HAL
open science

# Development of Audio Sensing Technology for Ambient Assisted Living: Applications and Challenges

Michel Vacher, François Portet, Anthony Fleury, Norbert Noury

## To cite this version:

Michel Vacher, François Portet, Anthony Fleury, Norbert Noury. Development of Audio Sensing Technology for Ambient Assisted Living: Applications and Challenges. International Journal of E-Health and Medical Communications (IJEHMC), 2011, 2 (1), pp.35-54. hal-00757407

**HAL Id: hal-00757407**

**https://hal.science/hal-00757407v1**

Submitted on 26 Nov 2012

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

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

# Development of Audio Sensing Technology for Ambient Assisted Living: Applications and Challenges*

Michel Vacher, Laboratoire d'Informatique de Grenoble, UJF-Grenoble, Grenoble-INP, CNRS UMR 5217, Grenoble, France
François Portet, Laboratoire d'Informatique de Grenoble, UJF-Grenoble, Grenoble-INP, CNRS UMR 5217, Grenoble, France
Anthony Fleury, University Lille Nord de France, France
Norbert Noury, INL-INSA Lyon Lab, France

ABSTRACT

One of the greatest challenges in Ambient Assisted Living is to design health smart homes that anticipate the needs of its inhabitant while maintaining their safety and comfort. It is thus essential to ease the interaction with the smart home through systems that naturally react to voice command using microphones rather than tactile interfaces. However, efficient audio analysis in such noisy environment is a challenging task. In this paper, a real-time audio analysis system, the AuditHIS system, devoted to audio analysis in smart home environment is presented. AuditHIS has been tested thought three experiments carried out in a smart home that are detailed. The results show the difficulty of the task and serve as basis to discuss the stakes and the challenges of this promising technology in the domain of AAL.

KEYWORDS: Ambient Assisted Living (AAL), Audio Analysis, AuditHIS System, Health Smart Homes, Safety and Comfort.

## 1 Introduction

The goal of the Ambient Assisted Living (AAL) domain is to enhance the quality of life of older and disabled people through the use of ICT. Health smart homes were designed more than ten

*Reference of the published paper of this draft:
@article{VACHER-JEHMC11,
author = "Michel Vacher and François Portet and Anthony Fleury and Norbert Noury",
title = "Development of Audio Sensing Technology for Ambient Assisted Living: Applications and Challenges",
journal = "International Journal of E-Health and Medical Communications",
publisher = "IGI Publishing, Hershey-New York, USA",
month = "march",
year = "2011",
volume = "2",
number ="1",
pages ="35 - 54",
keywords = "Daily Living, Independent Living, Life Expectancy, Support Vector Machine, Telemedecine",
issn ="1947-315X EISSN: 1947-3168", }

years ago as a way to fulfil this goal and they constitute nowadays a very active research area (Chan et al., 2008). Three major applications are targeted. The first one is to monitor how a person copes with her loss of autonomy through sensor measurements. The second one is to ease daily living by compensating one's disabilities through home automation. The third one is to ensure security by detecting distress situations such as fall that is a prevalent fear of the elderly.

Smart homes are typically equipped with many sensors perceiving different aspects of the home environment (e.g., location, furniture usage, temperature...). However, a rarely employed sensor is the microphone whereas it can deliver highly informative data. Indeed, audio sensors can capture information about sounds in the home (e.g., object falling, washing machine spinning...) and about sentences that have been uttered. Speaking being the most natural way of communicating, it is thus of particular interest in distress situations (e.g., call for help) and for home automation (e.g., voice commands). More generally, voice interfaces are much more adapted to people who have difficulties in moving or seeing than tactile interfaces (e.g., remote control) which require physical and visual interaction.

Despite all these advantages, audio analysis in smart home has rarely been deployed in real settings mostly because it is a difficult task with numerous challenges (Vacher et al., 2010b). In this paper, we present the stakes and the difficulties of this task through experiments carried out in realistic settings concerning sound and speech processing for activity monitoring and distress situations recognition. The remaining of the paper is organized as follow. The general context of the AAL domain is introduced in the first section. The second section is related to the state of the art of audio analysis. The third section describes the system developed in the GETALP team for real-time multi-source sound and speech processing. In the fourth section, the results of three experiments in a Smart Home environment, concerning distress call recognition, noise cancellation and activity of daily living classification, are summarized. Based on the results and our experience, we discuss, in the last section, some major promising applications of audio processing technologies in health smart homes and the most challenging technical issues that need to be addressed for their successful development.

## 2 Application Context

Health smart homes aim at assisting disabled and the growing number of elderly people which, according to the World Health Organization (WHO), is forecasted to reach 2 billion by 2050. Of course, one of the first wishes of this population is to be able to live independently as long as possible for a better comfort and to age well. Independent living also reduces the cost to society of supporting people who have lost some autonomy. Nowadays, when somebody is loosing autonomy, according to the health system of her country, she is transferred to a care institution which will provide all the necessary supports.

Autonomy assessment is usually performed by geriatricians, using the index of independence in Activities of Daily Living (ADL) (Katz and Akpom, 1976), which evaluates the person's ability to realize different tasks (e.g., doing a meal, washing, going to the toilets...) either alone, or with a little or total assistance. For example, the AGGIR grid (Autonomie Gérontologie Groupes Iso-Ressources) is used by the French health system. In this grid, seventeen activities including ten discriminative (e.g., talking coherently, find one's bearings, dressing, going to the toilets...) and seven illustrative (e.g., transports, money management, ...) are graded with an A (the task

can be achieved alone, completely and correctly), a B (the task has not been totally performed without assistance or not completely or not correctly) or a C (the task has not been achieved). Using these grades, a score is computed and, according to the scale, a geriatrician can deduce the person's level of autonomy to evaluate the need for medical or financial support.

Health Smart Home has been designed to provide daily living support to compensate some disabilities (e.g., memory help), to provide training (e.g., guided muscular exercise) or to detect potentially harmful situations (e.g., fall, gas not turned off). Basically, a health smart home contains sensors used to monitor the activity of the inhabitant. Sensor data are analyzed to detect the current situation and to execute the appropriate feedback or assistance. One of the first steps to achieve these goals is to detect the daily activities and to assess the evolution of the monitored person's autonomy. Therefore, activity recognition is an active research area (Duong et al., 2009)(Vacher et al., 2010a) but, despite this, it has still not reached a satisfactory performance nor led to a standard methodology given the high number of flat configurations. Moreover, the available sensors (e.g., infra-red sensors, contact doors, video cameras, RFID tags, etc.) may not provide the necessary information for a robust identification of ADL. Furthermore, to reduce the cost of such equipment and to enable interaction (i.e., assistance) the chosen sensors should serve not only to monitor but also to provide feedbacks and to permit direct orders.

As listed by the Digital Accessibility Team [1] (DAT), smart homes can be inefficient with disabled people and the ageing population. Visually, physically and cognitively impaired people will find very difficult to access equipments and to use switches and controls. This can also apply to ageing population though with less severity. Thus, except for hearing impaired persons, one of the modalities of choice is the audio channel. Indeed, audio processing can give information about the different sounds in the home (e.g., object falling, washing machine spinning, door opening, foot step...) but also about the sentences that have been uttered (e.g., distress situations, voice commands). This is in line with the DAT recommendations for the design of smart home which are, among other things, to provide hands free facilities whenever possible for switches and controls and to provide speech input whenever possible rather than touch screens or keypads. Moreover, speaking is the most natural way for communication. A person, who cannot move after a fall but being conscious, has still the possibility to call for assistance while a remote control may be unreachable. Callejas and Lopez-Coza (Callejas and Lopez-Cozar, 2009) studied the importance of designing a smart home dialogue system tailored to the elderly. The involved persons were particularly interested in voice commands to activate the opening and closing of blinds and windows. They also claimed that 95% of the persons will continue to use the system even if this one is sometimes wrong in interpreting orders. In Koskela and Väänänen-Vainio-Mattila (Koskela and Väänänen-Vainio-Mattila, 2004), a voice interface was used for interaction with the system in the realisation of small tasks (in the kitchen mainly). For instance, it permitted to answer the phone while cooking. One of the results of their study is that the mobile phone was the most frequently used interface during their 6-month trial period in the smart apartment. These studies showed that the audio channel is a promising area for improvement of security, comfort and assistance in health smart homes, but it stays rather unexplored in this application domain comparing with classical physical commands (switches and remote control).

[1] http://www.tiresias.org/research/guidelines/smart_home.htm

# 3 Sound and Speech Analysis

Automatic sound and speech analysis are involved in numerous fields of investigation due to an increasing interest for automatic monitoring systems. Sounds can be speech, music or more generally sounds of the everyday life (e.g., dishes, step. . . ). This state of the art presents firstly the sound and speech recognition domains and then details the main applications of sound and speech recognition in the smart home context.

## 3.1 Sound Recognition

Sound recognition is a challenge that has been explored for many years using machine learning methods with different techniques (e.g., neural networks, learning vector quantization, ...) and with different features extracted depending on the technique (Cowling and Sitte, 2003). It can be used for many applications inside the home, such as the quantification of water use (Ibarz et al., 2008) but it is mostly used for the detection of distress situations. For instance, (Litvak et al., 2008) used microphones and an accelerometer to detect a fall in the flat. (Popescu et al., 2008) used two microphones for the same purpose. Out of a context of distress situation detection, (Chen et al., 2005) used HMM with the Mel-Frequency Cepstral Coefficients (MFCC) to determine different activities in the bathroom. Cowling (Cowling, 2004) applied the recognition of non-speech sounds associated with their direction, with the purpose of using these techniques in an autonomous mobile surveillance robot.

## 3.2 Speech Recognition

Human communication by voice appears to be so natural that we tend to forget how variable a speech signal is. In fact, spoken utterances, even of the same text, are characterized by large differences that depend on context, the speaking style, the speaker's accent, the acoustic environment and many others. This high variability explains that the progress in speech processing has not been as fast as hoped at the time of the early work in this field.

The phoneme duration and the melody were introduced during the study of phonograph recordings of speech in 1906. The concepts of short-term representation of speech in the form of short (10-20 ms) semi-stationary segments were introduced during the Second World War and led to a spectrographic representation of the speech signal and emphasized the importance of the formants as carriers of linguistic information. The first spoken digit recognizer was presented in 1952 (Davis et al., 1952). Rabiner and Luang (Rabiner and Luang, 1986) published the scaling algorithm for the Forward-Backward method of training of Hidden Markov Model recognizers and nowadays modern general-purpose speech recognition systems are generally based on HMMs as far as the phonemes are concerned.

The current technology makes computers able to transcript documents on a computer from speech that is uttered at normal pace (for the person) and at normal loud in front of a microphone connected to the computer. This technique necessitates a learning phase to adapt the acoustic models to the person. Typically, the person must utter a predefined set of sentences the first time she uses the system. Dictation systems are capable of accepting very large vocabularies, more than ten thousand words. Another kind of application aims to recognize a small set of commands,

e.g., for home automation purpose or for Interactive Voice Response (of an answering machine for instance); this must be done without a speaker adapted learning step. More general applications are for example related to the context of civil safety. Clavel, Devillers, Richard, Vasilescu, and Ehrette (Clavel et al., 2007) studied the detection and analysis of abnormal situations through feartype acoustic manifestations.

Furthermore, *Automatic Speech Recognizer (ASR)* systems have reached good performances with close talking microphone (e.g., head-set), but the performance decreases significantly as soon as the microphone is moved away from the mouth of the speaker (e.g., when the microphone is set in the ceiling). This deterioration is due to a broad variety of effects including background noise and reverberation. All these problems should be taken into account in the home assisted living context.

### 3.3 Speech Recognition for the Ageing Voices

The evolution of human voice with age was extensively studied (Gorham-Rowan and Laures-Gore, 2006) and it is well known that ASR performance diminishes with growing age. The public concerned by the home assisted living is aged, the adaptation of the speech recognition systems to aged people in thus an important though difficult task. Experiments on automatic speech recognition showed a deterioration of performances with age (Gerosa et al., 2009) and also the necessity to adapt the models used to the targeted population when dealing with elderly people (Baba et al., 2004). A recent study has used audio recordings of lawyers at the Supreme Court of the United States over a decade (Vipperla et al., 2008). It showed that the performance of the automatic speech recognizer decreases regularly as a function of the age of the person but also that a specific adaptation to each speaker allow to obtain results close to the performance with the young speakers. However, with such adaptation, the model tends to be too much specific to one speaker. That is why Renouard, Charbit, and Chollet (Renouard et al., 2003) suggested using the recognized word in on-line adaptation of the models. This proposition was made in the assisted living context but seems to have been abandoned. An ASR able to recognize numerous speakers requires to record more than 100 speakers. Each record takes a lot of time because the speaker is quickly tired and only few sentences may be acquired during each session. Another solution is to develop a system with a short corpus of aged speakers (i.e., 10) and to adapt it specifically to the person who will be assisted.

It is important to recall that the speech recognition process must respect the privacy of the speaker, even if speech recognition is used to support elderly in their daily activities. Therefore, the language model must be adapted to the application and must not allow the recognition of sentences not needed for the application. An approach based on keywords may thus be respectful of privacy while permitting a number of home automation orders and distress situations being recognized. Regarding the distress case, an even higher level approach may be to use only information about prosody and context.

## 4 The Real-Time Audio Analysis System

According to the results of the DESDHIS project, everyday life sounds can be automatically identified in order to detect distress situations at home (Vacher et al., 2006). Therefore, the

AuditHIS software was developed to insure on-line sound and speech recognition. Figure 1 depicts the general organization of the audio analysis system; for a detailed description, the reader is refereed to (Vacher et al., 2010b)). Succinctly, each microphone is connected to an input channel of the acquisition board and all channels are analyzed simultaneously. Each time the energy on a channel goes above an adaptive threshold, an audio event is detected. It is then classified as daily living sound or speech and sent either to a sound classifier or to the ASR called RAPHAEL. The system is made of several modules in independent threads: acquisition and first analysis, detection, discrimination, classification, and, finally, message formatting. A record of each audio event is kept and stored on the computer for further analysis (Figure 1).

Data acquisition is operated on the 8 input channels simultaneously at a 16 kHz sampling rate. The noise level is evaluated by the first module to assess the Signal to Noise Ratio (SNR) of each acquired sound. The SNR of each audio signal is very important for the decision system to estimate the reliability of the corresponding analysis output. The detection module detects beginning and end of audio events using an adaptive threshold computed using an estimation of the background noise.

The discrimination module is based on Gaussian Mixture Model (GMM) and classifies each audio event as everyday life sound or speech. The discrimination module was trained with an everyday life sound corpus and with the Normal/Distress speech corpus recorded in our laboratory. Then, the signal is transferred by the discrimination module to the speech recognition system or to the sound classifier depending on the result of the first decision. Everyday life sounds are classified with another GMM classifier whose models were trained on an eight-class everyday life sound corpus.

## 4.1 The Automatic Speech Recognizer RAPHAEL

RAPHAEL is running as an independent application which analyzes the speech events sent by the discrimination module. The training of the acoustic models was made with large corpora in order to ensure speaker independence. These corpora were recorded by 300 French speakers in the CLIPS (BRAF100) and LIMSI laboratories (BREF80 and BREF120). Each phoneme is then modelled by a tree state Hidden Markov Model (HMM). Our main requirement is the correct detection of a possible distress situation through keyword detection, without understanding the person's conversation. The language model is a set of n-gram models which represent the probability of observing a sequence of n words. Our language model is made of 299 unigrams (299 words in French), 729 bigrams (2-word sequences) and 862 trigrams (3-word sequences) which have been learned from a small corpus. This small corpus is made of 415 sentences: 39 home automation orders (e.g., *"Monte la temperature"*), 93 distress sentences (e.g., *"Au secours"* (Help) and 283 colloquial sentences (e.g., *"Á demain"*, *"J'ai bu ma tisane"*...).

## 4.2 The Noise Suppression System in case of Known Sources

Sound emitted by the radio or the TV in the home $x(n)$ is a noise source that can perturb the signal of interest $e(n)$ (e.g., person's speech). This noise source is altered by the room acoustic through his transfer function: $y(n) = h(n) * x(n)$, where $n$ represents the discrete time, $h(n)$ the impulse response of the room acoustic and $*$ the convolution operator. It is then superposed to

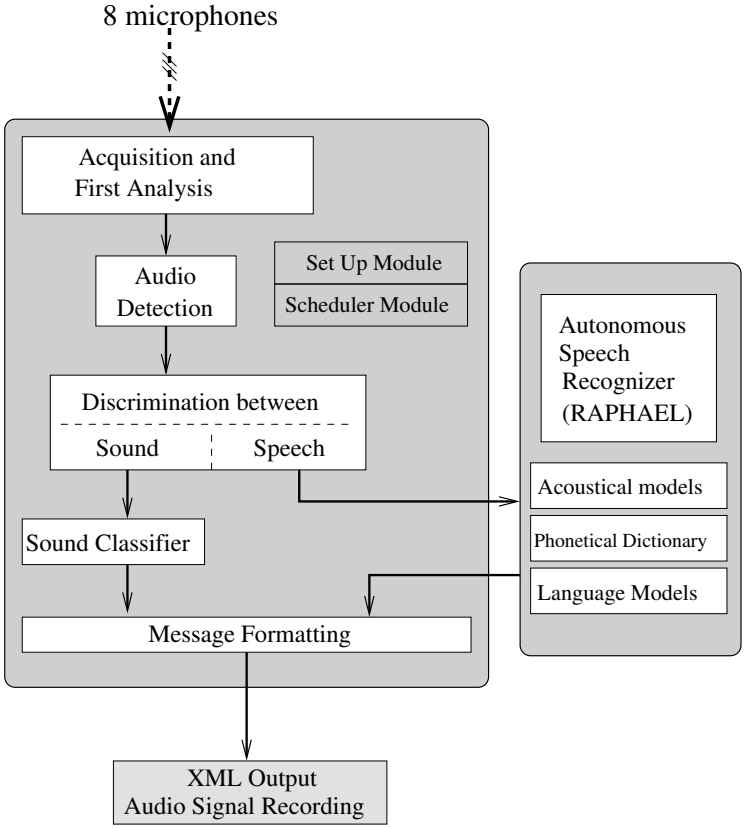


Figure 1: Global organisation of the AuditHIS and RAPHAEL softwares

the signal of interest $e(n)$ emitted in the room: speech uttered by the patient or everyday life sound. The signal recorded by the microphone in the home is then $y(n) = e(n) + h(n) * x(n)$. AuditHIS includes the SPEEX library for noise cancellation that needs two signals to be recorded: the input signal $y(n)$, and the reference signal $x(n)$ (e.g., TV). The aim is thus to estimate $h(n)$ ($\hat{h}(n)$) to reconstruct $e(n)$ ($v(n)$) as shown on Figure 2.

Various methods were developed in order to suppress the noise (Michaut and Bellanger, 2005) by estimating the impulse response of the room acoustic $\hat{h}(n)$. The Multi-delay Block Frequency Domain (MDF) algorithm is an implementation of the Least Mean Square (LMS) algorithm in the frequency domain (Soo and Pang, 1990). This algorithm is implemented in the SPEEX library under GPL License (Valin, 2007) for echo cancellation system.

One problem in echo cancellation systems, is that the presence of audio signal $e(n)$ (double-talk) tends to make the adaptive filter diverge. A method (Valin and Collings, 2007) was proposed by the authors of the library, where the misalignment is estimated in closed-loop based on a gradient adaptive approach; this closed-loop technique is applied to the block frequency domain (MDF) adaptive filter.

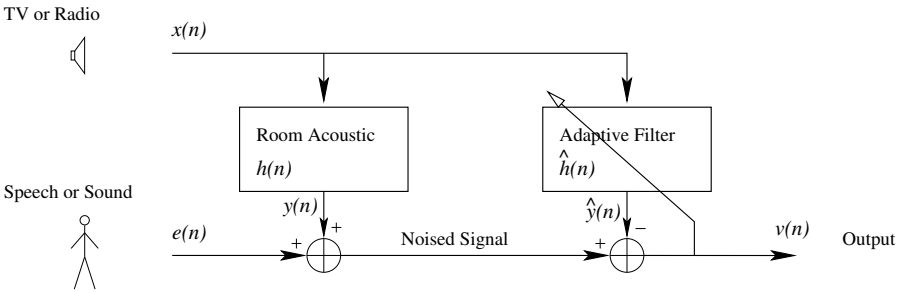


Figure 2: Block diagram of an echo cancellation system

This echo cancellation let some specific residual noise in the $v(n)$ signal and a postfiltering is requested to clean $v(n)$. The method implemented in SPEEX and used in AuditHIS is Minimum Mean Square Estimator Short-Time Amplitude Spectrum Estimator (MMSE-STSA) presented in (Ephraim and Malah, 1984). The STSA estimator is associated to an estimation of the a priori SNR. Some improvements have been added to the SNR estimation (Cohen and Berdugo, 2001) and a psycho-acoustical approach for post-filtering (Gustafsson et al., 2004). The purpose of this post-filter is to attenuate both the residual echo remaining after an imperfect echo cancellation and the noise without introducing "musical noise", i.e., randomly distributed, time-variant spectral peaks in the residual noise spectrum as spectral subtraction or Wiener rule (Vaseghi, 1996). This method leads to more natural hearing and to less annoying residual noise.

# 5 Experimentations in the Health Smart Home

AuditHis was developed to analyse speech and sound in health smart homes and, rather than being restricted to in-lab experiments; it has been evaluated in semi wild experiments. This section describes the three experiments that have been conducted to test the distress call/normal utterance detection, the noise source cancellation and the daily living audio recognition in a health smart home.

The experiments have been done in real conditions in the **HIS** (for *Habitat Intelligent pour la Santé* or *Health Intelligent Space*) of the TIMC-IMAG laboratory, excepted for the experiment related to distress call analysis in presence of radio. The HIS, represented in figure 3, is a flat of 47 $m^2$ inside a building of the faculty of medicine of Grenoble. This flat comprises a corridor (a), a kitchen (b), toilets (c), a bathroom (d), a living room (e) a bedroom (f) and is equipped with several sensors. However, only the audio data recorded by the 8 microphones have been used in the experiments. Microphones were set on the ceiling and directed vertically to the floor. Consequently, the participants were situated between 1 and 10 meters away from the microphones, sat down or stood up. To make the experiences more realistic they had no instruction concerning their orientation with respect to the microphones (they could have the microphone behind them) and they were not wearing any headset (Figure 3).

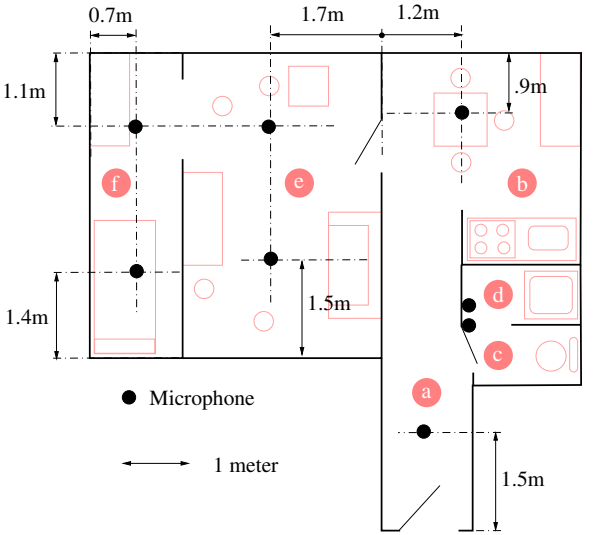


Figure 3: The Health Smart Home of the Faculty of Medicine of Grenoble and the microphone set-up

## 5.1 First Experiment: Distress Call Analysis

The aim of this experiment was to test whether ASR using a small vocabulary language model was able to detect distress sentences without understanding the entire person's conversation.

### 5.1.1 Experimental Set-Up

Ten native French speakers were asked to utter 45 sentences (20 distress sentences, 10 normal sentences and 3 phone conversations made up of 5 sentences each). The participants included 3 women and were 37.2 (SD=14) years old (weight: 69 ± 12 kg, height: 1.72 ± 0.08 m).The recording took place during daytime; hence there was no control on the environmental conditions during the sessions (e.g., noise occurring in the hall outside the flat).

A phone was placed on a table in the living room. The participants were asked to follow a small scenario which included moving from one room to other rooms, sentences reading and phone conversation. Every audio event was processed on the fly by AuditHIS and stored on the hard disk.

During this experiment, 3164 audio events were collected with an average SNR of 12.65 dB (SD=5.6). These events do not include 2019 ones which have been discarded because the SNR was inferior to 5 dB. This 5 dB threshold was chosen based on an empirical analysis (Vacher et al., 2006). The events were then furthermore filtered to remove duplicate instances (same event recorded on different microphones), non speech data (e.g., sounds) and saturated signal. At the end, the recorded speech corpus (7.8 minutes of signal) was composed of nDS = 197 distress sentences and nNS = 232 normal sentences. This corpus was indexed manually because some

speakers did not follow strictly the instructions given at the beginning of the experiment.

### 5.1.2 Results

These 429 sentences were processed by RAPHAEL using the acoustic and language models presented in previous sections. To measure the performances, the Missed Alarm Rate ($MAR$), the False Alarm Rate ($FAR$) and the Global Error Rate ($GER$) were defined as follows: $MAR = nMA|nDS$; $FAR = nFA|nNS$; $GER = (nMA + nFA)|(nDS + nNS)$ where $nMA$ is the number of missed detections; and $nFA$ the number of false alarms.

Globally the $FAR$ is quite low with 4% (SD=3.2) whatever the person. So the system rarely detects distress situations when it is not the case. The $MAR$ (29.2%±14.3) and $GER$ (15.8%±7.8) vary from about 5% to above 50% and highly depend on the speaker. The worst performances were observed with a speaker who uttered distress sentences like a film actor ($MAR$=55% and $FAR$ =4.5%). This utterance caused variations in intensity which provoked the French pronoun ”je” to not be recognized at the beginning of some sentences. For another speaker the MAR is higher than 40% (FAR is 5%). It can be explained by the fact that she walked when she uttered the sentences and made noise with her high-heeled shoes. This noise was added to the speech signal that was analyzed. The distress sentence ”help” was well recognized when it was uttered with a French pronunciation but not with an English one because the phoneme [h] does not exist in French and was not included in the acoustical model. When a sentence was uttered in the presence of an environmental noise or after a tongue clicking, the first phoneme of the recognized sentence was preferentially a fricative or an occlusive and the recognition process was altered.

The experiment led to mixed results. For half of the speakers, the distress sentences classification was correct ($FAR < 5\%$) but the other cases showed less encouraging results.

This experience showed the dependence of the ASR to the speaker (thus a need for adaptation), but most of the problems were due to noise and environmental perturbations. To investigate what problem could be encountered in health smart homes another study has been run in real setting focusing on lower aspect of audio processing rather than distress/normal situation recognition.

## 5.2 Second Experimentation: Distress Call Analysis in Presence of Radio

Two microphones were set in a unique room, the reference microphone in front of the speaker system in order to record music or radio news (France-Info, a French broadcasting news radio) and the signal microphone in order to record a French speaker uttering sentences. The two microphones were connected to the AuditHIS system in charge of real-time echo-cancellation.

For this experiment, 4 speakers (3 men and 1 woman, between 22 and 55 years old) were asked to stand in the centre of the recording room without facing the signal microphone. They had to speak with a normal voice level; the power level of the radio was set to be rather strong and thus the SNR was approximately 0 dB. Each participant uttered 20 distress sentences of the Normal/Distress corpus in the recording room. This process was repeated by each speaker 2 or 3 times.

The average missed alarm rate ($MAR$), for all the speakers, was 27%. These results depend on

the voice level of the speaker during this experiment and on the speaker himself. Also a big issue is the recognition of the frontiers between the 'silence' periods and the beginning and the end of the sentence. Because of this is not well assessed some noise are included in the first and last phoneme spectrum leading to false identification and thus to false sentence recognition. It may thus be useful to improve the detection of these 2 moments with a good precision to use shorter silence intervals.

## 5.3 Third Experimentation: Audio Processing of Daily Living Sounds and of Speech

An experiment was run to test the AuditHIS system in semi wild conditions. To ensure that the data acquired would be as realistic as possible, the participants were asked to perform usual daily activities. Seven activities, from the index of independence in Activities of Daily Living (ADL) were performed at least once by each participant in the HIS. These activities included: (1) Sleeping; (2) Resting: watching TV, listening to the radio, reading a magazine . . . ; (3) Dressing and undressing; (4) Feeding: realizing and having a meal; (5) Eliminating: going to the toilets; (6) Hygiene activity: washing hands, teeth. . . ; and (7) Communicating: using the phone. Therefore, this experiment allowed us to process realistic and representative audio events in conditions which are directly linked to usual daily living activities. It is of high interest to make audio processing performing to monitor these tasks and to contribute to the assessment of the person's degree of autonomy because the ADL scale is used by geriatricians for autonomy assessment.

### 5.3.1 Experimental Set Up

Fifteen healthy participants (including 6 women) were asked to perform these 7 activities without condition on the time spent. Four participants were not native French speakers. The average age was 32±9 years (24-43, minmax) and the experiment lasted in minimum 23 minutes 11s and 1h 35minutes 44s maximum. Participants were free to choose the order in which they wanted to perform the ADLs. Every audio event was processed on the fly by AuditHIS and stored on the hard disk. For more details about the experiment, the reader is refereed to (Fleury et al., 2010).

It is important to note that this flat represents a hostile environment for information acquisition similar to the one that can be encountered in real homes. This is particularly true for the audio information. The AuditHIS system (Vacher et al., 2010a) was tested in laboratory with an average Signal to Noise Ratio (SNR) of 27dB. In the smart home, the SNR felt to 11dB. Moreover, we had no control on the sound sources outside the flat, and there was a lot of reverberation inside the flat because of the 2 important glazed areas opposite to each other in the living room.

### 5.3.2 Data Annotation

Different features have been marked up using Advene [2] developed at the LIRIS laboratory. A snapshot of Advene with our video data is presented on figure 4. On this figure, we can see the

[2] http://liris.cnrs.fr/advene/

organisation of the flat and also the videos collected. The top left is the kitchen, the bottom left is the kitchen and the corridor, the top right is the living-room on the left and the bedroom on the right and finally the bottom right represents another angle of the living room. Advene allows organizing annotations elements – such as type of annotations, annotations, and relationships – under schemes defined by the user. The following annotation types have been set: location of the person, activity, chest of drawer door, cupboard door, fridge door, posture.

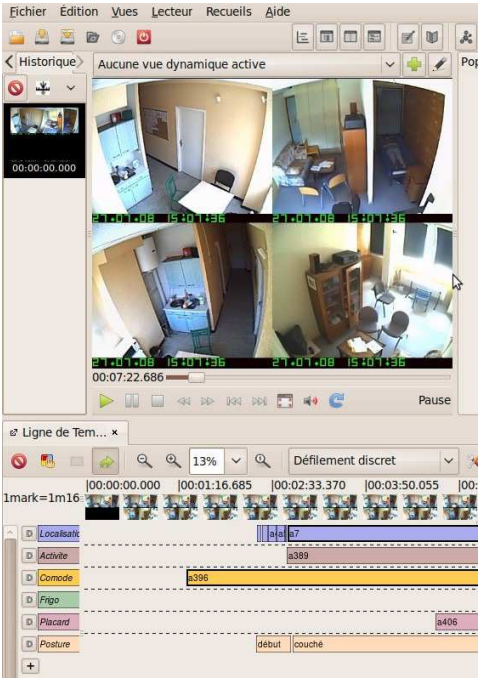


Figure 4: Excerpt of the dvene software for the annotation of the activities and events in the Health Smart Home

To mark-up the numerous sounds collected in the smart home, none of the annotation software we tested has shown to be convenient. It was mandatory to identify each sound by hearing and visual video analysis. We thus developed our own annotator in Python that plays each sound one at a time while displaying the video in the context of this sounds and proposing to keep the AuditHIS annotation or select another one in a list. About 2500 sounds and speech have been annotated in this way.

### 5.3.3 Results

The results of the sound/speech discrimination stage of AuditHIS are given on Table 1. This stage is important for two reasons: 1) these two kinds of signal might be analyzed by different paths in the software; and 2) the fact that an audio event is identified as sound or as speech indicates very different information on the person's state of health or activity. The table shows the high confusion between these two classes. This has led to poor performance for each of the classifiers

Table 1: Sound/Speech Confusion Matrix

| Target / Hits | Everyday Sound | Speech |
|---|---|---|
| Everyday Sound | 1745 (87%) | 141 (25%) |
| Speech | 252 (23%) | 417 (75%) |

(ASR and sound classification). For instance 'dishes' sound was very often confused with speech because the training set did not include enough examples and because of the presence of a fundamental frequency in the spectral band of speech. Falls of objects and speech were often confused with scream. The misclassification can be related to a design choice. Indeed, scream is both a sound and a speech and then the difference between these two categories is sometimes thin. For example "Aïe!" is an intelligible scream that has been learned by the ASR but a scream could also consist in "Aaaa!" which, in this case, should be handled by the sound recognizer.

However, most of the poor performances can be explained by the too small training set and the fact that unknown and unexpected classes (e.g., thunder, Velcro) were not properly handled by the system. These results are quite disappointing but the corpus collected during the experiment represents another result, precious for a better understanding of the challenges to develop audio processing in smart home as well as for empirical test of the audio processing models in real settings.

In total, 1886 individual sounds and 669 sentences were processed, collected and corrected by manual annotation. The detailed audio corpus is presented in Table 2. The total duration of the audio corpus, including sounds and speech, is 8 min 23 s. This may seem short, but daily living sounds last 0.19s on average. Moreover, the person is alone at home; therefore she rarely speaks (only on the phone).

The speech part of the corpus was recorded in noisy conditions (SNR=11.2dB) with microphones set far from the speaker (between 2 and 4 meters) and was made of phone conversations. Some sentences in French such as *"Allo"*, *"'Comment ça va"* or *"à demain"* are excerpts of usual phone conversation. No emotional expression was asked from the participants.

According to their origin and nature, sounds have been gathered into sounds of daily living classes. A first class is made of all the sounds generated by the human body. Most of them are of low interest (e.g., clearing throat, gargling). However, whistling and song can be related to the mood while cough and throat roughing may be related to health. The most populated class of sound is the one related to the object and furniture handling (e.g., door shutting, drawer handling, rummaging through a bag, etc.). The distribution is highly unbalanced and it is unclear how these sounds can be related to health status or distress situation. However, they contribute to the recognition of activities of daily living which are essential to monitor the person's activity. Related to this class, though different, were sounds provoked by devices, such as the phone. Another interesting class, though not highly represented, is the water flow class. This gives information on hygiene activity and meal/drink preparation and is thus of high interest regarding ADL assessment. Finally the other sounds represent the sounds that are unclassifiable even by a human expert either because the source cannot be identified or because too many sounds were recorded at the same time. It is important to note that 'outdoor sounds' plus 'other sounds' represent more than 22% of the total duration of the non speech sounds.

Table 2: Every Day Life Sound Corpus

| Category | Sound Class | Class Size | Mean SNR (dB) | Mean length (ms) | Total length (s) |
|---|---|---|---|---|---|
| **Human sounds:** | | **36** | **12.1** | **101** | **3.4** |
| | Cough | 8 | 14.6 | 79 | 0.6 |
| | Fart | 1 | 13 | 74 | 0.01 |
| | Gargling | 1 | 18 | 304 | 0.3 |
| | Hand Snapping | 1 | 9 | 68 | 0.01 |
| | Mouth | 2 | 10 | 41 | 0.01 |
| | Sigh | 12 | 11 | 69 | 0.8 |
| | Song | 1 | 5 | 692 | 0.7 |
| | Speech | 669 | 11.2 | 435 | 290.8 |
| | Throat Roughing | 1 | 6 | 16 | 0.02 |
| | Whistle | 5 | 7.2 | 126 | 0.6 |
| | Wiping | 4 | 19.5 | 76 | 0.3 |
| **Object handling:** | | **1302** | **11.9** | **59** | **76.3** |
| | Bag Frisking | 2 | 11.5 | 86 | 0.1 |
| | Bed/Sofa | 16 | 10 | 15 | 0.2 |
| | Chair Handling | 44 | 10.5 | 81 | 3 |
| | Chair | 3 | 9 | 5 | 0.01 |
| | Cloth Shaking | 5 | 11 | 34 | 0.1 |
| | Creaking | 3 | 8.7 | 57 | 0.1 |
| | Dishes Handling | 68 | 8.8 | 70 | 4.7 |
| | Door Lock&Shutting | 278 | 16.3 | 93 | 25 |
| | Drawer Handling | 133 | 12.6 | 54 | 7 |
| | Foot Step | 76 | 9 | 62 | 4 |
| | Frisking | 2 | 7.5 | 79 | 0.1 |
| | Lock/Latch | 162 | 15.6 | 80 | 12.9 |
| | Mattress | 2 | 9 | 6 | 0.01 |
| | Object Falling | 73 | 11.5 | 60 | 4.4 |
| | Objects shocking | 420 | 9 | 27.6 | 11.6 |
| | Paper noise | 1 | 8 | 26 | 0.03 |
| | Paper/Table | 1 | 5 | 15 | 0.01 |
| | Paper | 1 | 5 | 31 | 0.03 |
| | Pillow | 1 | 5 | 2 | 0 |
| | Rubbing | 2 | 6 | 10 | 0.02 |
| | Rumbling | 1 | 10 | 120 | 0.1 |
| | Soft Shock | 1 | 7 | 5 | 0 |
| | Velcro | 7 | 6.7 | 38 | 0.2 |
| **Outdoor sounds:** | | **45** | **9.0** | **174** | **7.9** |
| | Exterior | 24 | 10 | 32 | 0.77 |
| | Helicopter | 5 | 10 | 807 | 4.4 |
| | Rain | 3 | 6 | 114 | 0.3 |
| | Thunder | 13 | 7.5 | 208 | 2.7 |
| **Device sounds:** | | **72** | **8.0** | **209** | **15.1** |
| | Bip | 2 | 8 | 43 | 0.08 |
| | Phone ringing | 69 | 8 | 217 | 15 |
| | TV | 1 | 10 | 40 | 0.04 |
| **Water sounds:** | | **36** | **10.1** | **1756** | **63.2** |
| | Hand Washing | 1 | 5 | 212 | 0.2 |
| | Sink Drain | 2 | 14 | 106 | 0.2 |
| | Toilet Flushing | 20 | 12 | 2833 | 56.6 |
| | Water Flow | 13 | 7 | 472 | 6.1 |
| **Other sounds:** | | **395** | **9.5** | **94** | **37.1** |
| | Mixed Sound | 164 | 11 | 191 | 31.3 |
| | unknown | 231 | 8.5 | 25 | 5.8 |
| **Overall sounds except speech** | | **1886** | **11.2** | **108** | **203.3** |

# 6 Applications to AAL and Challenges

Audio processing (sound and speech) has great potential for health assessment, disability compensation and assistance in smart home such as improving comfort via voice command and security via call for help. Audio processing may be a help for improving and facilitating the communication of the person with the exterior, nursing auxiliaries and family. Moreover, audio analysis may give important additional information for activity monitoring in order to evaluate the person's ability to perform correctly and completely different activities of daily living.

However, audio technology (speech recognition, dialogue, speech synthesis, sound detection) still need to be developed for this specific condition which is a particularly hostile one (i.e., unknown number of speakers, lot of noise sources, uncontrolled and inadequate environments). This is confirmed by the experimental results. Many issues going from the treatment of noise and source separation to the adaptation of the model to the user and her environment need to be dealt with. In this section the promising applications of audio technology for AAL and the main challenges to address to make them realisable are discussed.

## 6.1 Audio for Assessment of Activities and Health Status

Audio processing could be used to assess simple indicators of the health status of the person. Indeed, by detecting some audible and characteristic events such as snoring, coughing, gasps of breath, occurrences of some simple symptoms of diseases could be inferred. A more useful application for daily living assistance is the recognition of devices functioning such as the washing machine, the toilet flush or the water usage in order to assess how a person copes with her daily house duty. Another ambitious application would be to identify human nonverbal communication to assess mood or pain in person with dementia. Moreover, audio can be a very interesting supplementary source of information in case of a monitoring application (in one's own home or medical house). For instance, degree of loneliness could be assessed using speech recognition and presence of other persons or telephone usage detection. In the case of a home automation application, sound detection can be another source of localisation of the person (by detecting for instance door shutting events) to act depending on this location and for a better comfort for the inhabitant (turning on the light if needed, opening/closing the shuttle...).

## 6.2 Evaluation of the Ability to Communicate

Speech recognition could play an important role for the assessment of person with dementia. Indeed, one of the most tragic symptoms of Alzheimer's disease is the progressive loss of vocabulary and ability to communicate. Constant assessment of the verbal activity of the person may permit to detect important phases in dementia evolution. Audio analysis is the only modality that might offer the possibility of an automatic assessment of the loss of vocabulary, decrease of period of talk, insulation in conversation etc. These can be very subtle changes which make them very hard to detect by the carer.

## 6.3 Voice Interface for Compensation of Disabilities and Comfort

The most direct application is the ability to interact verbally with the smart home environment (through direct voice command or dialog) providing high-level comfort for physically disabled or frail persons. This can be done either indirectly (the person leaves the room and shuts the door, then the light can be automatically turned off) or directly through voice commands. Such system is important in smart homes for the improvement of comfort of the person and the independence of people that have physical disabilities to cope with. Moreover, recognizing a subset of commands is much easier (because of the relatively low number of possibilities) than an approach relying on the full transcription of the person' speech.

## 6.4 Distress Situation Detection

Everyday living sounds identification is particularly interesting for evaluating the distress situation in which the person might be. For instance, window glass breaking sound is currently used in alarm device. Moreover, in case of distress situation during which the person is conscious but cannot move (e.g., a fall), the audio system offers the possibility to call for help simply using her voice.

## 6.5 Privacy and Acceptability

It is important to recall that the speech recognition process must respect the privacy of the speaker. Therefore the language model must be adapted to the application and must not allow the recognition of sentences not needed for the application. It is important too that the result of this recognition is used to be processed on-line (for activity or distress recognition) and not to be analyzed afterward as a whole text. An approach based on keywords may thus be respectful of privacy while permitting a number of home automation orders and distress situations being recognized. Our preliminary study showed that most of the interrogated aged persons have no problem with a set of microphones installed at home while they categorically refuse any video camera.

Another important aspect of acceptability of the audio modality is the fact that such system should be much more accepted if it is useful during all the person's life (e.g., home automation system) rather than only during a important but very short period (e.g., a fall). That is why we have oriented our approach toward a global system (e.g., monitoring, home automation and distress detection). To achieve this, numerous challenges need to be addressed.

## 6.6 Recognizing Audio Information in a Noisy Environment

In real home environment the audio signal is often perturbed by various noises (e.g., music, roadwork...). Three main sources of errors can be considered: 1) The measurement errors which are due to the position of the microphone(s) with regard to the position of the speaker; 2) The acoustic of the flat; 3) The presence of undetermined background noise such as TV or devices

The mean SNR of each class of the collected corpus was between 5 and 15 dB, far less than the in-lab one. This confirms that the health smart home audio data acquired was noisy and explained the poor results. But this also shows us the challenges to obtain a usable system that will not be set-up in lab conditions but in various and noisy ones. Also, the sounds were very diverse, much more than expected in this experimental conditions were participants, though free to perform activities as they wanted, had recommendations to follow.

In our experiment, 8 microphones were set in the ceiling. This led to a good coverage of the area but prevented from an optimal recording of speech because the individuals spoke horizontally. Moreover, when the person was moving, the intensity of the speech or sound changed and influenced the discrimination of the audio signals between sound and speech; the changes of intensity provoked sometimes the saturation of the signal (door slamming, person coughing close to the micro). One solution could be to use head set, but this would be a too intrusive change of way of living for aging people. Though annoying, these problems are mainly perturbing for fine grain audio analysis but can be bearable in many settings. The acoustic of the flat is another difficult problem to cope with. In our experiment, the double glazed area provoked a lot of reverberation. Similarly, every sound recorded in the toilet and bathroom area was echoed. These examples show that a static and dynamic component of the flat acoustic must be considered. Finding a generic model to deal with these issues adaptable to every home is a very difficult challenge and we are not aware of any existing solution or smart home. Of course, in the future, smart homes could be designed specifically to limit these effects but the current smart home development cannot be successful if we are not able to handle these issues when equipping old-fashioned or poorly insulated home. Finally, one of the most difficult problems is the blind source separation. Indeed, the microphone records sounds that are often simultaneous as showed by the high number of mixed sounds in our experiment. Some techniques developed in other areas of signal processing may be considered to analyze speech captured with far-field sensors and to develop a Distant Speech Recogniser (DSR) such as blind source separation, independent component analysis (ICA), beam-forming and channel selection. Some of these methods use simultaneous audio signals from several microphones.

## 6.7 Verbal vs. Non Verbal Sound Recognition

Two main categories of audio analysis are generally targeted: daily living sounds and speech. These categories represent completely different semantic information and the techniques involved for the processing of these two kinds of signal are quite distinct. However, the distinction can be seen as artificial. The results of the experiment showed a high confusion between speech and sounds with overlapped spectrum. For instance, one problem is to know whether scream or sigh must be classified as speech or sound. Moreover, mixed sounds can be composed of both classes. Several other orthogonal distinctions can be used such as voiced/unvoiced, long/short, loud/mute etc. These would imply using some other parameters such as sound duration, fundamental frequency and harmonicity. In our case, most of the poor results can be explained by the lack of examples used to learn the models and the fact that no reject class has been considered. But choosing the best discrimination model is still an open question.

## 6.8 Recognition of Daily Living Sound

Classifying everyday living sounds in smart home is a recent trend in audio analysis and the "best" features to describe the sounds and the classifier models are far from being standardized (Cowling and Sitte, 2003)(Fleury et al., 2010) (Tran and Li, 2009). Most of the current approaches are based on probabilistic models acquired from corpus. But, due the high number of possible sounds, acquiring a realistic corpus allowing the correct classification of the emitting source in all conditions inside a smart home is a very hard task. Hierarchical classification based on intrinsic sound characteristics (periodicity, fundamental frequency, impulsive or wide spectrum, short or long, increasing or decreasing) may be a way to improve the processing and the learning. Another way to improve classification and to tackle ambiguity is to use the other data sources present in the smart home to assess the current context. The intelligent supervision system may then use this information to associate the audio event to an emitting source and to make decisions adapted to the application.

In our experiment, the most unexpected class was the sounds coming from the exterior to the flat but within the building (elevator, noise in the corridor, etc.) and exterior to the building (helicopter, rain, etc.). This flat has poor noise insulation (as it can be the case for many homes) and we did not prevent participants any action nor stopped the experiment in case of abnormal or exceptional conditions around the environment. Thus, some of them opened the window, which was particularly annoying considering that the helicopter spot of the hospital was at short distance. Furthermore, one of the recordings was realized during rain and thunder which artificially increased the number of sounds.

It is common, in daily living, for a person, to generate more than one sound at a time. Consequently, a large number of mixed sounds were recorded (e.g., mixing of foot step, door closing and locker). This is probably due to the youth of the participants and may be less frequent with aged persons. Unclassifiable sounds were also numerous and mainly due to situations in which video were not enough to mark up, without doubts, the noise occurring on the channel. Even for a human, the context in which a sound occurs is often essential to its classification (Niessen et al., 2008).

## 6.9 Speech Recognition Adapted to the Speaker

Speech recognition is an old research area which has reached some standardization in the design of an ASR. However, many challenges must be addressed to apply ASR to ambient assisted living. The public concerned by the home assisted living is aged, the adaptation of the speech recognition systems to aged people in thus an important and difficult task. Considering evolution of voices with age, all the corpus and models have to be constructed with and for this targeted population.

# 7 Conclusion and Future Work

Audio processing (sound and speech) has great potential for health assessment and assistance in smart home such as improving comfort via voice command and security via distress situations

detection. However, many challenges in this domain need to be tackled to make audio processing usable and deployed in assisted living applications. The paper presents the issues in this domain based on three experiments conducted in a health smart home involving the audio processing software AuditHIS.

The 2 first experiments were related to distress detection from speech. Most of the encountered problems were due to noise or environmental perturbation. The third experiment was related to the audio analysis of usual daily activities performed by fifteen healthy volunteers. The dataset was recorded in realistic conditions and underlines the main challenges that audio analysis must address in the context of ambient assisted living. Among the most problematic issues were the uncontrolled recording condition, the mixing of audio events, the high variety of different sounds and the complexity to discriminate them.

Regarding the latter, we plan to conduct several studies to determine what the most interesting features for sound classification are as well as how hierarchical modelling can improve the classification. Moreover, regarding speech recognition, probabilistic models need to be adapted to the ageing population. We are currently recording seniors' voice to adapt our ASR to this population. Finally, an audio technology needs to be improved to be efficient in health related application. The nationally financed Sweet-Home project, which focuses on home automation through voice orders, will permit additional study on speech recognition in smart homes. Keyword recognition and signal enhancement through Independent Component Analysis methods are part of this project.

# References


Baba, A., Yoshizawa, S., Yamada, M., Lee, A., and Shikano, K. (2004). Acoustic models of the elderly for large-vocabulary continuous speech recognition. *Electronics and Communications in Japan, Part 2, Vol. 87, No. 7, 2004*, 87(2):49–57.

Callejas, Z. and Lopez-Cozar, R. (2009). Designing smart home interface for the elderly. *Accessibility and Computing*, 95:10–16.

Chan, M., Estève, D., Escriba, C., and Campo, E. (2008). A review of Smart Homes - Present state and future challenges. *Computer Methods and Programs in Biomedicine*, 91(1):55–81.

Chen, J., Kam, A. H., Zhang, J., Liu, N., and Shue, L. (2005). Bathroom activity monitoring based on sound. In Heidelberg, S. B. ., editor, *Pervasive Computing*, volume 3468/2005 of *Lecture Notes in Computer Science*, pages 47–61.

Clavel, C., Devillers, L., Richard, G., Vasilescu, I., and Ehrette, T. (2007). Detection and analysis of abnormal situations through fear-type acoustic manifestations. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, pages IV–21–IV–24, Honolulu, HI.

Cohen, I. and Berdugo, B. (2001). Speech enhancement for non-stationary noise environments. *Signal Processing*, 81(11):2403–2418.

Cowling, M. (2004). *Non-Speech Environmental Sound Classification System for Autonomous Surveillance*. PhD thesis, Griffith University.

Cowling, M. and Sitte, R. (2003). Comparison of techniques for environmental sound recognition. *Pattern Recognition Letter*, 24(15):2895–2907.

Davis, K. H., Biddulph, R., and Balashek, S. (1952). Automatic recognition of spoken digits. *The Journal of the Acoustical Society of America*, 24(6):637–642.

Duong, T., Phung, D., Bui, H., and Venkatesh, S. (2009). Efficient duration and hierarchical modelling for human activity recognition. *Artificial Intelligence*, 137(7-8):830–856.

Ephraim, Y. and Malah, D. (1984). Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. In *Proceedings of the IEEE Transactions on Acoustics, Speech, and Signal Processing*, volume 32(3), pages 1109–1121.

Fleury, A., Vacher, M., and Noury, N. (2010). SVM-Based Multi-Modal Classification of Activities of Daily Living in Health Smart Homes: Sensors, Algorithms and First Experimental Results. *Information Technology in Biomedicine, IEEE Transactions on*, 14(2):274–283.

Gerosa, M., Giuliani, D., and Brugnara, F. (2009). Towards age-independant acoustic modeling. *Speech Communication*, 16:499–509.

Gorham-Rowan, M. and Laures-Gore, J. (2006). Acoustic-perceptual correlates of voice quality in elderly men and women. *Journal of Communication Disorders*, 39(3):171–184.

Gustafsson, S., Martin, R., Jax, P., and Vary, P. (2004). A psychoacoustic approach to combined acoustic echo cancellation and noise reduction. *IEEE Transactions on Speech and Audio Processing*, 10(5):245–256.

Ibarz, A., Bauer, G., Casas, R., Marco, A., and Lukowicz, P. (2008). Design and evaluation of a sound based water flow measurement system. In *Smart Sensing and Context*, volume 5279/2008 of *Lecture Notes in Computer Science*, pages 41–54. Springer Verlag.

Katz, S. and Akpom, C. (1976). A measure of primary sociobiological functions. *International Journal of Health Services*, 6(3):493–508.

Koskela, T. and Väänänen-Vainio-Mattila, K. (2004). Evolution toward smart home environments: empirical evaluation of three user interface. *Personal and Ubiquitous Computing*, 8(3-4):234–240.

Litvak, D., Zigel, Y., and Gannot, I. (2008). Fall detection of elderly through floor vibrations and sound. In *Proceedings of 30th Annual Int. Conference of the IEEE-EMBS 2008*, pages 4632–4635.

Michaut, F. and Bellanger, M. (2005). *Filtrage adaptatif: théorie et algorithms [Adaptative Filtering: theory and algorithms]*. Paris: Lavoisier.

Niessen, M., van Maanen, L., and Andringa, T. (2008). Disambiguating sounds through context. In *Proceedings of Semantic Computing, 2008 IEEE International Conference on*, pages 88 – 95.

Popescu, M., Li, Y., Skubic, M., and Rantz, M. (2008). An acoustic fall detector system that uses sound height information to reduce the false alarm rate. In *Proceedings of 30th Annual Int. Conference of the IEEE-EMBS 2008*, pages 4628–4631.

Rabiner, L. and Luang, B. (1986). An introduction to hidden markov models. *IEEE ASSP Magazine*, 3(1):4–15.

Renouard, S., Charbit, M., and Chollet, G. (2003). Vocal interface with a speech memory for dependent people. *Independent Living for Persons with Disabilities*, pages 15–21.

Soo, J.-S. and Pang, K. (1990). Multidelay block frequency domain adaptive filter. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(2):373–376.

Tran, H. D. and Li, H. (2009). Sound event classification based on feature integration, recursive feature elimination and structured classification. In *Proceedings of IEEE Int. Conference on Acoustics, Speech, and Signal Processing*, pages 177–180.

Vacher, M., Fleury, A., Portet, F., Serignat, J.-F., and Noury, N. (2010a). *New Developments in Biomedical Engineering*, chapter Complete Sound and Speech Recognition System for Health Smart Homes: Application to the Recognition of Activities of Daily Living, pages 645 – 673. Intech Book. ISBN: 978-953-7619-57-2.

Vacher, M., Portet, F., Fleury, A., and Noury, N. (2010b). Challenges in the processing of audio channels for ambient assisted living. In *Proceedings of IEEE HealthCom 2010*, pages 330–337, Lyon, France.

Vacher, M., Serignat, J., Chaillol, S., Istrate, D., and Popescu, V. (2006). *Speech and Sound Use in a Remote Monitoring System for Health Care*, volume Lecture Notes in Artificial Intelligence, 4188/2006, pages 711–718. Springer Berlin/Heidelberg.

Valin, J.-M. (2007). On adjusting the learning rate in frequency domain echo cancellation with double talk. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 15(3):1030–1034.

Valin, J.-M. and Collings, I. B. (2007). A new robust frequency domain echo canceller with closed-loop learning rate adaptation. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2007)*, pages I–93–I–96, Honolulu, HI.

Vaseghi, S. V. (1996). *Advanced Signal Processing and Digital Noise Reduction*. New York: John Wiley and Sons.

Vipperla, R., Renals, S., and Frankel, J. (2008). Longitudinal study of asr performances on ageing voices. In *Proceedings of the 9th Annual Conference of the International Speech Communication Association, INTERSPEECH'08*, pages 2550–2553, Brisbane, Australia.

*Michel Vacher received the Ph.D. degree in acoustical science from the INSA of Lyon, in 1982. Since 1986, he has been Research Scientist at the "Centre National de la Recherche Scientifique". He firstly joined the LTPCM laboratory where he worked on high resolution electron microscopy (HREM) image simulation and analysis. He joined the Laboratory of Informatics of Grenoble (ex CLIPS lab) at the end of 2000 to develop new research direction for smart house applications. His research interest includes auditory scene analysis, keyword/speech recognition in Smart Environments and aging voice recognition. He is project coordinator of the ANR Sweet-Home project and published some 37 papers on various aspects of these topics.*

*François Portet obtained his PhD in computing science at the University of Rennes 1 in 2005 where he stayed as a short-term lecturer until late 2006. In autumn 2006, he joined, as Research Fellow, the department of computing science of the University of Aberdeen. Since October 2008, he is associate Professor at the Grenoble Institute of Technology and at the Laboratoire d'Informatique de Grenoble. His research interests lie in the areas medical decision support systems, signal processing, data mining, and reasoning with uncertainty.*

*Anthony Fleury received an Engineer (Computer Science) and a M.Sc. (Signal Processing) degree in 2005 in Grenoble and a PhD degree in Signal Processing from the University Joseph Fourier of Grenoble in 2008 for his work on Health Smart Homes and activity recognition. He joined then the LMAM team at Swiss Federal Institute of Technology and is now, since sept. 2009, Associate Professor at Ecole des Mines de Douai. His research interests include the modeling of human behaviors and activities, machine learning and pattern recognition with applications to biomedical engineering.*

*Norbert Noury has specialized in the field of "smart sensors" for Ambient Assisted Living environments and for Ubiquitous Health monitoring systems. He received the M.Sc Electronics from Grenoble Polytechnic Institute (1985) and the PhD Applied Physics from Grenoble University (1992). He spent 8 years in various industrial companies (1985-93), and then joined the Grenoble University (1993) where he founded a new research activity on Health Smart Homes (2000). Norbert Noury is now a full Professor at the University of Lyon. He has guided 16 PhD students, authored or co-authored about 150 scientific papers (including 30 journal papers, 80 communications in International Conferences, 8 patents, 14 book chapters) and is a recognized expert at the European Commission.*