Conference & Workshop on Assistive Technologies for People with Vision & Hearing Impairments
Assistive Technology for All Ages
CVHI 2007, M.A. Hersh (ed.)

HOUSEHOLD SOUND IDENTIFICATION SYSTEM FOR PEOPLE WITH HEARING DISABILITIES

H. Lozano, I. Hernáez, E. Navas, FJ. González, I. Idígoras

ABLE Technologies Group, INNOVA Department,
Robotiker - Tecnalia, Vizcaya, Spain.
Tel: +34 94 600 22 66. Fax: +34 94 600 22 99. Email: hlozano@robotiker.es

Abstract: This article concerns the classification of household sounds for the development of an application to help people with hearing disabilities to resolve everyday situations which may present them with serious problems: telephone calls, the sound of the door bell, an alarm clock, the completion of a domestic appliance programme, etc.
It presents the study carried out in order to obtain the best acoustical parameters by adopting training and evaluation techniques using GMM models with a varying number of Gaussian components.
Parameters such as MFCCs, ZCR, Roll-Off Point and Spectral Centroid are tested using the classifier.

Keywords: Sound Classification, Signal Processing, Deaf, Hearing Impaired systems

1. Introduction

A fire alarm or a telephone call at an ungodly hour warning of danger can be sounds which are perceived to be urgent when they are detected. Hearing problems or simply a high level of noise are factors which sometimes make it difficult for the human ear to identify these sounds.
Deaf people experience the issues that stem from not being able to detect or identify sounds on a daily basis. Studying the techniques and algorithms which enable this task to be performed automatically, not simply based on overcoming an intensity threshold like the majority of products available on the market, is being viewed as significant technological progress which will offer huge benefits to people with hearing disabilities.

It is about improving their autonomy and independence when performing everyday tasks which due to their limitations are often either extremely difficult or impossible. Developing an application which can detect and classify the various sounds which may emerge in a home is considered to be a fundamental requirement towards helping to improve the quality of life of people with hearing impairments.

This article presents the preliminary results of the study carried out in order to obtain a set of parameters relevant to the classification of impulsive sounds such as door bells, alarm clocks, a baby crying, with a high degree of accuracy and reliability.

In the experiments described below Gaussian Mixture Models (GMM) have been used trained with different acoustical parameters: Mel Frequency Cepstral Coefficients (MFCCs), Zero Crossing Rate (ZCR), Roll Off Point (RF) and Spectral Centroid.

2. Sound event detection and classification systems

The stages making up this kind of system can be analysed in three distinct modules (Dufaux, Besacier, Ansorge, Pellandini, 2000; Istrate, Vacher, Serignat, Castelli, 2004): sound detection, feature extraction and sound classification, as illustrated by figure 1.

Figure 1, Block Diagram of the Generic Structure of a Detection and Classification System

These modules are responsible of receiving and processing the sound with the aim of notifying the deaf person of the appropriate option everytime. This article will focus on the last module "Sound Classification" (Markou, Singh, 2003).

3. Sound classification

3.1. Sound database

In order to test and validate the recognition system, it is essential to have a sound database with representative size. In order to achieve this, part of the commercial database "Sound Scene Database in Real Acoustical Environment" (RWCP, 1998) has been used.

The tests were performed using three different types of sound (100 signals per type): telephones, alarm clocks and door locks, all obtained using a sampling frequency of 44100Hz. The selection was made on the grounds that these sounds are commonly found at home.

3.2. Selection of acoustical parameters

In order to extract the characteristics, the sounds in the DB were segmented into 20-millisecond frames prior to analysis (Clavel, Ehrette, Richard, 2005; Vacher, Istrate, Serignat, 2004).

Sound Type	Frames
Lock	1235
Alarm clock	2300
Telephone	1862

Table 1. Audio frames available for each sound type

The acoustical parameters used for the recognition were selected based on different studies (Istrate, Vacher, Serignat, Castelli, 2004; Vacher, Istrate, Besacier, Serignat, 2003) where the most common characteristics are MFCCs (common in speech recognition) as well as ZCR, Centroid and Roll-Off Point (common in the recognition of instruments and environmental sounds).

For each frame, 13 cepstral coefficients, the zero crossing rate, the Spectral Centroid and the Roll-Off Point were calculated, providing a 16-parameter vector. These parameters are briefly described below.

3.2.1. Mel Frequency Cepstral Coefficients (MFCCs): This is a perception parameter based on the FFT. After calculating the logarithm for the FFT magnitude of the signal, the bins (minimum units in the spectral domain) group together and soften in line with the Mel Frequency scale which is defined mathematically in formula 1:

$$Mel\ (f) = 2595\ \log_{10}\left(1 + \frac{f}{700}\right) \quad (1)$$

Lastly, a DCT (Discrete Cosign Transform) is performed in order to decorrelate the vector of the resulting parameters. The function defining this process is as follows:

$$c_q^{(p)} = \sum_{n=1}^{k=20} \log m_n^{(p)} \cos \frac{\pi q(2n+1)}{2K} \tag{2}$$

3.2.2. Zero Crossing Rate (ZCR): This is the number of zero crossings which occur in the analysis frame. In the signal spectrum, high (or low) frequencies imply high (or low) zero crossing rates. As such, the number of zero crossings is an indicator of the signal's high frequency content. The mathematical definition is shown in equation 3:

$$Z_t = \frac{\frac{1}{2}\sum_{n=1}^{N} \left| sign\,(x[n]) - sign\,(x[n-1]) \right|}{t} \tag{3}$$

where t is the frame length.

3.2.3. Spectral Centroid: This parameter measures how strong a sound is. The centre of gravity is assessed based on the information obtained from the Fourier Transform. It is defined as:

$$C_t = \frac{\sum_{k=1}^{N} |\, X_t[k] \,| \cdot k}{\sum_{k=1}^{N} |\, X_t[k] \,|} \tag{4}$$

where $X_t[n]$ represents the nth sample of the Fourier Transform for the t frame.

3.2.4. Roll Off Point (RF): This parameter represents the frequencies below which 85% of the energy in the audio spectrum resides. It is commonly used to identify sounds from musical instruments. In this context, the percussive sounds and note attacks usually have more energy, which is why this is a measure of the existence of abrupt signal changes. Equation 5 defines this characteristic, $M[f]$ being the energy of the signal in frequency bands higher than f. The maximum value of f is delimited by the size and the sample ratio of the acoustical band.

$$RF : \sum_{f=1}^{RF} M[f] = 0.85 * \sum_{f=1}^{N} M[f] \tag{5}$$

3.3. Classification method

There are several probabilistic classification techniques, but they are not all appropriate for recognising sounds that are not related to speech (Vacher, Istrate, Besacier, Serignat, 2003). The chosen classifier, GMM (Atrey, Maddage, Kankanhalli, 2006), is a simple model which can be described as a Hidden Markov Model (HMM) of a single state. Impulsive sounds which are difficult to separate into states are classified by observing the distribution followed by the extracted parameters. Its implementation cost is low and it offers good properties for identifying short and impulsive events. Equation 6 defines the model.

$$gm(x) = \sum_{k=1}^{K} w_k \cdot g(\mu_k, \Sigma_k)(x) \qquad \sum_{i=1}^{k} w_i = 1 \quad \forall \quad i \in \{1,.....,K\} \quad : w_i \geq 0 \tag{6}$$

An initial experiment was carried out with the aim of observing the complexity that the classification involves by only using the first two cepstral coefficients (MFCC1 and MFCC2) and the probability awarded by the Gaussian Mixture Model. Figure 2 shows how the separation performed by the classifier is reasonably high.

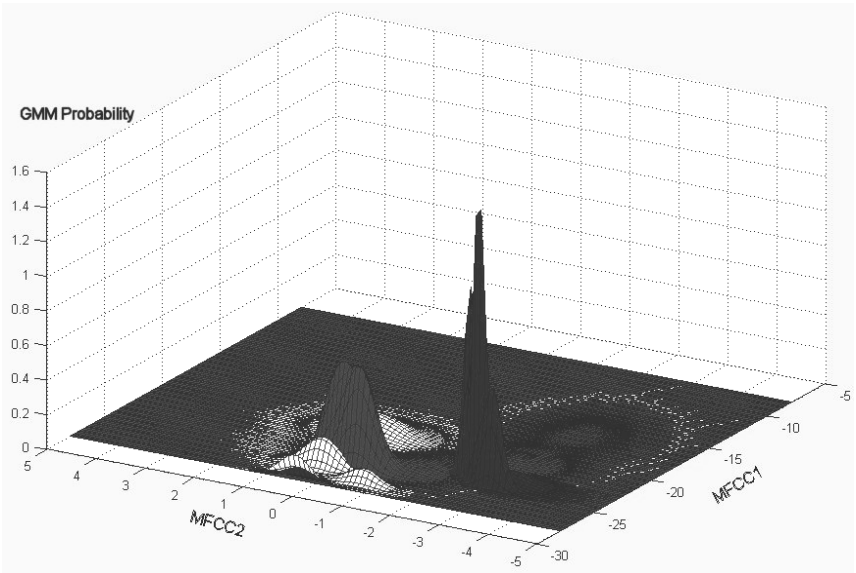

Fig. 2. GMM distribution with two cepstral parameters (MFCC1 and MFCC2)

The same test using the third and fourth coefficients, MFCC3 and MFCC4, produces different distributions with more overlapping areas, as shown by the graph in figure 3.

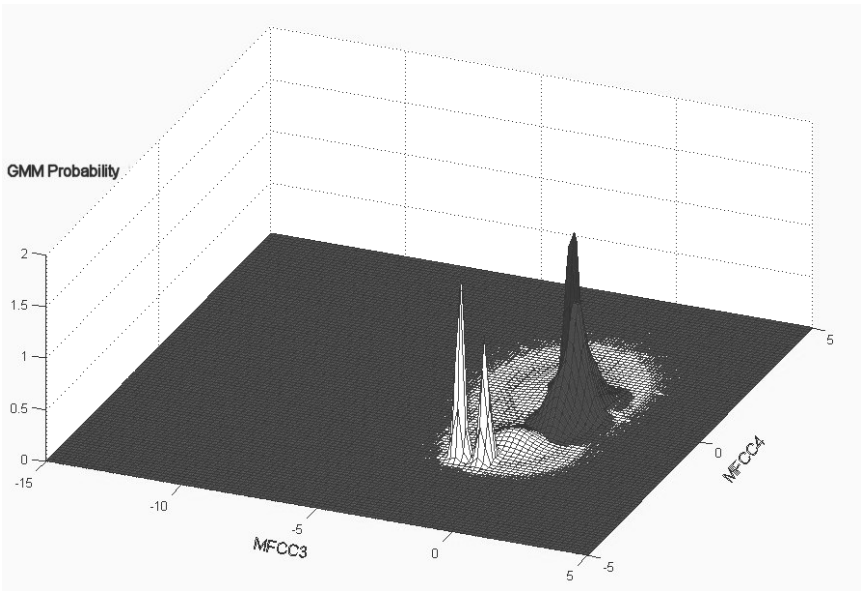

Fig. 3. GMM distribution with the third and fourth cepstral coefficients.(MFCC3 and MFCC4)

4. Experiments performed

A study has been carried out in order to obtain the optimum parameter combination, using different data sizes in training. Exhaustive tests have been performed, using mixture models of 3 to 8 Gaussians for 4 different training database sizes.

5. Results

The results obtained are presented using two measures for the success rate: by frame, calculating the average success of all frames, and by sound, by assigning each sound the most likely type according to the average success of the frames of which they are composed and by calculating the average success rate for all sounds of the same type.

The results obtained are divided into the three points below.

5.1. Influence of training database size.

Training the Gaussian Mixture Model with sizes of 66, 50, 33 and 15 sounds of each type, the number of correctly classified frames goes down slightly as the number of sounds used to train the model decreases. In terms of sound classification 100% of sounds are correctly classified in all cases. Figure 4 shows the best results.

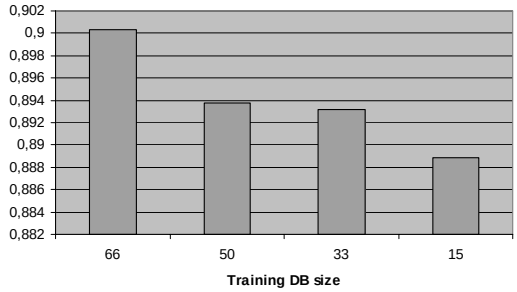

Fig. 4. Best results obtained in frame classification.

5.2. Influence of number of Gaussians

In a training session with 33 sounds of each type, varying the GMM parameter, the best results are obtained by using 5 or more Gaussians, as can be seen in figure 5. Also in this case, the percentage of correctly classified sounds is 100%.

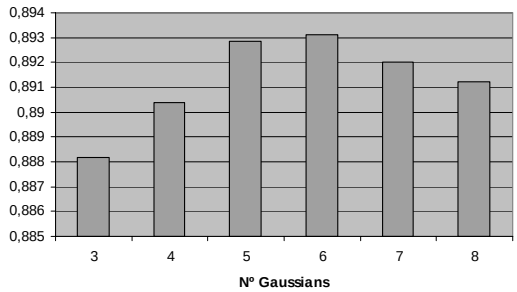

Fig. 5. Best results by number of Gaussians per frame

5.3. Relevance of parameters

Figures 6 and 7 show the relevant percentage of the different parameters. In order to measure relevance, for each test performed with the different groups of parameters, 1 to 5 points were awarded to the parameters used in the 5 combinations offering the best results (more points for a better result). The total score obtained by a specific parameter is normalised to the maximum possible score.

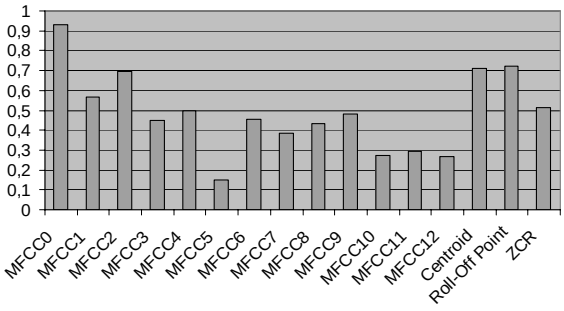

Fig. 6. Best parameters in evaluation by frame

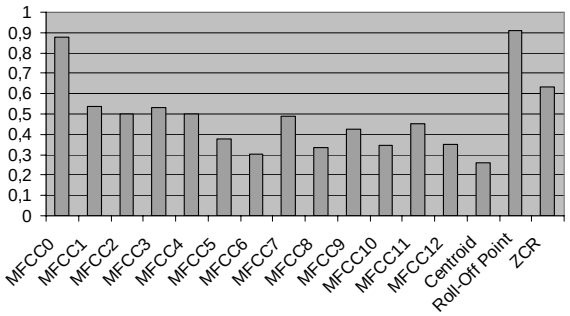

Fig. 7. Best parameters in evaluation by sound

The best result has an overall frame classification precision of 90.1% using a database containing 66 sounds, 8 Gaussians and a combination of 8 parameters: MFCC0, MFCC1, MFCC2, MFCC4, MFCC9, MFCC11, Centroid and Roll Off Point.

6. Conclusions and Future Work

At this work a first study of household sounds classification has been made. The used database has been used to evaluate the different acoustic parameters from the audio and the classification model. The used signals are clean and they don't contain any type of background noise. In the future we want to create an own database which not only will add a wider range of sounds, but it will also take into account the existence of different possible noises, such as electrical appliances, outdoor traffic noises,…

References

Atrey K., C. Maddage and S. Kankanhalli (2006). Audio based event detection for multimedia surveillance, *Proc ICCASP 2006*.

Dufaux A., L. Besacier, M. Ansorge and F. Pellandini F. (2000). Automatic sound detection and recognition for noisy environment, in *EUSIPCO 2000*, Tampere, Finland.

Istrate D., M. Vacher, J.F. Serignat and E. Castelli (2004) Multichannel smart sound sensor for perceptive spaces, *Complex Systems, Intelligence and Modern Technology Applications*, *(CSIMTA 2004)*, pp. 691-696, Cherbourg, France.

Markou M. and S. Singh (2003). Novelty detection: a review - part1: statistical approaches *Signal Processing*, vol. 83(12), p.2481-2497.

Vacher M., D. Istrate L. Besacier and J.F. Serignat (2003). Life sounds extraction and classification in noisy environment, in *Proceedings of the International Association of Science and Technology for Development, Signal and Image Processing IASTED'SIP*, Horiolulu, Hawaii, USA.

RWCP (1998). *Sound Scene Database in Real Acoustical Environments, Voice Activity Detection in Noisy Environments*, http://tosa.mri.co.jp/sounddb/nospeech/research/indexe.htm

Clavel C., T. Ehrette and G. Richard (2005) Events detection for an audio-based surveillance system, *Proceedings of the IEEE Int. Conf. on Multimedia and Expo (ICME 2005)*, Amsterdam

Vacher M., D.Istrate D. and J.F. Serignat (2004). Sound detection and classification through transient models using wavelet coefficient trees, *EUSIPCO*.

Acknowledgements: This research work is promoted by Fundacion Centros Tecnologicos – Iñaki Goenaga.