Automatic Detection and Classification of Wideband Acoustic Signals*

Alain Dufaux, Laurent Besacier, Michael Ansorge, and Fausto Pellandini

Institute of Microtechnology, University of Neuchâtel,
Rue A.-L. Breguet 2, CH-2000 Neuchâtel, Switzerland
(alain.dufaux,laurent.besacier)@imt.unine.ch

1 INTRODUCTION

Until now, very few contributions were published for wideband acoustic signal detection and recognition. However, this field of activities offers concrete potentialities in the next areas :

– *Surveillance and security applications,*
for alarm verification / validation using audio verification techniques, in particular for intrusion verification (e.g. in bank offices, stores, and private homes), or for the supervision of public sites (e.g. parking garages, with detection of person aggressions, etc.). A further application concerns alarm detection and recognition, relying on noise discrimination algorithms for the automated recognition of alarm conditions and triggering of surveillance installations.

– *Telematics applications for disabled and elderly persons,*
in view of providing assistance to disabled and elderly persons affected in their hearing capabilities, by catching their attention when needed, in particular in case of danger, and helping them to keep / recover some independence in their daily occupations. The concept is in detecting acoustic alarm signals encountered either indoors or outdoors (e.g. door ring, telephone ring, car/truck horn, tram bell, siren), and to inform the interested person by tactile or/and visual means that an acoustic alarm signal of a given class is active.

This paper particularly deals with the detection and recognition of sounds related to surveillance applications. In this case, impulsive noise signals such as glass breaking, detonations, or door slams, are considered, where the signals are strongly non-stationary and composed of higher frequency components. The remainder of the paper is structured as follows. An overview of the detection/recognition system is presented in *Section 2*, whereas *Section 3* describes the used sounds database. The detection and recognition algorithms are then discussed in *Sections 4* and *5*, respectively, where the latter can be efficiently tackled using pattern recognition methods relying either on Bayes classifiers, or on artificial Neural Networks (NN). After comparison of the achieved performance, *Section 6* draws the conclusions along with a sketch of future work.

2 SYSTEM OVERVIEW

The surveillance system (*Fig. 1*) is made up of a microphone, which records the sound activity on-line. Once the detection module finds discontinuities or anomalies in the input signal, the recognition module is activated. A time-frequency analysis of the signal is performed, and the class of the detected alarm is determined after comparison with different sounds of a database. Adequate human intervention can then be undertaken according to the automatic system verdict.

---

* Acknowledgments: This work was supported by the Swiss National Science Foundation under Grant FN 20-53'843.98 , and by Cerberus AG, Männedorf / CH.

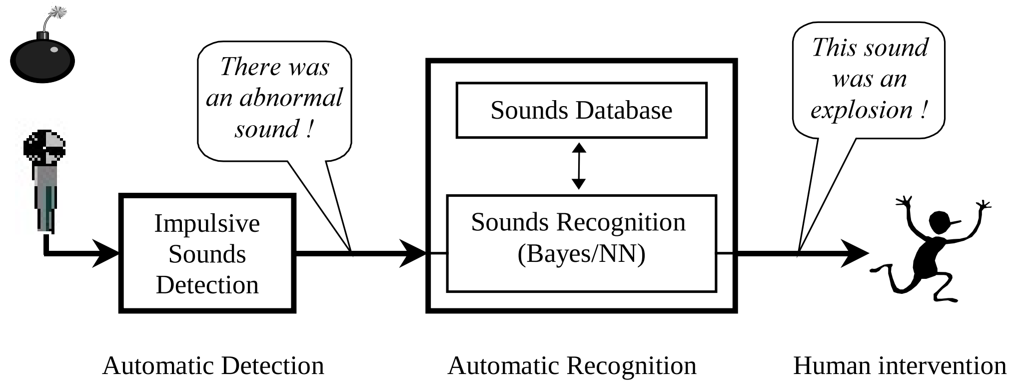

Figure 1 : Overview of the surveillance system

3 SOUNDS DATABASE

The database used for the experiments reported in this paper contains 3 different sound classes, namely : 69 door slams, 38 glass breaks and 58 explosions. Each signal has a duration of approximately one second, and a signal-to-noise ratio of 40 dB. The door slams and glass breaks were digitally recorded with a *Sony* Digital Audio Tape recorder, a *Sony ECM-MS907* microphone, and a digital I/O PC audio card. The explosion sounds were taken from different CDs like the “BBC sound FX library” or the “Digiffects” library, available at [1].

4 IMPULSIVE SOUND DETECTION

The detection module involves a median filter analyzing the energy variations in the 44.1 kHz-sampled input signal, with the effect of selectively amplifying the pulses occurring in the temporal energy sequence (*Fig. 2a*). An adaptive thresholding – depending on the standard deviation of a past long-term windowed energy sequence – is then applied. This method provides a very precise detection scheme for impulsive signals, where the pulses can be detected under quite adverse background noise conditions, with a signal-to-noise ratio becoming as low as –10 dB. *Fig. 2b* shows the achieved performance evaluated on the impulsive sounds database for a variable level of gaussian white noise.

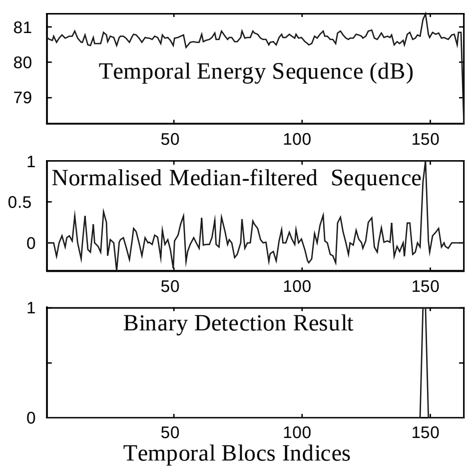

Figure 2a : Detection of an impulse, placed in a –8 dB SNR random white noise background environment

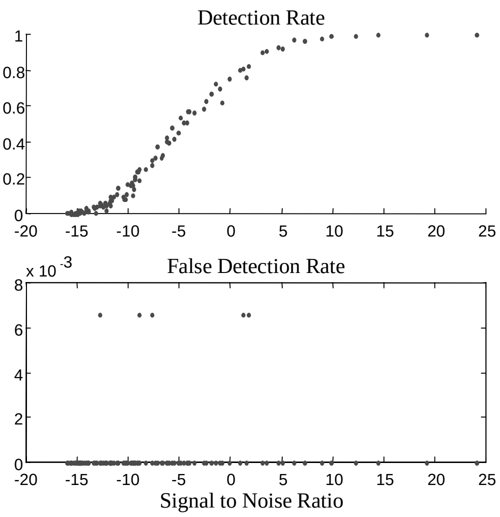

Figure 2b : Performance of the impulsive sound detection algorithm, evaluated on the whole database

5 IMPULSIVE SOUND RECOGNITION

5.1 Impulsive Sound Analysis

The first step of the recognition algorithm consists in an analysis of the signal to be classified, in view of extracting some typical features. In this work, a time-frequency analysis of the signal was selected, so as to take the non-stationary properties of impulsive sounds into account. Each second, an energy spectrogram of the signal is calculated, with a temporal resolution of 50 ms, whereas the frequency domain covering the range from 0 to 20 kHz is uniformly decomposed into 5 bands of 4 kHz each for every temporal frame. Typical features vectors are then achieved from the concatenation of these spectrogram energy values, in order to exploit dynamic temporal information. The redundancy induced by this dynamic modeling requires a reduction of the feature space dimension, performed with a conventional Principal Component Analysis (PCA). Six coefficients are finally kept after the parameter domain reduction (*Fig. 3*).

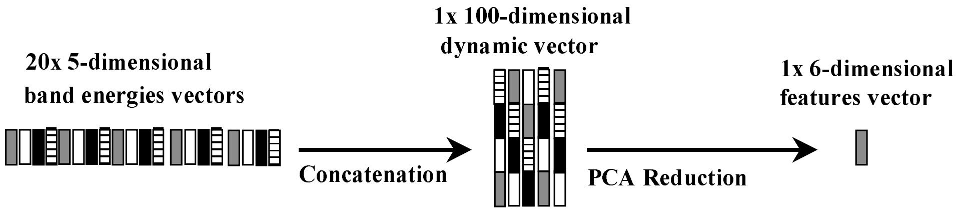

Figure 3 : Features reduction by Principal Component Analysis (PCA)

5.2 Recognition Methods

5.2.1 Bayes Classifier

The classical pattern recognition method (Bayes statistical classifier) assumes that each feature probability density function is a gaussian distribution. At the training stage, the maximum likelihood method [2] is used to determine for each class the mean and covariance parameters of the multidimensional gaussian. The classification of a signal is then performed by computing its “a posteriori probability” to belong to the different classes. Obviously, the signal will be attributed to the class characterized by the highest a posteriori probability. It is noticed that the statistical functions provided in the *Pattern Recognition Toolbox for Matlab* [3] were used both during the training phase to estimate the gaussian models of each class, and during the signal classification phase to determine the class featuring the maximum probability.

5.2.2 Artificial Neural Network Classifier

A 3-layered Perceptron feed-forward neural network is used in this case. The number of input neurons corresponds to the features dimension, while the number of neurons in the output layer fits the number of considered classes. Experimentally, 30 neurons in the hidden layer seems to be a good quality/complexity trade-off, so that the network architecture is 6:30:3 (*Fig. 4*). The units of the network are fully connected to ensure that the network is able to learn and tackle complex non-linear decision surfaces. Training is made with the backpropagation algorithm. Mathematically, backpropagation is a gradient descent algorithm, which minimises the recognition mean squared error to select the optimal weights and biases. The network is designed with the *Matlab’s Neural Network Toolbox* [4].

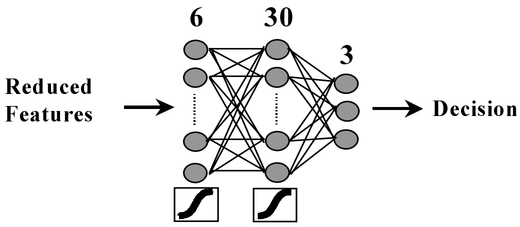

Figure 4 : Feed-forward 3-layer perceptron neural network architecture

5.3 Experiment Protocol

34 door slams, 19 glass breaks and 29 explosion sounds were used for the training of the class models, whereas 35 door slams, 13 glass breaks and 29 explosions were used for the testing of the algorithms. The signals dispatched between training and testing were randomly chosen from the database, and many iterations of the protocol were made, to produce as many training/test configurations as possible (cross-validation). A total of 60'000 tests were made for each tested classifier.

5.4 Recognition Performance

Bayes and NN classifiers were tested for different signal-to-noise ratios (white noise perturbations) of the sound signals. Training was made with the corresponding noise levels, so that the SNR was supposed to be *a priori* known during the test phase. *Tab. 1* shows recognition results obtained with each method for the three-class problem. Both methods are featuring a similar performance, but the NN classifier seems to be slightly better, particularly when the background noise level increases.

Signal-to-noise Ratio (dB)	Rec. Rate (%) Bayes Classifier	Rec. Rate (%) NN Classifier
No perturbations	98.5	99.2
40	97.2	97.9
20	96.0	96.7
10	94.9	96.3
1	93.1	95.0

Table 1 : Recognition performance of Bayes and NN algorithms (60'000 tests)

6 CONCLUSION AND FUTURE WORK

As shown in this paper, the problem of classifying wideband impulsive sounds can be efficiently solved, at least for a low number of sound classes. In future work, an increasing number of classes will be handled, in particular under unknown background environment conditions, requiring other techniques such as unsupervised learning and robust recognition methods, e.g. Gaussian Mixtures Models, Hidden Markov Models, or Statistical/Neural Hybrid Structures.

7 REFERENCES

[1] Leonardo Software, [http://www.leonardosoft.com], Santa Monica, CA 90401.
[2] C. Couvreur, "*Environmental Sound Recognition : a Statistical Approach"*, PhD thesis, Faculté Polytechnique de Mons, Belgium, June 1997.
[3] Ahlea Systems Corporation, "*Pattern Recognition Toolbox*", [http://www.ahlea.com], Nepean, Canada, 1996.
[4] H. Demuth, M. Beal, *Neural Network Toolbox for Use with Matlab,* Fifth printing, Version 3 – Jan. 1998.