An Elderly - Care System Based on Sound Analysis

Martin Frešer[1], Igor Košir[2], Violeta Mirchevska[1], Mitja Luštrek[1]
[1] Department of Intelligent Systems, Jozef Stefan Institute, Slovenia
[2] Smart Com d.o.o., Slovenia
martin.freser@gmail.com, igor.kosir@smart-com.si
{violeta.mircevska, mitja.lustrek}@ijs.si

ABSTRACT: *This paper proposes an elderly-care system, which uses a single sensing device installed in the user's home, primarily based on a microphone. We present preliminary results on human activity recognition from sound data. The recognition is based on 19 types of sound features, such as spectral centroid, zero crossings, Melfrequency cepstrum coefficients (MFCC) and linear predictive coding (LPC). We distinguished between 6 classes: sleep, exercise, work, eating, home chores and home leisure. We evaluated the recognition accuracy using 4 supervised learning algorithms. The highest accuracy, obtained using support vector machines, was 76%.*

Keywords: Microphone, Linear Predictive Coding, Support Vector Machines, Sound Data

Received: 12 January 2019, Revised 19 March 2019, Accepted 8 April 2019

DOI: 10.6025/stj/2019/8/2/54-59

1. Introduction

Predictions made by the Statistical Office of the European Communities state that the over-65 population in EU28 expressed as a percentage of the working-age population (aged between 15 and 64) will rise from 27% in 2014 to 50% in 2060 [1]. This demographic trend puts an immense pressure to change current health and care practices, which already accounts for around 10 % of EU's GDP spending [2]. Innovative remote care systems are emerging, which motivate and assist the elderly to stay independent for longer, thus reducing the costs for elderly care and the burden put on the working-age population.

This paper presents an elderly-care system, which uses a single sensing device installed in the user's home, primarily based on a microphone. A microphone may serve both as a sensor and as a communication device. As a sensor, a microphone may be used for detecting user's activity (e.g. sleep, eating, opening a door) and consequently reasoning about potential problems related to the user (e.g. the user did not eat whole day, the user is sleeping much more than usual). As a communication device, it allows the user to initiate specific services by simply saying a keyword (e.g. call for help). It is also needed for remote user-carer communication.

Elderly care based on microphone has not received a lot of attention, although technology acceptance studies show that most users would accept to have a microphone for home care services. Ziefle et al. [3] performed a user acceptance study comparing three home-integrated sensor types: microphone, camera and positioning system. According to this study, the microphone (plus speaker) is the most accepted technology, followed by the positioning system, while the camera is ranked last.

The paper is organized as follows. In Section 2, we describe the system architecture. Activity recognition based on sound analysis is presented in Section 3. Evaluation of the presented approach on real-world recordings follows in Section 4. Section 5 concludes the paper and presents future work.

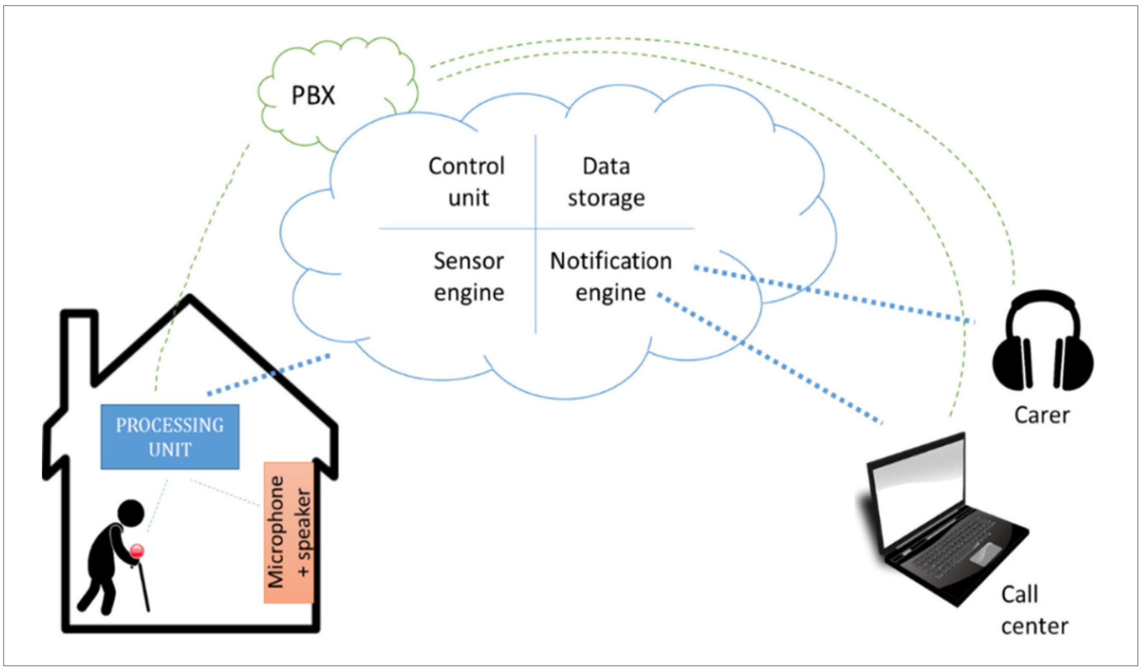

Figure 1. An elderly-care system architecture

2. System Architecture

Figure 1 presents the architecture of the elderly-care system.

Most of today's commercial elderly-care systems offer a so-called emergency call functionality. The user is wearing a red button using which he/she may call for help in case of emergency. By pressing the red button, the care-system establishes a phone connection to a carer or a call center through a telephone network. We use a private branch exchange network (PBX) for establishing such calls.

We extend this functionality with sound analysis in order to provide context to the emergency call (e.g. past user activity), as well as to provide higher safety – the system establishes an emergency call when certain types of sound, such as screaming, are detected. In order to do so, a cloud based system is established consisting of 4 main components: sensor engine, data unit, notification engine and control unit. The sensor engine analyses the sound in the apartment in order to detect user's activity (e.g. eating, sleep) or critical sound patterns (e.g. screaming, fire alarm). The output of this engine is kept in the data unit. In case of emergency detected through sound analysis, the control unit notifies a carer or a specialized call center through the notification engine about the user who needs help and why automatic emergency call is being established. When the carer responds, the control unit establishes a telephone connection with the user's apartment through the PBX network, enabling the carer to hear what is happening in the apartment and act accordingly.

3. Activity Recognition Based on Sound Data

People can distinguish quite well between some everyday activities just by listening to them. For example, if we hear a spoon hit a plate, we can say that the person is probably eating; if we hear the sound of pressing keyboard buttons, we can say that the person is either at work or at home and is using a computer. We developed a system that automatically detects everyday home activities based on sound.

Figure 2 presents the process of activity recognition from sound data. Firstly, we gather data using a recording device, such as a microphone. When recording, some privacy protection should be taken into account (e.g. we could record short sequences of time so we could not be able to recognize spoken words). We propose recording for 5 minutes in the following way: we record 200ms in every second for 1 minute and we do not record remaining 4 minutes.

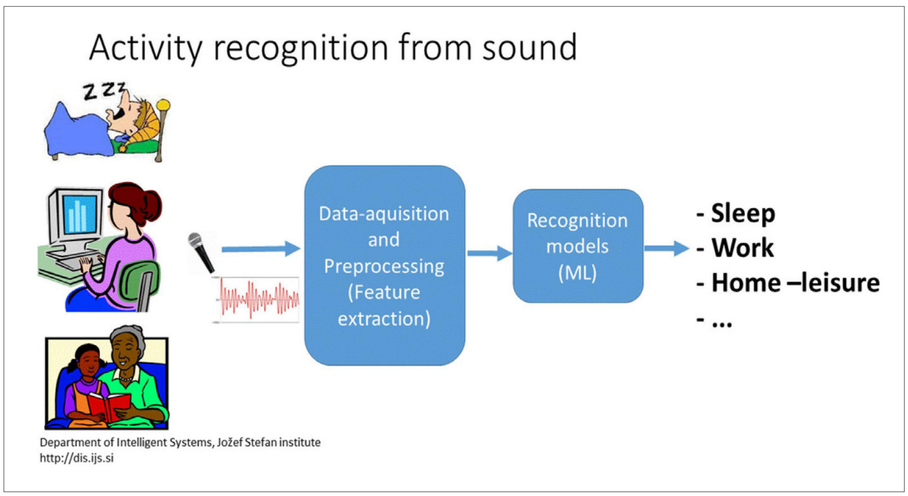

Figure 2. The activity recognition process

Then we extract sound features. Each feature is extracted in 20 ms long window. This is because sound signal is constantly changing and in such short window we assume that is not changing statistically much. We still have to have enough samples though, so shorter windows are inappropriate. We can also use window overlap so we lose less information. We used 20% of window overlap.

We extracted 19 types of features and we aggregated them in a window of one-minute length. We put together all recordings that were recorded in one minute, then we extracted each feature in 20 ms window and we aggregated it using mean and standard deviation, so we got a feature vector, which represented one minute. We also tried to aggregate for one second, but we got worse results.

Features were: Spectral centroid, Spectral rolloff point, Spectral flux, Compactness, Spectral variability, Root mean square, Fraction of low energy, Zero crossings, Strongest beat, Strength of strongest beat, Strongest frequency via FFT (Fast Fourier transform) maximum, MFCC's (Mel frequency cepstrum coefficients) (13 coefficients), Linear predictive coding (LPC) (10 coefficients), Method of moments (5 features), Partial based spectral centroid, Partial based spectral flux, Peak based spectral smoothness, Area method of moments (10 features) and Area method of moments of MFCCs (10 features). Those features were aggregated using mean and standard deviation. We also added 10 Area moments of Area method of moments of MFCC's. This sums up to 136 features. All features are explained in [4].

Since we have a lot of features, we use feature selection algorithms.

Finally, we use supervised machine learning techniques to build classifiers for our data.

4. Evaluation

In this section we present an evaluation of our experiment.

We gathered recordings from 3 persons in their everyday living with smart phone's microphone. They labeled data with the following activities: "Sleep", "Exercise", "Work", "Eating", "Home - chores" and "Home - leisure".

Data was firstly intended for monitoring chronic patients.

We split data into training and test set. We were recording each person for 2 weeks and we used the first week as training set and the second week as test set.

We extracted features using open-source library jAudio [4], [5].

For feature selection and machine learning we used the open-source library Weka [6]. We used the feature selection algorithm ReliefF implemented in Weka on every person. We used 4 machine learning algorithms: SMO, J48, RandomForest and iBK, all with default parameters. We measured accuracies of all algorithms and then we used the best-performing algorithm and we measured F-measures for all the activities.

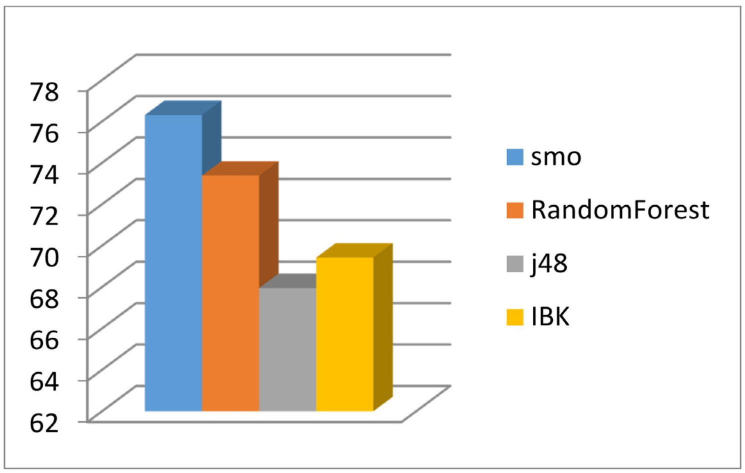

Figure 3. Average accuracy of all classifiers

As can be seen in Figure 3, the best performing algorithm was SMO, which produced the highest accuracies on all the tested persons. The average accuracy of SMO is 76 %. In the second and third place are RandomForest and iBK with the average accuracies of 73 % and 69 % respectively. The worst was j48 with the average accuracy of 68 %.

In Figure 4 we can see the average F-measure per activity for the best performing algorithm SMO. The best recognized activity is "Sleep" with the average F-measure of 0.96, following by "Work" with 0.85. SMO detected "Eating" and "Exercise" relatively well with the average F-measure of 0.46 and 0.43, respectively. The remaining average Fmeasures for "Home - leisure" and "Home - chores" were 0.38 and 0.26, respectively.

Since we had 3 persons, we trained one best-performing classifier (SMO) for each person. We got a confusion matrix for each person and we summed all 3 matrices in one. It can be seen in Table 1. We can see that activity "Sleep" is almost flawless. We can also see that "Home - chores" is usually misclassified as "Home - leisure", which could be a consequence of similar sounds produced in a person's home during various activities. Due to the high number of instances labeled as "Work", we got very good classification of "Work", but there are also many instances misclassified as "Work". We must take into account that recorded persons worked in the office, so many sounds are similar as in the home environment. We can conclude that for different activities, there can be many similar sounds, e.g. when person reads a book at home ("Home - leisure"), there can be silence as if the person took a nap ("Sleep"), so it is very challenging for classifiers to achieve high accuracies.

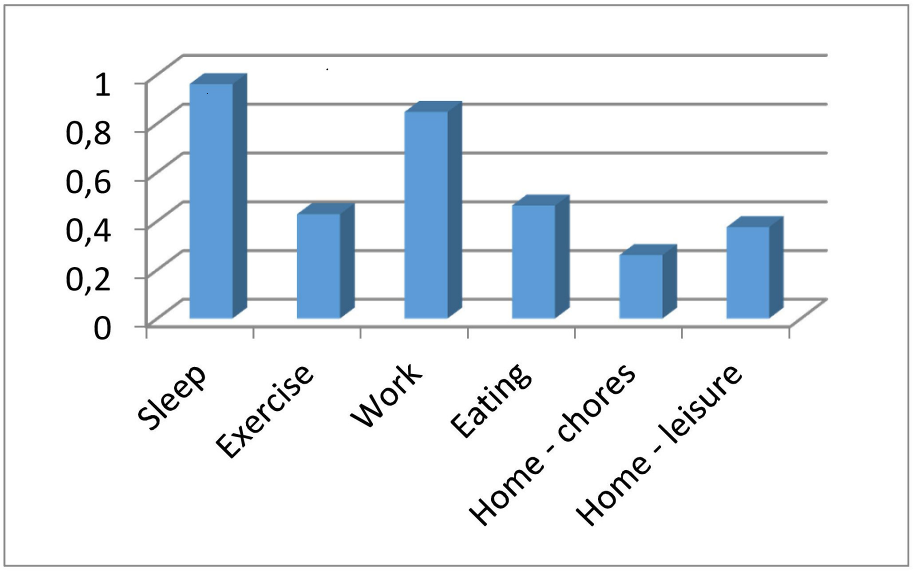

Figure 4. Average F-measure values of SMO

Sleep	Exercise	Work	Eating	Home - chores	Home - Leisure	classified as
198	0	1	0	0	5	**Sleep**
0	48	40	0	0	0	**Exercise**
0	18	129 0	43	20	95	**Work**
0	2	48	50	7	10	**Eating**
10	6	92	3	66	44	**Home - chores**
7	0	61	5	31	216	**Home - leisure**

Table 1. Summed confusion matrix of all persons

5. Conclusion

This paper presents a system and an approach to human activity recognition based on sound. The approach was tested on real-life recordings of three persons who annotated their activity for 2 weeks.

As outlined in Section 4 activity recognition from sound on 1 minute intervals may be challenging. There may be complete silence during different kinds of activities (e.g. sleep, work) or the recording may be dominated by speech. Therefore, it is difficult to achieve high accuracies in such settings.

Nevertheless, activity recognition from sound may be used for remote elderly-care. If we detect that the user was eating at usual times during the day, even though we do not have correct value about the eating period, we may conclude that user's state is normal. Having reliable sleep recognition, we may detect if the person is waking up during the night or if the period of sleep is lengthening, both of which may indicate a health problem. As future work, we need to record everyday living activities of the elderly, and test the system's capability to detect events that are critical for determining their health state.

References

[1] Eurostat, Retrieved September 2, 2014, from http://epp.eurostat.ec.europa.eu/tgm/table.do?tab=table&init=1&language=en&pcode=tsdde511&plugin=1.

[2] European commission, Horizon 2020, Health, demographic change and well being. Retrieved September 2, 2014, from http://ec.europa.eu/research/participants/portal/doc/call/h2020/common/1617611-part_8_health_v2.0_en.pdf

[3] Ziefle, M., Himmel, S., Wilkowska, W. (2011). When your living space knows what you do: Acceptance of Medical Home Monitoring by Different Technologies, Lecture Notes in Computer Science, 607 - 624.

[4] McEnnis, Daniel., Fujinaga, Ichiro., McKay, Cory., DePalle, Philippe. (2005). JAudio: A feature extraction library. ISMIR.

[5] McEnnis, Daniel., Fujinaga, Ichiro. (2006). jAudio: Improvements and additions. ISMIR.

[6] Hall, Mark., Frank, Eibe., Holmes, Geoffrey., Pfahringer, Bernhard., Reutemann, Peter., Ian, H., Witten. (2009). The WEKA Data Mining Software: An Update; SIGKDD Explorations, 11(1).