Indoor Personal Monitoring, Supervising and Assistance Sweet-Home and AmiHomeCare case studies

Michel Vacher*, François Portet*, Pedro Chahuara*
Simona Caramihai, Calin Munteanu**, Irina Mocanu**, Stefan Mocanu****

**Laboratoire d'Informatique de Grenoble UMR 5217,*
CNR /UJF/Grenoble-INP/UPMF, F-38041, Grenoble, France
(e-mails: Michel.Vacher@imag.fr;Francois .Portet@imag.fr,Pedro.Chahuara@imag.fr)
***Automatic Control and Computer Science Faculty,*
University POLITEHNICA of Bucharest, Bucharest, Romania
(e-mails: caramihai2002@yahoo.com;mc_aurel@yahoo.com;irinag.mocanu@gmail.com;smocanu@rdslink.ro)

Abstract: This paper proposes a new solution for an e-health system architecture related to personal assistance at home for aged people. This architecture takes into account the results of two ongoing projects on which the authors are involved: Sweet-Home and AmiHomeCare and the possibility of combining different parts of these projects in order to obtain a more versatile system. Each project has a distinctive way of gathering the required information: Sweet-Home uses sound processing, AmiHomeCare – video processing. The paper analyzes the advantages and drawbacks of both approaches in order to propose an integrative new architecture.

Keywords: ambient intelligence, voice recognition, wireless body sensors, body posture recognition, gesture recognition

1. INTRODUCTION

Ambient Intelligence (AmI) is a new paradigm in artificial intelligence that considers ICT environments to be fully aware of the user's needs and preferences. AmI is greatly influenced by the Ubiquitous Computing domain. As introduced by (Weiser, 1991), ubiquitous computing refers to the computing technology which disappears into the background, which becomes so seamlessly integrated into our environment that we do use it naturally without noticing it. One of the very active AmI applications is the Smart Home domain. A Smart Home is a habitation equipped with a set of sensors, actuators, automated devices and centralized software controllers specifically designed to respond and anticipate the user's needs (Chan *et al.*, 2009).

Another very active research application of ambient intelligence is Ambient Assisted Living (AAL) that aims to support elderly or disabled people in their daily lives, allowing them to live in their homes or their familiar community as long as possible instead of being moved to nursing homes. This means that the user's activity and different contexts must be analysed all day long in order to provide the relevant assistance.

The intersection of Smart home and AAL naturally gave rise to the notion of health smart home which is specifically designed for daily living task support, early detection of distress situations, remote monitoring and promotion of safety and well-being.

A large number of research projects have contributed to the field; some of them being: House_n (Intille, 2002), CASAS (Chen and Cook, 2012), ISpace (Holmes *et al.*, 2002), Aging in Place (Skubic *et al.*, 2009) or SOPRANO (Wolf *et al.*, 2008), etc. However, despite many contributions, there is still no consensus about the solution to adopt to address the AAL challenges in smart home.

The research presented in this paper develops and extends the work performed by the authors in two ongoing projects. The features of these projects (Sweet-Home and AmiHomeCare) are presented in this paper as well as the architecture of a system that combines those features. Both systems aim at assisting people in their homes by continuous analysis of the environment using different approaches: SWEET-HOME relies on a speech interface to allow the user to control his home while AmiHomeCare uses video processing.

In the SWEET-HOME project, the targeted users are elderly people who are frail but still autonomous. AmiHomeCare final users are elderly or disabled people though the technology can be used by anybody. This population has special needs that must be addressed, as it will be shown next.

SWEET-HOME and AmiHomeCare are presented in detail in Sections 2 and 3 respectively. The advantages and drawbacks of each technique along with the new developed architecture are presented in Section 4.

2. THE SWEET-HOME PROJECT

The main goal of the SWEET-HOME (http://sweet-home.imag.fr/) project is to develop an intelligent home automation system to provide *natural man-machine* *interaction* by using voice commands and *security reassurance* by detecting situations of distress, and to facilitate *social interaction* (Vacher, 2011a). In this section we focus on the context-aware voice interaction inside aspect of sweet-home.

2.1 Speech analysis challenges

The audio analysis (of sounds and speech) has a great potential for monitoring the person, disability compensation, assistance, and security enhancement. It can also be useful for improving and facilitating communication of the person with the outside. However, as previous experiments confirmed (Vacher, 2011b), this technology (recognition of speech, dialogue, speech synthesis, sound detection) must take into account the very difficult environment of the home. Speech recognition is already an established research area which, despite many progresses in close talking situations, must address numerous challenges in noisy distance speech context as in a home with microphones set in the ceiling. Sound recognition in domestic areas (washing machine, keys, door…) has been considered as an interesting research area in smart home since the 2000's. Though it is not directly useful to command a smart home, when well detected, it can support health monitoring (e.g., scream, cough, snoring…) and can serve for disambiguation and activity recognition purpose (e.g., dishes manipulation, water, glass break…).

Recent developments have produced significant results and enabled the Automatic Speech Recognizer (ASR) to be a component of many industrial products, but there are still many challenges to make this feature available for Smart Homes. For instance, ASR systems achieved good performance when the microphone is placed near the speaker (e.g., headset), but the performance degrades rapidly when the microphone was placed at several meters (e.g., in the ceiling) (Wölfel, 2009). This is due to different phenomena such as the presence of noise background and reverberation (Vacher, 2008).

Several experiments carried out in automatic speech recognition showed degradation of performances with 'atypical' people such as children or the elderly (Wilpon, 1996; Vipperla, 2008; Gerosa, 2009). Other studies (Gorham-Rowan, 2006) emphasized the effects of ageing on speech production and the implications this has on speech recognition. The elderly speakers are characterized by a trembling of the voice, hesitations, and the production of inaccurate consonants, a breaking voice, and a slower articulation. The speech recognition from elderly voice is an under explored area (Vacher, 2012a).

Given the increased trend in fitting houses with more and more sophisticated ICT devices, the question of privacy in ones own home is emerging (van Hoof, 2007; Sharkley, 2012). As any other technologies, speech recognition must respect the privacy of the speaker. Therefore, the language model must be adapted to the application and should not allow recognizing phrases whose meaning is not essential to the application; a keyword recognition system respects this constraint. Moreover, it is an open question whether the system can store the audio signals or not (and for how long).

Regarding the acceptability aspect, a system will be far better accepted if it is a useful daily living assistant rather than an occasional one (e.g., fall detector). A general system covering surveillance, home automation and detection of distress, would be more easily accepted than scattered one-purpose ICT devices. Acceptability is key aspect of the successful development of smart homes. In accordance with the user centered approach method, we conducted a specific study on the acceptability of a voice interface for older people.

The targeted users of the SWEET-HOME project are elderly people who are frail but still autonomous. There are two reasons for this choice. Firstly, a home automation system is costly and would be much more profitable if it is use in a life-long way rather than only when the need for assistance appears. Secondly, in the case of a loss of autonomy, the person would continue to use their own system with some adaptations needed by the new situation (e.g., wheelchair) rather than having to cope simultaneously with their loss of autonomy and a new way of life imposed by the smart home. To assess the acceptance of this new technology, a qualitative user evaluation was performed (Portet, 2013). 8 healthy persons between 71 and 88 years old, 7 relatives (child, grand-child or friend) and 3 professional carers were questioned in co-discovery in a fully equipped smart home alternating between interview and wizard of Oz (wizard of Oz, 2013) periods. The four important aspects of the project have been assessed: voice command, communication with the outside world, home automation system interrupting a person's activity, and electronic agenda. In each case, the voice based solution was far better accepted than more intrusive solutions such as video camera. Thus, in accordance with other user studies (Lopez-Cozar, 2010), audio technology appears to have a great potential to ease daily living for elderly and frail persons. To respect privacy, it must be emphasized that the adopted solution analyses the audio information on the fly and does not store the raw audio signal. Moreover, the speech recognizer is made to recognize only a limited set of predefined sentences which prevents recognition of intimate conversations.

2.2 Sweet-Home system architecture

The SWEET-HOME system is composed of an Intelligent Controller which analyses the streams of data and makes decision based on these. These streams are composed of all the usual home automation data sensors on the KNX home automation network (button, lights, blinds, etc.), multimedia control (uPnP), X2D (contact-door), and the audio events processed in real-time by the multi-channel audio analysis system: PATSH.

The different modules of the system are shown on Fig. 1. The PATSH real-time framework was developed to manage the acquisition/processing flow of the sound events detected. Several threads are plugged in and synchronized, such as the sound and speech analysis modules and the multichannel data acquisition card (NI-DAQ6220E). The system is composed of the following 4 stages which are fully described in (Sehili, 2012; Vacher, 2012b): Sound Acquisition and Detection stage, Sound/Speech Discrimination, Sound Classification and Automatic Speech Recognition (ASR) stages.

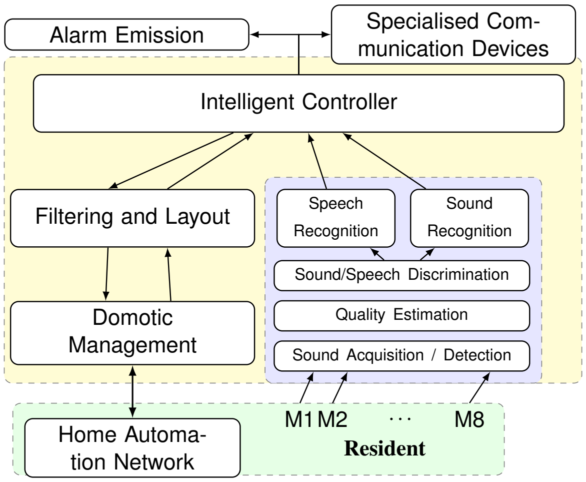

Fig. 1. The SWEET-HOME architecture (Vacher, 2011a).

First tests in a real Smart Home have shown 38% of home automation error rate, including missed detections and sentences classified as sound. Theses poor performances are explained by the fact that PATSH was not successful in selecting the best audio event among the set of simultaneous event and thus the event with low SNR introduced errors and were not properly segmented.

2.3 Multimodal Corpus recorded in a Smart Home

To provide data to test and train the different processing stages of the SWEET-HOME system, experiments were run in the DOMUS smart home that was designed by the Laboratory of Informatics of Grenoble to observe users' activities interacting with the ambient intelligence of the environment. Fig. 2 shows the details of the flat. It is a thirty square meters suite flat including a bathroom, a kitchen, a bedroom and a study, all equipped with sensors and effectors so that it is possible to act on the sensory ambiance, depending on the context and the user's habits. The flat is fully usable and can accommodate a dweller for several days. More than 150 sensors, actuators and information providers are managed in the flat (e.g., lighting, shutters, security systems, energy management, heating, etc.).

The flat has also been equipped with 7 radio microphones set into the ceiling that can be recorded in real-time thanks to a dedicated PC embedding an 8-channel input audio card.

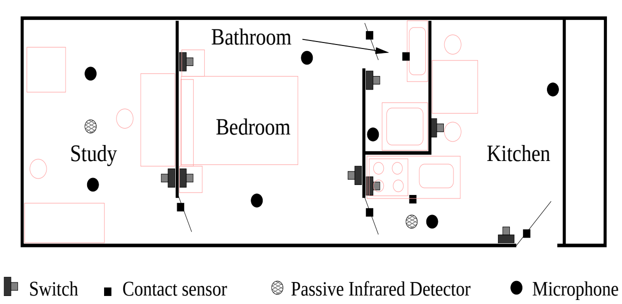

Fig. 2. Position of the sensors inside the DOMUS flat.

21 persons (including 7 women) participated to an experiment to record multimodal data in a daily living context. To make sure that the data acquired would be as close as possible to real daily living data, the participants were asked to perform several daily living activities in the smart home. The average age of the participants was 38.5±13 years (22-63, min-max). The experiment consisted in following a scenario of activities without condition on the time spent and the manner of achieving them (e.g., having a breakfast, simulate a shower, get some sleep, clean up the flat using the vacuum, etc.). During the experiment, event traces from the home automation network, audio and video sensors were captured. Video data were only captured for manual marking up and are not intended to be used for automatic processing. In total, more than 26 hours of data have been acquired.

2.4 Interaction in context

A context aware system can adapt its behaviour in accordance with the characteristics of the environment, the users, their localisation, the accessible devices, but with the variations of these characteristics along the time too (Schilit, 1994). According to (Dey, 2001), a context aware system must use the context to provide information or services pertinent for the user, the pertinence being depending on the task.

In a smart home based on voice command voice command must be interpreted by a decision module to generate the correct actions. But when the information is imprecise and incomplete, the decision module must rely on context to make the best decision. For instance, when the decision module receives the vocal "*Nestor allume la lumière*" (Nestor[1] turn on the light), this order give only partial information: the light must be put on, but generally there is not only one lamp in the flat, in what room must be the lamp put on? Therefore, it must be necessary to determine in what room the person is standing. This is the first information related to context. Moreover, there is no information about the intensity of the lighting. If the person is reading, a high intensity is required; but if she is waking on in the night to go to the toilets, a low intensity will be preferred. The second information related to context is then partial information on the activity of the person at this time.

2.5 Context Retrieval in the Smart Home

The input of the SWEET-HOME system is composed of the information from the home automation system transmitted via a local network and information from the microphones transmitted through radio frequency channels. Then there is no direct information about localization or activity and this information must be inferred from data given by sensors which are imprecise, uncertain and incomplete.

[1] *Nestor* represents a keyword that activates the system, letting it know that what follows represent a command that the system is expected to perform (see paragraph 2.6).

2.5.1 Localisation

The method developed for locating a person from multiple sources is based on the modeling of the links between observations and location assumptions by a two-level dynamic network (Chahuara, 2011). The dynamic network that we designed is organized in two levels: the first level corresponds to location hypotheses generated from an event; and the second level represents the occupation context for each room whose weight of activation indicates the most likely location given the previous events (as presented in Fig. 3). Location hypotheses correspond to area where the person can be at a specific time while occupation contexts correspond to rooms in which the person is over time.

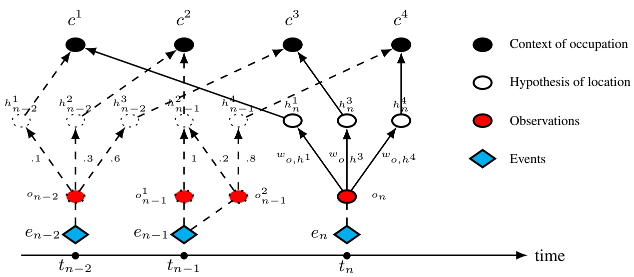

Fig. 3. Example of dynamic network (Chahuara, 2011).

A weight must be associated to each sensor, and the activation for each microphone is computed using the signal-to-noise ratio (SNR) estimated in real-time. However, conversely to contact sensors on furniture and windows which are always linked to a unique hypothesis, contact sensors on the indoor doors can be ambiguous regarding the location. The problem is to decide which of the two rooms around the door should have the highest weight. In that case, the weight is taken into account the localization at time t_{n-1}.

Using this method, the localization had improved the accuracy from 62.8 % to 84.0% using three sources of information (switch, PIR and speech occurrences).

2.5.2 Activity Recognition

Smart Homes and home automation impose constraints on the sensors and the technology used for recognition. Indeed, information provided by the sensor for activity recognition is indirect (no worn sensors for localization), heterogeneous (from numerical to categorical), transient (no continuous recording), noisy, and non visual (no camera). This application setting calls for new methods for activity recognition which can deal with the scarcity and unreliability of the provided information, process streams of data, and whose models can be checked by human and linked to domain knowledge. To this end, a method based on Markov Logic Network (MLN) (Richardson, 2006) that combines first-order logic and Markov Networks was used. A MLN is composed of a set of first-order formulae each one associated to a weight that expresses a degree of truth. This approach softens the assumption that a logic formula can only be true or false. A full presentation of this method used to recognise activities of daily living in a perceptive environment is available in (Chahuara, 2012). Using MLN, the accuracy of recognition of 7 activities of daily living involving 21 persons was 85.3% against 66.1% for naïve Bayesian network and 59.6% for an SVM approach.

2.6 Vocal orders grammar

Our user study showed that targeted users prefer precise short sentences over more natural long sentences. Possible voice orders were defined using a very simple grammar, a key-word followed by an initiate command or a stop command before the name of the device. Each order belongs to one of three categories: initiate command, stop command and emergency call. Except for the emergency call, every command starts with a unique key-word that permits to know whether the person is talking to the smart home or not. In the following, we will use '*Nestor*' as keyword, it may be '*maison*' (home):

set an actuator on	Example:	'Nestor ferme fenêtre'
	Format:	key initiateCommand object
stop an actuator	Example:	'Nestor arrête'
	Format:	key stopCommand [object]
emergency call	Example:	(e.g. 'au secours')

In case of emergency call, the key-word is omitted.

The sentence '*Nestor ferme la fenêtre*' in French means: Nestor close the window, '*Nestor arrête*': Nestor stop, and '*au secours*': help me.

2.7 Intelligent Controller

The complete SWEET-HOME system is managed by what we call the Intelligent Controller. The Intelligent Controller is responsible for the perception of the environment through the information coming from the home automation system and the audio system. Based on these data it interprets the voice orders, recognizes the user's activity and situation and make context-aware decision to modify the environment to answer the user's needs. Fig. 4 depicts the architecture of the controller.

It is composed of four main parts:

- Interface which handle on-line communication with the home automation system and the audio analysis system;
- Knowledge representation which contains the descriptive and procedural domain knowledge as well as serve as data repository;
- Context analysis which analyze the stream of data to extract information about the user (e.g., activity, location) and the environment (e.g., a window is open);
- Decision which based on a given recognized situation generate command to change the environment (e.g., turn on one light) or to warn the user (e.g., beware the door is not locked).

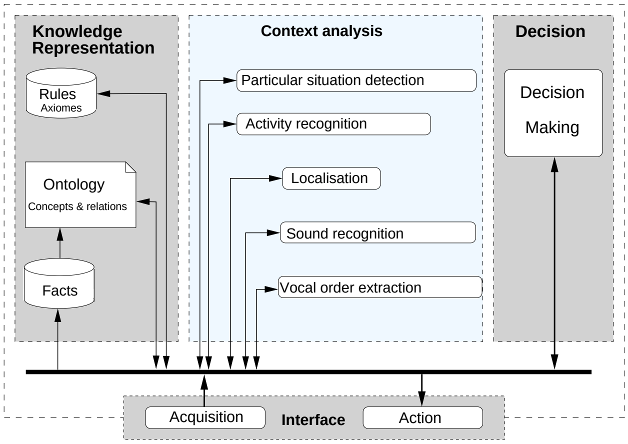

Fig. 4. Architecture of the SWEET-HOME intelligent Controller.

The knowledge representation contains two semantic layers: the *low-level* and the *high-level* ontologies. The former ontology is devoted to the representation of raw data and network information description while the high level ontology represents concepts being used at the reasoning level such as: Actions that can be performed in a home and the context in which a home can be (e.g., making coffee, being late). This separation between low and high levels makes possible a higher re-usability of the reasoning layer when the sensor network and the home must be adapted.

The analysis of the current context is carried out through the collaboration of several processors, each one being specialized in a certain context aspect, such as location detection or activity recognition. All processors share the knowledge specified in both ontologies and use the same repository of facts. Furthermore, the access to the knowledge base is executed under a service oriented approach that allows any processor being registered to be notified only about particular events and saving any inferred information to be available to other processors. This data and knowledge centered approach permits to ensure that all the processors are using the same data structure and that the meaning of each piece of information is clearly defined among all of them. In addition, the chosen architecture is more flexible than a classical pipeline of processors, making possible the easy insertion of new processors. Once the current context has been determined, the controller evaluates if an action must be taken. What supports the process of decision is a set of logic rules which are part of the knowledge base.

3. THE AmiHomeCare SYSTEM

The AmiHomeCare project's goal is to develop a system for home assistance and medical care. This is done by personal activity monitoring, personal assistance, medical care and environment monitoring and control. As presented in

(Mocanu *et al.*, 2011), in AmiHomeCare system's architecture (fig. 5) there is a specific module for each of these functions. These modules are:

- *Multi Agent System for Information Access* (MAS-IA) module, in charge with the personal activity monitoring. The main goal of this module is to monitor daily activities using video surveillance in order to detect emergency situations. The MAS-IA module has a secondary usage represented by the ability to analyze the environment and to provide information for retrieving different objects from it.
- *Elderly or Disabled People Recognition and Assistance* (*EDPRA)* module, in charge with user's recognition and assistance. The recognition part is needed for access to restricted area that can present a high risk to user's safety. The assistance is achieved through gesture recognition and gesture-based interaction with a robotic personal assistant.
- *Medical Home Surveillance Devices Monitoring and Coordination* (*MHSDMCS)* module. It's main goal is to monitor user's vital signs and send alerts in case of emergencies.
- *Intelligent Home Monitoring and Assistance* (*IHMAS)* module, which aims at providing a pleasant and safe environment in terms of monitoring and controlling important ambient parameters and to predict user's needs.

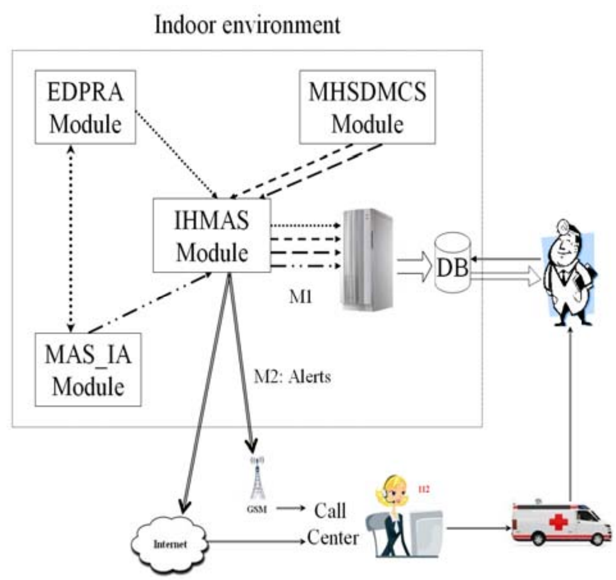

Fig. 5. The AmIHomCare system's structure.

The AmIHomeCare system integrates monitoring of the vital and non-vital signs together with the biometric identification, access control and restriction, customized environment and medical alerts.

Both the MAS_IA module and the EDPRA module will analyse human body posture and human gestures in order to integrate them in the daily activity recognition model or in the interaction with the "robot-like" personal assistant. These two components are grouped in what is called *the supervising system*.

3.1 The supervising system

The purpose of the supervising system is person's monitoring in his house, recognizing its daily activities and interacting with a "robot-like" personal assistant.

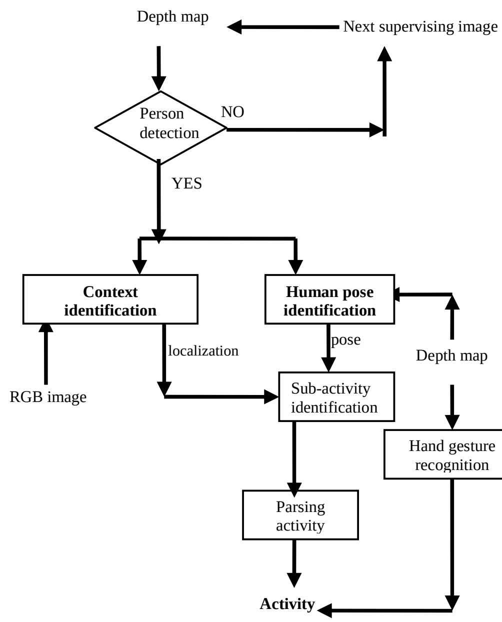

Fig. 6. The supervising system's logic.

The supervising system analyses the images captured by the supervision cameras (Fig. 6). The supervision camera is a Kinect sensor which provides an RGB image together with the depth map. Each activity is composed of a set of successive sub-activities. The context (the location in the room) together with the pose of the supervised person forms a sub-activity.
In order to detect daily activities, the context of the supervised person (its location in the room) together with its static body postures must be identified.

Also gestures made by the supervised person will be analysed in order to detect emergency situations.

For the moment only one person is considered in the smart environment. Also, only simple daily activities (not interleaved activities) are identified.

3.1.1 Daily Activity Recognition

The main steps of the human activity identification, based on the model described in (Mocanu *et al.*, 2013) are:

- Person detection from the depth map;
- Identification of the detected person's context from the RGB image;
- Person's pose identification from the Kinect skeleton, in the recognized context from the and
- Sub-activity identification based on the obtained context and on the person's pose.
- Activity recognition

The context of a person refers to the zone from the room in which the person was detected together with the surrounding objects (furniture objects) from the room. Thus, each room is divided in zones of interest. For example a context of a person is composed of the current zone (we call this zone R) and the nearest furniture objects. The RGB image is analyzed in order to detect the person's context. First the image is annotated so that image objects will get associated keywords. Image annotation is performed with a parallel genetic algorithm, described in (Mocanu, 2010), which determines the best match between each image region and the corresponding objects in the room. Secondly, from the depth map, the supervised person skeletal is computed. The bounding boxes of the objects from the image (the detected person and the furniture objects) are compared. Thus the objects close to the supervised person will be identified. Then these objects will be used to determine the zone of the room in which the person is located. This assumes that a model of the house is available. The house model is a domain ontology – the context ontology, which consists of the rooms in the house, the zones inside each room and the furniture objects. Each room has its component zones and each zone consists of the component furniture objects. For a better person's context identification, each furniture object from the house and the supervising person have associated a depth (the model for the room was computed by scanning the space with the Kinect sensor). Thus each furniture object from the ontology will have an associated list of one or more depth values. Each depth value in the list will be measured considering a known angle of the Kinect sensor. Next the ontology is queried with the objects close to the supervised person. The query result is zone R which contains the majority of objects close to the detected person (under a predefined threshold). The zone R together with the nearest furniture object(s) from R forms the person's context.

The human pose identification is described in (Mocanu *et al.*, 2013). The human body is modeled by its body components. It is used a list of known poses stored in an and-or graph. The pose is obtained using a set of rules modeled as a stochastic context-free grammar, which is transformed into the equivalent and-or graph. The human pose identification is performed probabilistically, by bottom-up constructing a list of parse trees in the grammar. The parse tree with the biggest probability is chosen from the result list.

Each *activity* is decomposed into a sequence of sub-activities. Each *sub-activity* consists of the human pose and its context. A set of successive sub-activities are assembled in an activity using a stochastic context-free grammar.

3.1.2 Gesture Recognition

In this case we extend the model for activity recognition by adding the component for gesture identification (Anton et al. 2012b). Thus an activity will be completed with the performed hand gestures.

Methods for human gestures analyses can be classified in two classes: static and dynamic detection methods, as it is described in (Moeslund *et al.*, 2001). This classification occurs due to various issues that need to treat algorithms used to identify gestures from each category. The dynamic gestures involve several aspects: (i) it is necessary to be able to represent a temporal dependency between actions; (ii) the learned patterns must be invariant in time and (iii) actions must not require alignment over time.

In (Yang, 1999) motion patterns are learned from the extracted trajectories from videos using a time-delay neural network. The experiments described in the paper demonstrate recognition of 40 hand gestures of the American Sign Language. Experimental results show that hand gestures can be extracted and identified with high recognition rate. In paper (Ivanov *et al.*, 2000) is described a model for recognizing general activities using Hidden Markov Models (HMM) and stochastic parsing. These activities are first identified as a series of low-level primitive actions, represented using the HMM and then they are recognized by string parsing using a context-free grammar. In (Aarob *et al.*, 2013) human activities are recognized by matching temporal templates with saved instances of images for known activity. Variable Length Markov Models to model human behavior in order to include temporal dependencies is used in (Galata *et al.*, 2001).

The above approaches require image processing to extract relevant data to detect human body postures. This process involves some limitations, such as depth information which is difficult to extract only from one image. Thus, in (Chu *et al.*, 2013) is adopted a system based on four synchronized cameras to recognize human postures.

Recognition of the human gestures is made by processing the human skeletal obtained from Kinect sensor (Anton *et al.*, 2012a). The information provided by Kinect is processed in order to obtain relevant data for the analyzed gesture, which will be used to train the TDNN. After the training phase we test the learned model and determine the error rate. The system architecture is presented in Fig. 7.

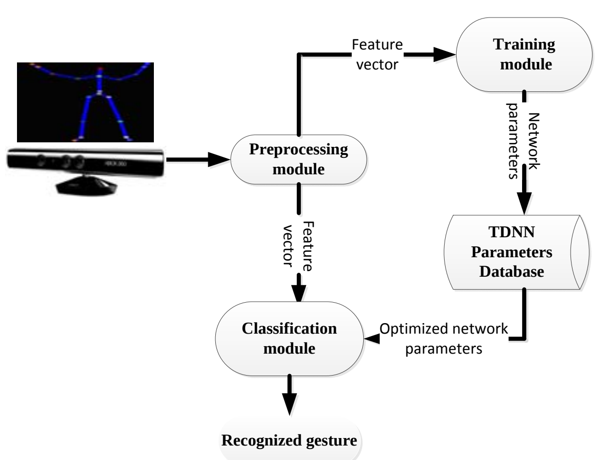

Fig. 7. Gesture recognition information flow.

The main parts of the gesture recognition system are:

- *the preprocessing module* - receives the human skeletal from the Kinect, computes a feature vector for each skeletal;
- *the training module* – trains the TDNN; the input data is composed by the feature vector computed by the preprocessing module. The parameters of the trained TDNN are saved in the TDNN Parameters Database.
- *the classification module* – classifies a new gesture; it receives a feature vector of a new human skeletal from the preprocessing module and classifies it using the learned parameters of the neural network from the TDNN Parameters Database.

The Kinect sensor produced by Microsoft, (Kinect, 2013), uses the PrimeSense's 3D sensing technology and it is composed by three main parts: an infrared laser projector, infrared camera, and a RGB color camera. The depth projector simply floods the surrounding space with Infra Red laser beams creating a depth field that can be seen only by the infrared camera. Due to infrared’s insensitivity to ambient light, the sensor can be used in any lighting conditions. The hardware is the basis for creating an image that the processor can interpret the software behind the sensor is what makes everything possible.

3.1.3 Communicating with a Robot Assistant

In order to use the gestures for communicating with a robot (or any other device) a communication session must be started. The communication session is defined by the following information:

- the communication parameters and connectivity;
- the set of gestures that must be recognized;
- the set of actions associated with the gestures.

The communication parameters and connectivity are required in order to establish a stable communication session with the robot. The most common and facile connectivity method is using a wireless network because the robot is a mobile device, so the components of this setup will be the kinect sensors connected to a PC which has a WiFi network card or an Ethernet card connected with a WiFi router.

The robot must also have a WiFi connection, if the robot doesn't have this feature (which was in our case) can be interfaced with an Arduino Mega 2560 Rev3 and an WiFi Shield. The Arduino board and Shield are connected in the WiFi network and run a communication server. The PC obtains the skeletal information form the Kinect sensors, analyses the information extracting the gestures, associates the gestures with actions and send them through the wireless network to the Arduino board which forward them using a serial connection to the robot. If the robot doesn't have an available serial connection the same command can be send using digital I/O lines.

For testing purposes we used an iRobot Roomba 560 robot, the Arduino board and Shield are powered directly from the battery of the robot (16Vcc) the Arduino board supporting a voltage up to 20Vcc. The serial line is configured to 115200 bps and uses specific robot commands for basic low level robot control iRobot (2005).

The second set of data required for the communication session is the set of gestures which must be recognized. In order to have a stable communication session and to ensure that the gestures will be interpreted correctly, the gestures must be compared and if there are gestures which are very similar and if the probability that they will be misinterpreted is over an imposed threshold, one of the gestures must be removed in order to overcome this problem.

In order to compute the distances between similarities of gestures a set of metrics are used (Bayes, Mahalanobis, Euclid, Chebychev, City_Block, S_Euclid, S_Chebychev). The metrics S_Euclid and S_Chebychev are scaled variants of the Euclid and Chebychev.

The set of actions associated with the gestures is represented by a user defined strings which are sent over the network to a central application when the gesture is recognized. The application logs these strings and/or forwards them to other devices or systems (for example the gestures recognition system sends the messages to the central system, which logs these messages and also forward them to the robot.)

3.2 The monitoring system

The monitoring system consists in two individual and independent modules as described in (Mocanu et al., 2011a). The first module is represented by the *Intelligent Home Monitoring and Assistance System*, or shortly, *Environment Monitoring System*, while the second is represented by the *Medical Home Surveillance Devices Monitoring and Coordination System* or, shortly, *Personal Monitoring System.*

3.2.1 Intelligent Home Monitoring and Assistance System

The IHMAS module is in charge with home monitoring and control of important environmental parameters (temperature and lighting) and with monitoring the life threatening events: CO and CO_2 accumulation, fire, smoke, flooding.

The environment monitoring and control uses wireless sensors for the monitoring part.

The IHMAS module architecture (presented in (Mocanu *et al.*, 2011a)) consists in four components:

- the temperature monitoring and control,
- the lighting monitoring and control,
- special events (life threatening ones) and
- communication interface (both with other AMI-Home-Care system modules and with the exterior represented by a call center or emergency response units).

For the prototype of the module, a wireless sensor network kit was chosen (Crossbow Memsic starter kit - memsic). The kit has (in its sensor nodes) embedded abilities for monitoring among other parameters temperature and lighting and also the capabilities of replacing (for a node) the standard board with a more general board to which external sensors (e.g. – the ones used for special events) can be attached.

3.2.2 Medical Home Surveillance Devices Monitoring and Coordination System

The purpose of this component is to monitor vital signs of the elderly or disabled person while he/she is inside of the AmiHomCare coverage area (house, apartment or even a medical room) and to send alerts to IHMAS component in case an abnormal situation is noticed. From that point, the alert will be relayed to the Call Centre.

The functional architecture of the personal monitoring system is presented in Fig. 8.

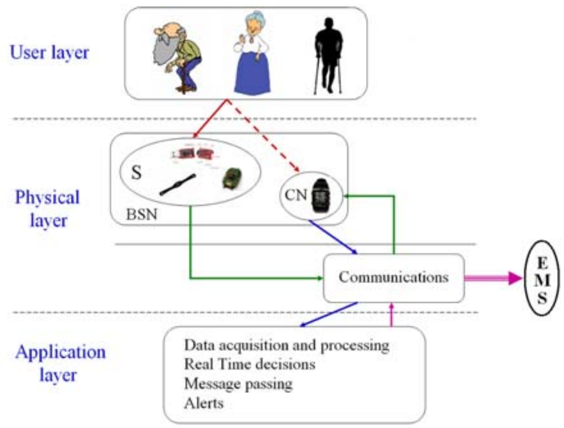

Fig. 8. The architecture of personal monitoring system.

In order to detect abnormal values of the monitored vital signs, the personal monitoring system is based on the concept of Body Sensor Network (BSN) (Chen *et al.*, 2010; Norgall *et al.*, 2006) or Body Area Network (BAN). The main difference stands in the communications between the Body Sensor Units (BSU) and the Body Central Unit (BCU). Unlike a typical BSN, in AmiHomeCare approach, the communications between the terminal nodes (represented by sensors, as one can see in figure 9) and the Central Unit (CN in figure 9) are not limited to wireless channels, wired channels can also be used. The communication between the Personal Monitoring System and the Environment Monitoring System is, on the other hand, exclusively wireless.

The sensor nodes can acquire information such as body temperature, heart rate, acceleration, oxygenation level, blood pressure, movement, etc. and send it to the Central Node where all the computation is made. The result is sent to the Environment Monitoring System under the form of a log message (in case no abnormal situation was detected) or an alert (if the system considers that monitored person is in immediate danger). Some of the parameters will not reveal an immediate crisis situation but can be used in order to detect and prevent sedentary habits that may lead to a slow but, nonetheless, dangerous deterioration of elderly or disabled people physical condition. Along with settling potential controversies this also proves the importance of log messages.

As shown above, the BSN is designed to have a hybrid architecture which combines the advantages of wired and wireless networks (low power consumption, safe communications, reduced sensitivity to electrical interferences, reasonable speed). Of course, there are some inconveniences for the human subject: reduced mobility, less ergonomic, extra weight. The flexible design of the BSN allows turning it into a fully wireless network depending on the technical specifications of the available hardware at the final implementation stage.

For testing purposes only, development kits from Texas Instruments were used: eZ430 - RF 2500 and TI Chronos development kits.

The eZ430 - RF 2500 kit (Texas Instruments, 2013) was used for testing the BSN architecture, the compliance with 802.15.4 standard and the functional characteristics in our indoor testing environment

The second development kit used for design and testing purposes in case of Personal Monitoring System of AmiHomCare project was TI – Chronos (Texas Instruments, 2013).

Presented as a hand digital sports watch, TI-Chronos embeds several internal sensors (such as temperature sensor, 3 axis accelerometer, pressure sensor) and allows wireless communications via a proprietary protocol namely BlueRobin. This allows bi-directional communications with a computer (desktop or laptop) and, also, with other devices. In our case, a simple chest strap was used for monitoring heart rates of the subject.

TI-Chronos is based on CC430 technology platform (Texas Instruments, 2013) which offers the industry's lowest power consumption, single-chip radio-frequency (RF) family for microcontroller (MCU)-based applications. CC430 platform helps advance RF networking applications in various domains (automation, tracking, medicine, sport, etc.) but it can’t and shouldn’t be used as reliable resource in critical applications.

For “proof of concept” purposes only body temperature, heart rate and fall detection were tested at this point. TI-Chronos plays the role of a relay while the Central Node was implemented in a laptop. The sensor nodes are represented by the temperature sensor, accelerometer and thoracic belt. For some applications (heart rate monitoring) the analysis was made offline due to the limitations in the equipments. TI-Chronos allows both sending and receiving data (through radio channels), however this can’t be done simultaneously.

Even if the Personal Monitoring System sends alert messages to the Environment Monitoring System, the values of monitored parameters will also be sent. This way, a medical history of the patient can be created and consulted either by the emergency team or by the personal physician during the periodical home visits.

4. INTEGRATIVE SYSTEM’S ARCHITECTURE

As mentioned before, the use of each technique comes with some advantages but also with some drawbacks.

The advantage of sound processing (the technique used in Sweet-Home project) is that the system is more likely to be embraced by its users. This is because of the fact that, especially for aged people, it is easier to give an oral command than to use an electronic device.

The drawback of sound processing technique comes from the fact that, along with ageing, the speech suffers major modifications and the use of common voice processing software is not an option for this type of users, so a different one must be developed. (Aman *et al.*, 2013)

Also, the audio information is easily influenced by other noises (both from the outside and from the inside of the apartment), so the task of a good identification of the commands in these conditions remains a research challenge.

The main advantage of using video processing is that it offers better localization results than the one obtained using presence sensors, thus improving the results of activity identification.

The drawback of this technique is that the use of video recording equipment is sometimes seen (especially by aged people) as a privacy violation.

The new system proposed architecture (fig. 9) consists in a combination of features from the two systems and in addition of some new features.

The proposed architecture is structured on three layers: Home Automation, Basic Personal Assistance and Additional Assistive Services.

As the name suggests, the *Home Automation* consists in basic equipment needed for monitoring and control of different indoor environment parameters. There are two types of equipment: sensors and actuators. Sensors are used for monitoring of temperature, lighting, detection of windows or doors opening and detection of dangerous events in the environment (fire, smoke, flooding, dangerous gases). Actuators are used for temperature and lighting control and command of windows and doors opening/closing.

Basic Assistance layer is made from the main assistive features:

- monitoring of personal health parameters;
- usual commands interpretation and performance
- distress situations detection
- abnormal activity detection

Apart from the personal health monitoring, for a greater accuracy of the results, all the other features involve both an audio processing part and a video processing part. The video processing part consists in two types of equipment: fixed (mounted cameras) and mobile (a personal assistant which moves along with the assisted person). In order to reduce the stress of monitoring, the personal assistant can be presented as a pet. Because the assistant moves along with the assisted person, the task of audio monitoring can also be spit in two parts also: the fixed microphones in the sealing will be used for activity detection and the assistant will be used for commands interpretation.

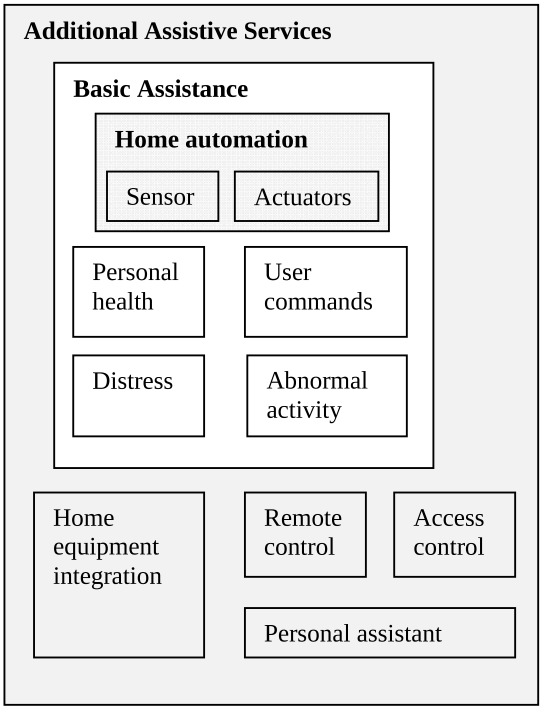

Fig. 9. Integrative system's architecture.

The distress situations can be grouped in two main categories: the ones announced by the user (either by voice commands, by gestures or using a special device – broche, watch, bracelet etc) and the ones detected by the system (in the absence of user's intervention). In the first case, the system must rightly interpret the voice information or user gestures. The second case consists in a more delicate task – the one of contextual interpretation.

Having in mind that the system's users are aged people and that the aged people have in most of the cases a well defined routine each day, a deviation of this routine can be useful in identifying the progress of some mental diseases.

Additional Assistive Services include:

- integration and control of different home-use equipment;
- access control for different areas/equipment;
- additional features for personal assistant;
- remote control or commands (via phone or internet) given to home automation.

The home-use equipment will be integrated in order to increase the safety and comfort of the user. This includes the cooking stove, the refrigerator and light control using voice commands. For user's safety, the access to harmful equipment may be restricted, as presented in Anton et al. (2013).

An example of additional features for personal assistant is represented by the possibility, in the distress situation announced by the user, of the personal assistant to transmit images at the call center in order to help the emergency evaluation process. This represents the only situation in which images from the apartment will be sent to another person. All the rest of video acquisition is used only for computational purpose, and thus will not represent a privacy violation.

5. CONCLUSIONS

Two ongoing projects (Sweet-Home and AmiHomeCare) related to home assistance for elderly people were presented in this paper. Each one fundaments its functioning on a different approach: Sweet-Home on sound signal processing, AmiHomeCare on video processing. A description for each project structure and features was presented in the beginning of the paper. The last part focuses on a new and more complex system description. The new system integrates features from two projects and adds new ones, aiming at increasing the accuracy of user's commands (either vocal or gestures) interpretation and of distress situation detections.

Future research topics that are opened by the work presented in this paper include:

- Fusion of audio and video information in order to increase the detection of distress situations;
- Voice and gesture based command;
- Enhanced localisation (when people are in the - Kinect area of the camera);
- Decision making based on sound and video;
- Development of personal assistant with speech recognition/understanding.

ACKNOWLEDGMENTS

The Sweet-home project is founded by the French National Agency (Agence Nationale de la Recherche / ANR-09-VERS-011).

The AmiHomeCare work has been co-founded by the Sectoral Operational Programme Human Resources Development 2007-2013 of the Romanian Ministry of Labour, Family and Social Protection through the Financial Agreement POSDRU/89/1.5/S/62557.

REFERENCES

Aarob, B., and Davis, J. W. (2013). Action Recognition using Temporal Templates. http://www.cse.ohio-state.edu/~jwdavis/CVL/Publications/motrec.pdf, accessed February 2013.

Aman, F., Vacher, M., Rossato, S. and Portet, F. (2013). Speech Recognition of Aged Voices in the AAL Context:

Detection of Distress Sentences. The 7th International Conference on Speech Technology and Human-Computer Dialogue, SpeD 2013, Cluj-Napoca, Romania, oct 2013, pp.177-184.

Anton F. and Anton S. (2012a). Robot-human Natural Interaction For Collision Detection And Basic Task Definition. 9th International Conference on Informatics in Control, Automation and Robotics - ICINCO 2012, Rome, Italy, 28-31 July, 2012.

Anton S., Mocanu I., Anton F., Borangiu T. and Ceccarelli M. (2012b). Gesture Recognition for Robot Assistance in Ambient Assisted Living Environments. In *Proceedings of the RAAD 2012 21st International Workshop on Robotics in Alpe-Adria-Danube Region*, September 10-13, 2012, Napoli, Italy

Anton S. and Anton F. (2013). Human Authentication Techniques in Ambient Intelligence Systems. In Proceedings of the International Workshop "Fostering Innovation in Healthcare Services", Brasov 14-15 March, 2012, Ed. Univ. "Carol Davila", 2013, ISBN: 978-973-708-659-4, pp:119-122.

Chahuara, P, Portet, F. and Vacher, M. (2011). Fusion of Audio and Temporal Multimodal Data by Spreading Activation for Dweller Localisation in a Smart Home. *In Proceedings of STAMI, Space, Time and Ambient Intelligence*, pp. 17-22.

Chahuara, P, Fleury, A., Portet, F. and Vacher, M. (2012). Using Markov Logic Network for On-line Activity Recognition from Non-Visual Home Automation Sensors. *In Proceedings of Ambient Intelligence*, pp. 177-192.

Chan, M., Campo, E., Estève, D. and Fourniols, J.-Y. (2009). Smart homes — current features and future perspectives. Maturitas, 64(2), pp. 90–97.

Chen, M., Gonzalez, S., Vasilakos, A., Cao H., and Leung, V. 2010). Body Area Networks: A Survey, ACM/Springer Mobile Networks and Applications (MONET), *Springer Science+Business Media*, LLC 2010, DOI: 10.1007/s11036-010-0260-8.

Chen, C. and Cook, D. (2012). Behavior-based home energy prediction. *In Proceedings of the International Conference on Intelligent Environments*, pp. 1-7.

Chu, C.W., and Cohen, I. (2013) Posture and Gesture Recognition using 3D Body Shapes Decomposition. http://iris.usc.edu/outlines/papers/2005/cwchu-v4hci05.pdf, accessed February 2013.

Dey, A. K. (2001). Understanding and Using Context. Personal and Ubiquitous Computing, 5(1), pp. 4–7.

Galata, A., Johnson, N., and Hogg, D. (2001). Learning Variable-Length Markov Models of Behavior. *In: Computer Vision and Image Understanding*; 81(3), pp. 398-413.

Gerosa, M., Giuliani, D. and Brugnara, F. (2009). Towards age-independent acoustic modeling. Speech Communication, 51(6), pp. 499–509.

Gorham-Rowan, M. and Laures-Gore, J. (2006). Acoustic-perceptual correlates of voice quality in elderly men and women. Journal of Communication Disorders, 39, pp. 171–184.

Holmes, A., Duman, H. and Pounds-Cornish, A. (2002). The idorm: gateway to heterogeneous networking environments. *In Proceedings of international ITEA workshop on virtual home environments*, Paderborn, Germany, pp. 20-37.

Intille, S. S. (2002). Designing a home of the future. IEEE Pervasive Computing, 1(2), pp. 76–82.

Ivanov, Y.A., Bobick, A.F. (2000). Recognition of Visual Activities and Interactions by Stochastic Parsing. *IEEE Trans. Pattern Analysis and Machine Intelligence*; 22(8), pp. 852-871.

iRobot, iRobot ® Roomba ® Serial Command Interface (SCI) Specification http://www.irobot.com/ images/ consumer/ hacker/Roomba_SCI_Spec_Manual.pdf

Kinect (2013). Microsoft Kinect for Windows SDK, Programming Guide. http://research.microsoft.com/redmond/kinectsdk/docs/programmingguide_kinectsdk.pdf, accessed February 2013.

Lopez-Cozar, R. and Callejas, Z. (2010). Multimodal dialogue for ambient intelligence and smart environments. Handbook of Ambient Intelligence and Smart Environments, pp. 559–579.

Memsic - http://www.memsic.com/wireless-sensor-networks.

Mocanu, I. (2010). From content-based image retrieval by shape to image annotation. *Advances in Electrical and Computer Engineering*, 10(4), pp. 49-56, (2010)

Mocanu I., Anton S. and Florea A.M. (2013). Human Posture Recognition and Pattern Discovery. In *Proceedings of the International Workshop "Fostering Innovation in Healthcare Services*", Brasov 14-15 March, 2012, Ed. Univ. "Carol Davila", 2013, ISBN: 978-973-708-659-4, pp:131-136

Mocanu, St., Mocanu, I., Anton, S., and Munteanu, C. (2011). AmIHomeCare: a complex ambient intelligent system for home medical assistance. *Proceedings of the 10th International Conference on Applied Computer and Applied Computational Science, Venice, March 2011*, pp. 181-186.

Modler, P. and Myatt, T. (2008). Recognition of Separate Hand Gestures by Time-delay Neural Networks Based on Multi-state Spectral Image Patterns from Cyclic Hand Movements. *IEEE International Conference on Systems, Man, and Cybernetics*.pp. 1539-1544.

Moeslund, TB., and Granum, E. (2001). A Survey of Computer Vision-Based Human Motion Capture. *In: Computer Vision and Image.* 81(3), pp. 231-268

Norgall, T., Blobel, B., and Pharow, P. (2006). Personal Health - The Future Care Paradigm, *in Medical and Care Components*, vol. 3, pp. 299-306, IOS Press, 2006, ISBN: 1-58603-620-3.

Portet, F., Vacher, M., Golanski, C. Roux, C. and Meillon, B. (2013). Design and evaluation of a smart home voice interface for the elderly: acceptability and objection aspects. Personal and Ubiquitous Computing, 17(1), pp. 127-144.

Ramos, C., Augusto, J., and Shapiro, D. (2008). Ambient intelligence - the next step for artificial intelligence. IEEE Intelligent Systems, 23, pp. 15-18.

Richardson, M., Domingos, P. (2006). Markov logic networks. Machine Learning 62(1-2), pp. 107-136.

Schilit, B. N., Heimer, M. M. (1994). Disseminating active map information to mobile hosts. Network, IEEE, 8(5), pp. 22–32.

Sehili, M., Lecouteux, B., Vacher, M., Portet, F., Istrate, D., Dorizzi, B. and Boudy, J. (2012). Sound Environnment Analysis in Smart Home, LNCS, 7683, pp. 208-223.

Sharkey A. and Sharkey, N. (2012). Granny and the robots: ethical issues in robot care for the elderly. Ethics and Information Technology; 14(1), pp. 27-40.

Skubic, M., Alexander, G., Popescu, M., Rantz, M., and Keller, J. (2009). A smart home application to eldercare: current status and lessons learned. Technology and Health Care, 17(3), pp. 183–201.

Texas Instruments, website, http://www.ti.com/, accessed February 2013

Vacher, M., Fleury, A., Serignat, J.-F., Noury, N. and Glasson, H. (2008). Preliminary evaluation of speech/sound recognition for telemedicine application in a real environment. *In Proceedings of Interspeech'2008*; pp. 496–499.

Vacher, M., Istrate, D., Portet, F., Joubert T., Chevalier, T., Smidtas, S., Meillon B., Lecouteux, B., Sehili, M., Chahuara, P. and Méniard, S. (2011a). The SWEET-HOME Project: Audio Technology in Smart Homes to improve Well-being and Reliance. *In: Proceedings of IEEE EMBC 2011;* pp. 5291-5294.

Vacher, M., Portet, F., Fleury, A. and Noury, N. (2011b). Development of audio sensing technology for ambient assisted living: Applications and challenges. International Journal of E-Health and Medical Communications, 2(1), pp. 35–54.

Vacher, M., Portet, F., Rossato, S., Aman, F., Golanski, C. and Dugheanu, R. (2012a). Speech-based interaction in an AAL context. Gerontechnology, ISG, Eindhoven, 11(2), p. 310.

Vacher, M., Lecouteux, B. and Portet, F. (2012b). Recognition of Voice Commands by Multisource ASR and Noise Cancellation in a Smart Home Environment. *In Proceedings of EUSIPCO'2012*; pp. 1663-1667.

van Hoof, J., Kort, H.S.M., Markopoulos, P. and Soede, M. (2007). Ambient intelligence, ethics and privacy. Gerontechnology, ISG, Eindhoven, 6(3), pp.155–163.

Vipperla, R., Renals, S. and Frankel, J. (2008). Longitudinal study of ASR performances on ageing voices. *In Proceedings of Interspeech'2008*, pp. 2550–2553.

Yang, MH, and Ahuja, N. (1999). Recognizing Hand Gesture Using Motion Trajectories. *In: Proc. IEEE Conf. Computer Vision and Pattern Recognition*; pp. 466-472

Weiser, M. (1991). The computer for the 21st century. Scientific American, 265(3), pp. 66–75.

Wilpon, J.G. and Jacobsen, C.N. (1996). A study of speech recognition for children and the elderly. *In Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 349–352.

Wizard of Oz (2013) http://en.wikipedia.org/wiki/Wizard_of_Oz_experiment

Wolf, P., Sschmidt, A. et Klein, M. (2008). Soprano - an extensible, open AAL platform for elderly people based on semantical contracts. In 3rd Workshop on Artificial Intelligence Techniques for Ambient Intelligence (AITAmI'08), 18th European Conference on Artificial Intelligence (ECAI 08), Patras, Greece.

Wölfel, M. and McDonough, J. (2009). Distant Speech Recognition. John Wiley and Sons, 573 pages.

World Health Organization (2007). WHO Global report on falls Prevention in older Age, WHO Press, 2007, ISBN 978-92-4-156353-6