

# Smart Audio Sensor for Telemedicine

Michel Vacher\*, Dan Istrate\*, Laurent Besacier\*,  
Eric Castelli\*\*, Jean-François Serignat\*

\*CLIPS-IMAG

BP 53 - 38041 Grenoble Cedex 9, France

\*\*International Research Center MICA

1, Dai Co Viet - Hai Ba Trung Hanoi - Vietnam

Michel.Vacher@imag.fr, Dan.Istrate@imag.fr, Laurent.Besacier@imag.fr,  
castelli@vn.refer.org, Jean-Francois.Serignat@imag.fr

## Abstract

In order to improve patients' life conditions and to reduce the costs of long hospitalization, medicine is more and more interested in telemonitoring techniques. We develop a smart audio sensor for a telemonitoring system. This sensor is equipped with microphones in order to detect a sound event (an abnormal noise or a call for help). The sound extracted information is sent through a CAN bus. The originality of our approach consists in replacing the video camera monitoring, which the patients are uncomfortable with, by microphones surveying the sounds. We present the hardware implementation, the software treatments and a first evaluation of the algorithms used.

## 1. Introduction

We describe in this paper a smart audio sensor for sound information extraction in telemedicine application. Telemedicine consists in associating electronic monitoring techniques with computers "intelligence" through network communications [1]. The smart sensor we work on is designed for the surveillance of the elderly, convalescent persons or pregnant women. Its main goal is to detect serious accidents as falls or faintness (which can be characterized by a long idle period of the signals) anywhere in the apartment. It was noted that the elderly had difficulties in accepting the video camera monitoring, considering it a violation of their privacy. Thus, the originality of our approach consists in replacing the video camera by a system of multichannel sound acquisition [2]. The smart sensor analyzes in real time the sound environment of the apartment and detects the abnormal sounds (objects or patient's falls) and the calls for help (or moans), that could indicate a distress situation in the habitat.

To respect patient privacy, no continuous recording or storage of the sound is made. However, if a sound event is detected, the last 20s of the audio signal are kept in a buffer and sent to the alarm monitor. Thus, a human operator can take a decision concerning the medical intervention.

The smart audio sensor analyzes the audio signals, detects events, recognizes the type of noise (presently we don't make "Word Spotting" for help call) and sends the information through a CAN bus. The smart sensor that we have designed is also capable to send information by TCP/IP network.

## 2. Sensor description

The smart audio sensor contains 8 microphones, 8 signal conditioning boards and a data acquisition board (see the figure 1). For the moment we use only 5 of the 8 available microphones, and the acquisition is made on all the 5 channels simultaneously to allow survey of the entire apartment.

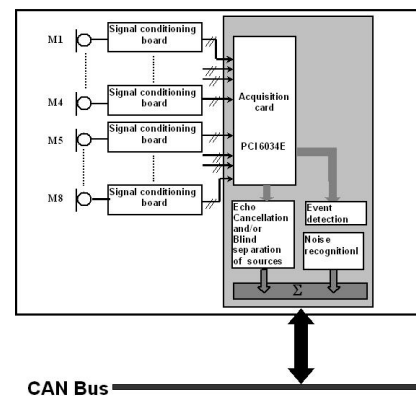


Figure 1: Smart Audio Sensor Diagram

### 2.1. Data Acquisition

The used microphones are omni-directional, condenser type, of small size and low cost. The signal conditioning card, associated with each microphone, consists of an amplifier and an anti-aliasing filter. The acquisition is made by a multi-channel National Instruments acquisition card PCI 6034E (8 differential channels), installed inside a computer. The acquisition is made at a sampling rate of 16 KHz, a frequency usually used in speech and audio applications. We have programmed the entire software which controls the real time acquisition under Lab-Windows/CVI [3] of National Instruments. In order to drive in real time the data acquisition board we have used the low-level functions. After digitalization, the sound data is either saved in real time on the hard disk of the host PC, or analyzed.

### 2.2. Data Processing

The sound event detection and classification are complex tasks since the audio signals are not clean and the everyday life

sounds are extremely diverse. This system makes a two-step analysis: firstly a sound event detection is made and secondly a sound classification. In the first step, signals from 5 channels are used to detect events and to localize the sound source. In the case of simultaneous detection, the most powerful signal is chosen. An event detected by the first step initiates the second step which begins with a segmentation speech/noise. If a speech has been detected, a Word Spotting system is launched in order to identify calls for help, while in the noise case a recognition system is started in order to recognize the noise class. The smart audio sensor will send an alarm if the Word Spotting detects a call for help or if the recognized noise class is a distress one. For the moment, the recognition system is not in use and the detected events are classified by a human operator.

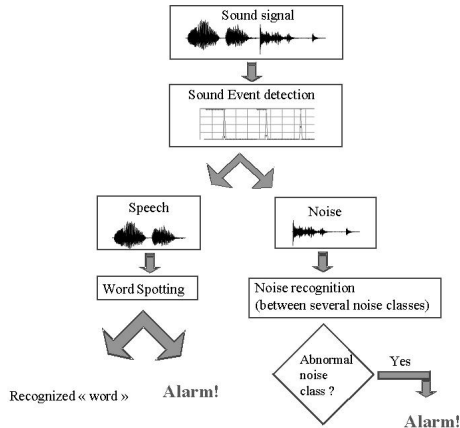


Figure 2: Sound analysis steps

### 2.2.1. Sound Event Detection

The detection performance [4] is very important, because a missed event can lead to dramatic consequences for patient. On the contrary, too many false alarms can saturate the recognition stage and hide new important events. Two detection algorithms based on cross-correlation and respectively on energy prediction, give good results in the experimental environment.

For the first one, the cross-correlation detection between two successive normalized windows of 2048 samples (128 ms) is initiated (see the figure 3). A threshold is applied on the maximum value of the cross correlation. A signal under the threshold reveals a significant statistical change of the sound and therefore an event.

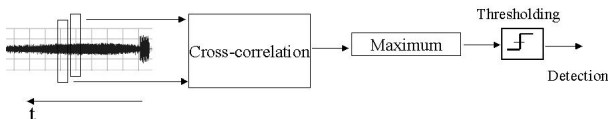


Figure 3: Flow-chart of the Cross-correlation algorithm

For the second one, the algorithm calculates the energy every 2048 sample windows across the time (see the figure 4). Initially, the future value of the next window is predicted with a Spline Interpolation Method using the last 10 known values. After 128 ms, the real value could be measured and compared

with the estimated one. A self-adjustable threshold is settled on the absolute difference between the two values, see (1) :

$$Threshold = \kappa + \frac{6.75}{\sigma} + \sqrt{2} \cdot \sigma + m \quad (1)$$

$m$  and  $\sigma$  denote the average, respectively the variance of the values needed for interpolation,  $\kappa$  is an experimentally adjustable coefficient.

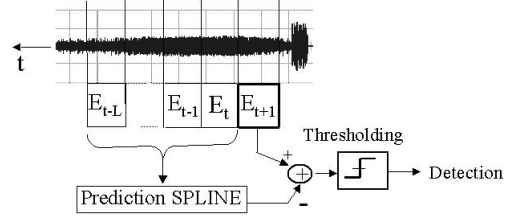


Figure 4: Flow-chart of the Prediction error algorithm

### 2.2.2. Sound Recognition

The sound classification method uses a Gaussian Mixture Model (GMM)[5], [6] with 4 gaussian components. The analysis window (frame of 256 samples) is set to 16 ms with an overlap of 8 ms. The training step of the classification is done on the Elisa platform [7]. For the testing step we use our own system.

Before classifying an audio signal, parameters must be calculated for each frame. The global performance of the classification depends on the choice of these parameters.

## 2.3. Experiments and Results

### 2.3.1. Sound Database for Our Tests

In order to test and validate the event detection system and the sound recognition system, we have recorded a sound corpus. It contains recordings made in the Clips laboratory (15% of the CD) [8], the files of "Sound Scene Database in Real Acoustical Environment" (70% of the CD) [9] and files from a commercial CD (film effects, 15% of the CD). There are 3354 files and every one is sampled at 16kHz and 44 kHz. The sound corpus contains : door slap sound (different type of doors), chair sound, step sound, electric shaver sound, hairdryer sound, door lock sound, dishes sound, glass breaking, objects fall sound (books, pencils, etc ...), screams, water sound, different ringing, etc. The sound corpus contains 20 types of sounds with 10 minimum repetitions per type (the maximum is 300 repetitions).

In order to validate the detection algorithms, we have generated a test set which is a mixture of environmental data noise (ordinary noise recorded in the apartment - noted as HIS noise, water noise and white noise) and useful sounds at different signal to noise ratios (SNR): 0dB, 10dB, 20dB and 40 dB. The database contains 2376 files ( $\approx$  3h of recording).

For the preliminary classification tests, we have generated 7 sound classes: dishes sound (163 files-7943frames), door clapping (523 files-47398frames), door lock (200 files-605frames), glass breaking (88 files-9338frames), screams (73 files-17509frames), telephone ringing (517 files-59188frames) and step sound (13 files-3648frames). The amount of recognition test set is 1577 files(145689 frames). Each frame (16ms) contains 256 samples. For the moment, the mixture environmental noise/useful sound has not been yet tested, therefore the

signals we are using for the classification tests can be considered as clean.

### 2.3.2. Detection Results

To evaluate the two algorithms for our application, we have calculated the Missed Detection Rate (MDR) and the False Detection Rate (FDR) on the test set. This allows us to determine the Equal Error Rate (EER) defined as the value of MDR for  $MDR=FDR$ . These rates are classically used in detection theory. The lower the EER is, the higher performances the algorithm has. The results are given in table (1).

Method	SNR	HIS Noise EER [%]	Water noise EER [%]
Cross-correlation	0	14.2	78.5
	+10	6.6	81.7
	+20	5	82.3
	+40	7.1	83.2
Prediction error	0	62	34
	+10	28	8
	+20	7	6
	+40	0	0

Table 1: Detection results

Both, the *cross correlation algorithm* and the *energy prediction algorithm* have one threshold to adjust. The *cross correlation algorithm* is very good in the case of HIS noise, but gives very bad results in the case of water noise because this type of noise is uncorrelated and gives many false alarms. We have tested another normalization (by squared root of the window energy instead of peak normalisation) for *cross correlation algorithm* which has best results for water noise but degrade the HIS noise results. The *energy prediction algorithm* is faster (10 times) than the *cross-correlation algorithm* and gives very good results, except for low SNR. In the future, the best solution appears to be the fusion of the two algorithms.

### 2.3.3. Sound Classification Results

The experimental results of sound classification, for different combinations of the parameters, are given in table (2). Classification performances are averaged for all the classes (the number of good classifications divided by the number of tests).

The test protocol is a "leave one out" protocol: the model of each class is trained on all files of the class, excepting one. Next, each model is tested on the remaining sounds of all classes. The whole process is iterated for all files (1577 tests). A file is recognized if the average of the all frames likelihood is the greatest for the model of the sound membership class.

The tested parameters are: MFCC (Mel-Frequency Cepstral Coefficients - classically used in speech recognition), LFCC (Linear Frequency Cepstral Coefficients), the energy coefficients of the linear (triangular or rectangular) filters, the energy coefficients on a MEL (logarithmic) scale, zero crossing rate (ZCR), roll-off (RF) point (a measure of spectral shape skewness), centroid (the barycenter of the spectrum), LPC (Linear Prediction Coefficients - classically used in speech compression) and LPCC (Cepstral LPC). We have also tested the first and the second derivative of the parameters.

We can observe that better results are done with the  $\Delta, \Delta\Delta$ MFCC parameters concatenation with zero crossing

Parameters	Recognition rate [%]
(16MFCC+E+ZCR+RF+Centroid) <sup>*</sup> , $\Delta$ ( <sup>*</sup> ), $\Delta\Delta$ ( <sup>*</sup> )	92.84
16 MFCC+E+ZCR+RF+Centroid	89.85
16 LFCC	86.69
16 MFCC+E	85.95
16 LPCC	85.67
16 LPC	79.84
16 Coef.Mel+ZCR+RF	74.70

Table 2: Classification results

rate, roll-off point, energy and centroid. Nevertheless, good results are also obtained with the Linear coefficients (LFCC). The addition of special parameters (like zero-crossing-rate, roll-off point and centroid) to MFCC coefficients done a 4% gain of performances. For instant the MFCC coefficients seems to be best adapted for sound recognition.

## 3. Sensor Output Interface

The sensor data is sent on an industrial CAN bus [10]. Its standard is ISO 11898, a serial bus. Any device on a CAN network can communicate with any other device using a common pair of wires. In a collision case, this bus gives a deterministic response, contrary to the one of the Ethernet network. Each node of CAN bus has its priority; in the collision case, the node with the most important priority continues to transmit. The CAN bus is low cost, has good resistance to harsh environments and high real-time capabilities. The CAN bus, used in our application, is a dedicated bus which provides a big security (only sensors are connected on the bus).

When a sound event is detected, the smart audio sensor sends a frame of information on the CAN bus. The frame contains: date and time detection (day, month, year, hour, minute, seconde, milliseconds), a flag to indicate the type of sound event (speech or noise) and a character field. This character field is composed by: the three most probable noise classes (or words) with their corresponding likelihoods and the localization of the sound event (the room).

Then the master PC in charge of all sensors can make a data fusion between data issued from the smart audio sensor and other sensors. Figure (5) shows the hardware configuration.

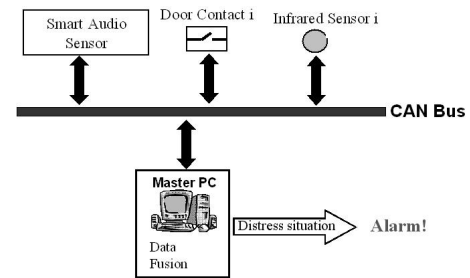


Figure 5: Hardware configuration

## 4. Application to Telemedicine

We have designed this sensor for a medical telemonitoring application. This sensor is one of the multiple sensors installed in

an experimental apartment (see figure 6). The habitat we used for experiments is a 30 m<sup>2</sup> apartment situated in the TIMC laboratory buildings, at the Michalon hospital of Grenoble [11].

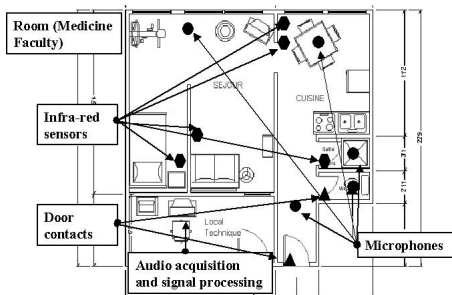


Figure 6: The experimental apartment

Until now we have realized a system of real time data acquisition and detection (see the panel from figure 7). The software carries out the 5 channels of data acquisition, the sound event detection for all the channels (with localization: room where the sound appeared), the transmission of the information through a CAN bus and the recording of the sound in case of an event detection. On the panel, we show the signal which generated the detected event, its localization (the room) and a list of detected events.

The detection is made simultaneously on each channel and in the case of a sound event detection, the sound signal is saved on the hard-disk ; for the moment the sound is analyzed by a person and the classification algorithm is not in use. We work on the implementation of the noise recognition algorithm in LabWindows/CVI. The noise recognition task should be executed in parallel with the acquisition-detection task. The sound extracted information is sent by CAN bus to the master PC for data fusion.

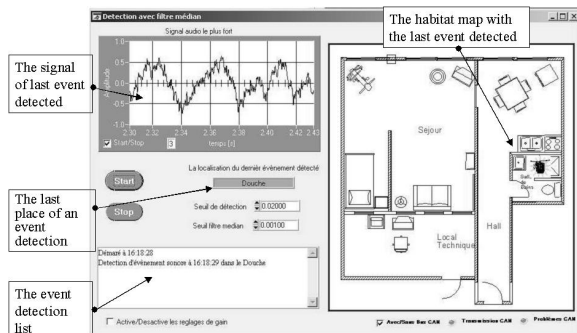


Figure 7: The panel of the real-time analysis software

## 5. Conclusions and Perspectives

In this paper we have presented a smart audio sensor which analyzes multiple audio channels in order to detect sound events and to recognize the noise class helping a medical telemonitoring system to survey elderly people. The smart sensor is composed for the moment by a data acquisition card and a PC. The present results of detection are good and we are working on the sound recognition.

Algorithms used in speech technology have encouraging performance for the sound recognition task. Current work is

done on finding adequate parameters to discriminate sounds more efficiently.

To make it a physically independent smart sensor, we shall implement the sensor system inside the digital signal processor card. We have chosen to test the sensor algorithms using a PC because of the facilities in terms of implementation and verification. We can use this sensor in telemedicine but also for a security alarm system or for a smart room (various domestic automation).

## 6. Acknowledgements

This system is a part of the RESIDE-HIS<sup>1</sup> project, a collaboration between the CLIPS<sup>2</sup> laboratory, in charge of the sound analysis, and the TIMC laboratory, charged with the medical sensors analysis and data fusion. This project is financed by IMAG<sup>3</sup>.

## 7. References

- [1] V. Rialle and N. Lauvernay and A. Franco and J.-F. Piquard and P. Couturier, "A Smart Room for Hospitalised Elderly People: Essay of Modeling and First Steps of an Experiment", *Technology and Health care*, Vol.7, pp343-357, 1999.
- [2] E. Castelli and D. Istrate, "Multichannel Audio Acquisition for Medical Supervision in an Intelligent Habitat", *EC-CTD 15th European Conference on Circuit Theory and Design*, Helsinki, Finlande, 28-31 August, Vol.II, pp1-4, 2001.
- [3] National Instruments Corporation, "LabWindows/CVI User Manual", December 1999.
- [4] A. Dufaux, "Detection and Recognition of Impulsive Sounds Signals", *PH.D Thesis, Faculté des sciences de l'Université de Neuchatel*, 2001.
- [5] D. Reynolds, "Speaker Identification and Verification using Gaussian Mixture Speaker Models", *Workshop on Automatic Speaker Recognition, Identification and Verification*, Martigny, Switzerland, April:27-30, 1994.
- [6] R.L. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", *Proceedings of the IEEE*, Vol.77, pp257-286, 1989.
- [7] I. Magrin-Chagnolleau, G. Gravier and R. Blouet, "Overview of the ELISA Consortium Research Activities", 2001 : a *Speaker Odyssey*, pp67-72, 2001.
- [8] Dan Istrate, "Base de données. Bruits de la vie courante CD", *CLIPS-IMAG Équipe GEOD*, Novembre 2001.
- [9] Real World Computing Partnership, "Sound Scene Database in Real Acoustical Environments CD", <http://tosa.mri.co.jp/sounddb/indexe.htm>, 1998-2001.
- [10] The Bosch site of CAN bus, <http://www.can.bosch.com>.
- [11] E. Castelli and D. Istrate, "Everyday Life Sounds and Speech Analysis for a Medical Telemonitoring System", *Eurospeech 2001*, Aalborg, Denmark, 3-7 September, Vol.E15, pp2417-2421, 2001.

<sup>1</sup>Reconnaissance de Situations de Détresse en Habitat Intelligence Santé (Distress situations recognition in a medical intelligent habitat)

<sup>2</sup>Communication Langagière et Interaction Personne-Système (Language communication and human-machine interaction)

<sup>3</sup>Institut d'Informatique et Mathématiques Appliquées de Grenoble (Institute of Informatics and Applied Mathematics of Grenoble)