based on END. These noise maps usually have been implemented by computing the averaged noise levels during one year and being reviewed (and revised, if needed) once every five years [3]. However, the implementation of dynamic noise maps that reflect more precise data about harmful sound sources in real time not only allows more detailed assessments,

but also checking the effectiveness of the conducted noise actions plans in short term, among other advantages. Following this goal, acoustic sensor networks that monitor the noise levels in urban areas have been recently proposed (e.g. see [4,5]).

The system proposed in this work locates within the European LIFE DYNAMAP project (Dynamic Acoustic Mapping - Development of low cost sensors networks for real time noise mapping). This project aims at easing the update of noise maps and at reducing their economic impact through the development of an automatic and integrated system for data acquisition and processing capable of detecting and representing, in real time, the acoustic impact due to road infrastructures by means of dynamic noise maps. The DYNAMAP system is composed of networks of low cost sensors measuring the sound pressure levels emitted by the traffic noise sources and of a software tool based on a geographic information system (GIS) platform performing real time noise maps updating. To increase its robustness, the DYNAMAP system will include an anomalous noise event detection (ANED) algorithm to avoid that non-traffic acoustic events alter the noise maps scaling.

The key contribution of this paper is a location-independent ANED algorithm based on a two-class ("road traffic noise", or RTN, vs. "anomalous noise event", or ANE) "detection-by-classification" approach. In a road traffic noise monitoring context like the DYNAMAP project, the presence of anomalous noise events can be *i)* highly local (e.g. sensors located in roads near airports will often capture aircraft noise while others will rarely be affected by this type of noise), *ii)* unpredictable and highly diverse (e.g. ambulance sirens or thundering), and *iii)* little likely to occur (e.g. a bird or a cricket that approaches the sensor). For these reasons, collecting a sufficient number of anomalous noise events samples that represent this high diversity of possible noise sources to accurately train the classifier would require a great effort. To circumvent this inconvenience, we propose a location-independent ANED algorithm based on a semi-supervised approach that avoids creating acoustic models for the minority ANE class. Instead, we employ distance-based classifiers and optimize a decision threshold based on distance distributions with respect to the predominant RTN class. In our experiments, we compare this approach to a classic two-class supervised classifier (used as a baseline) that creates acoustic models for both classes.

The algorithm has been evaluated on a dataset of synthetic mixtures of anomalous events and road traffic noise. The experimental results show that the proposed scheme outperforms the baseline detector in terms of recognition and detection rates especially in those scenarios in which ANE have higher pressure levels.

The paper is organized as follows. Section 2 describes several state-of-the-art approaches to noise event detection and recognition. Section 3 presents a description of the proposed system, and also its main implementation details: features, machine learning algorithms, and threshold decision optimization criteria. In Section 4, the conducted experiments and the obtained results are described and discussed. And finally, the Section 5 draws up the conclusions and future work.

2. Related work

Audio event detection is the task of finding the start and end points of a noise event of interest within a continuous audio stream. Focusing on the detection of environmental sound events, it is important to highlight that they are usually disconnected from one another (in contrast to what happens in speech or music, which present a strongly interconnected temporal "structure" composed of phonemes or notes, respectively [6]). For this reason, the detection of environmental sound events in a continuous audio stream is typically tackled by one of the two following approaches.

The first approach is based on using a novelty-detection system that considers any rapid change against the long-term background noise to be a sound event. This type of approach is usually referred to in the literature as "detection-and-classification". The second alternative consists in using a

sliding window detector that performs classification on each fixed-length segment in turn. This approach is commonly referred to as "detection-by-classification" [7].

The former type of approaches to audio event detection (i.e. "detection-and-classification") do not require being trained on labeled data, which makes them very adaptable to new auditory environments. One the first relevant examples of this type of audio event detection systems was presented in [8], based on the idea that abrupt changes in sound usually indicate a new event has occurred. To detect such abrupt changes, the authors computed a time series of temporal and frequency feature vectors over the audio stream, using the Mahalanobis distance to compare successive frames.

To compare adjacent frames in a more robust fashion, other metrics such as cross-correlation and energy spline interpolation were introduced [9]. Later, the same authors proposed the use of transient models based on dyadic trees of wavelet coefficients to clearly detect impulse noise events [10].

Lately, "detection-and-classification" approaches have shifted towards the use of sequential hypothesis testing [11]. For instance, Dessein and Colt [12] applied these techniques to real-time audio segmentation using the information geometry of exponential families. In [13], the authors showed how segmentations and similarities between neighboring frames can be computed in an information-geometric context by finding the centroids for each audio segment.

A very recent approach to audio event detection is based on the application of unsupervised learning techniques, such as clustering [14]. In this approach, the expressiveness of the model is exploited to discover the correct segmentation. To that end, several online learning algorithms are developed to apply Hidden Markov or semi-Markov Models based on incremental optimization schemes to the audio segmentation task.

On the other hand, the "detection-by-classification" approach performs classification of sequential audio segments, where the detection window shifts forwards over time. The output at each step is then a decision between noise and one of the trained sound events. Thus, here we need a classifier trained to detect the noise events of interest. According to [6], the advantage of this approach is that only one set of features needs to be extracted from the audio as the detection and classification modules are combined. The main disadvantage lies in choosing an appropriate window size and classification method capable of working well across a range of experimental conditions.

A relevant work in this type of detectors is the two-stage audio event detection system based on Support Vector Machine (SVM) classifiers described in [7]. In the first stage, silence/non-silence segmentation is performed. In the second stage, the sound occurring in the non-silence segments are classified into a series of predefined classes by means of SVM classifiers. The use of Hidden Markov Models (HMM) applied to audio event detection was the focus of [15]. In that work, the authors use the Kullback-Leibler distance to quantify the discriminant capability of speech feature components in acoustic event detection. Based on these distances, they use AdaBoost to select a discriminant feature set. More recently, Zhuang et al. proposed extracting discriminative features for audio event detection using a boosting approach [16], leveraging statistical models (a tandem connectionist-HMM plus an SVM–GMM-supervector approach) that better fit the audio event detection task. The use of part-based decompositions of the incoming audio stream is another recent approach to audio event detection. In [17], the authors proposed an approach to detect and model acoustic events that directly describes temporal context, using convolutive non-negative matrix factorization (NMF). The recent work by Schroder et al. on audio event detection has covered several alternatives of the "detection-by-classification" approach to tackle this task. For instance, the authors proposed in [18] an acoustic event detection system consisting of a noise reduction signal enhancement step, a Gabor filterbank feature extraction stage and a two layer HMM as back-end classifier.

The anomalous noise event detection presented in this work belongs to the latter type of detectors, i.e. "detection-by-classification". The next section presents a detailed description of its rationale and components.

3. System description

The ANED algorithm designed to automatically detect anomalous noise events follows a pattern recognition approach divided into two main steps: signal feature extraction and recognition. The recognition stage is tackled by supervised machine learning techniques. This requires training the system with noise samples with their corresponding labels in order to build acoustic models that allow recognizing different noise classes.

In the context of our problem, it would be sufficient to train the classifier with N=2 noise categories, as our goal is to detect the presence of noise events other than road traffic noise. Figure 1 depicts the block diagram of a generic 2-way "detection-by-classification" system, referred to as the *baseline* detector hereafter. Notice that two phases are envisaged: *i)* a *training+validation* phase, in which, firstly, one acoustic model per class is built after the parameterization of labelled *training* data (through windowing and feature extraction); and secondly, internal parameters of the classifier are tuned using labelled *validation* data of both classes; *ii)* an *operation* phase, in which the classifier assigns one of the two possible noise class labels to each frame of an unknown noise signal.

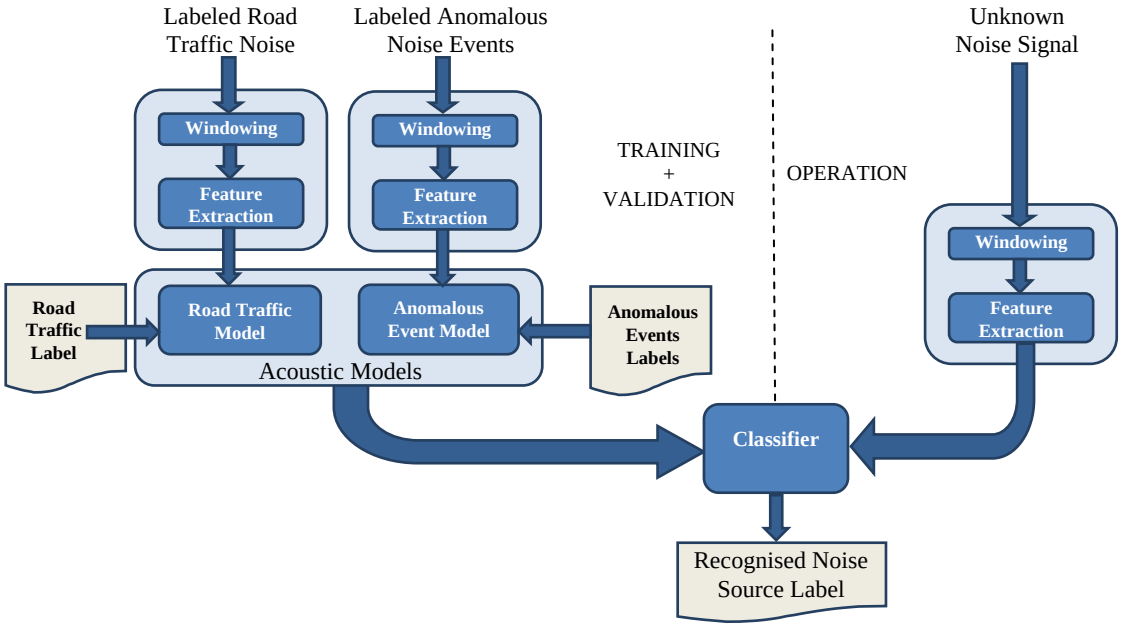

Figure 1. Block diagram of a baseline 2-way "detection-by-classification" system based on supervised learning.

Since DYNAMAP acoustic monitoring stations will be placed in a fixed location, it would be possible to train the ANED algorithm accurately by collecting a sufficient number of samples of both noise classes at each sensor's location. However, the highly local, occasional, diverse and unpredictable nature of most types of anomalous noise events makes sample collection a repetitive, difficult and burdensome task. For this reason, in this work we propose a location-independent semi-supervised 2-way "detection-by-classification" ANED system that minimizes the need for anomalous noise events samples collection, avoiding training with on-site collected noise samples.

3.1 Proposed ANED algorithm

Figure 2 shows the block diagram of the proposed anomalous noise event detection system. One of the main differences with regard to the baseline system depicted in Figure 1 is that anomalous events are now used only for adjusting a decision threshold, and no acoustic model is built for this class.

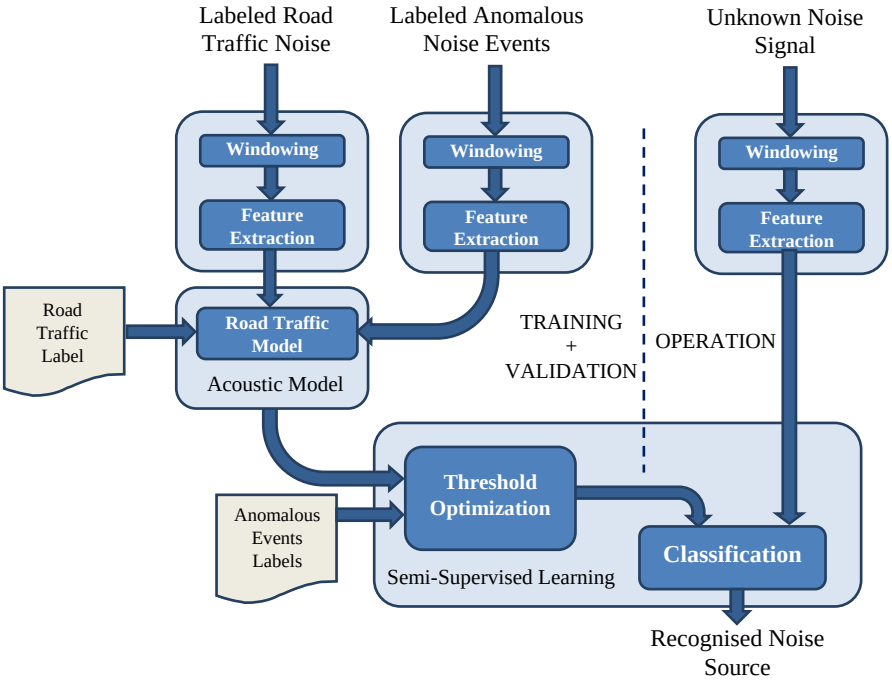

Figure 2. Block diagram of the proposed anomalous noise event detection system.

From an operational perspective, the proposed system uses *training* data corresponding to road traffic noise during the training phase to build the acoustic model of the RTN class. Next, a *validation* data set containing samples of both the RTN and ANE classes is employed to adjust the threshold that allows measuring the proximity of the signals of both classes to the learned RTN acoustic model. Although this decision threshold could be fixed heuristically, a more precise value can be obtained by a simple analysis of the two-class (RTN and ANE) distances distributions, obtained from the classifiers responsible for the detection. Finally, the system enters the *operation* mode, in which unseen noise samples are classified either as RTN or ANE.

The following paragraphs describe the constituting modules and implementation details of the proposed ANED system, namely: signal parameterization, classification algorithm and decision threshold optimization.

3.2 Signal parameterization

The signal feature extraction block of the ANED algorithm parameterises the noise signal by means of a fixed-size set of coefficients that model the spectro-temporal characteristics of the noise signals. To this end, first the input signal is segmented into 30 ms frames using a Hanning window. Subsequently, a feature set is extracted from each signal frame. In this work, we have selected the biologically-inspired Gammatone Cesptral Coefficients (GTCC), which have recently shown an improved performance in environmental sound recognition tasks [19]. Also, Mel-Frequency Cepstral Coefficients (MFCC) are compared to GTCC in the experiments for being a classic baseline benchmark. The number of computed coefficients is 13 for both GTCC and MFCC.

3.3 Supervised classification

In this work we consider two simple but effective classification techniques: K-Nearest Neighbour (KNN) and Fisher's Linear Discriminant (FLD). The choice of these two classification techniques is motivated by the fact that they both provide a certain distance measure that can be interpreted as a measure of similarity between the dominant class (i.e. RTN) and the input noise

frame, which will be the base for adjusting the detection decision threshold. For KNN, this is the distance between the input noise frame and the K closest training examples. In case of FLD, this measure corresponds to an estimation of the log probability that road traffic noise is the source of the input frame (thus, a value close to 0 shows high similarity to this class while high negative values show that the input could be an anomalous event).

As for the internal configuration of the KNN classifier, the validation set allowed deciding that the KNN would employ the Euclidean distance metric and consider the K=3 nearest neighbours.

3.4 Decision threshold optimization

The proposed technique for setting the decision threshold is based on obtaining an equally minimum value of both type I and type II errors (false positives and false negatives), as in [20], where the same criteria was adopted with the aim of obtaining an optimal speaker verification system. This threshold adjustment is performed by using samples from the validation dataset (see Section 4.1).

4. Experiments

4.1 Audio database

The audio database used for the experiments consists of real road traffic noise (RTN) recordings of the ring road surrounding the city of Barcelona, synthetically mixed with anomalous noise events (ANE) samples (containing up to 15 noise types, like horns, ambulance sirens, car collisions, church bells, birds, crickets, rain, thunders, etc.) gathered from free online repositories. Road traffic noise free field recordings were obtained using the Brüel & Kjaer 2250 sonometer, with 48 KHz sampling rate, 4.2 Hz - 22.4 kHz broadband linear frequency range and its own microphone (Type 4189). The total length of sound samples used for training is 250 seconds of RTN and 300 seconds of ANE. Finally, to test the system in scenarios in which ANE had different degrees of relevance, the level of each type of noise in the mixtures was adjusted to obtain RTN-to-ANE ratios of -6 and -12 dB.

4.2 Baseline techniques and experimental setup

The evaluation process is performed following a 4-fold cross validation scheme. In each repetition, *training* + *validation* and *test* subsets are changed so as to obtain statistically reliable results. As regards the baseline detector, *training+validation* data (75% of the total available data) contains both classes (RTN and ANE). In contrast, in the proposed ANED algorithm, *training* data (12.5% of the total available data) contains only RTN class, while the *validation* set (12.5% of the data) contains RTN and ANE samples.

4.3 Results

Table 1 presents the results of the conducted experiments in terms of two evaluation metrics: *i)* the F1 measure of the ANE class ($F1_{ANE}$), computed as the harmonic mean between *precision* (ratio between the true positives and the total amount of frames classified as ANE) and *recall* (ratio between the true positives and the total amount of true ANE frames); *ii)* the total classification rate R in % (averaged percentage of the testing samples correctly classified).

The performance of the proposed ANED algorithm is compared to that of the baseline detector for RTN-to-ANE ratios of -6 and -12 dB, using the FLD and KNN classifiers, as well as GTCC and MFCC features.

4.4 Discussion

It can be observed from Table 1 that the proposed ANED algorithm outperforms the baseline detector in terms of both evaluation metrics in most cases. Specifically, the proposed method attains better results in five of the eight experimental scenarios. In particular, the best recognition accuracy

Table 1. Results of the conducted experiments. The best results for each combination of RTN-to-ANE vs. Classifier vs. Features are highlighted in boldface.

		RTN-to-ANE = -12 dB				RTN-to-ANE = -6 dB			
		FLD		KNN		FLD		KNN	
		GTCC	MFCC	GTCC	MFCC	GTCC	MFCC	GTCC	MFCC
Baseline	$F1_{ANE}$	0,7877	0,7990	0,8305	**0,8266**	0.6983	**0.8233**	0.7738	**0.8478**
	R (%)	85.25	85.62	87.26	**86.42**	77.43	84.58	**83.77**	**87.12**
ANED	$F1_{ANE}$	**0,8976**	**0,8397**	**0,8440**	0,7710	**0.8252**	0.7906	**0.7810**	0.6718
	R (%)	**91.46**	**87.22**	**87.70**	83.35	**84.56**	**84.79**	79.97	76.82

and $F1_{ANE}$ are obtained for the configuration that considers using the proposed ANED scheme, FLD classifier, GTCC features and RTN-to-ANE = -12 dB (91.46% of recognition rate and 0.8976 value of $F1_{ANE}$).

It is important to note that the ANED performs better than the baseline detector especially in the most adverse scenario as regards the presence of anomalous noise events (RTN-to-ANE = -12 dB), which is particularly important in the context of the DYNAMAP system.

5. Conclusions

In this paper, a new strategy for conducting anomalous noise event detection in road traffic noise monitoring systems has been proposed. The technique is based on a distance-based classifier trained with RTN samples, and a subsequent decision stage in which a threshold is optimized by using both RTN and ANE samples. Preliminary experiments have been conducted using synthetic mixtures of RTN recordings and ANE samples obtained from the Internet, with the aim of validating the viability of the proposed approach. It is worth mentioning that the ANE dataset used for training and validation is small, especially considering the high diversity of possible non-traffic noise sources. The obtained results confirm that the proposed ANED method can outperform the baseline "detection-by-classification" algorithm in most of the simulated scenarios, especially in those with lower RTN-to-ANE ratio (i.e. those situations in which anomalous noise events are more likely to distort the pressure levels represented in the dynamic traffic noise maps, which is the case we need to address more specifically).

Further research will be oriented towards several directions, such as the exploitation of the temporal dynamics of the detection to increase its robustness, the evaluation of different system setups (e.g. different features sets and classifiers), the exploration of early and late fusion schemes, and the investigation of how the proposed location-independent approach can be adapted and optimized to perform in specific acoustic environments.

Acknowledgements

This research has been partially funded by the European Commission under project LIFE DYNAMAP ENV/IT/001254.

6. Bibliography

1 Babisch, W. Transportation noise and cardiovascular risk: Updated review and synthesis of epidemiological studies. *Noise&Health*, **8**(30): 1-29, (2006).

2 Hellmuth, T., Classen, T., Kim, R., Kephalopoulos, S. : *Methodological guidance for estimating the burden of disease from environmental noise*. The World Health Organization (2012).

3 *EU Directive: Directive 2002/49/EC of the European parliament and the Council of 25 June 2002 relating to the assessment and management of environmental noise*. Official Journal of the European Communities, L 189/12, July 2002.

4 Nencini, L., De Rosa, P., Ascari, E., Vinci, B., Alexeeva, N.: SENSEable Pisa – a wireless sensor network for real-time noise mapping. *Proc. Euronoise'12*, Prague (2012).

5 Valero, X., Nencini, L., Alías, F., Vinci, B.: Feasibility of automatic noise source recognition in collaborative wireless sensor networks. *Proc. AIA-DAGA*, Meran, Italy (2013).

6 Dennis J.W. *Sound Event Recognition in Unstructured Environments using Spectrogram Image Processing*, PhD thesis, School of Computer Engineering - Nanyang Technological University, Singapore, (2014).

7 Temko A. *Acoustic Event Detection and Classification*, PhD thesis, Universitat Politècnica de Catalunya, Spain, (2007).

8 Tzanetakis G., Cook, P. Multifeature Audio Segmentation For Browsing and Annotation, *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, (1999).

9 Vacher M., Istrate, D., Besacier, L., Serignat, J., Castelli, E. Life Sounds Extraction and Classification in Noisy Environment, *Proc. IASTED International Conference on Signal & Image Processing*, (2003).

10 Vacher M., Istrate, D., Serignat, J.F., Gac, N. Detection and Speech/Sound Segmentation in a Smart Room Environment, *Proc. International Conference on Speech Technology and Human-Computer Dialogue*, (2005).

11 Tartakovsky A., Nikiforov, I., Basseville, M. Sequential Analysis: Hypothesis Testing and Changepoint Detection, *Monographs on Statistics & Applied Probability*, **136** (2014).

12 Dessein A., Cont, A. An information-geometric approach to real-time audio segmentation, *IEEE Signal Processing Letters*, **20**(4):331–334 (2013).

13 Cont A., Dubnov, S., Assayag, G. On the information geometry of audio streams with applications to similarity computing *IEEE Transactions on Audio, Speech, and Language Processing*, **19**(4):837–846, (2011).

14 Bietti A. Online learning for audio clustering and segmentation, Technical Report HAL Id: hal-01064672 / IRCAM, (2014).

15 Zhou X., Zhuang, X., Liu, M., Tang, H., Hasegawa-Johnson, M., Huang, T. HMM-based acoustic event detection with AdaBoost feature selection. *Proc. Classification of Events, Activities and Relationships Evaluation and Workshop*, pp. 345–353, (2007).

16 Zhuang X., Zhou, X., Hasegawa-Johnson, M.A., Huang, T.S. Real-world acoustic event detection. *Pattern Recognition Letters*, **31**:1543-1551, (2010).

17 Cotton C.V., Ellis, D. Spectral vs. Spectro-temporal features for acoustic event detection. *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, (2011).

18 Schroder J., Cauchi, B., Schadler, M.R., Moritz, N., Adiloglu, K., Anemuller, A., Doclo, S., Kollmeier, B., Goetze, S. Acoustic Event Detection Using Signal Enhancement and Spectro-Temporal Feature Extraction. *Proc. IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events*, (2013).

19 Valero, X., Alías, F. Gammatone Cepstral Coefficients- Biologically Inspired Features for Non-Speech Audio Classification, *IEEE Transactions on Multimedia*, **14**(6):1684-1689, (2012).

20 Furui, S. Cepstral analysis technique for automatic speaker verification, *IEEE Transactions on Acoustics, Speech and Signal Processing*, **29**(2)254-272, (1981).