A MULTIMODAL PLATFORM FOR DATABASE RECORDING AND ELDERLY PEOPLE MONITORING

Hamid Medjahed, Dan Istrate

RMSE, ESIGETEL, 1 Rue du Port de Valvins, 77 215 Avon-Fontainebleau Cedex, France

Jerome Boudy, Jean-Louis Baldinger, Bernadette Dorizzi, Imad Belfeki, Vinicius Martins

EPH,INT, 9 rue Charles FOURIER,91011 Evry, France

François Steenkeste

U558, INSERM, Toulouse, France

Rodrigo Andreao

Departamento de Engenharia Eltrica, Universidade Federal do Espirito Santo, Vitoria, Brazil

Keywords: Medical Signal Acquisition, Biomedical multimodal database, Healthcare, Wearable Sensors and Systems, Acoustic Signal Processing, Telemedicine.

Abstract: This paper describes a new platform for monitoring elderly people living alone. An architecture is proposed, it includes three subsystems, with various types of sensors for different sensing modalities incorporated into a smart house. The originality of this system is the combination and the synchronization of three different televigilance modalities for acquiring and recording data. The paper focuses on the acquisition step of the system, usage and point out possibilities for future work.

1 INTRODUCTION

As the society is increasingly aging there is an important need to find an intelligent support system able to facilitate the maintenance at home of the disabled and/or old people with safety and providing their autonomy. The maintaining at home in safety of elderly people is a new major challenge to social and health government services: given limited resources, more and more elderly people living alone at home are particularly prone to accidents and falls in the home and can often lie injured and undiscovered for long periods of time. A statistical study indicates that 7% of elderly people have a home accident due to everyday life activity and in 84% of cases a fall occurs (B.Thélot, 2003). In practice all the industrialized countries are affected by this phenomenon.
Very few systems that support the home life and healthcare of elderly persons have been developed to improve quality of life and the alleviation of risk. Among established systems we can mention,

the TelePat project (French RNTS Program) (Boudy et al., 2006) where certain physiological data and the person's activity are measured by different sensors connected to a microcontroller based computing unit, are sent through radio connection to a remote central server application for exploitation and alarm decision. Now, within the Tandem project (French RNTS Program), accelerometer sensors are added to this system for the detection of falls. In the framework of DESDHIS project a medical home monitoring system which use an accelerometer based sensor, infrared sensor, an oxymeter and a blood pressure device has been developed at Grenoble (G.L.Bellego et al., 2006). A system of multi-channel sound acquisition is presented in (D.Istrate et al., 2006a), to analyze in real time the sound environment of the home to detect abnormal noises (i.e., call for helps or screams).
In this article a new multimodal platform for a home remote monitoring is proposed, using a large number of sensors in order to reinforce the secure detection of abnormal situations, in particular patient's

fall event. Our software implementation gathers three subsystems which have been technically validated from end to end, through their hardware and software. This specific platform is multimodal since it allows us to record physiological data via the RFpat (J.L.Baldinger et al., 2004) subsystem, audio information via Anason (D.Istrate et al., 2006a) subsystem and patient's localization through infra-red sensors via Gardien (S.Banerjee et al., 2003) subsystem. An additional simulation process is added and will be integrated to our platform as a way to overcome the lack of experimental data required to design the decision part of the system, such as the cardiac frequency during distress situations.

2 MONITORING SYSTEM HARDWARE ARCHITECTURE

We define an intelligent environment as one that is able to acquire and apply knowledge about its inhabitants and their surroundings in order to adapt to the inhabitant and to improve its comfort and efficiency. To record the multimodal medical database our first aim is focused on providing such an environment. We consider our system as an intelligent agent, which perceives the state of the environment using sensors and acts consequently using device controllers. The first part of this intelligent environment was realized within the framework of TelePat project, in order to study the secure detection of patient's fall event. The present work is developed in the framework of the Tandem project.

Our platform is a surface of 20 m2 in our laboratory which is arranged in two rooms with a technical area in order to evaluate and to supervise the experiments. It integrate smart sensors (infra-red, audio, physiological,) linked to a smart PC .The two microphones for audio monitoring are linked to the PC through an external sound card, and can be interpreted as a single smart audio sensor for the Anason software. Eight infra-red sensors are fixed on specific places of the house (walls and ceiling) and connected to an acquisition card (ADAM) (F.Steenkeste et al., 1999), which is linked to the serial port of the PC. The card output is RS485 which is converted in RS232 in order to allow Gardien software to acquire the patient position at any time. The RFpat subsystem is composed of two main components: (1) a wearable terminal carried by the patient, continuously recording his physiological data and urgency call, (2) an in-door reception base station linked to the PC via RS232 serial link providing the information usually every 30 seconds. The layout of our house environment is shown in Figure 1.
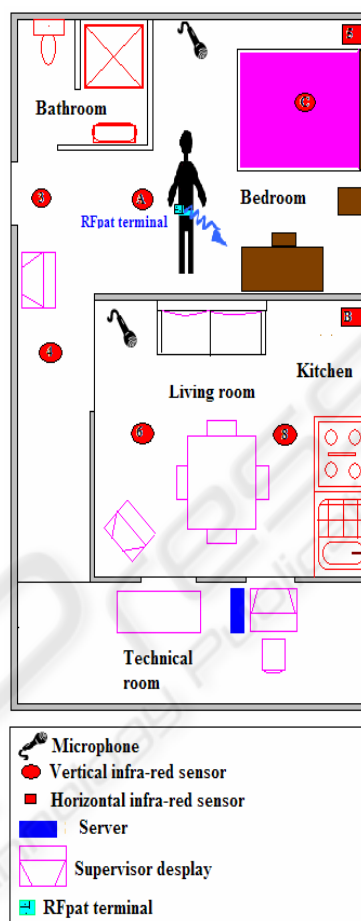
Figure 1: Layout of the house environment.

3 MONITORING SYSTEM SOFTWARE ARCHITECTURE

The multimodal system has three main subsystems like in Figure 2 and provides a general user interface which encapsulates the Anason subsystem. It is implemented under LabWindows/CVI software and communicates with RFpat subsystem and Gardien subsystem by client-server model using TCP/IP and appropriate application protocols. Gardien is implemented under C++ and recovers data every 500 milliseconds. RFpat is also implemented under C++ and receives data from receiver every 30 seconds. The use of the inter-module communication through TCP/IP socket allows each module (subsystem) to be run on a different computer, and to synchronize each televigilance modality channel. The user can interact with the system via internet navigator and supervises the different applications. For example, we use this web server to communicate with the person, who inter-
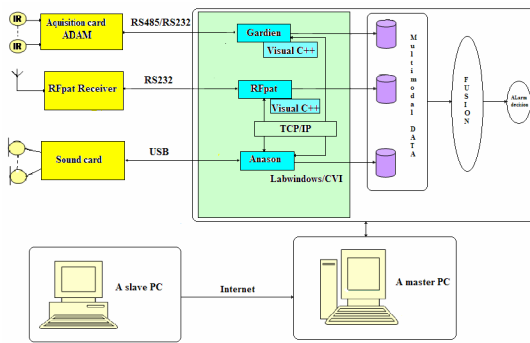
Figure 2: Software architecture of the system.

prets a patient's activity by displaying a reference scenario on the monitoring screen. This feedback can significantly help the system manipulation. The system flexibility obtained through TCP/IP socket communication allows to add others sensors like heart monitoring sensor (ECG).

Currently these three modalities work individually, we investigate multimodal data fusion methods by exploiting the measurements coming from the platform in order to increase reliability and to reinforce the secure detection of patient's distress events.
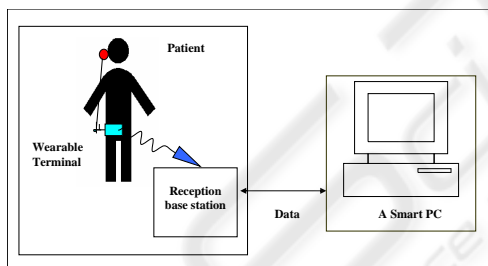
3.1 RFpat

Figure 3: Architecture of RFpat system.

RFpat subsystem is composed of two fundamental elements (Figure 3):

- Wearable terminal carried out by the patient continuously recording his physiological and activity data.

- A reception base station (receptor connected to a PC), which receives signals from the patient's terminal, analyses data in order to generate an alarm after identification of an emergency situations.

All the sensors data are processed within the wireless portable device by using low consumption electronic components in order to face autonomy problems which are also crucial in that application. The circuit architecture is based on different micro-controllers

devoted to acquisition, signal processing and emission. The wearable terminal includes chain of various physiological signals, their possible pre-processing in order to eliminate the power-line interference signal (50 Hz) and the various measurement noises, such as generated by the displacements of the sensors fixed on the patient's body. The latter type of noise is generally a factor limiting the use of such systems in ambulatory mode because the patient is often moving, even if slightly. In this system, the noise problem was solved in the acquisition stage of the portable device, by applying a digital noise reduction filter to the different sensors signals, movements, attitude and namely the pulse signal (heart rate). The performances of signal acquisition could be substantially improved when the patient performs movements. The noise reduction processing (J.L.Baldinger et al., 2004) reduces the variations of pulse measurement lower than 10%, even 5%, which remains in conformity with the recommendations of the health professionals.

The design of sensors and embedded processing has led to the realization of a remote wearable monitoring terminal, equipped with actimetry and physiological sensors, indicating the attitude of the patient (vertical/horizontal positions, activity) and his heart rate (pulse measurement); these specific sensors to recorded physical data type are, either integrated in the terminal, or externally fixed .

Data generated from the different sensors are transmitted, via an electronic signal conditioner, to a micro-controller based computing unit, embedded in the mobile terminal fixed on the patient's waist. Currently, a fall-impact detector sensor is added to this system for robustizing the detection of falls.

3.2 Anason

The sound remote monitoring subsystem analyzes the acoustical environment in real time and is made up of four main modules which are presented in the Figure 4 (D.Istrate et al., 2006b). The first module M1 continuously, supervises the sound environment in order to detect and extract useful sounds or speech from environmental noise. The signal extracted by the M1 module is classified like sound or speech by the M2 module. In the case of sound label, the sound recognition module M3.1 classifies the signal between eight predefined sound classes, while in the case of speech label, the extracted signal is analyzed by a speech recognition engine in order to detect distress sentences. For both cases, if an alarm situation has been identified (the sound or the sentence belong to an alarm class) this information is sent to the data fusion system.
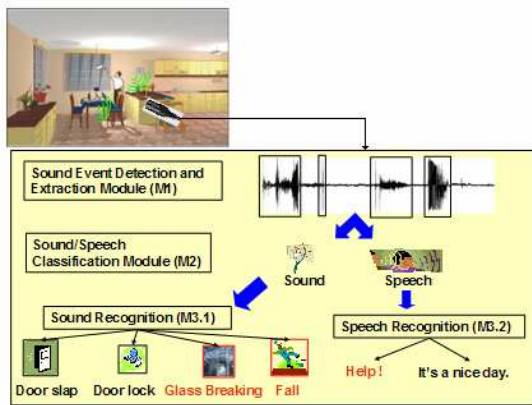
Figure 4: Sound monitoring architecture.

Sound Event Detection Module (M1). The sound flow is analyzed through a wavelet based algorithm aiming at sound event detection. This algorithm must be robust to noise like neighborhood environmental noise, water flow noise, ventilator or electric shaver. Therefore an algorithm based on energy of wavelet coefficients was proposed and evaluated in (D.Istrate et al., 2006a). This algorithm detects precisely the signal beginning and its end, using properties of wavelet transform.

Sound/Speech Classification Module (M2). The method used by this module is based on Gaussian Mixture Model (GMM) (D.A.Reynolds, 1995) (K-means followed by Expectation Maximisation in 20 steps). There are other possibilities for signal classification: Hidden Markov Model (HMM), Bayesian method, etc. Even if similar results have been obtained with other methods, their high complexity and high time consumption prevent from real-time implementation. A preliminary step before signal classification is the extraction of acoustic parameters: LFCC (Linear Frequency Cepstral Coefficients)-24 filters. The choice of this type of parameters relies on their properties: bank of filters with constant bandwidth, which leads to equal resolution at high frequencies often encountered in life sounds.

The BIC (Bayesian Information Criterion) is used in order to find the optimal number of Gaussians (G.Schwarz, 1978). The best performances have been obtained with 24 Gaussians.

Sound Recognition Module (M3.1). This module is based, also, on a GMM algorithm. The LFCC acoustical parameters have been used for the same reasons than for sound/speech module and with the same composition: 24 filters. The method BIC has been used in order to determine the optimum num-

ber of Gaussians: 12 in the case of sounds. A log-likelihood is computed for the unknown signal according to each predefined sound classes; the sound class with the biggest log likelihood is the output of this module.

Speech Recognition Module (M3.2). For Speech Recognition, the autonomous system RAPHAEL is used (M.Akbar and J.Caelen, 1998). The language model of this system is a medium vocabulary statistical model (around 11,000 words). This model is obtained by using textual information extracted from the Internet as described in (D.Vaufreydaz et al., 1999) and from "Le Monde" corpora. It is then optimized for the distress sentences of our corpus. In order to insure a good speaker independence, the training of the acoustic models of RAPHAEL has been made with large corpora recorded with near 300 French speakers (J.L.Gauvain et al., 1990): BREF80, BREF120 and BRAF100 corpora.
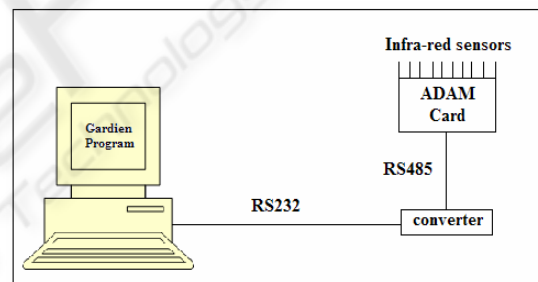
3.3 Gardien

Figure 5: The Gardien system.

The subsystem knows as Gardien (Figure 5) consists of passive infra-red sensors placed in a residence and connected to a remote computer. All the sensors are connected through cables to an Input/Output parallel card (ADAM 4053) which is connected to a master PC. The computer automatically captures and registeres data obtained from the different sensors, with the help of Gardien software. Data corresponding to movements are collected twice per second, and stored with the time of the event in a specific file. When several consecutive data are identical, only the first instance is stored.

The sensors are activated by passage of person underneath, and remained activated as long as there is movement under that sensor and for an additional time period of $1/2$ seconds after the movement end.

The results from the automatic processing of this data are displayed in the form of list with all movements noted together with the time and each movement's

duration. Gardien is also able to display the data either in the form of graph (activity duration versus days) or as three-dimensional histograms (each sensor activation versus time). To validate the system, the results from the automatic processing are compared with manual analysis by an expert.
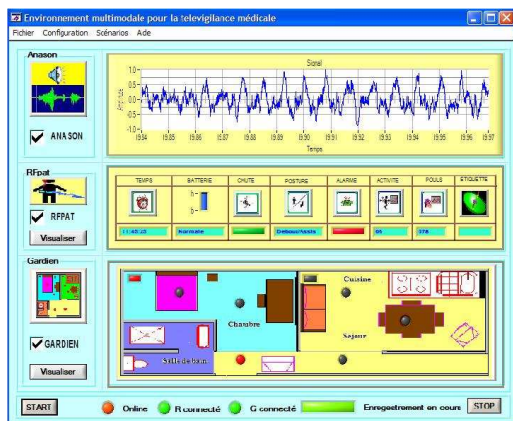
4 GENERAL INTERFACE OF THE SYSTEM

Figure 6: Main windows of the system during the acquisition step.

Figure 6 shows the front panel of the software system, where we can supervise the multimodal data acquisition step. The user must firstly select the modality to record and to configure its parameters. RFpat and Gardien need only to select the IP address and the TC/IP port number, while Anason requests the selection of the sound card (if two are present), the sampling rate and the location of the backup file.

5 THE BIOPHYSICAL SIGNALS SIMULATION STAGE (BSS)

The aim of this stage is to create pathological or critical situations for the patient at home. Indeed most of actual signals recorded on domotic platform are generally and hopefully in normal conditions. The simulator is based on the existing RFpat sensors device. The first main goal was to simulate cardiac pathological profiles such as in particular bradycardias: the design was done with the helpful collaboration of SAMU-92 (French emergencies service. In its implementation, are also foreseen functional stages for the actimetry simulation: patient's inclination (horizontal or vertical position), his body movement and
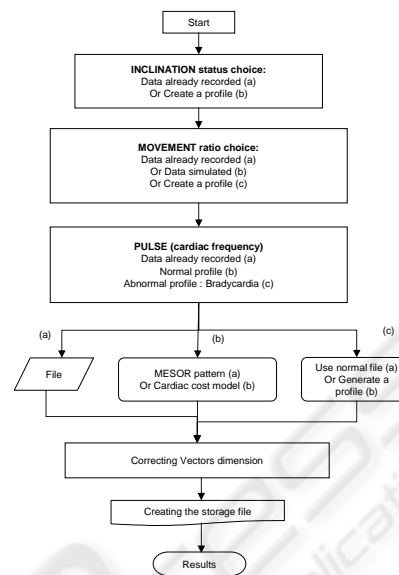

Figure 7: The Biophysical Signals Simulation.

in a larger extent patient's fall situations. The simulator software architecture is summarized in the Figure 7. For the cardiac frequency generation, three cases were proposed:

- A first normal cardiac category, based on the COSINOR method (F.Halberg, 1969), providing a global pulse variations trend within one day; this formula gives the cardiac frequency or pulse Fc in quiet situation under the following form: $Frest(t) = Fmoy + A\sin(2pi/24*t)$, where Fmoy (around 70 bpm), A are respectively the average pulse or MESOR value and its maximal amplitude variation (about 6 bpm) along one circadian cycle; the Akrophase or maximal amplitude is statistically located around 16 hour.

- A second normal situation, called "Cost model", providing a pulse variation model, still denoted Fc, depending on the patient's activity; the formula is based on the pulse in a quiet situation (Fcrest) and the delta-variation due to patient efforts, namely representing the cardiac cost: $Fc(t) = Frest(t) + DFc(t)$. This part is developing stage.

- Bradycardia model corresponding to a situation met with elderly persons, by assuming no specific medication having a cardiac impact: the model is either artificially inserted inside an existing pulse signal sequence by taking into account the actual pulse variance, or is completely substituting the actual pulse sequence.

This tool is still open to other simulation process, namely for the actimetry where presently are per-

formed investigations on the potential correlations between the Cardiac cost model and the body movemenent. The BSS stage has been designed to be fully interfaced to the multimodal patient database.

6 APPLICATION

6.1 Recording of a Multimodal Medical Database

Most of monitoring systems use some form of learning method to discriminate between different types of normal and abnormal events. This methodology requires large amounts of training data that can be difficult to obtain especially data describing abnormal events that are by definition rare occurrences. An important issue for this problem, is to record a multimodal medical database which is the first application of our platform.

Data acquired from the patient are stored on the Master PC in a folder named with a code number corresponding to the patient. Each recording is composed from five files corresponding to the different subsystems.

The first one, named "personnel.xml", contains the patient's identifier and some personal information like age, native language, usual drugs treatments, etc. The second, named "scenario.xml", describes the reference scenario. All these data relative to the tester are protected for his privacy and let to his agreement.

The sound data is saved in real time, in a wav file with 16 bit of resolution and a sampling rate of 16 KHz, a frequency usually used for speech applications.

The clinical data acquired from RFpat are saved in a separate file which contains information about patient's attitude (lied down or upright/seated), his agitation (between 0% and 100%), his cardiac frequency, fall events and emergency call in a binary type. The acquisition sample rate is 1/30 seconds.

The data acquired every 500 milliseconds by Gardien subsystem are saved in a separate text file, fully respecting the original storage format of the GARDIEN application. Each line of this file contains the infrared sensors which are excited (they are represented by hexadecimal numbers from 1 to D) plus the corresponding date and hour.

To tackle the problem of the variety of each data sampling rates, a synchronization prototype between the three subsystems is obtained through TCP/IP protocol. RFpat is the master and supervise Gardien and Anason by TCP/IP commands.

Thus, our multimodal database acquisition software provides a very helpful and well-targeted application to elaborate and assess the data fusion-based decision methods. The low level data recorded by our system will be useful for the development of processing algorithms of each modality.
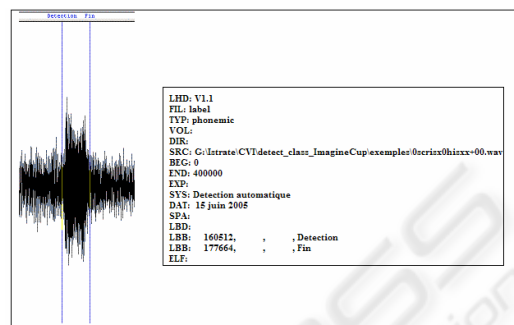

Figure 8: Sound file (*.wav) and its corresponding SAM file.

In order to index our multimodal database, we have retained the SAM standard indexing file (D.Well et al., 2004) generally used for Speech Databases descriptions. The SAM labeling of a sound file is shown in Figure 8; it indicates information about the sound file and describes this file by delimiting the useful part for analyzing and processing. For each modality of the database a corresponding indexing file is created, we have adapted this type of files to the specificities of each modality, and we have added another indexing file for the entire database.

6.2 The First Approach of Fusion: Bimodal Fusion between RFpat and Gardien

This work is a first step for a multi-modal experiment. Indeed, this was performed with only two of the tele-vigilance modalities presented in this paper: the fixed infra-red sensors based Gardien system and the mobile sensors RFpat device. Its conclusions have motivated the extension of these modalities to the combination with the sound detector AnaSon through the current investigations led by ESIGETEL, INT and IN-SERM.

In this first step we have used a PCA analysis in order to preliminary determine potential correlation between the combined data and in order to obtain a data reduction. After preliminary evaluations with the K-Nearest Neighbors algorithm, the Gaussian Mixtures Models and Neural Network on RFPAT data, a Bi-modal fusion was carried out by using the Neural Network .

6.2.1 Principal Components Analysis

The data resulting from RFpat subsystem and Gardien subsystem were analyzed simultaneously through their respective variables: posture, activity, cardiac frequency, exposure time under the infra-red sensor C3 (sensor indicating the input/output of the person in the home), and exposure time under the infra-red sensor C5 (sensor indicating the fall of the person). The analysis of the PCA algorithm results made it possible to propose a set of decision rules on several levels:

- To define an estimator in two levels: a "physiological" distance between two parameters (cardiac frequency, activity) normally correlated: normal state if they are close or pathological state if they are distant. Then, a "actimetric" distance (Slope, C5): normal if distant or pathological if close.

- There are a correlation between the cardiac frequency and the activity which will allow the fusion system to avoid a malfunction of one of the two sensors.

6.2.2 Application of the Neural Networks

The Neural Networks (NN) consist in an input layer, the sensors signals, several transition layers (denoted as hidden layers) and of an output layer delivering the classification of the data observed in situation either 'Normal', or critical 'detected Fall'. A classical NN structure was implemented by using a Multi-Layer Perceptron (MLP) based on only one hidden layer consisting in eight neurons after an optimal tuning.

Each neuron realizes a scalar product between its input vector and the weight vector, where a deviation is added, then operates an activation function in order to generate its output value y: $y = f(x.w + b)$.

The activation function must be strictly crescent and bounded. A classical function used in our experiment is the standard sigmoid function whose equation is reminded hereafter: $f(x) = \frac{tanh(x)+1}{2}$.

Two types of networks were compared, with respectively as input vector of the MLP first layer:

- Single actimetric data of RFpat in entry of the network, giving a rate of recognition of the order 84%.

- The actimetric data of RFpat and horizontal infared sensors of Gardien, providing a rate of 86%.

The improvement nevertheless remains quite limited. One improvement track is to increase the data corpus used for the learning phase, namely by recording more specific actual and simulated emergency situations thanks to the multi-modal recording tool previously described in this paper. Another main improvement track will be investigated by adding the AnaSon (abnormal sound detector) modality. Therefore that is why the need of a new multi-modal recording tool was considered as crucial for the follow-up. Thorough investigations will also be performed again on KNN and GMM techniques, namely by working on the data pre-processing (normalized, transformed input data).

7 CONCLUSIONS

This paper has focused on the technology used for implementing the acquisition step of the platform. Preliminary results are encouraging with the achievement a multimodal medical database including patient's clinical data, usual environment sounds and patient localization. The platform enables us to have a full and tightly controlled universe of data sets and to evaluate the decision part of remote monitoring systems.

Our platform is in the research phase targeting a prototype, the system will be completed and improved by adding a data fusion-based decision element exploiting the measurements coming from this platform in order to propose new processes to reinforce the secure detection of patient's distress events. In particular the fall situations are studied: indeed one or more televigilance modalities might be out of order, or a particular environmental situation (ambiant noise, bad wireless conditions, sensors disabilities …) can hide one particular modality or more. This is a very challenging issue for hospital emergency units such as for instance SAMU in France or Telecare services providers in general. Studies of usability are planned, in order to test the satisfactoriness of patients towards this system and to get a standardization prototype for our platform. This constitutes indeed, a first concrete step before a prototype deployment. In actual situation, evaluation and connection to smart home system are also planed to be performed in the framework of a new European project.

REFERENCES

Boudy, J., JL.Baldinger, F.Delavault, Farin, D., R.Andreao, S.Torres-Muller, A.Serra, F.Rocaries, C.Dietrich, F.Steenkeste, A., Dr.M.Baer, Dr.Ozguler, and D.Gaiti (June 2006). Telemedecine for elderly patients at home. In *the TelePat project.ICOST conference ulster Belfast*.

B.Thélot (June 2003). Resultats de l'enquete permanente sur les accidents de la vie courante. In *Reseau*

EPAC,Institut de Veille Sanitaire,Departement maladies chroniques et traumatismes.

D.A.Reynolds (January 1995). Speaker identification and verification using gaussian mixture speaker models. *Speech Comm*, 17:91–108.

D.Istrate, E.Castelli, M.Vacher, L.Besacier, and J.F.Serignat (April 2006a). Information extraction from sound for medical telemonitoring. *Transactions on Information Technology in Biomedicine*, 10:264–274.

D.Istrate, M.Vacher, and F.Serignat, J. (Aug 30-Sept3, 2006b). Generic implementation of a distress sound extraction system for elder care. In *28th IEEE EMBS Annual International Conference,New York City, pp.3309-3312*. IEEE.

D.Vaufreydaz, J.Rouillard, and M.Akbar (December 1999). Internet documents: a rich source for spoken language modelling. *Workshop ASRU'99, Keystone-Colorado*, pages 277–281.

D.Well, J.Barry, W.Grice, M.Fourcin, and A.Gibbon (2004). Sam esprit project 2589-multilingual speech input/output assessmen,methodology and standardization. In *Final report. Technical Report SAM-UCL-G004, University CollegeLondon*.

F.Halberg (1969). Chronobiology. In *Ann Rev Physiol, vol.31,pp.675-725*.

F.Steenkeste, H.Bocquet, M.Chan, and B.Vellas (December 1999). Remote monitoring system for elders in a geriatric hospital. In *Promoting Independence and quality of life for older persons:an international conference on aging Arlington*.

G.L.Bellego, N.Noury, G.Virone, M.Mousseau, and J.Demongeot (January 2006). Measurement and model of the activity of a patient in his hospital suite. *TITB*, 10:92–99.

G.Schwarz (1978). Estimating the dimension of a model. *Annals of Statistics*, 6:461–464.

J.L.Baldinger, J.Boudy, B.Dorizzi, J.P.Levrey, R.Andreao, C.Perpre, F.Devault, F.Rocaries, and A.Lacombe, C. (Juillet 2004). Tele-surveillance system for patient at home: The medeville system. In *Congrs ICCHP 2004,Paris*.

J.L.Gauvain, L.F.Lamel, and M.Eskenazi (1990). Design considerations and text selection for bref, a large french read-speech corpus. In *ICSLP '90, pp. 1097-1100*.

M.Akbar and J.Caelen (1998). Parole et traduction automatique : le module de reconnaissance raphael. In *COLING-ACL'98, Montral, Quebec, vol.2, p. 36-40*.

S.Banerjee, F.Steenkeste, P.Couturier, M.Debray, and A.Franco (Septembre 2003). Telesurveillance of elderly patients by use of passive infra-red sensors in a smart room. In *Telemed Telecare*.