

HABITAT TELEMONITORING SYSTEM BASED ON THE SOUND SURVEILLANCE

Eric Castelli, Michel Vacher, Dan Istrate, Laurent Besacier, Jean-François Sérignat

▶ **To cite this version:**

Eric Castelli, Michel Vacher, Dan Istrate, Laurent Besacier, Jean-François Sérignat. HABITAT TELEMONITORING SYSTEM BASED ON THE SOUND SURVEILLANCE. 1st International Conference on Information Communication Technologies in Health, Jul 2003, Samos, Greece, Greece. pp.141-146. hal-01085262

HAL Id: hal-01085262
https://hal.science/hal-01085262

Submitted on 21 Nov 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

HABITAT TELEMONITORING SYSTEM BASED ON THE SOUND SURVEILLANCE

Eric Castelli[*], Michel Vacher[**], Dan Istrate**, Laurent Besacier[**] and Jean-François Sérignat[**]
[*]International Research Center MICA
Hanoi - Vietnam
Tel: +84 4 868 30 87
E-mail: castelli@vn.refer.org
[**]CLIPS-IMAG
Grenoble - France
E-mails: Michel.Vacher@imag.fr, Dan.Istrate@imag.fr, Laurent.Besacier@imag.fr, Jean-Francois.Serignat@imag.fr

Abstract:

This paper presents a telemonitoring system in an habitat equipped with physiological sensors, position encoders of the person, and microphones. The originality of our approach consists in replacing the video camera monitoring, not well accepted by the patients, with microphones acquiring the sounds. The sounds are analyzed and not stored in order to maintain the person privacy. We present the entire telemonitoring system which makes the data fusion between medical information and sound information and particulary the sound processing algorithms to detect a distress situation. The first step of sound processing is the sound event detection in a noisy everyday life environment. Sound event detection is necessary to extract the significant sounds before initiating the classification step. Sound classification system and its performances are presented in this paper, too.

Introduction

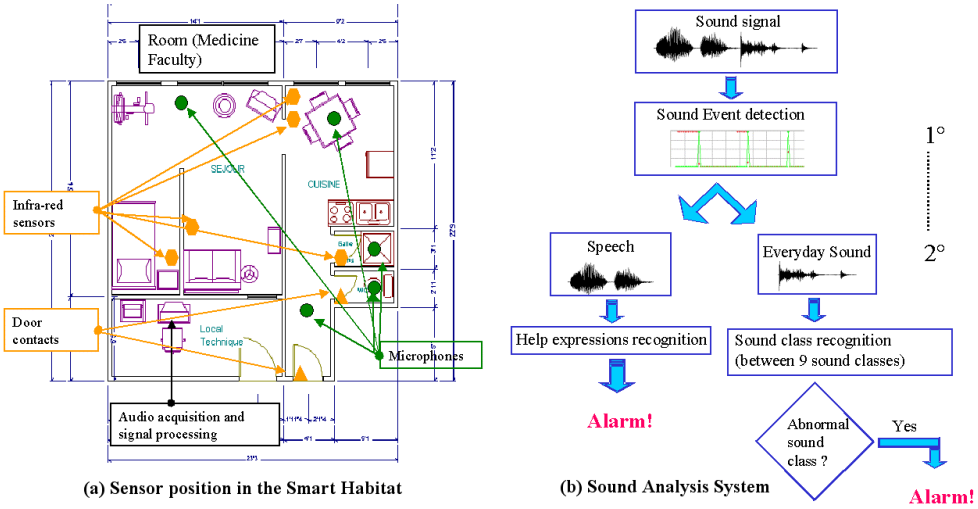
Medical monitoring is more and more frequently used in order to reduce hospitalisation costs. There are many researches in telemedicine, but few of them are sound based. In this paper, we present a medical telemonitoring system with a smart audio sensor. The system we work on is designed for the surveillance of the elderly, convalescent persons or pregnant women [1]. Its main goal is to detect serious accidents as falls or faintness at any place 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. The system analyzes in real time the sound environment of the apartment and detects the abnormal sounds (falls of objects or patient) and the calls for help, that could indicate a distress situation in the habitat. Again, to respect privacy, no continuous recording or storage of the sound is made, since only the last 5s of the audio signal are kept in a buffer and sent to the alarm monitor if a sound event is detected. The sound information extraction

is a complex task because the audio signals occur in a noisy environment and the everyday life sounds are extremely diverse.
We shall start with an overview of the medical telemonitoring system, next we shall expose the detection algorithm necessary to extract the sound event and we shall finish with the classification system.

The Sound Analysis System

The habitat we used for experiments is a 30 m^2 apartment situated in the TIMC laboratory buildings (See the Figure 1 (a)), filled with various sensors, especially microphones. A microphone is placed in every room (toilet, kitchen, shower-room, hall and living-room). This allows a continuous sound surveillance in the entire apartment. Each of the 5 microphones is connected to the slave computer. The sound or speech source can be localized.

Figure 1 Smart Habitat plan and analysis system



(a) Sensor position in the Smart Habitat

(b) Sound Analysis System

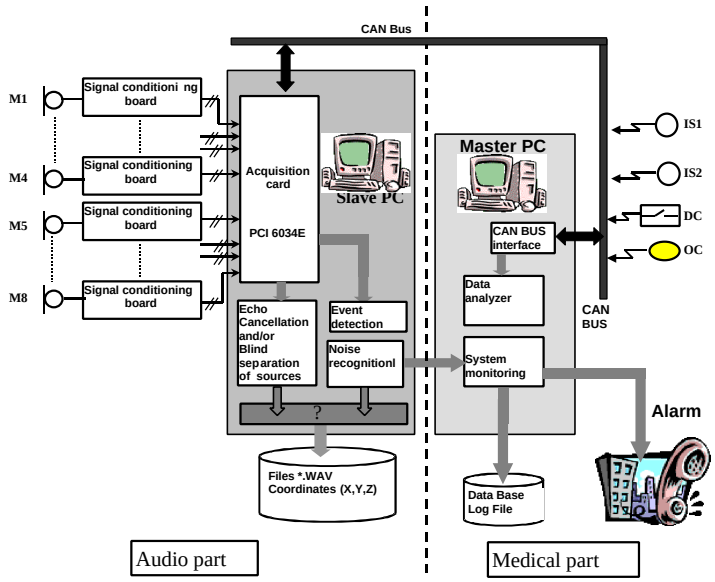
The microphones used are omni-directional, condenser type, small size and low cost. A signal conditioning card, consisting in an amplifier and an anti-aliasing filter is associated to each microphone. The acquisition system consists in a multi-channels acquisition card PCI 6034E of National Instruments, used with a 16KHz sampling rate (usual frequency in speech applications).
The sound analysis has two steps: the first step concerns the detection of a sound event and the second one the sound classification (see Figure 1 (b)).The second step includes an automatic sound classification and a recognition of calls for help expressions.
In the first step, signals from all the 5 channels are used to detect events. It is a difficult task because of the environmental noise. If a sound event is detected, extracted signal is transmitted to the second step and sound classification is initiated.
The everyday sounds are divided in 9 classes. The criteria used for this repartition were :

- ?? Statistical probability of occurrence in everyday life
- ?? Possible alarm sounds (scream, person fall) are priority
- ?? The duration of the sound: significant sounds are considered to be short and impulsive

The 9 sound classes are divided in 2 categories: normal sound classes (door clapping, phone ringing, step sound, human sounds (cough,

sneeze,...), dishes sound, door lock) and sound classes that generate an alarm (breaking glasses, screams, fall sounds).
In conclusion, if an abnormal sound class is detected or a call for help is recognized the sound analysis system transmits an alarm to the data fusion system (which fusions sound and medical sensor information). The decision to call the emergency is taken by the data fusion system.
The entire telemonitoring system is composed of two computers which exchange information through a CAN bus (see Figure 2). The CAN bus is low cost and provides a high level of security (only sensors are connected on the bus).

Figure 2 The acquisition and analysis system



The master computer is in charge of data fusion and analyses data coming from medical sensors and information coming from the slave computer, which is continuously surveying the microphones. The sound analysis system should function as follows: each time a sound event is detected, a message is sent to the master computer, notifying occurrence time of the detection, type of the event (speech or other sound), localization of the emitting source ; it also needs to indicate either most probable sound classes (with the corresponding confidence index), or most probable words (calls for help), with their confidence index. Using this data, the master computer could send an alarm if necessary.

The Sound Database

As no everyday life sound database was available in the scientific area, we have recorded a sound corpus. The corpus contains recordings made in the Clips laboratory, the files of "Sound Scene Database in Real Acoustical Environments" (RWCP Japon) and files from a commercial CD: door slap sound (different types of doors), chair sound, step sound, electric shaver sound, hairdryer sound, door lock sound, dishes sound, glass breaking, objects fall sounds, screams, water sound, different ringing, etc. To summarize, the sound corpus contains

20 types of sounds with minimum 10 repetitions per type (the maximum is 300 repetitions).

The Test Set for the Detection Algorithms. In order to validate the detection algorithms we have generated a test set which is a mixture of environmental noises and useful sounds. In the framework of the RESIDE-HIS project, we consider to be *useful* (impulsive and short) sounds such as: door slap, glass breaking, objects fall, etc... ; and to be *environmental* (long and stationary) noises like: water flow, hairdryer, electric shaver, etc... For every mixture sound-noise, there are several signal to noise ratios (SNR). To summarize, the test signal database contains a total of 2376 files ? 3h of signal.

Table 1 Sound classes

Sound Class	Alarm	Sound Class	Alarm
C1 – Door Slap	No	C6 – Fall sound	**YES**
C2 – Breaking glasses	**YES**	C7 – Dishes sound	No
C3 – Ringing phone	No	C8 – Human sounds	No
C4 – Step sound	No	C9 – Door lock	No
C5 - Scream	**YES**		

The Test Set for the Recognition Task. The test set used for the sound recognition task is composed of 9 sound classes (see Table 1).

The Detection Algorithms

The detection of a signal (useful sound) is very important because if an event is lost by this first system, it is lost forever and that can lead to severe consequences for the patient. On the other side, if there are too many false alarms the recognition system is saturated. Therefore, the performance of the detection algorithm is very important for the entire system. We have tested 5 algorithms: 3 adapted detection algorithms [3] and 2 original algorithms. The description of algorithms is in [2].

Experimental Results of Detection Algorithms

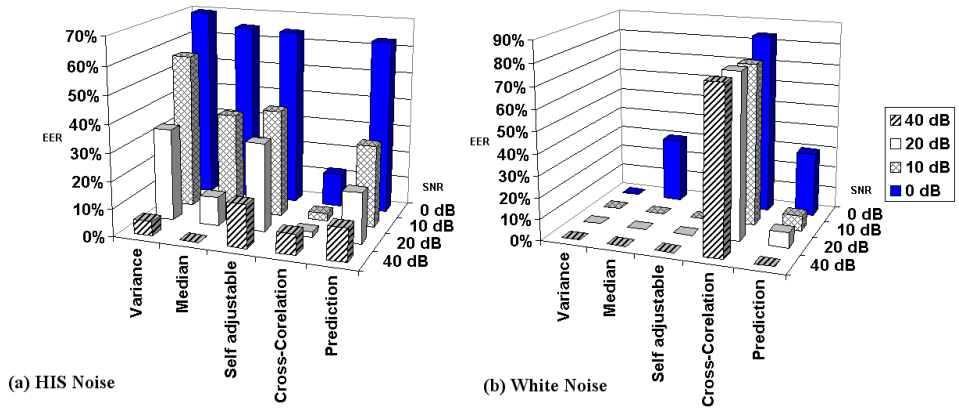
To find the best algorithm for our application, we have calculated the Missed Detection Rate (R_M) and the False Detection Rate (R_F) on the test set. To compare the algorithms we have determined the equal error rate (EER), defined as value of R_M for R_M=R_F.

Detection Results with Our Test Set. The EER of the different algorithms tested on our test set are given in Figure 3 for HIS and white noise at several signal to noise ratio (HIS noise is the environmental noise of our experimental habitat in Grenoble). Because of the numerous calculus necessary for the entire sound system and of the necessity of a real time processing (medical conditions) we must make a trade-off between the performance and the complexity of the algorithm. Besides, medical care imposes a very small missed detection rate.

Most algorithms are very efficient in case of white noise (EER=0% for SNR?+10dB), only the *cross-correlation* (EER>70%) give bad results because it is not adapted to uncorrelated noise. But in real conditions, white noise is not realistic for our application. Therefore to analyze the results, we must compare their corresponding performances only for HIS noise for a 10dB SNR (our real environmental conditions).

The *median filtering*, *variance* and *self-adjusting threshold* are not suited because EER>10% for a 20 dB SNR. We can state that the *energy prediction* algorithm is fast and gives good results (EER=7% at +20dB) except for HIS noise at low SNR. The *cross-correlation* algorithm is better for the HIS noise but requires long calculus.

Figure 3 Sound detection results for HIS and White noise



Detection Results in Real Conditions. We have recorded 60 files inside our test-apartment (real conditions) at SNR ?15 dB. We have used the sounds of the test base played with a speaker. Results obtain for the best detection algorithms are: *cross-correlation* with an EER of **4.4%** and *prediction error* with an EER of **10%** which are better than the previous results (Figure 3).

Sound Classification

For these preliminary tests only the pure sounds (without environmental noise) of the data-base have been used. For the classification task we do not use directly the signal samples, but a vector of acoustical parameters calculated on the analysis windows [5]. The acoustical parameters are determined for each analysis window of 16ms with an overlap of 8ms.

We are using a **G**aussian **M**ixture **M**odel (GMM) method with 4 gaussians in order to classify the sounds [4]. This method evolves in two steps: a training step and a recognition step. We use the classically acoustical parameters used in speech recognition like: MFCC, LFCC, LPC and non-classically like zero crossing rate, roll-off point, centroïd.

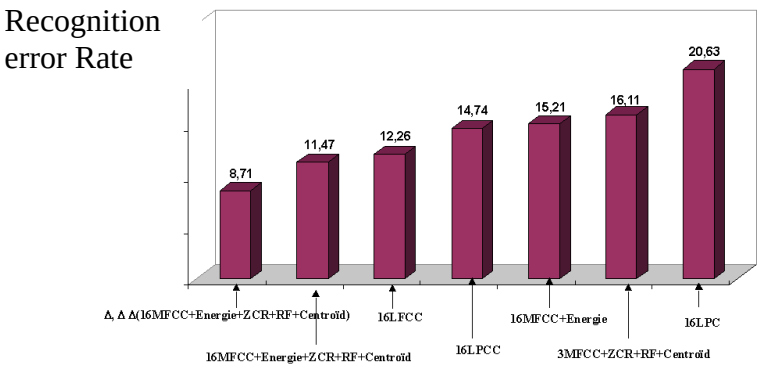
Results with Sound Classification

The training/test protocol is a "leave one out" protocol: the model of each class is trained on all the files of the class, excepting one. Next, each model is tested on the remaining sounds of all classes [6]. The whole process is iterated for all files (1577 tests).

The experimental results are presented in Figure 4. For each parameter, we calculate the average of the error value on all classes. We can observe that the best results are obtained with the MFCC parameters added with zero crossing rate, roll-off point and centroïd. We have tested the combination of three MFCC coefficients with the zero crossing rate, roll-off point and centroïd, suggested by a statistical study. We have noticed that the parameters considered to be irrelevant after statistical study can be eliminated with practically no negative influence over the performances of the system; reducing the number of

parameters by a factor of 3 produces only 4.5% increase of the error classification rate .

Figure 4 Sound Recognition Results



Conclusions and Perspectives

We have presented a telemonitoring system for an habitat equipped with audio and infrared sensors. This is meant to replace video cameras. We have collected a corpus of everyday life sounds. After study of several detection algorithms we have obtained good results which allows us to detect a sound event in the habitat and its position in space (the room). Concerning the sound classification we must make a trade-off between the number of acoustical parameter(calculus time) and performances. The actual performances of ?10% error rate are encouraging for our future work.

Acknowledgements

This system is a part of the RESIDE-HIS(REconnaissance de SItuations de DEtresse en Habitat Intelligence Santé) project, financed by IMAG, a collaboration between CLIPS and TIMC laboratories. This project is also a part of a collaboration between CLIPS laboratory and the International Research Center MICA (Hanoi – Vietnam) a partly financed by the French National Center of Scientific Research (CNRS) and by the French Embassy to Hanoi-Vietnam.

References

[1] G.Virone, N.Noury and J.Demongeot, “A system for automatic measurement of circadian activity in telemedicine,” *IEEE Transactions on Biomedical Engineerin*g, vol.49, no.12, pp.1463–1469, 2002

[2] M.Vacher, D.Istrate, L.Besacier, E.Castelli and J.F.Sérignat, "Smart Audio Sensor for Telemedicine", accepted in *Smart Objects Conference* , Grenoble, France, May 2003, 4 pages

[3] A.Dufaux, *Detection and Recognition of Impulsive Sounds Signal*s, Ph.D. thesis, Faculté des sciences de l’Université de Neuchatel, 2001.

[4] D.Reynolds, “Speaker identification and verification using gaussian mixture speaker models”, in *Workshop on Automatic Speaker Recognition, Identification and Verification,* Switzerland, 1994, 4pages

[5] M.Cowling and R. Sitte, “Analysis of speech recognition techniques for use in a non-speech sound recognition system,” in *Digital Signal Processing for Communication Systems, Sydney-Manl*y, January 2002.

[6] G.Gravier I.Magrin-Chagnolleau and R.Blouet, “Overview of the ELISA consortium research activities,” *2001 : a Speaker Odysse*y, pp. 67–72, June 2001.