Generic Implementation of a Distress Sound Extraction System for Elder Care

Dan Istrate, Michel Vacher, Jean-François Serignat

Generic Implementation of a Distress Sound Extraction System for Elder Care

Dan Istrate, Michel Vacher and Jean François Serignat

***Abstract*— Medical remote monitoring at home is an alternative to improve the patient's comfort, to detect distress situation rapidly and reduce hospitalization costs. Physiologic and position sensors give already numerous information, but sound classification can give interesting additional information. A Real-Time implementation of a smart sound system capable of detecting and identifying sound events in noisy conditions is presented in this paper. The advantage of this implementation is the use of a generic PC station: the hardware requirements are only a sound card, a microphone and an internet link used to transmit alarm. In the case of an alarm, the information can be sent through network to a remote monitoring center and/or to a close person by email or SMS. The system is composed of 2 modules: detection and classification. The event detection module is carried out in real time in order to extract possible alarm sounds. The sound classification module is launched in a parallel task; it carries out a first segmentation between sound and speech. In the speech case, a speech recognition system is launched (not described in this paper) and in the sound case, a classification between predefined classes is carried out.**

I. INTRODUCTION

The elderly and cardiac persons need a constant monitoring in their everyday environment in order to detect rapidly a distress situation. Home telemonitoring proposes a solution to decrease the presence of a third person, a clearly expressed wish, but only in her utility aspect and not in term of human relation. The purpose of this technical solution is not to remove any human presence around the patient but just to offer an efficient remote medical monitoring. Remote monitoring preserves patient privacy because only medical information are saved in secure data bases.

Efficient detection of distress situation is the goal of various systems. Most of them aims to detect falls of elderly patient at home, since this is the cause of the most urgency cases. Such an implementation requires a reliable alarm system able to warn a medical remote monitoring center. Several approaches are proposed in the literature based on position, inclination measurements [1] and accelerometers sensors [2]. The MEDIVILLE system proposes a fusion between patient position, agitation measurements and heart rate frequency [3], [4]. Other systems use infrared sensors to obtain a patient activity information and position localization, to merge with medical sensors. Moreover, various wearable sensors for medical monitoring are available [5].

We have already proposed to add the sound monitoring in a such system in order to increase its reliability [6]. The developed sound analysis system is capable of detecting distress sounds or speech in a noisy environment using specific acquisition card. The medical monitoring part of this work has installed at the Department of Medical Oncology of the Grenoble University Hospital to produce preliminary clinical data to study patients under chemotherapy [7].

This paper proposes an optimized implementation of previous sound analyzing system on a more flexible and low cost equipment: just a PC with a sound card and microphone. This equipment proposes alternative solutions for alarm transfer: email or SMS to a close person.

D. Istrate is with RMSE-ESIGETEL, 1, Rue du Port de Valvins, 77215 Fontainebleau-Avon Cedex, France, *dan.istrate@esigetel.fr*

M. Vacher and J.F. Serignat are with CLIPS-IMAG (UMR CNRS-UJF-INPG 5524),385, Rue de la Bibliothèque - BP 53, 38041 Grenoble Cedex 9, France, {*Michel.Vacher, Jean-Francois.Serignat*}*@imag.fr*

II. SYSTEM HARDWARE

In Figure 1, we can observe that the proposed system does not need specific components: only a PC (Windows NT/2000/XP) with a sound card (existing device in almost PC), a microphone and Internet link. The software implementation proposes an assisted sound level calibration in order to adapt acquisition gain. The sampling frequency was fixed at 16 kHz, usual value in speech recognition.

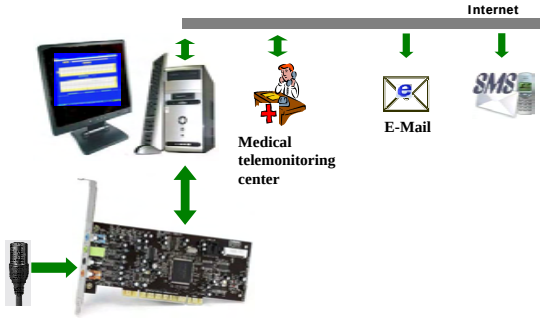

Fig. 1. System hardware

The system can communicate with a medical remote monitoring center but it can send an email or a SMS to a close person too.

For sound sample acquisition, Windows SDK functions are used in order to be independent from hardware. The detected events considered as alarms are saved on PC hard-disk and sent through the network to medical remote monitoring center or through email to a close person. An history of detected events (detection time, detection type) is recorded in a text file. The abnormal detected signal is recorded in a standard Wave format (without compression).

III. REAL-TIME SOFTWARE ARCHITECTURE

For a real time working purpose, the sound analysis is divided in two parts: sound event detection and classification. The classification module is composed of 3 modules:

a segmentation module which labels the extracted signal: speech or sound, a classification module which identifies everyday sounds between 7 predefined classes and a speech recognition one. The distress expression recognition is carried out using a classical speech recognition system which is not described in this paper. In the case of an alarm, the extracted information and the sound file are sent through network either to monitoring center or by email, to a close person.

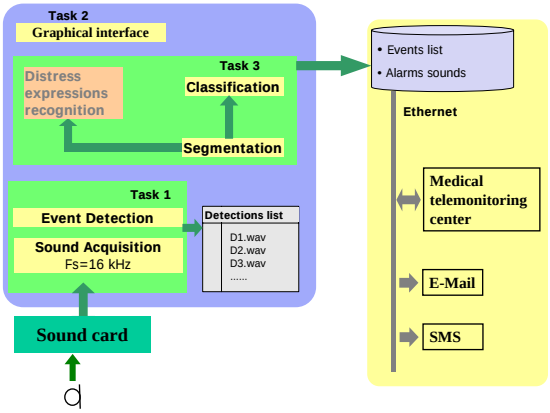

Fig. 2. Real-Time Software Flow-Chart

The sound analysis system has been divided into three parallel tasks, as shown in Figure 2. The first task controls the acquisition process and event detection; it has a high priority. The event detection module outs a list with detected events and corresponding "wav" files.

The third task receives the list with sound event extracted and launches a 2 classes classification (segmentation): speech or everyday sounds. In the case of a sound, the most probable sound class is estimated between the seven predefined sound classes. In the case of speech detection, a classical recognition system looks to identify distress expressions ("Help!", "A doctor please!", etc.).This task manages also the alarm and sound files sending through network.

The second parallel task is the graphical user interface (GUI) which displays sound signal and a list with detected events. This interface is used also to set up the application (medical remote monitoring center parameters and/or email and phone number of a close person).

A. *Sound event detection and capture*

The event detection aims at finding impulsive signals in the noise and to extract them from the signal flow. It must be capable to identify impulsive signals like door clapping, dishes sounds but also speech in a noisy environment.

There are many techniques possible to be used for sound detection: energy threshold, statistical model [8], energy processing [9] or wavelet processing [10].

A wavelet based event detection algorithm has been proposed in [6] and improved for this application. Unlike Fast Fourier Transform, Wavelet Transform is well adapted to signals that have very localized features in the time-frequency space. This transform is frequently used for signal detection and audio processing. We have chosen Daubechies wavelets with 6 vanishing moments to compute DWT. A complete orthonormal wavelet basis consists of scalings and translations of the mother wavelet function.

The DWT consist of applying a wavelet coefficient matrix hierarchically, first to the full data vector of length N, then to the down-sampled vector of length N/2, then to the down-sampled vector of length N/4 and so on until the vector length becomes 2. This procedure is called pyramidal algorithm.

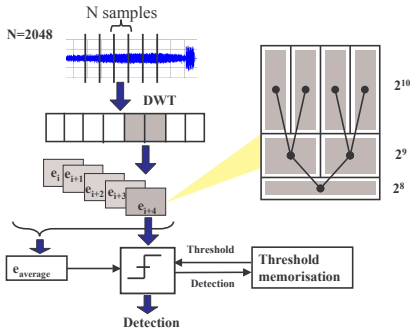

Fig. 3. Flowchart of the wavelet based algorithm

The DWT window duration is 128 ms (2048 samples), imposed by real time constraints. The algorithm (flowchart in figure 3) calculates the energy of the 8, 9 and 10 wavelet coefficients (the three higher order coefficients), because the significant wavelet coefficients of the sounds to be detected are rather high order. A better time resolution being necessary, the output vector of DWT transform is cut up in 4 analysis frames of 32 ms each one. Thus the analysis frame has 32 ms with an overlapping of 16 ms (50%).

The detection is achieved by applying a threshold on the sum of energies of the three highest order wavelet coefficients. The threshold is self-adjustable and depends on the average of the 10 last energy values.

The method to detect the signal end starts at the beginning of signal detection with the threshold memorization. If the signal wavelet energy is lower then the frozen threshold for M successive frames, the signal end is reached. A value of 16 frames (0.25 s) has been chosen for M parameter in order to avoid the silence between words and to make possible the use of the algorithm in the speech case. The algorithm needs to avoid phrase cut because in speech case the system is designed to detect distress expressions and not only keywords. The M parameter value has been chosen using a statistical study of the BRAF100 French speech corpora. This corpora contains 100 speakers, 10000 sentences, 20000 words and has 28 hours of speech. In this corpora we found an average of 3 words by second, with an average word duration of 0.33s which implies an average silence duration of 0.2s.

An example of sound detection achieved by the presented algorithm is shown in Figure 4. The amplitude of the sound signal that contains a mixture between a ringing phone at 3.2 second and a water flow noise at 0 dB of SNR can be seen in the first window, while the second window shows the wavelet energy outlined in black and the self-adjustable threshold in grey. The detection signal presented in the third

window shows clearly that the algorithm detects the signal from noise despite their close amplitudes.

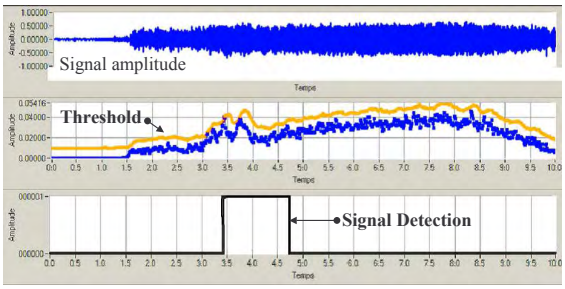

Fig. 4. Detection of a ringing mixed with water flow noise at 0 dB of SNR

1) Algorithm Evaluation: In order to test and validate the sound analysis system we have generated a *sound corpus*. It contains recordings made in the Clips laboratory (15% of the CD), the files of "Sound Scene Database in Real Acoustical Environments" [11] (70% of the CD) and files from a commercial CD (film effects, 15 % of the CD). Entire corpus is composed of 3354 files. The detection algorithm has been evaluated on a test set which contains 11 types of sound mixed with 2 types of noise (really apartment noise - HIS and white noise) at 4 SNR levels: 0, 10, 20 and 40 dB. Each sound and noise are repeated 3 times. The detection test set is made up of about 2000 files.

In order to evaluate detection algorithm the Missed Detection Rate (MDR), the False Alarm Rate (FAR) and Equal Error Rate (EER) are calculated using ROC curves. The results of proposed DWT based algorithm in the case of HIS noise and white noise are presented in Figure 5.

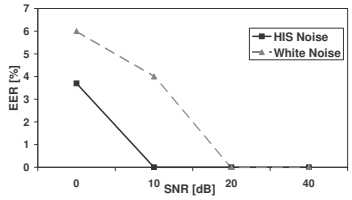

Fig. 5. Detection algorithm performances

B. Sound classification - second module

The classification module looks to identify the sound event between predefined sound classes. This module uses a **G**aussian **M**ixture **M**odel (GMM) method [12]. There are other possibilities for the classification: HMM, Bayesian method and others but GMM classification can be implemented more easily, procures comparable performances and requires low processing time, an important constraint of this application.

This module carries out firstly, a signal segmentation between sound and speech and secondly, in the case of sounds, a classification among predefined sound classes. These two steps use the same GMM method with different parameters.

GMM method like all statistical recognition methods needs a training step before the identification procedure. The training is initiated for each class of signals from sound corpus and gives a model containing the characteristics of each Gaussian of the class after 20 iterations of an "EM" algorithm (Expectation Maximization) following a K-means algorithm. The unkown signal belongs to the class for which likelihood is maximum.

The Bayesian Information Criterion (BIC) was used in order to determine the optimal number of Gaussian. The minimum value of BIC indicates the best model order.

After BIC calculation for the sound respectively for the speech class in noiseless conditions, a value of 24 Gaussian was choosen (good compromise between segmentation performances and calculus consumption). The same calculus has been made for sound classification module, in the case of the sound class with the smallest number of files. 4 Gaussian distributions has been chosen for this module.

1) Acoustical parameters: Classification does not use directly signal samples, but a vector of acoustical parameters computed on analysis windows. The acoustical parameters are determined for each analysis window of 16 ms with an overlap of 8 ms.

There are many types of acoustical parameters like MFCC (Mel Frequencies Cepstral Coefficients), LFCC (Linear Frequencies Cepstral Coefficients), LPC (Linear Predictive Coefficients), ZCR (zero crossing rate), RF (Roll-off point), Centroid, etc, but only few of them are appropriate to the sound classification.

After a statistical study based on Fisher Discriminant Ratio (FDR) and a validation on test set, a combination of 16 MFCC with ZCR, RF and Centroid has been chosen for sound classification (error classification rate was 10% on a 1577 test set).

2) Classification evaluation: We have carried out *7 sound classes* from the sound corpus, in order to train the classification algorithm. The 7 sound classes are: door slaps, ringing phone, step sounds, dishes sounds, door locks, breaking glasses, screams.

The speech distress expressions corpus has been recorded in the CLIPS laboratory by 21 speakers (11 men and 10 women). It is composed of 126 sentences in French: 64 are characteristic of a normal situation for the patient and 64 are distress sentences. It has a total duration of 38 minutes and is constituted by 2646 audio files.

The sound classification performances are evaluated through the error segmentation rate which represents the ratio between the bad classified sounds and the total number of sounds to be classified. In Figure 6, the segmentation results are presented for 16 MFCC acoustical parameters coupled with normalized energy.

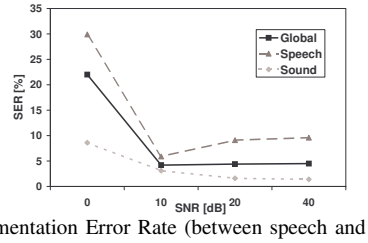

Fig. 6. Segmentation Error Rate (between speech and everyday sounds)

Concerning the sound classification in apartment noise, the results are presented in Figure 7. We can observe that for SNR $\geq$ 20 dB the classification module has a classification error rate about 11% but for SNR of 0 and 10 dB is unusable.

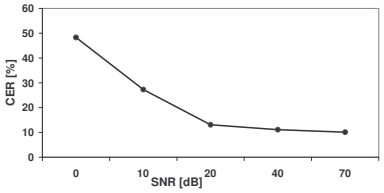

Fig. 7. Classification error in HIS noise between 7 sounds classes

IV. Real Time Implementation

The whole sound analysis flow-chart (figure 2) is implemented by a software written with LabWindows/CVI on PC. This software drives the real time sound acquisition at 16 kHz sample rate using a double buffering of 2048 samples. The event detection is launched between the acquisition of 2 consecutive frames (2048 samples).

The extracted event signal is recorded on the PC hard disk and a list with extracted sound events is updated; this list is used by classification algorithms in the second parallel task. The sound classification is carried out in two steps: first the event is labelled with sound or speech and secondly, if a sound has been detected the most probable sound class is identified. If a speech has been detected, the speech recognition module is launched.

When an alarm sound or a distress expression are identified, the corresponding sound file is kept and an alarm is sent, with respect to the configuration, either to the remote medical monitoring center trough TCP/IP protocol and/or to a close person trough email/SMS. The sent information contains: date and time detection, the three most probable sound classes with their corresponding likelihoods, patient identification and the corresponding sound file.

The software configuration panel allows the microphone calibration and the choice of alarm type: communication with remote medical center (parameters are IP address and used port) or with a close person (parameters are email address and/or telephone number).

On the software front panel (Figure 8) there are 2 tabs: **Monitoring** one, which allows to launch the real-time monitoring system or to stop it and **Configuration** one, which allows to set up the alarm configuration and to launch microphone level calibration. On the first tab, the real-time signal is shown with a chronological list of detected events and the alarm configuration (type of alarm) is reminded.

V. Conclusions and Perspectives

This paper presents a generic implementation of sound monitoring system which aims to detect alarm sounds (falls, screams, glass breaking,...) and distress speech expressions (”Help”, ”A doctor please”,...). This system monitors in real time the sound and in the case of an alarm situation detection it sends the information through network. The proposed solution does not need a specific hardware only

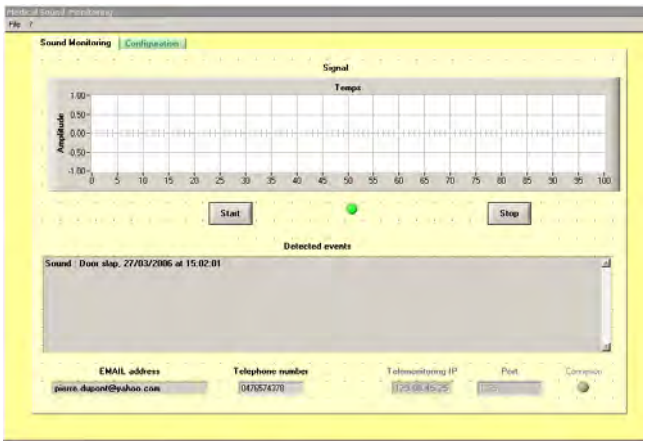

Fig. 8. Front panel of implemented sound monitoring system

a PC station with a sound card and a microphone. Another system characteristic is the possibility to send an email/SMS to a close person and if needed, it can work without a remote medical center. This possibility can be used in the case of healthy person which has reduced accident risks.

Currently, efforts are done to transfer this implementation to WindowsCE platform in order to make possible the use of sound monitoring system on PDA or smart phones. The WindowsCE has native speech recognition module and all PDA or smart phones has complete sound acquisition module. Another possibility to be investigated further is the use of this WindowsCE implementation on an embedded PC which can be used to monitor simultaneously another medical parameters.

References

[1] G. Williams, K. Daougthy, *et al.*, “A smart fall and activity monitor for telecare applications,” in *20th IEEE EMBS*, October 1998.

[2] N. Noury, A. Tarmizi, *et al.*, “A smart sensor for the fall detection in daily routine,” in *SICICA 2003*, Aveiro, Portugal, July 2003.

[3] J. L. Baldinger, J. Boudy, *et al.*, “Tele-Surveillance System for Patient at Home: the MEDIVILLE System,” in *ICCHP 2004*, France, 2004.

[4] A. Lacombe, F. Rocaries, *et al.*, “Open technical platform prototype and validation process model for patient at home medical monitoring system,” in *BioMedsim*, Linkping, Sweden, May 2005.

[5] A. Dittmar and G. Delhomme, “Living tissue mechanisms and concepts as models for biomedical microsystems and devices,” in *1st Annual International IEEE-EMBS*, Lyon, France, October 2000.

[6] D. Istrate, E. Castelli, M. Vacher, L. Besacier, and J. Serignat, “Information extraction from sound for medical telemonitoring,” *IEEE Transactions on TITB*, vol. 10, pp. 264–274, April 2006.

[7] G. L. Bellego, N. Noury, G. Virone, M. Mousseau, and J. Demongeot, “Measurement and model of the activity of a patient in his hospital suite,” *IEEE Transactions on TITB*, vol. 10, pp. 92–99, January 2006.

[8] T. Yamada and N. Watanabe, “Voice activity detection using non-speech models and HMM composition,” in *Workshop on Hands-free Speech Communication, Tokyo, Japan*, 2001.

[9] A.Dufaux, “Detection and recognition of impulsive sounds signals,” Ph.D. dissertation, Université de Neuchatel, 2001.

[10] L. Daudet, “Représentations structurelles de signaux audiophoniques - méthodes hybrides pour des applications à la compression,” Ph.D. dissertation, Université de Provence, Marseille, 2000.

[11] R. W. C. Partnership, “CD - Sound scene database in real acoustical environments,” http://tosa.mri.co.jp/sounddb/indexe.htm, 1998-2001.

[12] D. Reynolds, “Speaker identification and verification using gaussian mixture speaker models,” in *Workshop on Automatic Speaker Recognition, Identification and Verification*, Martigny, Switzerland, April 1994, pp. 27–30.