

Automatic Monitoring of Activities of Daily Living based on Real-life Acoustic Sensor Data: a preliminary study

Lode Vuegen^{1,2,3}, Bert Van Den Broeck^{1,2,4}, Peter Karsmakers^{1,2,4}, Hugo Van hamme³,
Bart Vanrumste^{1,2,4}

¹MOBILAB, TM Kempen, Kleinhoefstraat 4, 2440, Geel, Belgium

²iMinds, Future Health Department, Kasteelpark Arenberg 10, 3001, Leuven, Belgium

³ESAT-PSI, KU Leuven, Kasteelpark Arenberg 10, 3001, Leuven, Belgium

⁴ESAT-SISTA, KU Leuven, Kasteelpark Arenberg 10, 3001, Leuven, Belgium

Lode.Vuegen@kuleuven.be

Abstract

This work examines the use of a low-power Wireless Acoustic Sensor Network (WASN) for the observation of clinically relevant activities of daily living (ADL) (e.g. eating, personal hygiene, toilet usage, etc.) from elderly. The sensors used in the WASN are both audio and ultrasound receivers. To the best of our knowledge, the combination of audio and ultrasound as a basis for ADL monitoring has not been investigated yet. This paper describes a baseline approach for ADL classification based on Gaussian mixture models. Preliminary results in this work indicate that classification accuracies up to 85.0 % \pm 14.6 for audio and 61.7% \pm 11.3 for ultrasound are already achievable on realistic real-life recorded data.

Index Terms: acoustic scene analysis, audio, ultrasound, acoustic scene classification, activities of daily living, automatic monitoring

1. Introduction

Because of the retirement of the baby-boom generation and the increasing life expectation, the ratio of dependent elderly to working people is rising sharply. Research predicts that in 2020, 19% (extrapolated to 26% in 2060) of the Belgian population will be older than 65 years [1, 2]. This aging brings important challenges for our society. One of these challenges is to facilitate a safe functioning of elderly people in their own home environment. Even solutions which only allow a small additional fraction of care-requiring elderly to live longer safely at home, for a reasonable investment, make economical sense.

The functioning of elderly at home is often limited by underlying physical or cognitive dysfunctions, which are often the cause of diseases. Nowadays, these changes in functioning are often unrecognized, or recognized too late. One of the reasons for under-detecting these changes is that they are often minimal and not noticed or ignored by the patient or family. Still, early detection could lead to early intervention and prolong the possibility to live safely at home. As technological support, a monitoring system aims to detect and analyze relevant changes and create a safe environment to the elderly at home. More specifically, the aim of this research is to provide the caregiver objective information, compiled in a summary report, concerning the daily activities of the elderly.

The present solutions found in the literature can be split into two main categories based on the type of sensors used. A first category requires the use of sensors that make contact with

the human body such as Radio Frequency Identification (RFID) readers [3, 4, 5], accelerometers [6, 7, 8, 9, 10] and gyroscopes [11]. A second category uses contact-less sensors. These have the advantages over wearable sensor systems in that they do not affect the normal behavior of the users, do not require human interaction (e.g. such as a push button system), and cannot be forgotten to wear. In [12] a survey of different approaches to detect human activities using video images is discussed. Aside video cameras also other contact-less sensors were explored in this context such as infra-red sensors [13], door contacts [13], radars [14], sensors for monitoring the use of domestic utilities [13, 15, 16] and microphones [17, 18, 19, 20, 21]. Compared to the other modalities, acoustic (specifically audio) technology has received little attention. Few research groups have considered using daily living acoustics in their systems.

Our research investigates the use of a wireless acoustic sensor network (WASN) that extracts information from both the audio and ultrasound frequency range. Such networks contain multiple so-called nodes each holding one or more acoustic sensors. These WASNs have advantages over other kinds of setups. For instance, the nodes can be small while maintaining large spatial sampling [22]. The nodes can be placed in a room without inconvenient cables, which is a desirable property in a home environment. Additionally, the workload (which can be significant) can be distributed among nodes so that cheaper hardware can be selected [22]. A WASN allows to estimate the source position from the acoustic signal and increase the quality of the recorded signals through spatial filtering [22, 23]. To our knowledge such a WASN setup that extracts acoustic information from audio and ultrasound signals for the purpose of home monitoring has not yet been reported in the literature. Aside the clinically relevant information that is present in the audio frequency range it is also investigated which useful information could be extracted from the ultrasonic frequency range. More in detail, it will be examined if typical ADLs (e.g. eating, personal hygiene, walking, etc.) can be detected and recognized in the ultrasound spectrum. The combination of audio and ultrasound might have the following advantages over existing contact-less alternatives: a) occlusions might have less impact compared to a video-based system, b) less processing power might be needed compared to video-based approaches, c) ultrasound and audio signals might provide complementary information, d) is expected to be easily integrated with dialog systems (notably for virtual assistants or robots), with emergency and security systems (mainly fall detection and distress

situation recognition), is expected to be augmented with human machine interaction (e.g. voice commands, conversational systems), e) can be extended with ultrasound transmitters which allows to estimate object distances to detect changes in the living environment.

This paper describes a baseline architecture for daily activity observation via audio and ultrasonic measurements and discusses the preliminary results obtained on realistic real-life recorded data. The work will serve as a starting point for further improvements of the classification accuracy and required annotation efforts for model estimation.

In section 2 we will briefly discuss the used experimental setup. Topics such as hardware configuration, living environment, and the performed Activity of Daily Living (ADL) scenarios will be clarified. Section 3 describes the baseline system architecture and clarifies a functional block diagram of the proposed solution. Section 4 discusses the feature extraction process and how these features are implemented in a classifier. The conducted experiments and obtained results are presented and clarified in section 5. In Section 6 the findings will be discussed and is followed by the conclusions in Section 7.

2. Experimental setup

2.1. Hardware

The acoustic sensor network used during the recordings consisted of two different types of nodes, i.e. audio and ultrasound, and are briefly explained in the following two paragraphs.

Each audio node was equipped with three linearly spaced electret microphones with an inter-sensor distance of 6.8 cm. The microphone signals are sent to preamplifiers with a cut-off frequency of 20 kHz to improve the signal level.

The ultrasound node consisted of three 40 kHz centered ultrasound receivers with an inter-sensor distance of 10 cm. Next, the captured ultrasound spectrum is downshifted to the audible frequency range to make it recordable with standard audio hardware. The latter is done by analog mixing the ultrasound signal with a square block wave of 31.5 kHz which results into a transformed center frequency of 8.5 kHz. Next, the downshifted signal is filtered by a 10 kHz low pass filter to cancel out the higher order harmonic product terms. The motivation for downshifting to a center frequency of 8.5 kHz instead of direct to DC (and using a square wave of 40 kHz) is merely because it is expected that the lower ultrasound frequencies (range from 31.5 till 40 kHz) also contain valuable information.

All captured acoustic signals were recorded using a 4 channel 24-bit soundcard sampling at 32 kHz. Each soundcard additionally recorded a reference signal that was received from a single transmitter through a RF channel. These reference signals were used for the purpose of off-line synchronization of the different inter node channels as described in [24].

2.2. Living environment

The domestic environment used in this work was a room of size 6 m by 4 m with a combined kitchen and living space as shown in Figure 1. Each of the 4 corners was equipped with an audio node to ensure full coverage of the acoustic sensor network. Additionally, the use of multiple node reduces the maximum possible distance between source and node and thereby increasing the SNR of the received signals as well. In contrast with the audio nodes, only a single ultrasound node was available for installation in the domestic environment. The most suitable position for this node with respect to the maximum possible

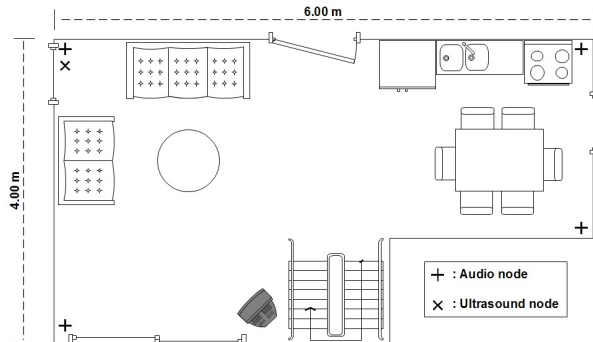


Figure 1: Floor plan of the domestic environment.

coverage was the the corner in the living room.

2.3. Recording scenario and data

The collection of audio and passive ultrasound data from clinically relevant domestic events is required to analyze and explore the proposed observation system. Therefore, during this data collection session eight different people performed some typical ADLs over a time span of three days in the domestic environment. Table 1 gives a detailed overview of the collected data in terms of both audio and ultrasound.

Table 1: A detailed overview of the collected audio and ultrasound data.

Activity class	Number of examples	Recording Duration (minutes)
Cooking and eating	7	287.01
Reading	2	22.04
Using laptop	2	21.16
Vacuum cleaning	4	34.15
Walking around	6	59.12
Watching TV	3	73.44

3. System architecture

The proposed system architecture is shown in Figure 2. Each node first estimates whether or not the input contains acoustic information by using a sound activity detector (SAD). Since a wide range of acoustics can be useful in recognizing an activity it is difficult to select a certain model-based sound activity detector. Therefore, a simple energy based threshold is implemented as SAD. First each sample is squared and thresholded. Then a hangover scheme labels each sample within 25 ms of a sample which passed the threshold as a positive detection. If acoustic information is detected, the raw waveform data will be further processed into acoustic and position features (as described in 4.1 and 4.2). Both acoustic and position features are calculated on 25 ms blocks of data with a time shift of 10 ms. The SAD is also used to estimate the signal-to-noise ratio (SNR) at which the acoustic information is received. Only this low dimensional information (SNR, acoustic and position features) is sent to a central processor. This strategy reduces the necessary bandwidth and power consumption.

The central processor combines the position features of all nodes and only selects the acoustic features from the micro-

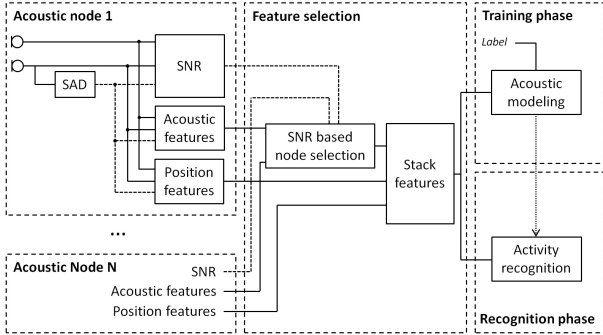


Figure 2: Block schematic of system architecture.

phone that receives the acoustic signal with the highest SNR. Once the features are combined, these will form the basis for the training and testing phase. It is worth mentioning that Figure 2 depicts a simplified architecture. In practice the block node selection will notify each node whether or not its acoustic features are needed such that no unnecessary CPU time nor bandwidth will be wasted.

4. Feature extraction and modeling

As discussed in previous sections, this work aims to reveal ADLs from the associated acoustic information. In order to optimize the classification objective, the raw stream of acoustic data is transformed into more consistent features. Therefore, two types of features will be extracted from collected sensor data, i.e. acoustic source localization features and acoustic features. It is expected that these two feature sets contain complementary information which will boost the classification performance. For instance, running water detected in the bathroom is more associated to personal hygiene than to cooking.

4.1. Acoustic features

Most of the presently available acoustic feature extraction approaches find their origin in speech applications and are often based on the properties of human speech production and perception. A well-known and often used feature extraction approach in the domain of speech- and speaker recognition applications are the so-called Mel-Frequency Cepstral Coefficients (MFCCs) [25]. Despite the fact that MFCCs are initially developed for speech applications, research indicates that MFCCs are also a successful choice for processing non-speech acoustic signals as well [26, 27]. Therefore, this work will use the MFCC approach as a baseline for computing the acoustic features from the collected audio data. The feature extraction process for the downshifted ultrasound signals is slightly different. Here, linearly spaced filter banks make more sense since there is no reason to assume that the frequency resolution should be significantly different at the low versus the high end of the spectrum. This changes the Mel-Frequency Cepstral Coefficients into a linear alternative which is denoted as Linear-Frequency Cepstral Coefficients (LFCCs) [28].

Both the MFCC- and LFCC-features will be extracted using the same parameter setting. Literature indicates that window sizes of 25 ms with an overlap of 10 ms are typically used [25]. The number of filterbanks is set to 40. The filtering operation is followed by a 13th order Discrete Cosine Transform. Finally, the Δ and $\Delta\Delta$ are computed and added as acoustic fea-

tures. This leads into a 42-dimensional acoustic feature vector for both audio and the downshifted ultrasound. Finally, each feature dimension is normalized by applying a standard mean and variance normalization algorithm.

4.2. Position features

Since sound travels at a finite speed, information about the direction of arrival (DOA) can be found in the time differences of arrival (TDOA) between different microphones in a node. The most simple way of measuring a TDOA is by a cross correlation, but this approach has a time resolution of a single sampling period. This problem can be resolved by using the so-called steered response power (SRP) algorithm [29]. SRP is based on a delay and sum beamformer, which is steered in multiple directions at once (ranging from -90° to $+90^\circ$ with a resolution of 2°) for one block of data and measures the retrieved energy in each direction. In this work, an enhancement on SRP is used, namely SRP phase transform (SRP-PHAT). PHAT basically pre-transforms the microphone frames to have an unity spectral density. This operation decorrelates the different microphone signals over time, making the directional energy peaks corresponding to a source narrower. The SRP-PHAT algorithm is described further in [29].

SRP-PHAT finally produces 91 points of the directional energy curve per node. These points were not directly used as features for the classification model. Since only a limited amount of training data was available it is desirable to keep the feature space low dimensional. Therefore, it was chosen to split-up the directional energy curve into two regions with broadside as the boundary. The energy in each region was integrated and the resulting two measures were combined into a single feature per node by taking the logarithmic ratio. The logarithm is taken to reshape the ratio intervals from $]0, 1]$ and $[1, \infty[$ to $]-\infty, 0]$ and $[0, \infty[$ to equalize the importance of both sides. Despite the resulting low resolution, using a combination of nodes allows to partition the living environment into several areas. Therefore, the position feature vectors used for the classification model are formed by concatenating the node specific position features. This differs from the acoustic features where only features from a single node are selected.

4.3. Gaussian Mixture Models

ADLs are classified by training a GMM with diagonal covariance per class. A sequence of feature vectors, possibly originating from multiple nodes and from both modalities, is assigned to the class that produces the maximal log-likelihood. In earlier work on similar problems [30], it was found that classification accuracy did not depend critically on the number of mixture components in the range from 5 to 20. For parsimony, five mixture components were used.

5. Experiments and results

Due to the limited amount of training data, a 10-fold cross-validation approach was used to evaluate the classification performance of the proposed system. The data was not first permuted before partitioning to leave the acoustic variation between training and validation partition as realistic as possible. In order to preserve the class balance in training- and validation set, each activity class was first partitioned into 10 folds followed by the combination of corresponding class specific folds.

5.1. Audio based ADL classification

The first conducted experiment in this work analyzes the audio based classification performance of the sensor network. In order to examine the additional value of the position information in terms of classification accuracy this experiment is carried out twice, i.e. once without and once including the position features. The corresponding confusion matrices are shown in Table 2 and Table 3 respectively. As one can see, promising results are obtained. The obtained average accuracy is $81.7\% \pm 14.6$ when only the acoustic features are taken into account. This value increases by 3.3% to an average accuracy of $85.0\% \pm 14.6$ when the position information is added as a feature. The following observations can be made by analyzing the corresponding confusion matrices more in detail:

1. The ADL *Cooking and eating*, *Vacuum cleaning*, and *Watching TV* have the best classification results. This is easy to comprehend because: 1) these activities are characterized by their own typical recurring characteristic acoustic information and 2) the energy of the acoustic sounds in these events is sufficiently high which results into higher SNRs.
2. The increase in accuracy by adding position information is due to the higher classification results obtained at *Reading* and *Walking around*. For these activities, the energy in the acoustic waves is low (e.g. page turn or a footstep) making the acoustic features unreliable.

Table 2: Confusion matrix of the obtained audio classification results without the position information.

True label:	Classified label:					
	Cooking and eating	Reading	Vacuum cleaning	Walking around	Watching TV	Working laptop
Cooking and eating	9	-	-	1	-	-
Reading	1	5	-	3	-	1
Vacuum cleaning	-	-	10	-	-	-
Walking around	2	-	-	7	-	1
Watching TV	-	-	-	-	10	-
Working laptop	-	1	-	1	-	8

5.2. Complementarity of audio and ultrasound

This experiment examines the complementarity of the audio and ultrasound signals. As discussed in section 2, only 1 ultrasound node was placed in the domestic environment. Therefore, in order to have a fair comparison between both modalities only the audio data from the corresponding audio node is used in this experiment. This differs from 5.1 where the node selection depends on the node's SNR.

Table 4 and 5 represent the obtained audio and ultrasound results. The average accuracy of the audio classification drops to $81.7\% \pm 12.3$, as expected. The average score of ultrasound is $61.7\% \pm 11.3$ which is lower than that when using audio but still very promising. By analyzing the results more in detail the following conclusions can be made:

Table 3: Confusion matrix of the obtained audio classification results when the position features are included.

True label:	Classified label:					
	Cooking and eating	Reading	Vacuum cleaning	Walking around	Watching TV	Working laptop
Cooking and eating	9	-	-	1	-	-
Reading	1	6	-	3	-	-
Vacuum cleaning	-	-	10	-	-	-
Walking around	1	1	-	8	-	-
Watching TV	-	-	-	-	10	-
Working laptop	-	-	-	2	-	8

1. Also in the ultrasound modality, the ADLs *Cooking and eating*, *Vacuum cleaning*, and *Watching TV* have the best accuracy. For these, the same explanation as in 5.1 is valid.
2. The classification accuracy of the activities *Reading* and *Working laptop* with the down-shifted ultrasound signals is inferior to the performance when using the audio data. Listening to the recording confirms that these activities are harder to recognize from the ultrasound recordings than from the audio recordings.
3. Ultrasound signals will be more attenuated than audio over a same distance since the attenuation of acoustic waves depends on the frequency [29]. This in combination with a sub-optimal ultrasound node (i.e. analog downshifting introduces a significant amount of noise) makes the ultrasound part of the sensor network less sensitive compared to audio and thereby also affecting the results.

Table 4: Confusion matrix of the obtained audio classification accuracies (only the acoustic and position features from the audio node corresponding to the ultrasound node position is used).

True label:	Classified label:					
	Cooking and eating	Reading	Vacuum cleaning	Walking around	Watching TV	Working laptop
Cooking and eating	9	-	-	1	-	-
Reading	1	5	-	3	-	1
Vacuum cleaning	-	-	10	-	-	-
Walking around	1	1	-	8	-	-
Watching TV	-	-	-	-	10	-
Working laptop	-	-	-	3	-	7

Table 5: Confusion matrix of the obtained ultrasound classification results (position features included).

True label:	Classified label:	Cooking and eating	Reading	Vacuum cleaning	Walking around	Watching TV	Working laptop
Cooking and eating		10	-	-	-	-	-
Reading		-	0	-	2	8	-
Vacuum cleaning		-	-	10	-	-	-
Walking around		-	-	-	6	4	-
Watching TV		-	-	-	1	9	-
Working laptop		-	-	-	3	5	2

Table 6 shows the confusion matrix of the classification results when audio and ultrasound are combined. The latter is done by summing the audio and ultrasound class posteriors together. The obtained average classification score is $80.0\% \pm 10.5$ which is slightly less accurate compared to the audio results from Table 4 but nevertheless still promising.

Table 6: Confusion matrix of the combination audio and ultrasound.

True label:	Classified label:	Cooking and eating	Reading	Vacuum cleaning	Walking around	Watching TV	Working laptop
Cooking and eating		10	-	-	-	-	-
Reading		-	4	1	3	1	1
Vacuum cleaning		-	-	10	-	-	-
Walking around		-	1	1	6	-	2
Watching TV		-	-	-	-	10	-
Working laptop		-	-	-	2	-	8

6. Discussion

The preliminary results in Section 5 indicate promising ADL classification results for both the audio and ultrasound modalities. Although the ultrasound based classification results were inferior to those obtained using audio data, one must be careful before drawing conclusions.

1. The SNR of the ultrasound signals was significantly lower than that of the audio signals. Further investigation is required to check which hardware improvements can be used to increase the SNR.
2. Although a simple combination of both the audio and ultrasound modalities resulted in a decrease of overall performance, other types of combination, such as a class-dependent weighted combination of both outcomes, might be more successful.

Therefore, future work will focus on the development of more sensitive and less-noisy ultrasound nodes and the optimal

integration of both modalities with respect to the classification performance of the WASN. Moreover, the added value of active observation will be investigated as well by extending the sensor network with ultrasound transmitters. This can lead to a better observation of dynamic ADLs (e.g. walking around) and thereby can also lead to an improved overall classification accuracy.

Furthermore, real-life data collection sessions over a longer time span in homes of elderly living alone are also planned in the near future. This allows the creation of larger acoustic datasets which will improve the estimation of acoustic models.

7. Conclusions

This work presents a distributed acoustic sensor network for the observation of activities of daily living from elderly on the basis of the corresponding audio and ultrasound data. The baseline system that is proposed was validated on realistic real-life data that was recorded in a domestic environment equipped with a prototype of the sensor network. The conducted classification experiments on the acquired data revealed promising preliminary classification accuracies, i.e. $85.0\% \pm 14.6$ and $61.7\% \pm 11.3$ for audio and ultrasound respectively. Combining both modalities by posterior summation did not yield an improvement over the audio-only modality. Other classifier combination methods will be studied in the near future. Despite these promising preliminary results, further work on a larger scaled dataset collected with multiple and more sensitive ultrasound nodes is required to increase the significance of the obtained results.

8. Acknowledgements

This work was performed in the context of following projects: ALADIN (IWT-SBO project contract 100049) and IWT doctoral scholarships (contract 111433 and 121565).

9. References

- [1] N. Consortium. (2010) The business of aging: Commercial challenges and opportunities in ambient assisted living. [Online]. Available: <http://www.netcarity.org>
- [2] F. planbureau: Economische analyses en vooruitzichten. (2011) Bevolkingsvooruitzichten 2010-2060. [Online]. Available: <http://www.plan.be>
- [3] D. J. Patterson, D. Fox, H. Kautz, and M. Philipose, "Fine-grained activity recognition by aggregating abstract object usage," in *Proceedings of the Ninth IEEE International Symposium on Wearable Computers*, ser. ISWC '05. Washington, DC, USA: IEEE Computer Society, 2005, pp. 44–51. [Online]. Available: <http://dx.doi.org/10.1109/ISWC.2005.22>
- [4] J. Wu, A. Osuntogun, T. Choudhury, M. Philipose, and J. M. Rehg, "A scalable approach to activity recognition based on object use," in *In Proceedings of the International Conference on Computer Vision (ICCV), Rio de, 2007*.
- [5] M. Stikic, T. Huynh, K. Van Laerhoven, and B. Schiele, "Adl recognition based on the combination of rfid and accelerometer sensing," in *Proceedings of the 2nd International Conference on Pervasive Computing Technologies for Healthcare (Pervasive Health 2008)*, IEEE Xplore. Tampere, Finland: IEEE Xplore, January 2008, pp. 258–263.
- [6] X. Long, B. Yin, and R. M. Aarts, "Single-accelerometer-based daily physical activity classification," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, Sep. 2009, pp. 6107–6110. [Online]. Available: <http://dx.doi.org/10.1109/iembs.2009.5334925>

- [7] O. Amft, H. Junker, and G. Troster, "Detection of eating and drinking arm gestures using inertial body-worn sensors," in *Proceedings of the Ninth IEEE International Symposium on Wearable Computers*, ser. ISWC '05. Washington, DC, USA: IEEE Computer Society, 2005, pp. 160–163. [Online]. Available: <http://dx.doi.org/10.1109/ISWC.2005.17>
- [8] W.-Y. C. Amit Purwar, Young-Dong Lee, "Triaxial mems accelerometer for activity monitoring of elderly person," *Sensor Letters*, vol. 6, no. 6, pp. 1054–1058, 2008.
- [9] A. Jin, B. Yin, G. Morren, H. Duric, and R. M. Aarts, "Performance evaluation of a tri-axial accelerometry-based respiration monitoring for ambient assisted living," *Conf Proc IEEE Eng Med Biol Soc*, vol. 1, pp. 5677–80, 2009. [Online]. Available: <http://www.biomedsearch.com/nih/Performance-evaluation-tri-axial-accelerometry/19964139.html>
- [10] C. Zhu and W. Sheng, "Multi-sensor fusion for human daily activity recognition in robot-assisted living," in *HRI, 2009*, pp. 303–304.
- [11] A. K. Bourke and G. M. Lyons, "A threshold-based fall-detection algorithm using a bi-axial gyroscope sensor," *Medical Engineering and Physics*, vol. 30, no. 1, pp. 84–90, 2008. [Online]. Available: <http://dx.doi.org/10.1016/j.medengphy.2006.12.001>
- [12] P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea, "Machine recognition of human activities: A survey," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 18, no. 11, pp. 1473–1488, nov 2008. [Online]. Available: <http://dx.doi.org/10.1109/tcsvt.2008.2005594>
- [13] A. Fleury, N. Noury, and M. Vacher, "Supervised classification of activities of daily living in health smart homes using svm," in *31th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC09)*, Minnesota, USA, 2009, pp. 6099–6102.
- [14] P. Karsmakers, T. Croonenborghs, M. Mercuri, D. Schreurs, and P. Leroux, "Automatic in-door fall detection based on microwave radar measurements," in *Proceedings of the 9th European Radar Conference*, B. Arbesser-Rastburg, Ed. ESA-ESTEC, TEC-EE, Oct. 2012, pp. 202–205. [Online]. Available: <https://lirias.kuleuven.be/handle/123456789/353333>
- [15] F. G.-B. M. Berenguer, M. Giordani and N. Noury, "Automatic detection of activities of daily living from detecting and classifying electrical events on the residential power line," in *10th IEEE Int. Conf. e-Health Netw., Appl. Serv. HealthCom*, ser. -, 2008, pp. 29–32.
- [16] S. Tsukamoto, H. Hoshino, and T. Tamura, "Study on indoor activity monitoring by using electric field sensor," *Gerontechnology*, vol. 7, no. 2, 2008. [Online]. Available: <http://www.gerontechnology.info/index.php/journal/article/view/gt.2008.07.02.163.00>
- [17] D. Istrate, M. Vacher, and J.-F. Sernignat, "Embedded implementation of distress situation identification through sound analysis," in *the 5th ICICTH International Conference on Information Communication Technologies in Health*, Greece, Jul. 5-7 2007, pp. 226–231. [Online]. Available: "http://www.ineag.gr/docs/Scientific_Programme_of_5th_ICictH_Samos_2007.pdf"
- [18] M. Sehili, B. Lecouteux, M. Vacher, F. Portet, D. Istrate, B. Dorizzi, and J. Boudy, "Sound Environment Analysis in Smart Home," in *Ambient Intelligence*, ser. Lecture Notes in Computer Science, F. Paternò, B. de Ruyter, P. Markopoulos, C. Santoro, E. van Loenen, and K. Luyten, Eds., vol. 7683. Pisa, Italy: Springer, nov 2012, pp. 208–223.
- [19] A. Temko, C. Nadeu, D. Macho, R. Malkin, C. Zieger, and M. Omologo, "Acoustic event detection and classification," in *Computers in the Human Interaction Loop*, 2009, pp. 61–73.
- [20] H. Lozano, I. Hernáez, A. Picón, J. Camarena, and E. Navas, "Audio classification techniques in home environments for elderly/dependant people," in *Proceedings of the 12th international conference on Computers helping people with special needs: Part I*, ser. ICCHP'10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 320–323. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1886667.1886725>
- [21] H. Kun-Yi, H. a Chi-Chun, T.Ming-shih, and Y. G.-L. C.Yu-Hsien, "Activity recognition by detecting acoustic events for eldercare," in *6th World Congress of Biomechanics*, ser. WCB 2010, 2010, pp. 1522–1525.
- [22] A. Bertrand, "Applications and trends in wireless acoustic sensor networks: a signal processing perspective," in *Proc. IEEE Symposium on Communications and Vehicular Technology (SCVT)*, November 2011. [Online]. Available: ftp://ftp.esat.kuleuven.be/pub/sista/abertran/papers_website/11-197.html
- [23] A. Brutti, M. Omologo, and P. Svaizer, "Oriented global coherence field for the estimation of the head orientation in smart rooms equipped with distributed microphone arrays," in *INTER-SPEECH'05*, 2005, pp. 2337–2340.
- [24] I. K. R. Lienhart and S. Wehr, "Universal synchronization scheme for distributed audio-video capture on heterogeneous computing platforms," in *eleventh ACM international conference on Multimedia*, ser. -, 2003, pp. pp. 263–266.
- [25] H. Beigi, *Fundamentals of Speaker Recognition*. Springer, 2011.
- [26] X. Zhuang, X. Zhou, T. Huang, and M. Hasegawa-Johnson, "Feature analysis and selection for acoustic event detection," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, 2008, pp. 17–20.
- [27] S. Ntalampiras, I. Potamitis, and N. Fakotakis, "Automatic recognition of urban soundscapes," in *New Directions in Intelligent Interactive Multimedia*, 2008, pp. 147–153.
- [28] X. Zhou, D. Garcia-Romero, R. Duraiswami, C. Espy-Wilson, and S. Shamma, "Linear versus mel frequency cepstral coefficients for speaker recognition," in *Automatic Speech Recognition and Understanding (ASRU), 2011 IEEE Workshop on*, 2011, pp. 559–564.
- [29] I. J. Tashev, *Sound Capture and Processing: Practical Approaches*. Wiley Publishing, 2009.
- [30] L. Vuegen, B. Van Den Broeck, P. Karsmakers, J. Gemmeke, B. Vanrumste, and H. Van hamme, "An mfcc-gmm approach for event detection and classification," in *AASP Challenge: Detection and Classification of Acoustic Scenes and Events*, 2013.