Analyzing Sounds of Home Environment for Device Recognition

Svilen Dimitrov, Jochen Britz, Boris Brandherm, Jochen Frey

DFKI, Saarbrücken, Germany
{<forename>.<surname>}@dfki.de

Abstract. Home environments are one of the subjects of study regarding ambient intelligent systems for various purposes, including development of assistance systems for the elderly and energy consumption optimization. Sensing the environmental state via different sensors is the first and crucial component of every ambient intelligent system. In this work we investigate the use of environmental sounds for touch-free audio-based device recognition in a home environment. For this purpose, we analyzed sound characteristics of typical home appliances using different processing techniques. We are using the acquired knowledge to develop a flexible set of features, which can be set manually or determined automatically. To classify the device-specific acoustic fingerprints – consisting of a significant subset of our features – we use established supervised learning techniques, whereby we optimized the straightforward ones. After building a recognition basis for the recognition of fixed length sound buffers on demand, we implemented a live recognition mode for real-time environment monitoring, providing runtime setup adjustments. We then extended our work with the recognition of untrained, simultaneously working, known devices by mixing their records, utilizing semi-supervised learning. We then anticipated promising results in our evaluation in various aspects, including recognition rate, performance for the different combinations of features, as well as to study the reliability of an automatic mixing of trained data.

Keywords: Ambient Intelligence, Smart Home, Sound-based Device Recognition

1 Introduction

In our modern way of life we are surrounded by an increasing number of devices, which we use to perform a large variety of activities. Some of those activities are not always straightforward and we often need some assistance to perform them. To make this happen, one has to give some intelligence to the devices to make them able to understand our intentions and to fit into our needs. In other words: making those devices sensitive and responsive to our presence, instead of relying on us to learn how to operate them. Making the devices more sensitive to human actions is one of the goals in the notion of activity recognition. This is the first step of designing a so called ambient intelligent system, which at first anticipates human actions with their purpose in a given environment, and then acts in an intelligent manner by predicting and assisting future actions. This should hold especially in the case, where humans are experiencing difficulties in performing those actions, but there are many further applications, such as optimizing electrical energy consumption.

In this work we study the sensing component of an ambient intelligent system. Such a component utilizes sensors and techniques to process their data in order to extract the desired information about the environment. Most studied techniques for this task, in respect to the human perception, are using visual and haptic sensors. Considering our perception, an additional way to sense the environment is by the perceived sounds. For this purpose, we introduce our Sound-based Device Recognition Framework – a fully developed system for device recognition based on analyzing environmental sounds. Our environment consists of a normal home. Its devices are commonly used for performing daily tasks, like the electrical toothbrush, the shaver, or the washing machine. Most of those devices create or disperse sounds, while being used to perform different activities. We study the most frequently used devices and the nature of the sounds, which accompany their usage. We then use this knowledge to transform those sounds to different acoustic representations in order to extract their most telling characteristics for the purpose of sound-based device fingerprinting. For the gathering of acoustic fingerprints we build a database, which is later used as a knowledgebase for further classification tasks. The latter are performed by trying out different machine learning algorithms and evaluating their performance in terms of complexity, recognition accuracy and adaptation capability. We then expand our work by adding further system capabilities, like live recognition using buffers of variable length and automatic mixing of different sounds for then the recognition of untrained combinations of known devices. Finally, we evaluate different aspects of the implemented recognition techniques in a smart home setup.

2 Related Work

Ambient intelligence has become a trending field in computer science as a natural consequence of high-instrumented environments, where each device is a target to embedding a microchip with increasing computational power. However, not all devices possess some sort of intelligence, nor do they need to. Furthermore, the so called intelligent devices are often not meant to be intelligent in a way besides accomplishing their function in a constant manner, regardless of environmental effects and regardless of potential improvement possibilities. From this standpoint, ambient intelligence is about to provide an intelligent interaction between different environmental parts, to integrate them in a holistic intelligent system, which automatically adapts to further environmental changes and incrementally increases the knowledge about the users [22, 4].

2.1 Sensor-based Environment Monitoring

The first component of such a system is the environment-sensing component, which recognizes all types of activities, ranging from long to short term and from large-scale to small-scale activities. Video cameras are a popular choice for a sensor when it comes to recognizing user activities, because they can provide a detailed knowledge about the ongoing activities in a home environment. On the other hand, cameras have some fallacies such as being obtrusive for its inhabitants regarding their presence [2], and usually suffer from bad recognition in sub-optimal light conditions. In addition, cameras are expensive and require computationally intensive algorithms for recognition [11].

Another function of the environment-sensing component is to recognize the different parts of the environment itself. Those include different devices, which are commonly used for performing different activities, or those devices performing periodic miscellaneous tasks by themselves. Since we investigate the device recognition part, we first get familiar with activity recognition using sounds and translate their results into our case (see Section 2.2). Then we make a comparison with the current research progress the concrete case of device recognition, mostly using power-based sensors (see Section 2.3).

2.2 Sound-based Activity Recognition

Another touch-free technique of recognition, regarding the human perception, is based on analyzing the audible sounds in a given environment. On the on hand, most of the sound-based recognizers are limited in recognizing human speech, together with some of its characteristics like speaker recognition and his emotional state in order to obtain detailed information about their subject of interest. On the other hand, there are very few studies, which aim to examine in an abstract way the daily human activities in a home environment according to their acoustic characteristics [18, 20, 9, 21, 12, 10, 16]. Despite their generalized way of analyzing sounds, they are all developed in a healthcare perspective and often make the implication that certain sound implies certain activity, which is not necessarily true. This slightly differs from our perspective of building up a set of audibly distinguishable entities, most of them being devices in an active state, without attempting to interpret their further meaning. Furthermore from a sound-processing standpoint, all of the mentioned studies use very similar techniques, which represent a small range of the available sound transformation techniques for recognition [15]. In this context, this study aims to integrate and evaluate also further recognition methods, based on refining and tuning of existing sound processing techniques and various machine learning algorithms, for the task of device recognition.

Among the most related works to our study, the most competing one has been done by Stäger et al. They have studied into detail different aspects of an activity recognition system, which we consider as important, too, like performing an optimization friendly installation [19] with carefully selected feature sets. However their biggest difference to our intended framework is that their recognition setup relies on wearable sensors, which is an uncomfortable way of sensing information. The second most competing work has been done by Istrate et al. [9]. They perform a large variety of sound processing techniques for their recognition, but in their tests they used data from multiple environments and trained the recognizer with 90% of it. In both points our work intends to do exactly the opposite. We first intend to create a personalized setup, and second to use less training data. Other interesting capabilities, which are out of the scope of this paper, but are implemented, include sound event detection and an attempt to recognize rare short-time events like glass breaking. The third related work collective also made an excellent job in placing multiple microphones and exploiting their installation, but we consider their setup as being too overwhelming for our purposes.

2.3 Power-based Device Recognition

Usually device recognition stands for recognizing different devices according to their interface. For instance, a recognizer scans the environment using different communication protocols like WiFi or Bluetooth and relies on the devices to support those protocols. However, not all devices had to support such communication protocols. Another recent research trend is to recognize electrical devices via a series of methods like using energy monitoring sockets, power analyzer [6, 1], and electromagnetic interference [5]. The problem is that all of those devices have to be electrical and connected to the power supply. Nevertheless, not all of the devices are electrical, nor should they be. Examples include toilet flush or a toothbrush. Some toothbrushes are electrical but they rely on batteries for their function, which makes them "invisible" for the mentioned technologies in their acting time. A natural way of recognizing devices, according to the human sense would be by analyzing their sounds. This includes the case where someone is using them to perform some activities, as well as the case, where they perform some periodic miscellaneous tasks without being operated directly by humans.

3 Sound-based Device Recognition

Figure 1 below illustrates our proposed process from an activity to its corresponding device recognition.

Fig. 1. Illustration of our process from an activity to the corresponding recognition

It starts with the environment (see Section 3.1) where some activity occurs (Section 3.2). Its sounds are captured by a microphone (see Section 3.3), and used to extract a desired set of features (see Section 3.4) to classify the devices (see Section 3.5). Furthermore we introduce sound mixing for recognizing untrained combination of devices (see Section 3.6). Finally we present how we cover the described process in a Sound-based Device Recognition Framework (see Section 3.7).

3.1 Environment

Since the main goal of this study is to provide device recognition in the smart home, the architecture, the implementation and the evaluations are centralized on this infrastructure. For the setup we have a home environment consisting of a single room (see Figure 2). For the sound monitoring a single microphone is used. We assume there is only one activity running at a time, with a single user that performs it. However as mentioned in the introduction, during certain activities, there may be multiple devices running at the same time, like shaving while showering. Activities consisting of simultaneously occurring actions are called complex activities. Our goal is to recognize all used devices during such a complex activity.

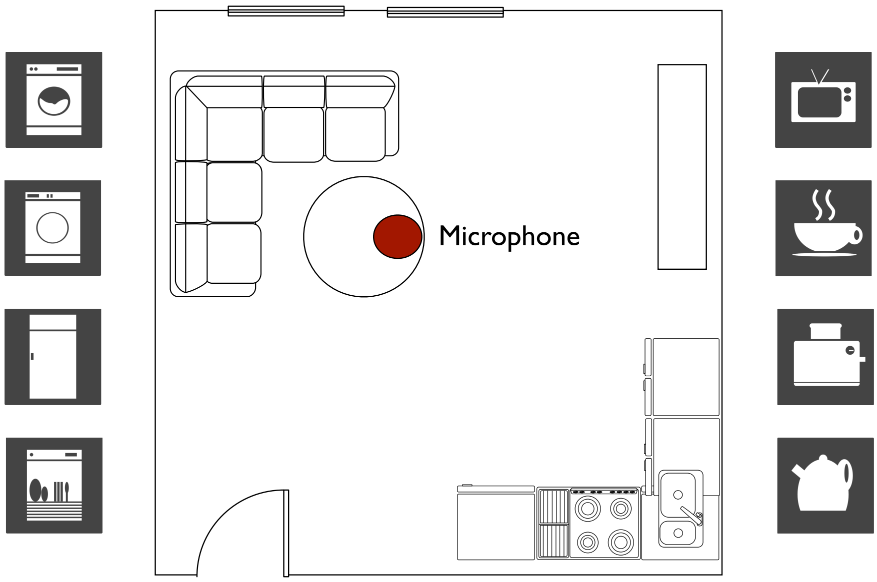

Fig. 2. Illustration of a one room home environment with its typical devices and their corresponding locations. The red dot represents a sample placement of the microphone. One should note that we are interested only in those devices that produce sound.

Compared to differently sensor equipped home environments our approach enables very simple and low cost installation. We eventually aim to optimize the sound processing and machine learning element components to complete the objective of creating a personalized low-cost recognition system.

3.2 Activity and Device Types

In this section we define the activity types, which determines the different devices and their corresponding actions in a home environment. The latter are of interest in this paper, due to the various device types used to perform them. It has to be noted that some devices perform miscellaneous tasks on a periodical basis. Such a task may or may not be part of an activity. Thus we assume that all recorded sounds are caused by tasks that correspond to at least one activity type.

In our framework the activity type encapsulates all types of activities, which produce sounds for time intervals longer than our smallest recognition window of 0.1s. Since we are interested in recognizing the devices used in activities we introduce a device type for devices used to perform the according activity. As we mentioned, not all of the activities are performed by humans. For example, a heating element starts heating by itself, when the temperature falls beyond a given threshold. We define the state of the device as active, if it disperses sounds. According to the heater example, it is on standby or inactive while measuring the temperature and active while heating.

To enable deeper knowledge about the used devices the user can determine whether the activity is performed with an electrical device while training the recognizer. On this basis one can make statements in recognition whether the recorded activity is performed with an electrical device. For example if we have five activities like brushing teeth with an electrical toothbrush, cleaning using a vacuum cleaner, coffee making using a moka pot, showering and speaking, we can separate those activities into three categories – performed with an electrical device (first two), performed with a non-electrical device (second two) and the last one performed without any device at all (see Figure 3). So after the recognition request the recognizer assigns different probabilities for the occurrence of each activity, and by those probabilities, one can make statements as to which of the currently defined three activity categories this was.

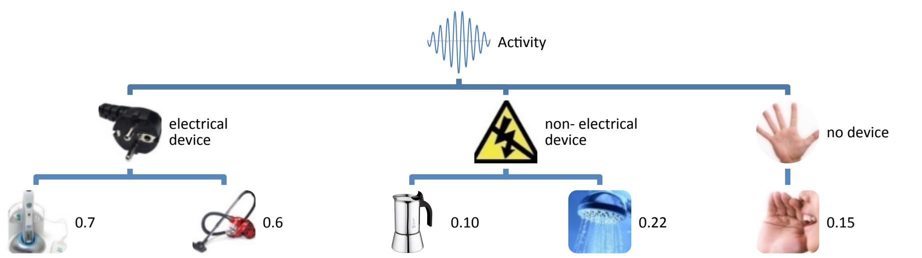

Fig. 3. This figure shows an example of deducting the sound source of an underlying activity.

3.3 Sound Recording and Buffering

Besides setting up the hardware there is a need to setup the software and its recording parameters. While choosing flexible recording settings, one still has to convert those unified metrics for the later transformation, which always results in data loss. So instead of bothering with conversion, we fix the recording parameters to unified supported constants, which satisfy our needs. It has to be noted that we make no assumptions about the capabilities of the recording device and we aim to use a single generic sound recording device.

The sampling frequency of the recorded sound is set to the most common used 44100 Hz (also used in compact disks). It has been chosen because every hardware supports this rate. Not all of the microphones support stereo recording though, which might produce some deviations during the hardware setup as well, so the number of channels is reduced to one (mono recording). For the sound transformations it is useful to have high precision variables, so the bit depth is set to 32 bit signed integer per sample, which doubles the standard bit depth value of 16 and is uniformly supported as well.

To extract the spectrum of the audio signal we performed a Fast Fourier Transformation (FFT). To increase the spectral resolution we set the FFT buffer to 4096 samples. For record frames we have chosen a variable time length, which is at least twice the FFT buffer size. The latter is important considering our choice of Hamming window function, where window overlapping makes sense in order to avoid missing recording information. Furthermore, to standardize the test data we chose a record length of 10 seconds. After performing the FFT we cropped the frequency window between 80 Hz and 5000 Hz. This frequency range contains the most distinguishable features in the spectrum, according to our manual study and automated evaluation of different signals. Further reasoning for this choice is the unsteady behavior of the frequency response, which different microphones provide. To handle further disturbances in the frequency response of our chosen range we implemented various filters at spectrum level.

3.4 Feature Extraction

In order to make a decision which features we should select, we first studied the nature of the sounds produced by devices using speech and music analysis software (see Figure 4). The first notable difference from the mentioned domains was that our signals were most noise-like, similar to the environmental sounds studied by [3]. So we had to consider a specific feature choice, different from the one used by speech and music recognition fields.

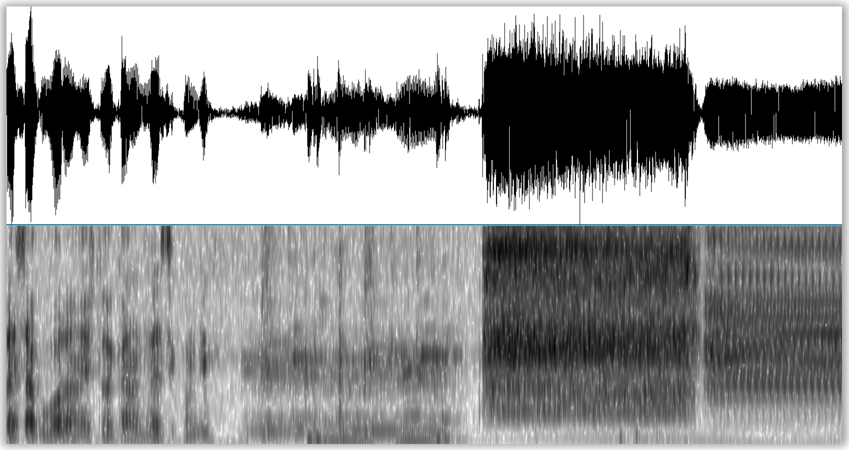

Fig. 4. Plot of 10 second recording of speech (first quarter), music (second quarter), epilator (third quarter) and hair trimmer (last quarter). Above we see the waveform of the recording and below its spectrogram between 80 Hz and 5000 Hz.

For activities performed with electrical devices it is typical that most of the defining part of the sound comes from its electrical motor, which is the actual sound source. An interesting finding was that other activities performed with non-electrical devices, like showering, have similar spectrograms compared to electrical devices.

We present the features in the order of their addition to the framework over the implementation process, which was influenced by our perception of sound and the conclusion from the last subsection. For example, the first perceivable feature of a sound is its loudness, so we chose to start with it. Then, in a mathematical perspective, zero crossings are one of the most important characteristics of a function, together with its maximums and minimums. Subsequently, we implemented a set of 8 features, for the task of audio-based device fingerprinting.

- **Loudness (LA)** is the average cumulative energy of the spectrum over the recognition interval. Note that this is a relative measure and is very dependent on filters and especially noise cancellation algorithms. Due to the nature of the feature it is also very important whether the activities occur at the same place, because the loudness is very dependent on their distance to the microphone.
- **Zero Crossing Rate (ZCR)** is the only feature derived from the time/amplitude domain (e.g., without processing the raw signal), after deciding to compute loudness after the filtering. To count zero crossings we check whether we have a zero crossing after each received sample.
- **Pitch (PA) and First Formant (FF)** is the pitch is also called fundamental frequency and represents the lowest frequency of a sound wave. It can be measured

by looking at the global maximum in the spectrum, while the second global maximum after the pitch in the spectrum is the first formant.

- **Pitch Span (PS)** is a temporal feature, which refers to the span of values, which the pitch takes over time. For some devices, like a vacuum cleaner, the pitch is steady and does not vary over time, while for some other devices, like a toothbrush, the pitch varies over time. In some cases, pitch span refers to the distance between the pitch and the first formant over time, due to the ambiguity of automatic distinction between local maximums (see Figure 5).
- **Pitch Energy (PE)** is the amount of energy as part of the whole energy, which surrounds the pitch in a 10% rectangular window (e.g. the 10% of the signal around the pitch as middle point).
- **Spectral Flatness (SF)** is an important measure, which is very useful to distinguish meaningful sound from noise (see Figure 5).
- **Spectral Roll Off (SRO)** is the point where the spectral function falls down. It provides important information about the main energy concentration over the frequencies.

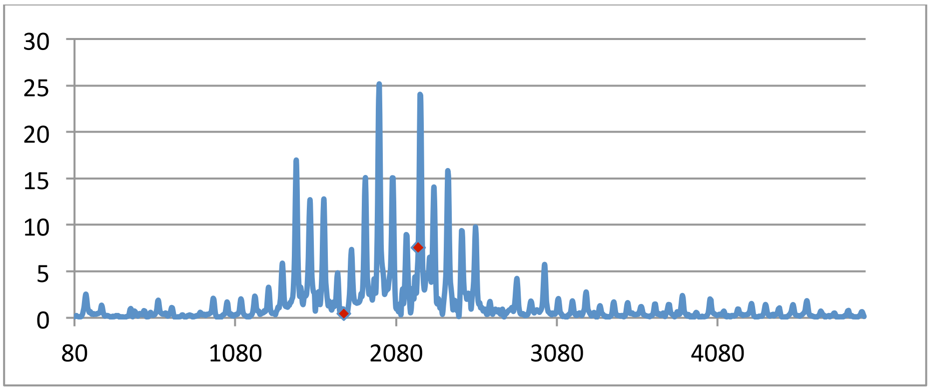

Fig. 5. Plot of spectrum of a hair trimmer illustrating that determination of the pitch and the first formant are a hard task. For the current snapshot the pitch would be calculated as 1975 Hz, while the first formant, would be 2234 Hz. One should note that over time those peaks switch, making the pitch vary between the mentioned two values, which is the way we compute the pitch span. Red points represent the beginning and the end of the interval for computing the pitch energy.

3.5 Device Classification

For a straightforward device classification we use the one-dimensional nearest neighbor algorithm. To enable fast runtime we use a single reference value. This reference value can be manually chosen either as the first, the last or the average of all features in the training set. The implemented algorithm provides a runtime with complexity in O(df) with number of devices (d) and number of enabled features (f).

To realize a more sophisticated solution, we use Infer.NET [14] as a state-of-the-art machine-learning library. Infer.NET implements the bayes point machine [8] in a standard supervised learning setting. The algorithms are trained via expectation propagation [13].

We also made a comparison between Infer.NET classifiers and our implementation of the multi-dimensional nearest neighbor for our test corpus with all selected features. Our results showed that for testing with single training, both recognizers achieved the same recognition accuracy. The drawback of Infer.NET was that it ran about 10 times slower, which is reasonable, considering the much larger number of computations it has to perform. However, with the increasing number of training data Infer.NET steadily increases its recognition rate, while our optimized implementation had a nearly constant recognition rate. Consequently, we anticipate a tradeoff between the runtime and the recognition rate, where one might choose the best option for the underlying setup.

3.6 Mixing

Mixing sounds is a novel approach in the field of sound-based activity recognition. It has been discussed in the field of music recognition for mixing different instruments in order to attempt their combined recognition [23]. However the technique used here is slightly different and avoids volume normalization, which is important for musical instruments, since they can play at different intensity, but mostly irrelevant for devices, which often have steady loudness. We are aware that defining which activities can occur simultaneously is a non-trivial task. In our framework we implemented up to three mixing mechanisms. Each mix consists of at most three records.

In order to test the mixing component we recorded two activities and their combination, and then we mixed automatically the recorded activities and compared them according to the obtained features (see Figure 6).

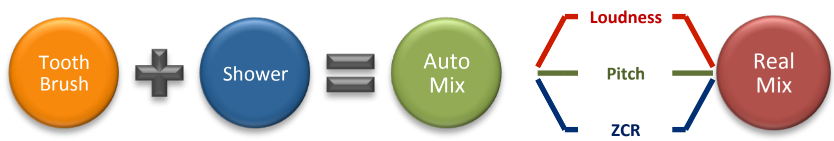

Fig. 6. Illustration of the comparison between automatically generated mix of records and a real mix according to their features.

To test whether mixing works or not, we compared 6 records of tooth brushing and showering together with their automatic and real mixes to illustrate their similarity in terms of the first three implemented features (see Figure 7).

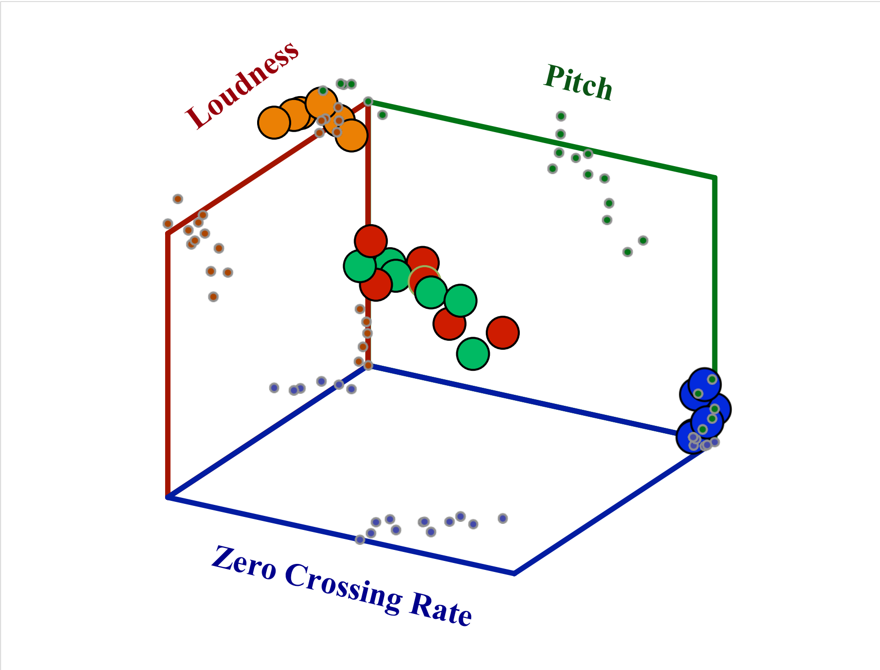

Fig. 7. Three-dimensional plot consisting of 6 experiment results with real (red) and automatic (green) mix of showering (blue) and tooth brushing (orange) according to the distribution over loudness, zero Crossing rate and Pitch.

We obtained similar results when mixing other devices, except while mixing the record of a blender and a vacuum cleaner. This mixing shared almost the same feature values between the automatic mixing and real mixing, except their average pitch energy. For instance, in the mixed version the pitch of the blender was perceived to be stronger, so it dominated and produced similar values throughout the tests. While in the real mixes of both devices we measured an average pitch ranging between 586 Hz and 915 Hz with one occurrence of 1531 Hz. Such deviations can be explained by the occurrence of acoustic resonance, which can alter the pitch frequency. Another possible explanation is that the sum of the energies at some frequency bin might be the strongest energy in the mix and thus regarded as a pitch by the sound-processing unit.

3.7 Sound-based Device Recognition Framework

We implemented a graphical user interface (GUI), which provides fast access to all functions and combinations of our sound-based device recognition framework. It is also important to visualize some of the sound derivations and to enable live tracking of relevant features.

In the following, we introduce the architecture of our sound-based device recognition framework. Its main part is the developer GUI, which contains all of the functionality of the system, while there is also a client, which can connect with the main program to ask for recognition results (see Figure 8). An example of a client can be a smartphone, which can provide basic tasks like training the recognizer and requesting recognition, without providing the user with all possible configuration steps.

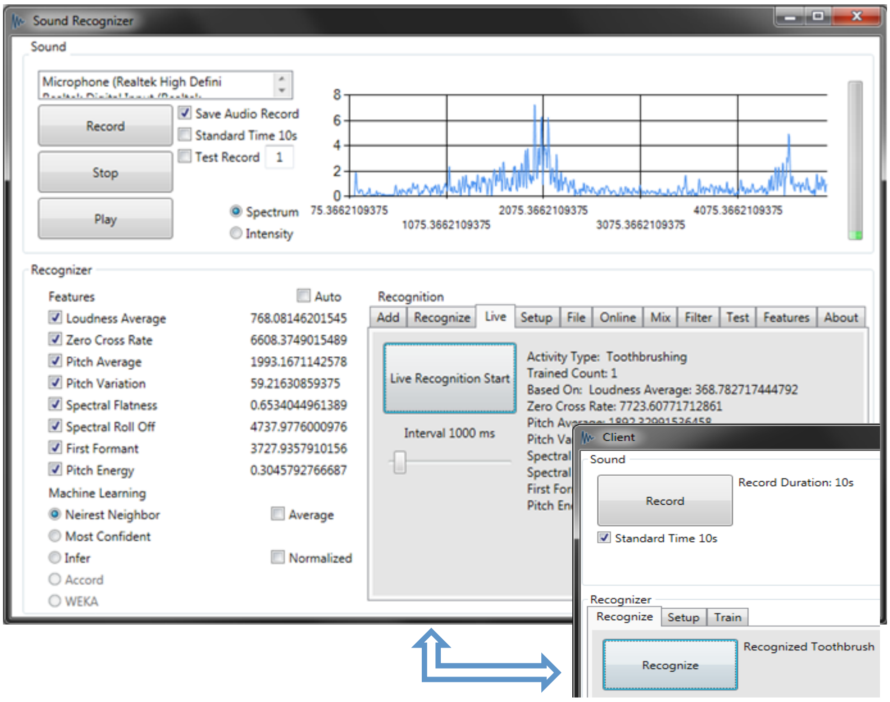

Fig. 8. Screenshot of our sound-based device recognition framework interacting with a simple client application to provide a mobile recognition service.

The developer GUI and the client application are both separated in two components, responsible for the sound processing and machine learning parts. From the user perspective, the user first creates records and then attempts to recognize them. Training the recognizer is realized by adding an activity with its corresponding device if it exists. A further way of increasing the knowledge of the recognizer is realized by providing a user feedback mechanism after presenting the recognition result. The last two properties are the basis of building an incremental learning system, which aims to make better recognitions over time in terms of a growing number of recognizable devices and recognition accuracy. The latter means that the system is capable to start from scratch after being installed in a personalized setup and rely on the user to build its knowledge. A further important capability is changing the settings during a live-recognition mode, which enables in depth control during the development process.

4 Results

We ran an evaluation in a home environment with a static microphone setup and recorded 150 records to build up our test corpus. It consists of 25 classes problem, out of which 20 classes represent devices, 3 classes are mixes of two devices, and the other two classes are speaker and silence. We ran the implemented multi-dimensional nearest neighbor algorithm with single training for the task of recognizing 125 devices consisting of 5 occurrences from each of the 25 classes. We tested the power set of all features. This means, with our implemented 8 features we tested 255 combinations, excluding the empty set. The results of best and worst performing feature combinations together with the average recognition results are shown in Table 1.

Feature Count	Best Set	Result	Worst Set	Result	Average Result
1	LA, (ZCR, PA)	52%	FF	30.4%	44.6%
2	LA, SRO	81.6%	SRO, FF	49.6%	67.49%
3	LA, SF, FF	93.6%	PS, FF, PE (PS,SRO,PE)	62.4%	76.69%
4	LA,PA,SF,FF (LA,SF,SRO,FF)	97.6%	PS,SRO,FF,PE	64%	79.97%
5	LA,PA,SF,SRO,FF	97.6%	ZCR,PS,SF,SRO,FF PA,PS,SRO,FF,PE	68.8%	80.66%
6	LA,ZCR,PA,SF,SRO,FF LA,ZCR,SF,SRO,FF,PE LA,PA,SF,SRO,FF,PE	94.4%	ZCR,PA,PS,SF,SRO,FF,PE (ZCR,PA,PS,SRO,FF,PE) (ZCR,PA,PS,SF,SRO,FF)	72%	80.06%
7	LA,ZCR,PA,SF,SRO,FF,PE	93.6%	PA,PS,SF,SRO,FF,PE (ZCR,PA,PS,SRO,FF,PE)	74.4%	79%
8	All Features	77.6%	All Features	77.6%	77.6%
AVG		85.9%		62.4%	73.26%

Table 1. Best, worst and average true positive recognition rate for all different feature combinations of different set sizes. The results in brackets were up to 1 recognition close to the provided result

With a single feature for recognition we obtained best results for LA, directly followed by ZCR and PA. They were also the first three implemented features. It is interesting, that the combination of those three was nowhere near to matching the performance of the winners in the next two categories.

For feature coupling, we anticipated also an interesting result having SRO in the best combination as well as in the worst combination. In addition, the FF is present in both, the best and the worst result sets. This is a clear evidence that the combination of features is crucial for the recognition process, rather than having one strong feature, supporting our claims in Section 2.2.

We can see also that LA performs well and could not be found in any of the worst results. This is due to the nature of the electrical devices to have a certain loudness level, especially, due to the fixed position of the microphone. It is important to note that most of the devices are used at fixed locations.

For best recognition set we identified two combinations, which beat the 97% rate and one not far behind – <LA,PA,SF,FF> (97.6%), <LA,PA,SF,SRO,FF> (97.6%), and <LA,SF,SRO,FF> (96.8%). We identified the reason for these exceptional good results being the sound processing setup for the environment, as well as most of the devices being tested throughout the development, thus enabling the precise extraction of their characteristics.

Our average results between 4 and 7 feature sizes was about 80%, which is also a same feature count, where the best results peaked. We tested our automatic feature selection algorithm and it chose a set of 6 features to obtain 91.2% recognition accuracy. Thus we conclude that the feature count range between 4 and 7 features is the best performing.

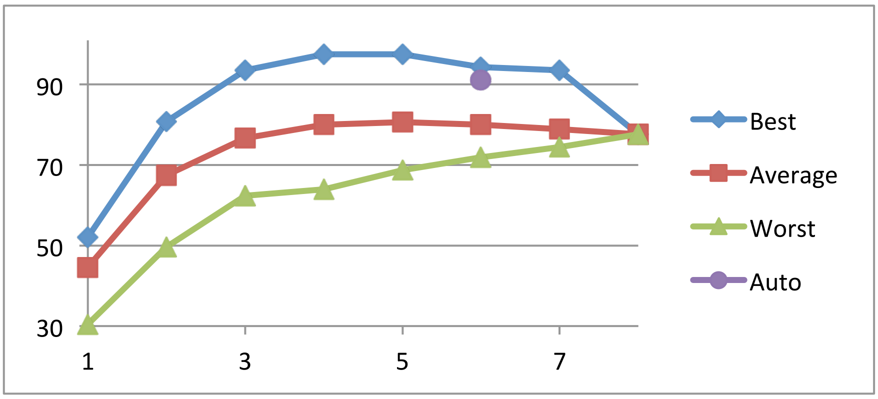

Fig. 9. A plot of best (blue) vs. average (green) vs. worst (red) results in terms of the different recognition rates (y-axis) according to the different feature set size (x-axis). The violet point represents the automatic feature selection, which selected 6 features and obtained 91.2% recognition accuracy.

Figure 9 shows a visualization of the results from Table 1. Both, the best and the average cases increase their accuracy for feature count up to 4 and 5, and from that point on there is a declining. Thus we observe that the increasing number of features doesn't necessarily mean better recognition, as mentioned in the introduction. However, more features should be implemented, since the worst-case recognition rate is increasing with the number of used features.

5 Conclusion

There are many contributions in the field of sound-based device recognition provided in this work. We have first shown that we can significantly reduce the complexity of a sound recognition system in a personalized home setup, as well as implement different ways to achieve this at both sound processing and machine learning levels. Our second contribution is in mixing automatically activities for their untrained recognition, where we obtained good results. We made also a detailed comparison between automatically mixed records of some activities and their real simultaneous occurrences.

We performed a manual study of specific characteristics of sounds produced by devices and made a full testing of all combinations of features to identify the best performing set. Most of those characteristics were not regarded by the majority of related works for the case of general activity recognition, since they are not applicable for speech recognition, which is their conventional research starting point. However, according to our evaluation, combinations of our chosen features are definitely important for classification of activities in a home environment. In addition, we implemented an automatic feature selection mechanism. Both, the chosen feature set for implementation, and their automatic selection for recognition, performed well in our evaluation. We adopted different machine learning techniques and optimized a couple of them for our purpose. We also introduced one of the first systems in the field of sound-based activity recognition, designed to learn over time using a feedback from the user and adapting its recognition settings, such as automatically choosing the best feature set.

References

1. Belley, C., Gaboury, S., Bouchard, B., & Bouzouane, A. (2013). Activity Recognition in Smart Homes Based on Electrical Devices Identification. Proceedings of the 6th International Conference on PErvasive Technologies Related to Assistive Environments (pp. 7:1--7:8). Island of Rhodes, Greece: ACM.
2. Brey, P. (2005). Freedom and Privacy in Ambient Intelligence. Ethics and Information Technology, 7(3), 157-166.
3. Chu, S., Narayanan, S., & Kuo, C.-C. (2009, Aug). Environmental Sound Recognition With Time–Frequency Audio Features. IEEE Transactions on Audio, Speech, and Language Processing, 17(6), 1142 - 1158.
4. Cook, D. J., Augusto, J. C., & Jakkula, V. R. (2009, August). Ambient Intelligence: Technologies, Applications, and Opportunities. Pervasive and Mobile Computing, 5(4), 277-298.
5. Gupta, S., Reynolds, M. S., & Patel, S. N. (2010). ElectriSense: single-point sensing using EMI for electrical event detection and classification in the home. Proceedings of the 12th ACM international conference on Ubiquitous computing (pp. 139-148). New York, NY, USA: ACM.
6. Hart, G. (1989, June). Residential energy monitoring and computerized surveillance via utility power flows. Technology and Society Magazine, 8(2), 12 - 16.

7. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning (2nd ed.). Springer.
8. Herbrich, R., Graepel, T., & Campbell, C. (2001, 9 1). Bayes point machines. The Journal of Machine Learning Research, 1, 245-279.
9. Istrate, D., Vacher, M., & Serignat, J.-F. (2008). Embedded Implementation of Distress Situation. Identification Through Sound Analysis. The Journal on Information Technology in Healthcare, 6(3), 204-211.
10. Karbasi, M., Ahadi, S., & Bahmanian, M. (13-16 Dec 2011). Environmental sound classification using spectral dynamic features. 8th International Conference on Information, Communications and Signal Processing (pp. 1 - 5). Singapore: ICICS.
11. Ke, S.-R. a.-J., Hwang, J.-N., Yoo, J.-H., & Choi, K.-H. (2013, 6 5). A review on video-based human activity recognition. Computers, 2(2), 88-131.
12. Lozano, H., Hernáez, I., Picón, A., Camarena, J., & Navas, E. (2010). Audio Classification Techniques in Home Environments for Elderly/Dependant People. Computers Helping People with Special Needs(6179), 320-323.
13. Minka, T. P. (2001). Expectation Propagation for Approximate Bayesian Inference. Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence. UAI '01, pp. 362-369. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.
14. Minka, T., Winn, J., Guiver, J., & Knowles, D. (2012). Infer.NET 2.5. Microsoft Research Cambridge.
15. Müller, M., Ellis, D. P., Klapuri, A., & Richard, G. (2011, February 4). Signal processing for music analysis. IEEE Journal on Selected Topics in Signal Processing, 5(6), 1088-1110.
16. Sehili, M. A., Lecouteux, B., Vacher, M., Portet, F., Istrate, D., Dorizzi, B., et al. (2012). Sound Environment Analysis in Smart Home. Ambient Intelligence: Third International Joint Conference, 208-223.
17. Stäger, M. (2006). Low-Power Sound-Based User Activity Recognition. Swiss Federal Institute of Technology Zurich, Information Technology and Electrical Engineering. Zürich: ETH Zürich.
18. Stager, M., Lukowicz, P., & Troster, G. (2004). Implementation and Evaluation of a Low-Power Sound-Based User Activity Recognition System. Proceedings of the Eighth International Symposium on Wearable Computers (pp. 138--141). Washington, DC, USA: IEEE Computer Society.
19. Stager, M., Lukowicz, P., & Troster, G. (2007, June). Power and accuracy trade-offs in sound-based context recognition systems. PerCom - Pervasive and Mobile Computing, 3(3), 300-327.
20. Temko, A., Malkin, R., Zieger, C., Macho, D., & Nadeu, C. (2006). Acoustic Event Detection and Classification in Smart-Room Environments: Evaluation of CHIL Project Systems. Cough, 5-11.
21. Wang, J.-C., Lee, H.-P., Wang, J.-F., & Lin, C.-B. (2008, January 4). Robust Environmental Sound Recognition for Home Automation. IEEE Transactions on Automation Science and Engineering, 5(1), 25 - 31.
22. Weber, W., Rabaey, J., & Aarts, E. H. (March, 2005). Ambient Intelligence. Springer.
23. Wieczorkowska, A., Kolczyńska, E., & Raś, Z. W. (2008). Training of Classifiers for the Recognition of Musical Instrument Dominating in the Same-Pitch Mix. New Challenges in Applied Intelligence Technologies, III(134), 213-222.