Sensors & Transducers, Vol. 212, Issue 5, May 2017, pp. 31-38

www.sensorsportal.com

Sensors & Transducers

Published by IFSA Publishing, S. L., 2017
http://www.sensorsportal.com

Home Sound: A GPU-based Platform for Massive Data Acquisition and Processing for Acoustic Ambient Assisted Living Applications for Behavior Monitoring

[1] Joan NAVARRO, [2] Rosa Mi ALSINA-PAGÈS and [2, *] Marcos HERVÁS

[1] GRITS - Grupo de Recerca en Internet Tecnologies & Storage (La Salle - Universitat Ramon Llull), Quatre Camins 30 (Barcelona), 08022, Spain

[2] GTM - Grup de recerca en Tecnologies Mèdia (La Salle - Universitat Ramon Llull), Quatre Camins 30 (Barcelona), 08022, Spain

[2] Tel.: +34-932902445, fax: +34-932902385

[*] E-mail: mhervas@salleurl.edu

Received: 11 April 2017 /Accepted: 29 May 2017 /Published: 31 May 2017

Abstract: Human life expectancy has grown over the last century, which has driven governments to increase the efforts on caring about the eldest population. Therefore, modern trends take advantage of latest advances in technology to remotely monitor those people with special needs at their home, increasing their life quality and with less impact on their social lives. This paper presents an acoustic event detection platform for assisted living that tracks patients' status by automatically identifying and analyzing the acoustic events happening in a house. Specifically, we have taken benefit of a Jetson TK1, with its NVIDIA Graphical Processing Unit, to process the acoustic data and identify a closed number of events in order to inform the care system. This is a proof of concept conducted with data of only one acoustic sensor, but we plan in the future to deploy a sensor network in several places in the house.

Keywords: Ambient assisted living, Sensor network, Machine hearing, Acoustic feature extraction, Machine learning, Graphics processor unit.

1. Introduction

Human life expectancy is increasing in the modern society [1]. Our society has to face new challenges in terms of health care because the number of patients to attend is increasing according to [2-3] the people ageing who need support [4]. Nowadays, public and private health services try to avoid long term hospitalizations and, instead, foster the elderly to remain at home for two reasons: on the one hand, it is better for their health to keep them – while not suffering from severe deterioration – in their own environment and, on the other hand, it is much cheaper for health services and care systems. However, nowadays there is still a quality gap between the service provided at medical facilities and the service provided at patients' home.

Technology is a powerful tool that can contribute to address this problem by enabling medical staff to monitor and attend patients while they are at home. Ambient Assisted Living (AAL) [5] can reduce the personnel costs in health assistance. AAL consists of monitoring the preferred living environment of the patients with intelligent devices that can track their status and improve their life quality, as well as obtain information about their behavior, which in the future

can lead the doctors to conclusions that with hospital visits could not identify. Some incipient illnesses show symptoms occasionally in time, so a short medical visit is not able to identify preliminary signals. Acoustic Ambient Assisted Living, by means of acoustic event detection, and focusing on behavioral monitoring can help early diagnoses of severe diseases.

To address this hot research topic, several engineering projects have been proposed to discuss the feasibility of deploying smart robots at the home of elderly not only to cover routine tasks, but also to remind them to have their medication or interact with them through serious games [6]. One of the main challenges that these proposals open is the huge amount of data that these robots have to collect in order to provide a meaningful response for patients. Typically, these robots have limited computing capabilities and, thus, are able to process data from a reduced number of sensors.

This paper explains the proof of concept of a software and a chosen hardware platform designed to recognize a set of the predefined events from the environmental sound in a house [7]. This information can be later used to infer the in-home context and detect some situations of risk. To process data from several sources (e.g., microphones) and conduct the computations associated to audio event identification in parallel, the system implements a recognition scheme using a NVIDIA Jetson TK1 [8] Graphical Processing Unit (GPU). This platform can reach to several decisions depending on the situation and home, and the final conclusion can be activating some kind of alarm or just track the patient's behaviour for health purposes. Overall, the purpose of this work is to present an approach to the implementation of an acoustic event recognition platform based on a GPU and the obtained results when classifying a limited corpus of events.

The reminder of this paper is organized as follows. Section 2 reviews the related work on environmental sound recognition; it is specially focused on ambient assisted living environments. Section 3 elaborates on the technical details of the proposed algorithm to solve the problem, which corresponds to a basic implementation. Section 4 gives details about the selected platform and its convenient features to process audio data. Section 5 describes the algorithm used to classify the events and shows the obtained results when running on the chosen platform. Finally, Section 6 details the conclusions and future work of this project.

2. Related Work

There are several approaches in the literature that aim to extract features from the sound. From these features, it is possible to create a corpus of a close universe of different sounds and train a machine learning system to classify the source of the sound. Therefore, environmental sound recognition has emerged as a hot research topic today, which has led to some interesting applications [9]; from animal recognition [10] to surveillance [surveillance], including ambient assisted living use cases.

Interest in detecting in-home sounds started from the beginning of this technology in 2005. Chen, *et al.* [11] were monitoring the bathroom activity using only the sound information. Afterwards, with research not detailed in this work, robust environment sound recognition motors were designed in 2008 [12]. One of the most challenging problems to be solved in this field, is to take into account the varying acoustic background, the noise sources. In this regard, the project SonicSentinel [13] uses noise-robust model-based algorithms to evaluate the noise sources. Evolving this technology, Valero, *et al.* [14] succeeded on classifying audio scenes. Additionally, several works can be found about audio analysis in a smart home to help doctors on the early diagnose of dementia diseases for the elder [15]. Also, it is worth mentioning that conditional random fields have been used to build an event detection framework in a real-world environment of eight households [16], which led the system to be sometimes unreliable.

From the applications point of view, one of the most popular use-cases nowadays of audio event recognition is its use in the smart home [17], especially when conceiving systems to meet the needs of the elderly people. The constraints around the design of a smart home for health care [5] based on audio event classification are as follows:

1) Degree of dependency of the disabled person,

2) Quality of life to be improved by means of automatizing the processes,

3) Distress situations recognition and the activation of the preassigned protocols, including reducing the false alarm situations [18].

Even though there are several solutions in the literature [19] that consider these three constraints, the primary goal of the platform presented in this paper is to accurately address the third one. Additionally, our proposal aims to meet the needs of ambient assisted living, which are the following [20]:

1) Increasing the comfort of living at home,

2) Increasing the safety, through detecting dangerous events,

3) Supporting health care by professionals, through detecting emergencies and monitoring vital signs.

3. System Description

When designing and deploying an Acoustic Wireless Sensor Network, the main parameter to be considered is power consumption. Generally, the power consumption of a node in a wireless network strongly depends on its assigned duties (e.g., data acquisition, data storage, data computing, data forwarding). Therefore, system architects have to carefully select which devices conduct every task. In this regard, with the growth of the number of mobile

user equipment (UE) devices and the advent of the Internet of Things, latest advances on the distributed systems field have proposed several approaches and reference models to offload the duties of each device by means of the Mobile Edge Computing (MEC) paradigm [21]. Indeed, MEC consists of deferring the power consuming activities to specialized and dedicated devices close to the wireless sensor network. Therefore, it can be best seen as a particular case of cloud computing where the storage and computation infrastructure are physically close to where data are generated, which brings appealing advantages such as data security, reduced communication delay, and energy efficiency [21]. This section enumerates the latest contributions in the MEC field, justifies the selected alternative for the Ambient Assisted Living use case proposed in this paper, and details its deployment.

So far, when designing a MEC architecture four main approaches have emerged whose details are further elaborated in [21]. These alternatives are summarized in what follows:

1. Small Cell Cloud (SCC): It consists of extending the capabilities of the UE by include a Small Cell Manager committed to forward the UE requests to a storage and computation cluster.

2. Mobile Micro Clouds (MMC): It consists of deploying a network of device managers each one committed to forward the UE requests from a small set of UEs. These device managers are interconnected between themselves and also connected to a storage and computation cluster. This approach minimized the load of the device manager, which makes it suitable for scenarios with a high number of UE devices.

3. Fast Moving Personal Cloud (FMPC): It consists of using Software Defined Networks and Network Function Virtualization to build a dynamic and adaptable set of forwarding devices to link the UE with the storage and computation cluster. Therefore, it uses a SDN enabled transport network typically residing in a cloud. This is a feasible approach for those application that change their traffic patterns frequently and are not very sensible to delay.

4. Follow Me Cloud (FMC): It consists of embedding the UE devices inside the cloud storage and computing cluster. In this way, the perception that the cloud following the roaming UE is given.

For the sake of our Ambient Assisted Living proposal in which an acoustic wireless sensor network based on microphones (i.e., UE) is deployed inside a home environment, the aforementioned four alternatives have been considered. FMPC and FMC approaches are designed for environments where UE are moving, which is not the case of AAL where the microphones are permanently installed in the same place.

MMC might be a feasible option if the home environment was big enough to deploy a high number of microphones (i.e., hundreds). Considering that there will be two microphones per room, the overall number of UE per home environment is still below the threshold imposed by MMC.

Therefore, we have selected the SCC option that is also convenient taking into account the low latency (i.e., AAL is committed to operate in real-time) and budget constraints requirements of the AAL paradigm. As a result, the proposed system diagram to monitor audio events in Ambient Assisted Living environments is shown in Fig. 1.

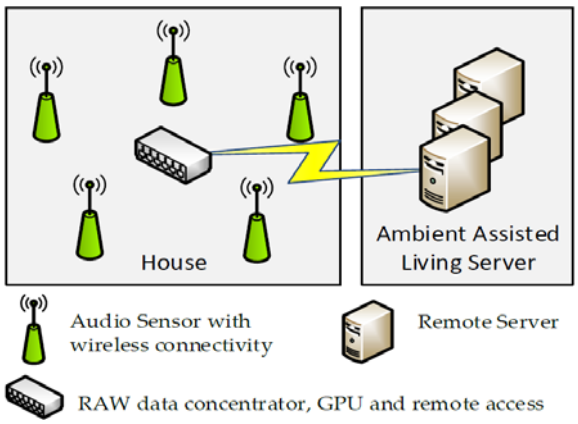


Fig. 1. Block diagram of the network elements of this system.

As far as the proof of concept herein presented is concerned, the system relies on a network of microphones consistently deployed around the house (see Fig. 2). The microphones are installed in such a way that they provide the maximum entropy of a given event (i.e., it is not necessary to analyze together different audio sources).

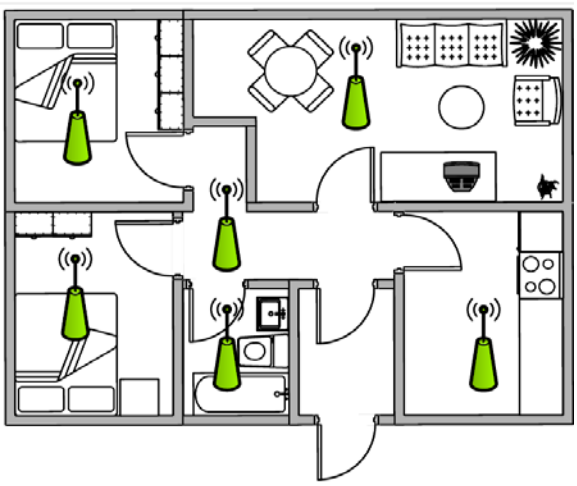


Fig. 2. Example of the proposed audio sensors network deployed in a house.

The microphones used in this application to sense the environmental sound should present a good trade-off between the frequency response and cost, for this reason tests are being conducted with the electret condenser microphone CMA-4544PF-W [22] of the manufacturer CUI inc. with a very low price.

In this way, each microphone transmits sounds to this device that acts as a concentrator – the core element of our proposal. As a matter of fact, this concentrator

1) Collects all the audio sounds of the house,

2) Processes them to extract their features,

3) Infers the source of the audio event,

4) Sends this information to a remote server that monitors the needs of the people living in the house.

The concentrator platform used in this work is the NVIDIA Jetson TK1 developer kit. This platform is based on the Tegra K1 SoC, which is composed of

1) NVIDIA Kepler GPU with 192 CUDA cores,

2) Quad core ARM cortex-A15 CPU.

The Tegra family is the proposal of the NVIDIA manufacturer for mobile processors in which you need GPU-accelerated performance with low power consumption.

This GPU is able to process up to 192 threads in parallel. Kepler architecture offers an improvement of performance up to 3 times more than the previous version, Fermi, [23]. This level of concurrency allows us to process audio events of several sources in real-time.

Therefore, to exploit the parallel capabilities of the concentrator, it opens a thread to process each audio source and infer the event that generated every sound.

4. Machine Learning

Endowing machines with the ability of hearing the acoustic environment to detect and recognize an event as humans do, is known as machine hearing. The algorithm used in this work is based on

1) Feature extraction using mel-frequency cepstral coefficients (MFCC) [24];

2) Pattern recognition using the k-Nearest Neighbors classifier (KNN) [25], see Fig. 3.

Fig. 3. Block diagram of a Hearing Machine algorithm.

4.1. Feature Extraction

Feature extraction aims to obtain a representation of audio events in which the dimensionality of this parametrization is much lower than the original samples [26]. This parametrization will be the input data of the classifier. The parametrization used in this work, MFCC [24], uses an approach based on perceptual-based frequency using the Mel scale [27] as shown in Fig. 4.

The incoming audio stream is divided into blocks of 30 ms with a sliding window. These frames are transformed into frequency domain using the DFT to measure the power of different bands of the spectrum. The power measures are conducted with a bank of 48 filters using the Mel scale (see Fig. 5).

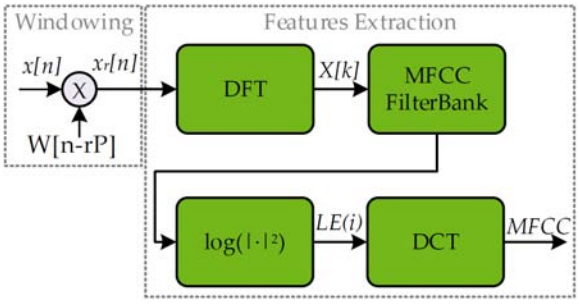


Fig. 4. Block diagram of the feature extraction based on the Mel coefficients used in this work.

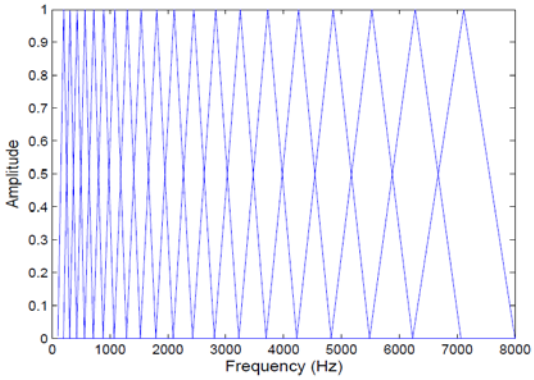
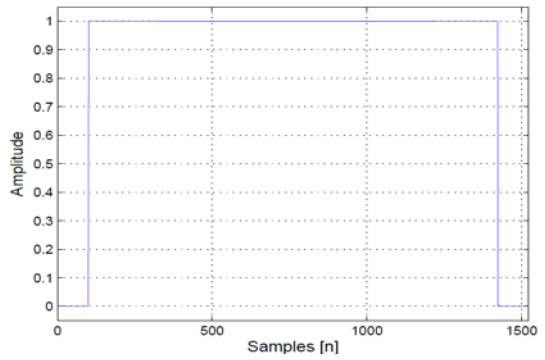


Fig. 5. Example of a Mel scale with a filter-bank of 20.

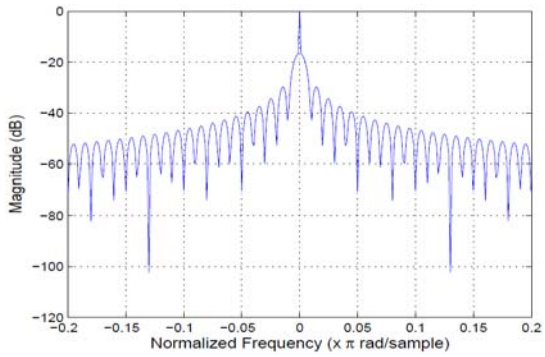
The MFCC coefficients are obtained from the Discrete Cosine Transform (DCT) of the logarithm of these 48 values. The higher order coefficients of the DCT are discarded to obtain a reduced dimensionality characterization of the sound event, this compression can be done because the main information is in the low frequency components of the signal's spectral envelop. The final number of MFCC coefficients is 13.

Window lengths between 10 and 50 ms are usually used to detect transient audio events [28]. A Hamming windowing is also applied to this frame of samples to improve the frequency resolution in the Discrete Fourier Transform (DFT) – as we can see comparing the differences between square and Hamming windows in Fig. 6. This sliding block has an overlap of 50 % of samples to compensate the power reduction of the data blocks due to the laterals of the Hamming window, see Fig. 6.

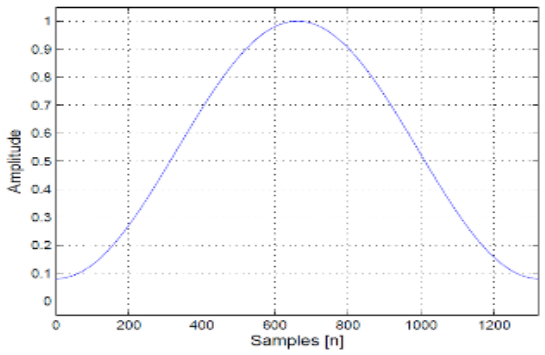
The Mel scale is a perceptual scale which aims to emulate the behaviour of the human hearing. As we can observe in Fig. 5, Mel scale is a bank of triangular filters.



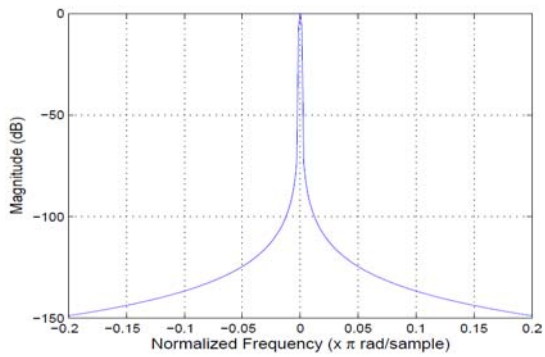
(a)



(b)



(c)



(d)

Fig. 6. Comparison between a squared and Hamming windows in time and frequency domain: (a) is a squared window, (b) is the spectrum of the squared window, (c) is a Hamming window and (d) is the spectrum of the Hamming window.

4.2. Automatic Audio Classification

Machine learning algorithms are widely used in the literature of speech technologies to automatically classify audio samples. In fact, most of the audio recognition systems settle the use of the MFCC coefficients as baseline in terms of feature extraction [26]. Then, when the signal is processed and the features are already extracted, a k-Nearest Neighbors (kNN) [25] system can be run [29].

Hence, we have followed this approach and trained a kNN classifier as follows. We have built a training data set composed by 2850 audio samples belonging to 14 in-home events lasting a total number of 20 hours. We have split every sample in several sub samples as detailed in the previous section, and for every sub sample we have computed the MFCC coefficients. This results in a vector of 13 components (each one corresponding to its associated MFCC) for every sound sub sample. As a result, a sound sample is characterized with a set of 13-component vectors.

For the sake of this paper we have implemented two classification strategies: raw-kNN and SVM.

The raw-kNN attempts to obtain a lower bound of up to what extent it is feasible to classify the training data set. In this regard, a simple 13-dimensions k-NN has been built and a grid search to come out with the best *k* parameter has been run. Actually, this classification strategy can be best seen as a worst case scenario in which the entropy provided by previous and subsequent subsamples is deliberately neglected. Thus, in order to classify a given subsample we only consider the closest *k* samples to it. As far as the computation cost is concerned, it is worth mentioning that the kNN has a linear cost (i.e., the input subsample has to be compared against all the vectors of the dataset. However, this overhead has been greatly alleviated by implementing a map-reduce inspired strategy to compare different subsets of the dataset in parallel. Specifically, we have assigned a thread to a segment of the dataset that conforms the kNN and, next, each thread shares its *k* nearest neighbors. Then, the output of a group of threads is collected by another thread in charge of selecting again the *k* nearest neighbors. This process is done recursively until the wining *k* neighbors have survived the whole recursively map and reduce process.

The second classification strategy has been designed to improve the results of the kNN classifier by considering the information of the subsamples belonging to the same sample. As the number of vectors that characterize a given sound depends on the length of the training sound, there is an inherent class imbalance in the formulation of this problem, which limits the classifier accuracy (i.e., shorter sounds of the same sound type would probably be misclassified). Therefore, to address this issue, we have built a bag of words with all the vectors belonging to the same sample using the k-means algorithm [30]. The resulting vector has a fixed length of K components. This gives an idea of how many portions of the training sound set belong to each centroid of the k-

means, which at the same time removes the temporal dimension of the sound event. Next, we normalize all these resulting vectors to make the suitable for a fair comparison. With this fixed size set of normalized vectors, we finally train a Support Vector Machine (SVM) that is running on the concentrator platform and uses a one-against-all strategy to deal with this multiclass problem. That is, a SVM has been built for every class in which the class of interest has been labelled as positive and all other classes as negative. Then the output of each SVM will be considered as a confidence factor. When this confidence factor is above a heuristic threshold, the output of the SVM will be considered as positive. To come out with this heuristic threshold we have held out a 10 % of the dataset and conducted a 4-fold cross-validated grid search. Analogously, we have found the best offset parameter for each SVM using the same strategy.

Finally, when our system is in exploitation mode, the concentrator platform extracts the audio sub samples and builds the fixed size vector accordingly. Then, this vector is delivered to all the SVMs that had been previously tested to predict the event. In order to obtain a positive outcome, a single SVM has to provide an output above the aforementioned threshold. If more than one SVM provides an output above the threshold, no class is assigned to that sample.

5. Results

With the dataset and the techniques described in the previous section we have conducted our experimentation to detect the following events: someone falling down, slice, screaming, rain, printer, people talking, frying food, filling water, door knocking, dog bark, car horn, glass breaking, baby crying, water boiling. We have used 60 % of instances to train the classifiers and the other 40 % to test them. In order to obtain statistically significant results we have run the classification in 1000 runs, performed a 10-fold cross validation, and averaged the output.

The obtained confusion matrix for the kNN classifier is shown in Table 1 with an overall accuracy of 50.24 %. In this confusion matrix we can see how often the kNN misclassifies a given class and, thus, assigns a wrong event to an audio sample. It is shown that in general, the best results for each sample are obtained when testing the sound event against itself. Also, it depicts the skill of the classifier on distinguishing one audio event from the others. The optimal value of this confusion matrix should be an Identity Matrix with the value 100 on its diagonal.

We can see that although some events are identified with a reasonable degree of accuracy (e.g., screaming), some others (e.g., dog barking) cannot be classified properly.

Table 1. Confusion Matrix of the kNN classifier. Events are ordered from left to right as follows: falling down, slice, screaming, rain, printer, people talking, frying food, filling water, door knocking, dog bark, car horn, glass breaking, baby crying, water boiling.

	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
0	58.54	0.00	2.44	0.00	4.88	2.44	7.32	4.88	1.22	2.44	0.00	10.98	3.66	1.22
1	0.00	60.26	0.00	3.85	0.00	11.54	3.85	0.00	6.41	2.56	7.69	0.00	1.28	2.56
2	0.00	0.00	89.33	0.00	0.00	0.00	0.00	3.33	0.00	4.00	1.33	0.00	0.67	1.33
3	0.00	1.54	0.00	76.92	0.77	2.31	3.08	2.31	0.00	4.62	6.92	0.00	0.00	1.54
4	9.52	1.19	0.00	2.38	55.95	4.76	3.57	0.00	2.38	7.14	0.00	8.33	1.19	3.57
5	2.91	8.74	1.94	4.85	3.88	18.45	1.94	0.00	6.80	18.45	7.77	1.94	3.88	18.45
6	7.89	6.58	1.32	3.95	3.95	3.95	15.79	15.79	1.32	15.79	1.32	2.63	13.16	6.58
7	1.08	0.00	3.23	2.15	0.00	0.00	5.38	60.22	0.00	9.68	10.75	0.00	4.30	3.23
8	1.41	4.23	0.00	0.00	1.41	9.86	0.00	0.00	76.06	1.41	0.00	5.63	0.00	0.00
9	2.65	0.88	4.42	7.96	1.77	10.62	6.19	4.42	3.54	14.16	25.66	3.54	3.54	10.62
10	0.00	5.33	0.00	10.67	0.00	8.00	0.00	13.33	0.00	32.00	21.33	0.00	1.33	8.00
11	8.70	1.74	0.87	0.00	6.96	1.74	1.74	0.00	3.48	5.22	0.87	68.70	0.00	0.00
12	3.08	0.00	10.77	0.00	3.08	4.62	10.77	9.23	0.00	6.15	1.54	1.54	30.77	18.46
13	1.63	3.25	1.63	1.63	0.81	11.38	1.63	4.88	0.00	7.32	4.07	0.00	4.88	56.91

For instance, on row 6 in Table 1, door knocking, people talking and frying food have similar MFCC vector patterns and, thus, the SVM features a low accuracy in these specific situations Such a poor performance can be explained because (1) a single subsample is only considered to decide to which class it belongs to (i.e., no information from previous nor subsequent subsamples is considered and (2) the MFCCs associated to some subsamples of these events

are pretty similar to other subsamples from other events.

On the contrary, Table 2 shows the confusion matrix obtained when using the bag of words approach and the aforementioned SVM classifier. We can see that this classification strategy reaches an overall accuracy of 72.88 %, which is much better than the kNN. Also, we can see that with this strategy, the classification accuracy is more consistent for all the events (i.e., the worst value 56.23 %). However, despite considering information from previous and subsequent subsamples thanks to the bag of words approach, this approach still gets confused on some sound events. To address this concern, we plan to (1) complement the training vector set with other features in addition to MFCCs, and (2) use a more sophisticated classifier such as a deep net.

6. Conclusions

Preliminary results of our paper encourage us to keep on working on the analysis of the events happening in the house. We will work with the feature extraction improvement with other methods, as well as we will test more machine learning algorithms to increase the accuracy of the system with just one acoustic measurement.

Next steps after this proof of concept using the Jetson TK1 are the expansion of the platform, by means of using a wider sensor network, where several autonomous acoustic sensors sending data to the GPU to be processed. In this stage, an important part of the work will be focused on the optimization of the acoustic event detection algorithm to take advantage of the parallelization of the GPU unit.

Acknowledgements

The authors would like to thank the Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement (Generalitat de Catalunya) under grant refs. 2014-SGR-0590 and ref. 2014-SGR-589.

Table 2. Confusion Matrix of the SVM classifier. Events are ordered from left to right as follows: falling down, slice, screaming, rain, printer, people talking, frying food, filling water, door knocking, dog bark, car horn, glass breaking, baby crying, water boiling.

	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
0	89.10	0.11	1.78	0.35	0.00	0.12	0.12	1.60	2.39	0.00	1.26	0.80	1.48	0.89
1	1.53	85.48	2.18	0.17	0.79	2.71	0.00	0.41	0.00	1.31	3.05	0.47	1.35	0.55
2	2.16	5.18	63.22	0.04	2.89	0.00	5.21	4.83	3.37	1.16	4.98	5.65	0.00	1.30
3	0.27	1.29	0.00	87.14	0.59	2.15	0.00	1.75	0.83	3.36	0.20	0.81	0.62	0.98
4	0.03	0.87	16.37	0.75	69.60	0.00	0.79	0.87	3.23	0.00	0.06	2.02	1.45	3.97
5	0.00	1.44	1.67	1.72	4.44	85.62	0.00	1.18	0.33	0.01	0.92	1.69	0.00	0.99
6	3.17	0.81	1.32	1.08	2.92	3.79	71.05	6.28	3.55	5.04	0.00	0.14	0.00	0.85
7	0.00	0.00	2.10	0.00	0.14	0.86	3.80	89.46	0.20	0.71	0.98	0.38	0.04	1.33
8	0.84	1.91	1.75	0.00	0.14	6.71	2.05	3.31	65.36	15.61	0.00	0.06	1.34	0.92
9	1.08	3.82	4.56	12.18	0.00	0.00	5.45	4.52	3.67	56.78	1.34	3.67	2.93	0.00
10	5.27	5.36	5.70	2.20	6.09	0.80	0.00	1.75	4.15	5.19	56.23	1.63	2.98	2.65
11	0.00	12.15	1.53	0.00	0.24	0.13	2.09	0.22	2.20	0.28	1.42	78.14	0.89	0.70
12	10.13	0.71	1.19	2.21	1.97	1.55	0.43	1.09	1.04	11.57	0.00	1.33	66.78	0.00
13	0.00	0.92	7.83	0.00	3.15	5.46	4.08	5.12	4.93	0.00	6.90	2.15	3.10	56.36

References

[1]. R. Suzman, J. Beard, Global health and aging – Living longer, *National Institute on Aging*, 2015.

[2]. Karp F. (Ed.), Growing Older in America: The Health and Retirement Study, *U.S. Department of Health and Human Services*, 2007.

[3]. S. Chatterji, P. Kowal, C. Mathers, N. Naidoo, E. Verdes, J. P. Smith, R. Suzman, The health of aging populations in China and India, *Health Affairs*, Vol. 27, Issue 4, 2008, pp. 1052-1063.

[4]. G. Lafortune, G. Balestat, Trends in Severe Disability Among Elderly People, in Assessing the Evidence in 12 OECD Countries and the Future Implications. OECD Health Working Papers 26, *Organization for Economic Cooperation and Development*, Paris, 2007.

[5]. M. Vacher, F. Portet, Challenges in the processing of audio channels for ambient assisted living, in *Proceedings of the IEEE 12th International Conference on E-health Networking Applications and Services (Healthcom)*, Vol. 12, 2010, pp. 330-337.

[6]. Y. S. Morsi, A. Shukla, Optimizing Assistive Technologies for Aging Populations, *IGI Global*, 2015.

[7]. M. Hervás, R. M. Alsina-Pagès, J. Navarro, homeSound: a High Performance Platform for

Massive Data Acquisition and Processing in Ambient Assisted Living Environments, in *Proceedings of the 6th International Conference on Sensor Networks (SENSORNETS'17)*, Porto, Portugal, 2017, pp. 182-187.
[8]. JETSON TK1. Unlock the power of the GPU for embedded systems applications (http://www.nvidia.com/object/jetson-tk1-embedded-dev-kit.html).
[9]. S. Chachada, J. Kuo, Environmental sound recognition: a survey, in *APSIPA Transactions on Signal and Information Processing*, Vol. 3, 2014, pp. 14-20.
[10]. Stowell D., Wood M., Stylianou Y., Glotin H., Bird detection in audio: a survey and a challenge, in *Proceedings of the IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*, 2016, pp. 1-6.
[11]. J. Chen, A. H. Kam, J. Zhang, N. Liu, L. Shue, Bathroom activity monitoring based on sound, in *Proceedings of the International Conference on Pervasive Computing*, 2005, pp. 47-61.
[12]. J. C. Wang, H. P. Lee, J. F. Wang, C. B. Lin, Robust environmental sound recognition for home automation, *IEEE Transactions on Automation Science and Engineering*, Vol. 5, Issue 1, 2008, pp. 25-31.
[13]. D. Hollosi, S. Goetze, J. Appell, F. Wallhoff, Acoustic Applications and Technologies for Ambient Assisted Living Scenarios, in *Proceedings of the Ambient Assisted Living Forum*, 2011, pp. 337-342.
[14]. X. Valero, F. Alías, Classification of audio scenes using narrow-band autocorrelation features, in *Proceedings of the 20th European Signal Processing Conference*, Bucharest, Romani, August 2012, pp. 2012-2019.
[15]. P. Guyot, J. Pinquier, X. Valero, F. Alías, Two-step detection of water sound events for the diagnostic and monitoring of dementia, in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, San Jose, California, USA, 2013, pp. 1-6.
[16]. D. Matern, A. Condurache, A. Mertins, Adaptive and automated ambiance surveillance and event detection for ambient assisted living, in *Proceedings of the 35th Annual International Conference of the IEEE EMBS*, Osaka, Japan, July 2013, pp. 3-7.
[17]. M. Chan, D. Estève, C. Escriba, E. Campo, A review of Smart homes – present state and future challenges, *Computer Methods and Programs in Biomedicine*, Vol. 91, Issue 1, 2008, pp. 55-81.
[18]. S. Goetze, J. Schroder, S. Gerlach, D. Hollosi, J. Appell, Acoustic Monitoring and Localization for Social Care, *Journal of Computing Science and Engineering*, Vol. 6, Issue 1, 2012, pp. 40-50.
[19]. M. Vacher, F. Portet, A. Fleury, N. Noury, Development of audio sensing technology for ambient assisted living: Applications and challenges, *Digital Advances in Medicine, E-Health, and Communication Technologies*, Vol. 2, Issue 1, 2011, pp. 35-54.
[20]. P. W. J. van Hengel, J. Anemuller, Audio event detection for in-home care, in *Proceedings of the International Conference on Acoustics (NAG/DAGA)*, Rotterdam, Netherlands, 2009, pp. 618-620.
[21]. P. Mach, Z. Becvar, Mobile Edge Computing: A Survey on Architecture and Computation Offloading, *IEEE Communications Surveys & Tutorials*, Vol. PP, Issue 99, 2017, pp. 1-1.
[22]. CMA-4544PF-W (http://www.cui.com/product/resource/pdf/cma-4544pf-w.pdf).
[23]. NVIDIA's next generation cuda compute architecture: Kepler TM GK110/210 (http://international.download.nvidia.com/pdf/kepler/NVIDIA-Kepler-GK110-GK210-Architecture-Whitepaper.pdf).
[24]. P. Melmstein, Distance measures for speech recognition, psychological and instrumental, *Pattern Recognition and Artificial Intelligence*, 1976, pp. 91-103.
[25]. T. M. Cover, P. E. Hart, Nearest neighbor pattern classification, *IEEE Transactions on Information Theory*, Vol. 13, Issue 1, 1967, pp. 21-27.
[26]. F. Alías, J. Claudi, X. Sevillano, A Review of Physical and Perceptual Feature Extraction Techniques for Speech, Music and Environmental Sounds, *Applied Sciences*, Vol. 6, Issue 5, 2016, pp. 143-187.
[27]. S. Liang, X. Fan, Audio Content Classification Method Research Based on Two-step Strategy, *International Journal of Advances in Computer Science Applications*, Vol. 5, Issue 3, 2014, pp. 57-62.
[28]. Z. Fu, G. Lu, K. M. Ting, D. Zhang, A survey of audio-based music classification and annotation, *IEEE Transactions on Multimedia*, Vol. 13, Issue 2, 2011, pp. 303-319.
[29]. M. L. Zhang, Z. H. Zhou, A k-nearest neighbor based algorithm for multi-label classification, in *Proceedings of the IEEE 1st International Conference on Granular Computing*, 2005, pp. 718-721.
[30]. Zhang Y., Jin R., Zhou Z. H., Understanding bag-of-words model: a statistical framework, *International Journal of Machine Learning and Cybernetics*, Vol. 1, No. 1-4, 2010, pp. 43-52.