2009 International Symposium on Wearable Computers

Can a Mobile Phone in a Pocket Reliably Recognize Ambient Sounds ?

Tobias Franke, Paul Lukowicz, Kai Kunze and David Bannach
Embedded Systems Lab
University of Passau, Germany
Email: franke@fim.uni-passau.de, {paul.lukowicz, kai.kunze, david.bannach}@uni-passau.de

Abstract

We investigate how different locations inside clothing influence the ability of a system to recognize activity relevant sounds. Specifically, we consider the recognition of sounds from 9 household and office appliances recorded using an iPhone placed in 2 trouser pockets, 2 jacket pockets, a belt holster and the users' hand. The aim is not to demonstrate good recognition rates on the above sounds (which has been done many times before) but to compare recognition rates from the individual locations and to understand how to best train the system to be location invariant.

Introduction

Mobile phones are an attractive platform for activity recognition in real world applications [3], [6]. So far, mostly accelerometers and location sensors (GPS, WLAN fingerprintig) built into phones have been used. In this paper we aim at the use of mobile phones for sound recognition. Previous work has demonstrated the usefulness of audio information in many scenarios [7], [2], [1]. Obviously, a mobile phone has a microphone. We have also previously demonstrated that reliable, relevant recognition can be achieved with computational resources far below those of a modern smart phone [5]. The same work has shown that sampling rates and frequency ranges comparable to what is used in telephony are sufficient for many recognition tasks. However, all the above work has been done with the microphone well exposed to the sound (worn on the wrist near the source of the sound or attached to the outside of clothing). A mobile phone on the other hand, is mostly inside a pocket or a bag. Even if worn on a belt, it is often covered by a jacket or a sweatshirt. This paper investigates how being at different locations inside clothing influences the recognition capability. Specifically we answer the following questions:

1) Can meaningful sound recognition be achieved with signals recorded from a microphone of a mobile phone carried at typical locations *inside clothing* ?
2) How far will the performance degrade when we train and test while at different locations
3) Can noise caused by the phone rubbing on the pocket be distinguished from external (interesting) sounds ?

We have based our choice of phone locations on [4] which found out that people mostly carry mobile phones in trouser pockets, upper body pockets, a belt holster or a bag. For our investigation we concentrate on the first three and leave out the bag because we are interested in indoor activity recognition, where people are unlikely to be carrying a bag.

Experimental Setup

For the recordings we used a pair of golfing trousers, jeans, a leather jacket, a sports jacket, and a belt holster, which was covered by a light jersey. In order to get a baseline we also made recordings under ideal circumstances which means holding the phone directly at the source of the sound. Whenever recording from the inside of a pocket we placed the phone in a way that would make it most usable (e.g. to answer an incoming call) right after it's been pulled out of the pocket (see Figure 1).

We consider sounds that have been used in previous work and are relevant to indoor (household, office, workshop) activity monitoring: printer, copier, coffee machine (more specifically the grinding phase),drill,hot air gun, microwave,toilet flush,bathroom water tap and water boiler. We also recorded the noise caused by the fabric of the clothing rubbing on the phones' microphone when moving and the background noise in an otherwise silent office room.

The sounds were recorded with an iPhone using our own data logging software and evaluated off-line in MATLAB. Altogether we made 990 recordings with a sampling frequency of 8 kHz - 15 for each sound and location. Out of those 15 recordings 5 were used for training the classifier and 10 were used for testing purposes. In case of sounds such as the coffee grinder we recorded the noise at full length. However, in case of longer and more repetitive sounds such as the water boiler we only recorded for a period of approximately 10 seconds. Since we always operated the devices during the recording the typical distance from the object was an arms length or about half a meter. As mentioned earlier we also made some reference recordings under ideal conditions where the recording distance was about 20 cm. The classification was based on methods we have previously used for similar problems [5]. A 512 point FFT was performed on frames 4000 samples (around 0.5 seconds) long. The feature space dimension was reduced with an LDA and a KNN with k=3 was used for classification. In addition

1550-4816/09 $26.00 © 2009 IEEE
DOI 10.1109/ISWC.2009.34

Figure 1. Locations (A: Belt, B: Sports Jacket, C: Golf Trousers, D: Leather Jacket, E: Jeans, F: Ideal Recording)

to the frame by frame classification we also performed a majority decision over all frames of a single recording (5 to 15 depending on sound).

Results and Conclusions

When trained and tested on the same location the results were between 90% (leather jacket) and 97% (golfing trousers and hand) for frame by frame recognition and between 98% (interestingly for the clean data) and 100% (for all the others) for the majority decision. This clearly shows that the answer to our first question is yes: it is feasible to do good quality recognition with the phone inside clothing.

Figure 2 shows the results of testing on different locations when the system was trained on the 'clean' (in the hand) data and on a mix of data from all locations (except the one on which it was tested). The results for 'clean' training are poor for both frame by frame and majority decision (around 60%). This shows that the damping induced through clothing has a significant influence on the spectral composition of the sound and has to be appropriately modeled to achieve reasonable performance. The effect of the damping depends on the sound. While microwave, printer, toilet flush, coffee grinder and background had a recognition rate of 90% to 100%, the rates for the remaining sounds dropped to 0 (hot air gun) 10% (water tap and water boiler) and 5% (drill).

The following results indicate that the damping model is reasonably independent of the type of clothing. When trained on a mix of locations the results are close to what we have seen when testing and training on the same location (80%-90% on frame by frame and 100% in majority decision). When trained on one and tested on another location (Figure 3) the recognition rates (majority decision) are between 80% and 100% with an average around 90%.

With mixed training data we had a 100% recognition rate with majority decision. This included the noise of the phone rubbing against the pocket when walking. On a frame by frame base the recognition rate for the 'rubbing' was 91%. Thus the answer to the third question is also yes: the noise caused by the phone moving around in the pocket can be well separated from relevant external sounds.

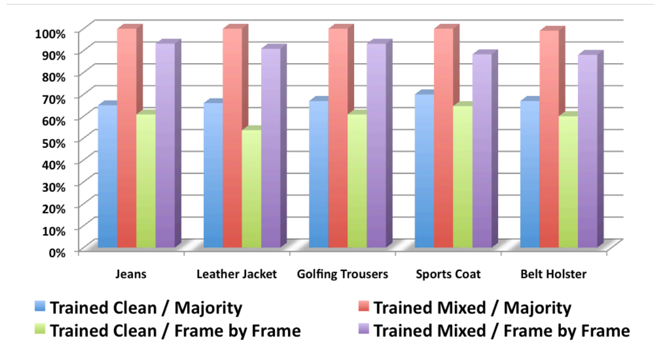

Figure 2. Classification rates for both classification methods with different training data

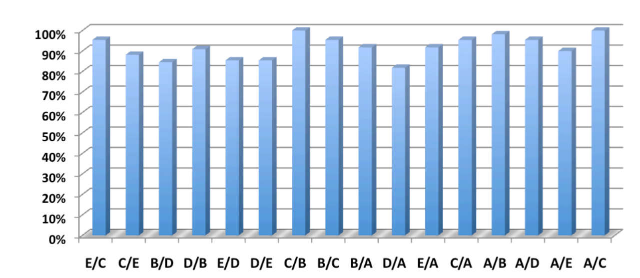

Figure 3. Classification rates when training data and test data are not from the same location (*training data / test data* - see Fig. 1 for explanation of abbreviations)

References

[1] J. Chen, A.H. Kam, J. Zhang, N. Liu, and L. Shue. Bathroom activity monitoring based on sound. pages 47–61. Springer, 2005.

[2] B. Clarkson, N. Sawhney, and A. Pentland. Auditory context awareness via wearable computing. *Proc. 1998 Workshop on Perceptual User Interfaces (PUI'98)*, pages 4–6, 1998.

[3] H.W. Gellersen, A. Schmidt, and M. Beigl. Multi-sensor context-awareness in mobile devices and smart artifacts. *Mobile Networks and Applications*, pages 341–351, 2002.

[4] F. Ichikawa, J. Chipchase, and R. Grignani. Where's the phone? a study of mobile phone locations in public spaces. *Proc. IEEE Mobility Conference 2005*, 2005.

[5] M-Stäger, P. Lukowicz, and G. Tröster. Power and accuracy trade-offs in sound-based context recognition systems. *Pervasive and Mobile Computing*, pages 300–327, 2007.

[6] D. Siewiorek, A. Smailagic, J. Furukawa, A. Krause, N. Moraveji, K. Reiger, and J. Shaffer. Sensay: a context-aware mobile phone. *Wearable Computers, 2003 Proc., Seventh IEEE International Symposium on*, pages 248–249, 2003.

[7] J.A. Ward, P. Lukowicz, G. Tröster, and T.E. Starner. Activity recognition of assembly tasks using body-worn microphones and accelerometers. *IEEE Trans. PAMI*, pages 1553–1567, 2006.