Exploiting Environmental Sounds for Activity Recognition in Smart Homes

Sebastien Tremblay, Dany Fortin-Simard, Erika Blackburn-Verreault, Sebastien Gaboury, Bruno Bouchard and Abdenour Bouzouane
Universite du Quebec a Chicoutimi (UQAC) Chicoutimi, G7H 2B1, Canada
{sebastien.tremblay8, dany.fortin-simard, erika.blackburn-verreault1, s1gabour, bruno.bouchard, abdenour.bouzouane}@uqac.ca

Abstract

The number of elderly and frail individuals in need of daily assistance increases and the available human resources will certainly be insufficient. To remedy this situation, smart habitats are considered by many researchers as an innovative avenue to help support the needs of elders. It aims at providing cognitive assistance in taking decisions by giving hints, suggestions, and reminders with different kinds of effectors to residents. To implement such technology, the first challenge we need to overcome is the recognition of the ongoing activity. In the literature, some researchers have proposed solutions based on cameras, binary sensors, radio-frequency identification and load signatures of appliances but all these types of approaches have certain limitations to perform a complete recognition. In order to provide additional and useful information, a complementary activity recognition system, based on environmental sounds and able to detect errors related to cognitive impairment, is presented in this paper. The entire system relies on a discrete wavelet transform, the zero-crossing rate and C4.5 algorithm. This system has been implemented and deployed in a real smart-home prototype. This paper also present the results of a first set of experiments conducted on this system with real cases scenarios.

Introduction

Developed countries are actually facing an important population ageing challenge, due to falling birthrates and rising of life expectancy (Nations 2009). This issue leads to significant social and economic problems, including medical staff shortages for home-care services, and an increasing number of people suffering from cognitive impairment (e.g. Alzheimer's disease) (Abdulrazak et al. 2011). With the advances of pervasive computing, ambient intelligence and the miniaturization of technology, many believe that promoting longer life at home by enhancing the residence of frail people is the most promising solution to these challenges. The idea of these technologically enhanced homes, called "smart home" (Pigot et al. 2003, Augusto et al. 2006), has led researchers to develop adapted algorithms for security, services delivery, opportunistic networking, etc. One problem that was largely studied in Artificial Intelligence (AI) is the problem of plan recognition or more precisely in that case, the key idea of Human Activity Recognition (HAR). The main idea is to take raw data from detection technologies, filter and transform them into relevant information to be able to associate it with basic activities of daily living performed by residents. Recognizing the ongoing activity is the first step to provide assistance to a frail or cognitively impaired resident. Multiple solutions were presented using cameras (Mihailidis A. 2008) because they allow obtaining very accurate information. Although the information provided is very rich and interesting, the cameras are intrusive and they infringe on the privacy of residents. On the other hand, teams of researchers have also approached activities daily living (ADL) recognition problem by combining binary sensors. Cook et al. (2007) presented solid works that performed well despite the low-quality information provided by this type of sensor. However, these works are still limited in the abilities to recognize complex scenarios. Alternatives such as RFID technology (Bouchard et al. 2013) and analysis of electrical signatures (Belley et al. 2013) are also considered. These types of technology have already proven themselves to be very effective and not intrusive. RFID can acquire rich information such as the position of objects and spatial relationships. On the other side, the electrical analysis systems are innovative and useful to collect relevant information.

In order to provide additional and useful information, in this paper, we present a complementary recognition system based on environmental sounds and that can detect errors related to cognitive impairment. Our sounds recognition system is based on wavelet decomposition model and the zero-crossing rate. Besides, it uses a decision tree constructed by an automatic learning mechanism using the well-known C4.5 algorithm. Specifically, our model is not intended as a complete system for HAR. However, the objective is to provide precise information that is missing in other systems. Usually, the approaches proposed in the literature do highest level of recognition (e.g. prepare a lunch), but the goal of the present system is to provide inaccessible and low-level information (e.g. brew coffee). We propose an original real time approach which not infringe on privacy by not analyzing the sound of the voice. In the literature, only few works exploit this technology in smart home because unstructured sound analysis is a difficult branch of sound recognition (Chachada and Kuo 2013).

This paper is divided as follows. Section 2 describes the existing approaches for recognition of activity in smart homes. Section 3 presents in details our environmental sounds activity recognition model. Section 4 presents the experiments that were carried out and the analysis of the results. Finally, Section 5 provides a conclusion and presents the future work that will be done soon.

Related work

The issue of HAR for smart homes is addressed by many researchers because the benefits of such effective systems are both economic and social. Even though many approaches already exist, there are many challenges remaining. Conditions and contexts are different and there is no ideal solution that will solve the whole problem. Mihailidis et al. (2008) presented the cognitive assistant system COACH dedicated to patients suffering from Alzheimer's disease. The objective of this system is to conduct surveillance of a patient performing a specific task of everyday life. For instance, cleaning his hands and offers them assistance (i.e. a voice warning or guidance) in the most appropriate way only when the situation requires. To do so, COACH uses a single camera as sensor and it is one of the most famous assistance systems. However, it has several limitations. Using cameras is ethically challenging and it significantly undermines the robustness of the system. Besides, cameras are expensive and fragile. In their work, Cook et al. (2007) have obtained good results with only binary sensors. In fact, the binary data collection system consists of an array of motion sensors, which collect information using small devices. In their laboratory, there are over 100 sensors deployed including light, temperature, humidity, and reed switches. Specifically, their model of learning relies heavily on time to make temporal recognition of activity. In fact, it requires an analysis of large volumes of data to detect automatically interesting patterns or relationships that allow for better understanding. These types of approaches have been used for many years and they have proven their robustness. However, they are still limited in the abilities to recognize elaborate scenarios. Furthermore, it is important to note that the deployment of this technology is very complex, since it requires the use of a large number of sensors. On the other hand, RFID technology has regained in popularity among the scientific community in recent years. At first glance, RFID systems do not have much interest in smart homes. However, many research teams developed localization algorithms based on this technology. A large part of them introduces the concept of location tags references placed at strategic location. Vorst et al. (2008) and Joho et al. (2009) built solution that rely on the large-scale tag references. However, this type of approach is not very suitable, nor always possible in the context of smart home. Other teams of researchers (Bouchard et al. 2013) have examined the issue and have developed new approaches of activity recognition based on accurate RFID localization of objects using the strength of received signals. Although these approaches are very promising, but it was limited by a lack of information that is not able to provide. Furthermore, although non-intrusive appliance load monitoring is used in many areas, this approach is not very widespread in the field of smart homes. Only certain teams worked to analyze the variation of electrical signals and try to recognize the signatures of devices. Among these, Belley et al. (2013) and Camier et al. (2013) have implemented such a system in a real smart home infrastructure. Although this approach is robust and perfectly suitable for smart homes, it is limited by the amount of information it provides. In fact, this approach is limited only to the recognition of ADLs involving electrical appliances. On the other hand, approaches based on sounds are not sufficient to perform complete activity recognition. However, sound analysis provides information that is interesting and often missing with other approaches. For example, if a patient uses a mixer, without sound analysis, it is hardly possible to detect if the mixer is on and at which level of intensity. In fact, that is just an example among many others that proves the usefulness of this additional technology in smart home context.

Activity recognition model based on environmental sounds

To address the limitations of the systems presented, this paper presents a new complementary approach for activity recognition based on the analysis of the environmental sounds. Indeed, in the context of smart home, using this technology has many advantages. First, microphones are very robust and easy to deploy. Besides the majority of smart homes are mostly already equipped with this technology. Furthermore, information provided by sounds is complementary to other technologies which will allow recognizing a wide variety of additional scenarios. It is also interesting to note that using microphones in our way is less intrusive than speech recognition which is essential in this context. In our case, we do not analyze the spectrum of voice. The figure 1 gives an overview of the intelligent system and its components. Specifically, the following sections discuss in detail the theoretical and practical foundations of each of these elements.

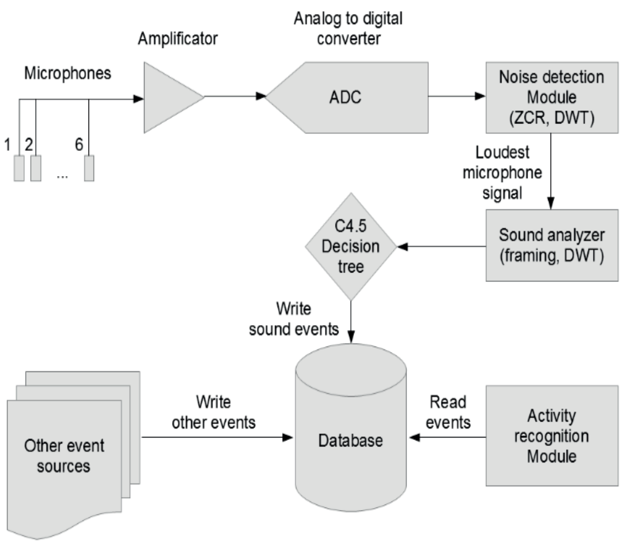

Figure 1: Environmental sound recognition schema

The detection and analysis of digital noise signals is a broad field (Wu and Kung 2004). In the literature, there are several options with which to define the measured characteristics; for example, energy, the probability of a statistical model, and the high order statistics (Nemer et al. 2001). Most existing systems attempt to detect the human voice and not impulsive sound (Marzinzik and Kollmeier 2002). Speech detection is based on the properties of the words such as spectral characteristics (Gazor and Zhang 2003), and the linear prediction coefficients (LPC) (Tanyer and Ozer 2000). On detection of impulsive behavior, there is not a lot of work. Among these, Dufaux et al. (2001) proposed three detection algorithms with impulse noise that has good results: one based on the variance of the signal energy, and the other two algorithms based on the median filter conditioning energy. Moreover, Vacher et al. (2010) use Linear-Frequency Cepstral Coefficients (LFCC) in their work. These features are used because environmental sounds are most discriminated with constant bandwidth filters. Besides, it would also have been possible to use the Mel-Frequency Cepstral Coefficients (MFCC) that are most commonly used for voice recognition.

For our part, to be able to analyze the environmental sounds correctly, we based our analysis on discrete wavelets transform (DWT) and the zero crossing rate (ZCR).

All signal *s(t)* can be decomposed in a sum of functions $\psi_{u,s}(t)$ localized and weighted by $\kappa_{u,s}$:

$$s(t) = \sum_{u,s} \kappa_{u,s} \psi_{u,s}(t) \qquad (1)$$

where u is the time shift (a constant for Fourier Transform) and s is the scale factor. The difference between the Short Time Fourier Transform and the Wavelet Transform rely on the choice of the function $\psi_{u,s}(t)$. The DWT is often preferred because it allows detecting impulsive signal. The wavelet base is obtained by translation and dilatation of the mother wavelet ψ. In our case, we selected, as mother wavelet, the Daubechies wavelets with 4 vanishing moments in computing the DWT (Akansu and Haddad 2000, Dragotti and Vetterli 2003).

In particular, we used the method proposed by Mallat et al. (1989) which computes the Fast Discrete Wavelet Transform FWT for frame length of 128ms. During the FWT, the signal is filtered by a cascade of low-pass and high-pass filters which the cutoff frequency is artificially reduced by half at each decomposition level. In our case, we took level 3 decomposition. Thus, the FWT returns four wavelet transform coefficients: three detailed coefficients (detailed level 1, detailed level 2 and detailed level 3) and an approximation coefficient (approximation level 3) for each frame of analysis. Thereafter, the energy of each coefficient ($E(d_1), E(d_2), E(d_3), E(a_3)$) is computed. We can see in Figure 2, an example of a short signal and its corresponding analysis which consists in the energy of each coefficients ($E(d_1), E(d_2), E(d_3), E(a_3)$) at each frame.

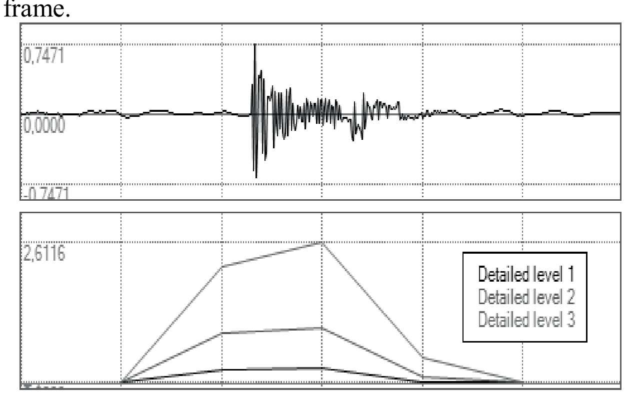

Figure 2: Audio signal processing

As mentioned above, we also used the zero-crossing rate which is the rate of sign-changes along a signal, i.e., the rate at which the signal changes from positive to negative or back (Chen 1988). The ZCR was widely used in sound recognition to recognize if a perceived change is relevant. It is defined by:

$$ZCR = \frac{1}{T}\sum_{t=0}^{T-1}\prod\{s_t s_{t-1} < 0\} \qquad (2)$$

where s is a signal of length T and $\prod\{A\}$ is an indicator function which is 1 if its argument A is true and is 0 otherwise.

After making the preliminary signal processing, we conducted a sampling period to obtain several data. Data were sampled according to the different characteristics of the sounds (ZCR, DWT) to create distinctive vectors. Then we used the C4.5 algorithm on the data collected to create a decision tree for the recognition of different sounds as you can see in Figure 3. Precisely, C4.5 is an algorithm used to generate a decision tree developed by Ross Quinlan (1993). C4.5 builds decision trees from a set of training data using the concept of information entropy. This automatic method of tree generation makes it very easy to upgrade our system by adding new relevant sounds. This is a real advantage of our model. In addition, we conducted a distinction between continuous and punctual sounds. This distinction has been made because a continuous sound has a variable duration and therefore it has a beginning and an end. In contrast, a punctual event has a relatively fixed duration and it is more chaotic than a continuous event. This distinction was made based on the need to treat them differently.

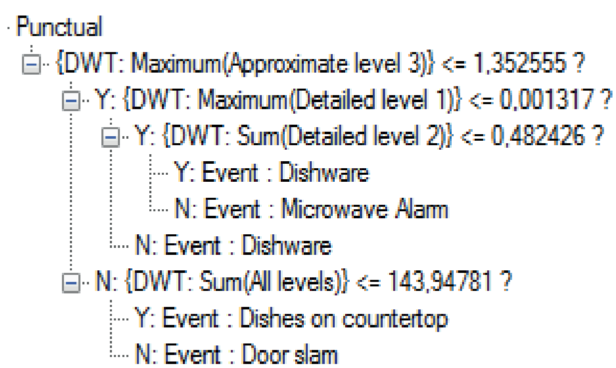

Figure 3: C4.5 decision tree for punctual sounds

It is essential to build a robust computational intelligent system having a maximum of scalability, because it is sometimes difficult to adjust or modify the system. In fact, the activity recognition module has been implemented with a symbolic algorithm. Specifically, it is a module based on rules that makes the link between sounds recognized and the activities library. The figure 4 briefly presents the entire HAR algorithm. For our part, since the objective of this paper was to evaluate the effectiveness of such a system, we have not implemented an advanced algorithm for HAR.

```
Output : Monitoring report
Loop
  For each microphone
    Acquire sound sample for a given time frame
    Compute zero crossing rate
    Compute discrete wavelets transform
  End
  If features meet the minimum thresholds values
    Select microphone having highest sum of detailed level 3
    DWT
    Loop
      Acquire another sound sample for a given time frame
      allowing 50% overlap with previous frame
      Compute zero crossing rate
      Compute discrete wavelets transform
    Until thresholds are no longer met
    Match acquired sample's features against C4.5 decision
    trees
    Write matched sound event to database
    Read 5 last events from database
    Match these events against activity descriptions
    If an activity is matched
      Raise activity event
    End
  End
Until the monitoring is stopped
```

Figure 4: Activity recognition pseudo code

Experimentation

To validate our model, we used the new cutting-edge smart-home infrastructure of the LIARA laboratory. Our prototyping space comprises over a hundred sensors. These are hidden as much as possible in order to keep the environment similar to a real apartment. Among the sensors, there is RFID technology, electromagnetic sensors, accelerometers, audio sensors, ultrasonic sensors, and more. We integrated all these different technologies for prototyping and developing algorithms. We also have many effectors that are strategically placed around the smart home in order to provide prompt assistive services to the resident whenever needed. For instance, there are a few screens to show guidance video as there are IP speakers installed in every corner. The experimental space was the entire apartment where six microphones were installed at different positions. Also, we used an amplifier and an Advantech USB-4716 ADC module for better indoor coverage. Microphones were strategically placed as recognition can be achieved. The entire apartment was also the ideal choice for our tests. In fact, ADLs are varied and analysis of all the sounds of the apartment allows us to monitor properly.

The library was created with a total of 1219 samples of sounds to be able to create a representative and effective decision tree. Following the generation of the tree, we obtained classification rates of 92.945% which is excellent. Then, to validate the effectiveness of our system of sound recognition, another human subject performed 200 tests reproducing 8 typical sounds representing step of an activity. The algorithm was able to correctly identify each of them 79.88% of the time (recognition rate). Based on our experience, the error rates of our system were mainly caused by the difficulty of identifying the impulsive sounds, which sometimes have an important fluctuation in reading. Occasionally, the sounds are not well captured and they are not properly identified which cause recognition errors. In our experiments, as shown in figure 6 and 7, we also tested the effectiveness of our system on punctual and continuous sounds. To this end, the adjustment done was very effective.

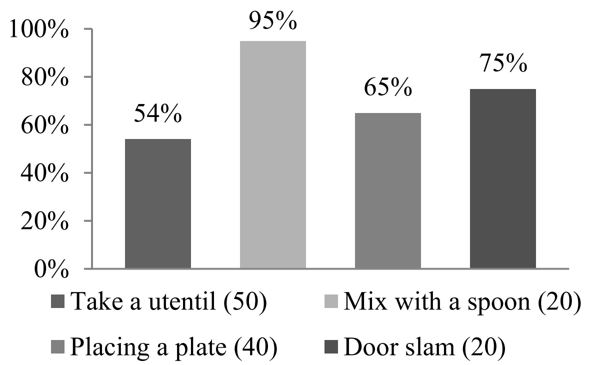

Figure 6: Punctual sound recognition rates

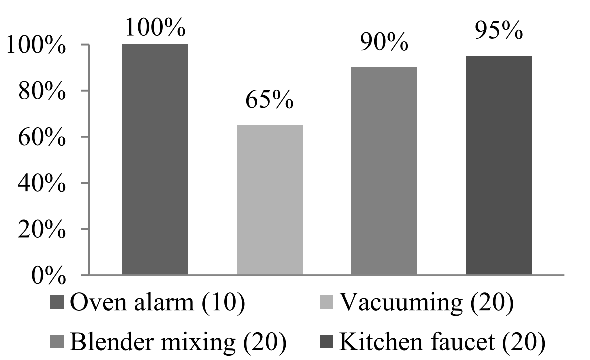

Figure 7: Continuous sound recognition rates

Secondly, we selected several ADLs (Breakfast, Household) that have been modeled in series of step (e.g. step1: fill the kettle, step2: grab a cup…) and the system tries to identify the activity correctly by following the different steps. The average rate of recognition obtained was 70.50% which is still not bad for a development system. Obviously, it is our first testing phase. In near future, we plan to test our system with much more scenarios and we want to integrate it with our complete multi-sensors HAR systems. Following that, the next step will be to test the entire systems with real participants.

To confirm the contribution of our recognition system, we conducted some additional experiments. In fact, we realize several tests that allowed us to confirm that the system is useful to recognize certain steps that are difficult to detect with other technologies. For example, if we take the classic scenario of achieving a coffee with the sound recognition system, we will be able to confirm that the coffee has been brewed, the cabinet has been closed and that the kettle was filled with a sufficient amount of water. This may seem unnecessary, but if we really want to attend a patient on a daily basis, it is necessary to monitor every movement.

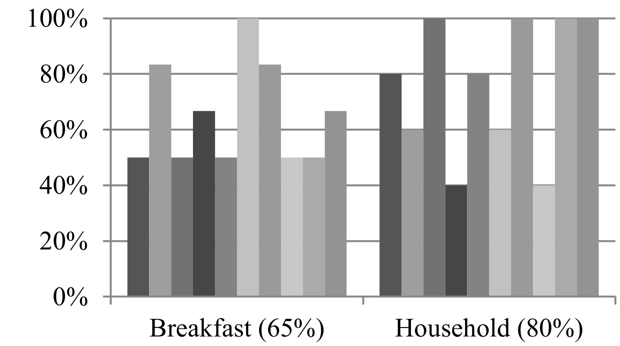

Figure 8: Step recognition rates on two ADLS

Conclusion and future works

In this paper, we described our recent progress towards the development of our complementary HAR system based on the analysis of environmental sounds. We have implemented and tested it in an infrastructure on the cutting edge of technology. Furthermore, we have described in detail the theoretical foundations of our first version of the algorithm. An approach based on discrete wavelet transform, the zero-crossing rate and the C4.5 algorithm. In fact, this model has several advantages for smart home support, because it does not rely on expensive, invasive or difficult to deploy technology. We have shown that the analysis of the sound spectrum can recognize usually imperceptible events proving the potential of this method. Otherwise, our approach is innovative because only few works in the literature have addressed the sound recognition in the context of HAR in smart home.

Despite our results, there are still some limitations that we have to work in the near future. First, we must improve our sound recognition algorithm. To do this, we will soon integrate to our model some additional spectral features such as the spectral centroid and the roll-off point which demonstrate to be very effective in the analysis of unvoiced sounds. We also want to add MFCC analysis that is commonly used in the treatment of voice but we believe that this technique could be as highly effective in the treatment of unvoiced sounds. In addition, in our next experimentation, it would be appropriate to vary the signal noise ratio to successfully analyze in case of disruption. Finally, in our future work, we want to integrate the sound analysis system to our complete multi-sensors HAR systems and we want to test our complete system on people with cognitive disabilities because we believe that the information provided by this complementary system could be strategically exploited.

Acknowledgments

We would like to thank our main financial sponsors: the Natural Sciences and Engineering Research Council of Canada, the Quebec Research Fund on Nature and Technologies, the Canadian Foundation for Innovation and Bell Canada.

References

Abdulrazak B.; Giroux S.; Bouchard B.; Pigot H.; Mokhtari M. 2011. *Toward Useful Services for Elderly and People with Disabilities.* Proc. of International Conference on Smart Homes and Health Telematics, ICOST, Montreal, Springer LNCS: 327.

Akansu A.N.; Haddad P.R. 2000. Multiresolution signal decomposition: transforms, subbands, and wavelets, Academic Press.

Augusto J., and Nugent C. 2006. *Designing Smart Homes: the role of artificial intelligence, State of the Art Survey.* Lecture Notes in Artificial Intelligence (LNAI 4008), Springer-Verlag: 1-183.

Belley C., Gaboury S., Bouchard B., and Bouzouane A. 2013. An efficient and inexpensive method for activity recognition within a smart home based on load signatures of appliances. Pervasive and Mobile Computing.

Bouchard K., Fortin-Simard D., Gaboury S., Bouchard B., and Bouzouane A. 2013. *Accurate RFID Trilateration to Learn and Recognize Spatial Activities in Smart Environment.* International Journal of Distributed Sensor Networks: 1-16.

Camier, T.R., Giroux S., Bouchard B., and Bouzouane A. 2013. *Designing a NIALM in Smart Homes for Cognitive Assistance*" Procedia Computer Science 19: 524-532.

Chachada, S., and Kuo, C.-C.J. 2013. *Environmental sound recognition: A survey.* Signal and Information Processing Association Annual Summit and Conference (APSIPA), Asia-Pacific, IEEE.

Chen, C.-H. 1988. *Signal processing handbook*, CRC Press.

Dragotti, P.L., and Vetterli M. 2003. "Wavelet footprints: theory, algorithms, and applications." Signal Processing, IEEE Transactions on 51(5): 1306-1323.

Dufaux, A. 2001. *Detection and recognition of impulsive sounds signals*. Institute de Microtechnique Neuchatel, Switzerland.

Gazor S., and Zhang W. 2003. *A soft voice activity detector based on a Laplacian-Gaussian model.* Speech and Audio Processing, IEEE Transactions on 11(5): 498-505.

Jakkula V.R., and Cook D.J. 2007. *Mining Sensor Data in Smart Environment for Temporal Activity Prediction.* Poster session at the ACM SIGKDD, San Jose, CA.

Joho D., Plagemann C. and Burgard W. 2009. *Modeling RFID signal strength and tag detection for localization and mapping.* Proceedings of the IEEE international conference on Robotics and Automation. Kobe, Japan, IEEE Press: 1213-1218.

Mallat S.G. 1989. *A theory for multiresolution signal decomposition: the wavelet representation.* Pattern Analysis and Machine Intelligence, IEEE Transactions on 11(7): 674-693.

Marzinzik M., and Kollmeier B. 2002. *Speech pause detection for noise spectrum estimation by tracking power envelope dynamics.* Speech and Audio Processing, IEEE Transactions on 10(2): 109-118.

Mihailidis A., Boger J., Canido M., and Hoey J. 2008. The COACH prompting system to assist older adults with dementia through handwashing: An efficacy study. BMC geriatrics 8(1): 28.

Nations U. 2009. *World Population Ageing 2009*. Department of Economic and Social Affairs: Population Division: 129.

Nemer E., Goubran R., and Mahmoud S. 2001. *Robust voice activity detection using higher-order statistics in the LPC residual domain.* Speech and Audio Processing, IEEE Transactions on 9(3): 217-231.

Pigot H., Mayers A., Giroux S. 2003. *The intelligent habitat and everyday life activity support*. 5th international conference on Simulations in Biomedicine.

Quinlan J.R. 1993. *C4. 5: programs for machine learning*, Morgan kaufmann.

Tanyer S.G., and Ozer H. 2000. *Voice activity detection in nonstationary noise.* IEEE Transactions on Speech and Audio Processing 8(4): 478-482.

Vacher M., Fleury A., Portet F., Serignat J.-F., and Noury N. 2010. Complete sound and speech recognition system for health smart homes: application to the recognition of activities of daily living. New Developments in Biomedical Engineering: 645-673.

Vorst P., Schneegans S., Bin Y., and Zell A. 2008. *Self-Localization with RFID snapshots in densely tagged environments*. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Nice, France.

Wu Y., and Kung S.-Y. 2004. *Signal detection for MIMO-ISI channels: An iterative greedy improvement approach*. Signal Processing, IEEE Transactions on 52(3): 703-720.