The SWEET-HOME Project: Audio Technology in Smart Homes to improve Well-being and Reliance

Michel Vacher, Pedro Chahuara, Benjamin Lecouteux, Dan Istrate, François Portet, Thierry Joubert, Mohamed El Amine Sehili, Brigitte Meillon, Nicolas Bonnefond, Sébastien Fabre, et al.

The Sweet-Home Project: Audio Processing and Decision Making in Smart Home to Improve Well-being and Reliance

Michel Vacher, Pedro Chahuara, Benjamin Lecouteux, Dan Istrate, François Portet, Thierry Joubert, Mohamed Sehili, Brigitte Meillon, Nicolas Bonnefond, Sébastien Fabre, Camille Roux and Sybille Caffiau

***Abstract*— The SWEET-HOME project aims at providing audio-based interaction technology that lets the user have full control over their home environment, at detecting distress situations and at easing the social inclusion of the elderly and frail population. This paper presents an overview of the project focusing on the implemented techniques for speech and sound recognition as context-aware decision making with uncertainty. A user experiment in a smart home demonstrates the interest of this audio-based technology.**

I. INTRODUCTION

Demographic change and ageing in developed countries imply challenges for the society to continue to improve the well being of its elderly and frail inhabitants. Since the dramatic evolution of Information and Communication Technologies (ICT), one way to achieve this aim is to promote the development of smart homes. In the health domain, a *health smart home* is a habitation equipped with a set of sensors, actuators, automated devices and centralised software controllers specifically designed for daily living task support, early detection of distress situations, remote monitoring and promotion of safety and well-being [1]. Among all the interaction and sensing technologies used in smart homes (e.g., infra-red sensors, contact doors, video cameras, RFID tags, etc.), audio processing technology has a great potential to become one of the major interaction modalities. Voice interfaces can be much more adapted to disabled people and the ageing population who have difficulties in moving or seeing than tactile interfaces (e.g., remote control) which require physical and visual interaction. Moreover, audio processing is particularly suited to distress situations [2].

The SWEET-HOME project which started in 2010 to address several of these issues is shortly introduced in Section II. The first task of the project was to acquire a large sample of situations for training and analysis purpose. Then, the core audio technologies were developed, namely, every day sound detection and distant speech recognition and incorporated into the real-time framework PATSH ("**P**lateforme d'**A**ccueil de **T**raitement **S**WEET-HOME) which is presented in Section III together with the intelligent controller designed to interpret the voice command and make decision on the smart environment. The evaluation and the results are presented in Section IV before conclusion in Section V.

II. THE SWEET-HOME PROJECT

The SWEET-HOME project (`sweet-home.imag.fr`) is a French national supported research project aiming at designing a new smart home system based on audio technology focusing on three main aspects: to provide assistance via *natural man-machine interaction* (voice and tactile command), to ease *social inclusion* and to provide *security reassurance* by detecting situations of distress. If these aims are achieved, then the person will be able to pilot, from anywhere in the house, their environment at any time in the most natural way possible. The targeted users are elderly people who are frail but still autonomous.

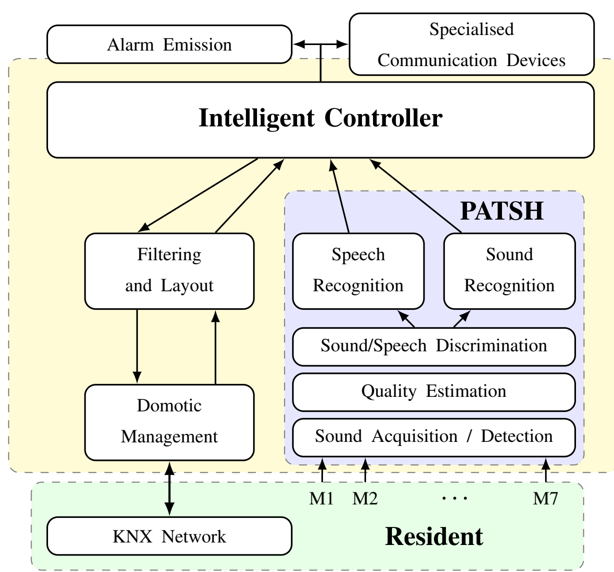

Fig. 1. System architecture.

The architecture of the system is depicted in Figure 1. The input is composed of the information from the home automation network and information from 7 microphones (M1, M2,. . . , M7) transmitted through radio frequency channels. The audio signals are analysed by the PATSH framework described in Section III. Thus, information can be provided

This work is supported by the French National Research Agency (Agence Nationale de la Recherche / ANR-09—VERS-011)

M. Vacher, P. Chahuara, B. Lecouteux, F. Portet, B. Meillon, N. Bonnefond and S. Caffiau are with the Laboratory of Informatics of Grenoble, UMR CNRS/UJF/G-INP 5217, Grenoble, France `firstname.lastname@imag.fr`

D. Istrate and M. Sehili are with ESIGETEL, 77210 Avon-Fontainebleau, France `dan.istrate@esigetel.fr`

T. Joubert is with THEORIS, 103 rue La Fayette - 75010 Paris, France `thierry.joubert@theoris.fr`

S. Fabre is with TECHNOSENS, 15 rue d'Alsace, 69150 Décines, France `sebastien.fabre@technosens.fr`

C. Roux is with FLORALIS, 6, allée de Bethléem, 38610 Gières, France `camille.roux@imag.fr`

directly by the user (e.g., voice order) or via environmental sensors (e.g., temperature). The information coming from the home automation system is transmitted on-line to the intelligent controller (through several preprocessing steps). This controller captures all streams of data, interprets them and executes the required actions. the controller can send alert or information messages to a speech synthesiser in case of emergency or on request of the user (*"Beware, the front door is not locked"*, *"The temperature is 21 degrees"*).

III. AUDIO ANALYSIS AND INTELLIGENT CONTROLLER

This section describes the audio analysis system and the intelligent controller.

A. Sound and speech analysis

For real-time audio analysis the PATSH framework (see Figure 1) was developed to manage the acquisition/processing flow of the sound events detected. In PATSH, several processors are plugged in and synchronised, 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:

1) the Sound Acquisition and Detection stage, detecting sound events (daily sounds or speech) from input streams;
2) the Sound/Speech Discrimination stage, discriminating speech from other sounds to extract voice commands;
3) the Sound Classification stage, recognizing daily living sounds; and
4) the Automatic Speech Recognition (ASR) stage, applying speech recognition to events classified as speech.

The Sound/Speech Discrimination stage has a very important role: firstly, vocal orders must not be missed, secondly, daily living sounds must not be sent to the ASR because undesirable sentences could be recognized. In order to recognize only vocal orders and not all sentences uttered in the flat, all sentences shorter than 150 ms and longer than 2.2 seconds are ignored. These values were choosen after a statistical study on our data bases. In the same manner, the Signal to Noise Ratio (SNR) must be greater than 0 dB. For a complete description of the system the reader is referred to [3].

B. Intelligent controller

The intelligent controller must not only analyse the recognized sentences but also monitor the smart home to detect risky situation. Its architecture and functioning are typical of the context-aware system model in pervasive environment.

The knowledge of the controller is defined using two semantic layers: the *low-level* ontology, devoted to the representation of raw data and network information description (e.g., state of switches and actuators); and the *high-level* ontology which represents concepts being used at the reasoning level. This separation between low and high levels makes it possible a high re-usability of the reasoning layer when the sensor network and the home have to be adapted [4]. Outputs of the system include home automation orders but also interactions with the user in case of alert messages (e.g. warn the person about an unclosed front door).

The decision making module is based on an influence diagram extended to handle uncertain information and to take its decision based on the context. This decision is implemented through a Markov Logic Network which include the utility measure [5] (i.e., to which extend a decision is relevant given the situation) and a post-processing strategy to balance its decision given the uncertainty of the context. A detailed explanation about the decision method can be found in [6]. The context provision is carried out through the collaboration of several processors executed under a service oriented approach, each one being specialized in a certain context aspect, such as location detection [7] or activity recognition [8]. This data and knowledge centred approach ensures 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.

The decision module has two modes of action: 1) reactive when a vocal command is uttered and transmitted by PATSH, and 2) pro-active, when a risky situation is detected. The proactive mode is performed through situation recognition which triggers the decision module. The situation recognition is modelled by rules of the Event Condition Action (ECA) type implemented directly into the high-level ontology as Semantic Web Rule Language (SWRL) rules [9]. A typical example of such rule is showed below:

Situation. *Event*: The person open the front door to go out. *Condition*: one window is not closed. *Action*: Send a warning message using voice message.

SWRL rule

DeviceEvent(?d), has_associated_object(?d, door),
takes_place_in(?d, kitchen),state_value(?d, open),
Window(?w), located_in(?w, bedroom),Application(?a),
has_application(?w,?a),curret_state(?a,on)
swlrb:moreThan(sqwrl:count(?w), 1)
$\rightarrow$ *current_state(BedroomWindowsOpen, detected)*

C. Vocal orders grammar

Our previous user study showed that targeted users prefer precise short sentences over more natural long sentences [10]. As shown in Figure 2, each order belongs to one of three categories: initiate command, stop command and emergency call. Except for the emergency call, all 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 key-word:

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

The sentences of the grammar are in French. The ASR uses a 3-gram language model (LM) with a 10K lexicon. It results from the interpolation of a generic LM (weight 10%) and a specialized LM (weight 90%). The generic LM was estimated on about 1000M of words from the French newspapers Le Monde and Gigaword. The specialized LM

```
basicCmd         = key initiateCommand object |
                   key stopCommand [object] |
                   key emergencyCommand
key              = "Nestor" | "maison"
stopCommand      = "stop" | "arrête"
initiateCommand  = "ouvre" | "ferme" | "baisse" | "éteins" | "monte" |
                   "allume" | "descend" | "appelle"
emergencyCommand = "au secours" | "à l'aide"
object           = [determiner] ( device | person | organisation)
determiner       = "mon" | "ma" | "l'" | "le" | "la" | "les" | "un" |
                   "des" | "du"
device           =  "lumière" | "store" | "rideau" | "télé" |
                    "télévision" | "radio"
person           = "fille" | "fils" | "femme" | "mari" | "infirmière" |
                   "médecin" | "docteur"
organisation     = "samu" | "secours" | "pompiers" | "supérette" |
                   "supermarché"
```

Fig. 2. Excerpt of the grammar of the voice orders

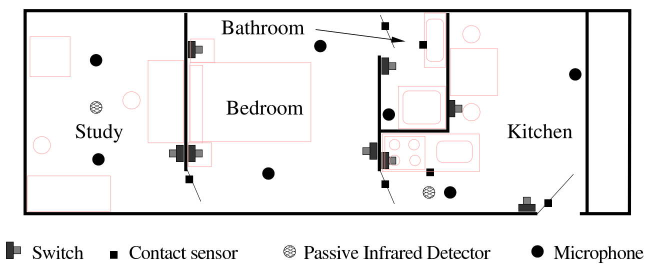

Fig. 3. Position of the sensors inside the MULTICOM-DOMUS smart home.

was estimated from all the sentences of the grammar. The recognized sentences are filterer at the input of the smart controller to determine their status, sentences not being in concordance with the grammar are rejected.

IV. EVALUATION

The system was tested in a smart home during an experiment involving 13 persons. This section describes the pervasive environment, the experimental set-up, the experiment and the preliminary results.

A. Materials

Experiments were run in the DOMUS smart home that was designed by the Laboratory of Informatics of Grenoble [11]. Figure 3 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. The flat is fully usable and can accommodate a dweller for several days. The technical architecture of DOMUS is based on the KNX bus system (standard ISO/IEC 14543). More than 150 sensors, actuators and information providers are managed in the flat (e.g., lighting, shutters, security systems, energy management, heating, etc.). A residential gateway architecture has been designed, supported by a virtual KNX layer seen as an Open Services Gateway Initiative service (OSGI) to guarantee the interoperability of the data coming and to allow the communication with virtual applications, such as activity tracking. The flat has also been equipped with 7 radio microphones for audio analysis with PATSH. A specialized communication device *e-lio* (`www.technosens.fr/`) from the Technosens company is used to initiate a communication with a relative.

B. Protocol

To validate the system in realistic conditions, we built scenarios in which every participant were asked to perform, at least once, some daily living activities chosen among the index of independence in Activities of Daily Living (ADL) [12]. These activities include: (1) Sleeping; (2) Resting: listening to the radio; (3) Feeding: realizing and having a meal; and (4) Communicating: going out of the flat to do shopping and chatting with a relative thanks to e-lio. Therefore, this experiment allowed us to process realistic and representative audio events in conditions which are directly linked to usual daily living activities. Moreover, to evaluate the decision making, some specific situations were planned in the scenarios. For instance, for the decision regarding the activation of the light, given that the bedroom contains two lights (the ceiling and the bedside one) as well as the kitchen (above the dinning table and above the sink), the four following situations were planed:

1) **Situation 1**. The person is having a meal on the kitchen table. The most appropriate light is the one above the table.
2) **Situation 2**. The person is cleaning up the bedroom. The most appropriate light is the ceiling one.
3) **Situation 3**. The person is cleaning the sink and doing the dishes. The most appropriate light is the one above the sink.
4) **Situation 4**. The person has just finished a nap. The most appropriate light is the bedside one.

The participant had to use vocal orders to make the light on or off, open or close blinds, ask about temperature and ask to call his or her relative. The instruction was given to the participant to repeat the order up to 3 times in case of failure. If after 3 times the system did not react, the activation was performed remotely by Wizard of Oz.

C. Experiments

Thirteen healthy participants (including 4 women) were asked to perform the scenarios without condition on the duration. As in our previous experimentations [13], a visit, before the experiment, was organized to ensure that the participants will find all the items necessary to perform the seven ADLs. It was necessary to explain the right way to utter vocal orders and to use the e-lio system, too. The average age was 35 years (19-62, min-max) and the experiment lasted between 23mn 49s and 46mn 51s (total for all participants = 7h 1mn 42s). The scenario includes at least 15 vocal orders for each participant but more sentences were uttered because the person was asked to repeat the order in case of malfunctioning.

D. Results

Regarding the results of the global audio analysis system in the case of the participant S02, 73 events were classified as speech (including 6 sounds wrongly classified as speech), 10 speech events were classified as sounds. Among these sentences, 32 were vocal orders in accordance with the grammar (some were duplicated because recorded simultaneously by more than one microphone), 2 were bad recognized and only 2 were missed (not detected or classified as sound). Other sentences were system messages or part of the talk with the relative through the e-lio devices, a part of these sentences were too long and not analysed. Some sentences were uttered in presence of noise (curtain motor or radio).

Regarding the decision, it was highly dependent on the activity recognition. Indeed, the correct inference of the current activity is necessary to choose, e.g., the most appropriate light and its intensity. Among the 52 instances of the activity, 13 were sleeping, 13 were having a meal and 26 were cleaning. The accuracy of the activity recognition was 63.4% on this data set using models that consider 7 activities in total (meal, cleaning up, dressing, sleeping, leisure, hygiene, and communication) [8]. This is an interesting score given that the activities were not learned using the same data.

Table I show the accuracy of the decision making in each of the four situations. The first column 'with uncertainty' present the accuracy of the decision when taking the probability of all the activities into account to weight the utility of each decision in each situation in order to compute the expected utility. The second column 'no uncertainty' consisted in taking only the most probable activity into account.

TABLE I

ACCURACY OF THE DECISION MAKING.

Accuracy:	with uncertainty	without uncertainty
Situation 1	46%	54%
Situation 2	100%	100%
Situation 3	84%	69%
Situation 4	54%	62%
Total	71%	71%

The lowest performance is in Situation 1. This can be explained by the fact that `meal` activity is often classified as `clean-up`. By contrast, the `clean-up` activity has a good recall mostly in the bedroom and thus leads to a correct decision. The results with and without uncertainty are similar as the cases in which the decision is different are only 4 out of 52. However, in one instance of Situation 3 the 'with uncertainty' case made it possible to find the correct solution. In that case, the highest activity probabilities were $P(A = hygiene) = 0.2$, $P(A = dressing) = 0.16$, $P(A = sleeping) = 0.28$, $P(A = meal) = 0.06$, $P(A = communication) = 0.05$ and $P(A = leisure) = 0.17$ whereas the ground truth was $P(A = clean - up) = 0.08$ in the kitchen. In this highly ambiguous situation, a correct decision was made because the most probable activities were the one with the highest degree of agitation. This helped the decision to choose the light with the highest intensity despite the most probable activity was sleeping.

V. CONCLUSIONS AND FUTURE WORKS

This paper presents an overview of the SWEET-HOME project which aims at designing a new smart home system based on audio technology. Several steps have been completed to provide real-time detection of sound of everyday living and speech recognition in the house. These modules were integrated inside the PATSH software to work in real-time. An intelligent controller makes the interface between the home automation network and PATSH in order to decide about the best action to make to adapt the smart home to the current situation. The experiment made in the smart home to evaluate the system showed promising results and validate the approach. This technology can benefit both the disabled and the elderly population that have difficulties in moving or seeing and want security reassurance.

Next steps in the project include the improvement of the current audio processing algorithms in order to detect in real-time the same vocal order recorded on several channels. This would improve the time response of the system in decreasing the number of sounds to analyse. Moreover, the time of analysis of the ASR is around 4s for each vocal order, which is far too long and biased the experiment. We are working in the optimisation of the ASR stage to perform new experiment with the targeted population with minimal technical bias.

VI. ACKNOWLEDGMENTS

The authors would like to thank the participants who took part to the different experiments. Thanks are also extended to D. Guérin to his support for PATSH development and to S. Pons for her support during the experiments inside the smart home.

REFERENCES

[1] M. Chan, E. Campo, D. Estève, and J.-Y. Fourniols, "Smart homes - Current features and future perspectives," *Maturitas*, vol. 64, no. 2, pp. 90–97, 2009.

[2] M. Vacher, F. Portet, A. Fleury, and N. Noury, "Development of audio sensing technology for ambient assisted living: Applications and challenges," *International Journal of E-Health and Medical Communications*, vol. 2, no. 1, pp. 35–54, 2011.

[3] M. Sehili, B. Lecouteux, M. Vacher, F. Portet, D. Istrate, B. Dorizzi, and J. Boudy, "Sound Environnment Analysis in Smart Home," in *Ambient Intelligence*, Pisa, Italy, 2012, pp. 208–223.

[4] M. Klein, A. Schmidt, and R. Lauer, "Ontology-centred design of an ambient middleware for assisted living: The case of soprano," in *30th Annual German Conference on Artificial Intelligence (KI 2007)*, 2007.

[5] A. Nath and P. Domingos, "A language for relational decision theory," in *Proceedings of the Sixth International Workshop on Statistical Relational Learning*, Leuven, Belgium, 2009.

[6] P. Chahuara, F. Portet, and M. Vacher, "Context aware decision system in a smart home : knowledge representation and decision making using uncertain contextual information," in *The 4th International Workshop on Acquisition, Representation and Reasoning with Contextualized Knowledge (ARCOE-12)*, Montpellier, France, 2012, pp. 52–64.

[7] ——, "Location of an Inhabitant for Domotic Assistance Through Fusion of Audio and Non-Visual Data," in *Pervasive Health*, Dublin, Ireland, 2011, pp. 242–245.

[8] P. Chahuara, A. Fleury, F. Portet, and M. Vacher, "Using Markov Logic Network for On-line Activity Recognition from Non-Visual Home Automation Sensors," in *Ambient Intelligence*, Pisa, Italy, 2012, pp. 177–192.

[9] I. Horrocks, P. F. Patel-schneider, H. Boley, S. Tabet, B. Grosof, and M. Dean. (2004) SWRL: A semantic web rule language combining OWL and RuleML. W3C. Last access on Dec 2011 at: http://www.w3.org/Submission/SWRL/.

[10] F. Portet, M. Vacher, C. Golanski, C. Roux, and B. Meillon, "Design and evaluation of a smart home voice interface for the elderly: acceptability and objection aspects," *Personal and Ubiquitous Computing*, vol. 17, no. 1, pp. 127–144, 2013.

[11] M. Gallissot, J. Caelen, F. Jambon, and B. Meillon, "Une plate-forme usage pour l'intégration de l'informatique ambiante dans l'habitat : Domus," *Technique et Science Informatiques (TSI)*, vol. 32, 2013.

[12] S. Katz, "Assessing self-maintenance: Activities of daily living, mobility, and instrumental activities of daily living," *Journal of the American Geriatrics Society*, vol. 31, no. 12, pp. 721–727, 1983.

[13] A. Fleury, M. Vacher, F. Portet, P. Chahuara, and N. Noury, "A french corpus of audio and multimodal interactions in a health smart home," *Journal on Multimodal User Interfaces*, vol. 7, no. 1, pp. 93–109, 2013.