Towards Grip Sensing for Commodity Smartphones Through Acoustic Signature

Namhyun Kim
Sungkyunkwan University (SKKU), Republic of Korea
csci6108@skku.edu

Jinkyu Lee
Sungkyunkwan University (SKKU), Republic of Korea
jinkyu.lee@skku.edu

UbiComp/ISWC'17 Adjunct, September 11–15, 2017, Maui, HI, USA

ACM ISBN 978-1-4503-5190-4/17/09...$15.00
https://doi.org/10.1145/3123024.3123090

Abstract

While hand grips are important to understand the intent of smartphone users, existing studies on hand grip detection either require additional hardware or exhibit limitations on the type/number of grips. In this paper, we propose a novel grip sensing system that enables a smartphone to detect various user-defined hand grips without any additional hardware. Our system emits a carefully-designed (inaudible) sound signal, and records the sound signal modified by an individual grip. The recorded sound signal is transformed into a unique sound signature through feature extraction process, and then SVM (Support Vector Machine) classifies the sound signature so as to identify the signature as one of pre-defined grips. With six representative grips, we demonstrate that our system exhibits 93.0% average accuracy for ten different users. Beyond this feasibility demonstration, our ongoing work is not only to improve the accuracy, but also to adapt our system to various real environments.

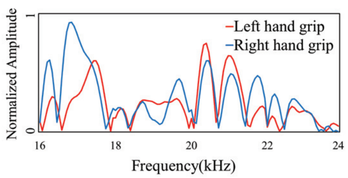
Author Keywords

Grip sensing; hand posture; sound signal; acoustic signature; smartphones.

ACM Classification Keywords

H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous.

(a) left hand **(b)** right hand



(c) corresponding sound signature

Figure 1: Sound signature comparison for two different grips

Introduction

Since users have their own distinct ways of grasping, utilizing hand postures as user interfaces necessitates the detection of various user-defined hand postures. While there have been many studies to address this grip sensing problem, they either require additional hardware (e.g., [3, 2, 7]), or differentiate a limited number/type of grips as well as operate only when there is a touch event on a smartphone's screen (e.g., [6, 5]).

In this paper, we propose a novel grip sensing system, which allows a smartphone to detect various hand postures without any additional hardware and screen touch event. To this end, our system utilizes the change of a sound signal according to different grasping hands for a smartphone. Once our system emits a sound signal that is carefully designed in terms of its frequency range, structure and duration, the sound signal from the speaker to the microphone can be divided into the structure-borne sound (that is propagated through the smartphone body) and the air-borne sound (that is propagated through air). Since these two sounds make a unique spectrum for each grip, the combined sound recorded by the microphone can be an unique acoustic signature of each grip. Then, the system classifies the sound signature into one of pre-defined grips, using SVM (Support Vector Machine).

To demonstrate the feasibility, we implemented the proposed system as an Android application, and experimented it on Samsung Galaxy Note 5. With six representative grips (shown in Figure 3), the proposed system exhibits 93.0% average accuracy for ten different users in an office environment. While succeeding to show the feasibility of the proposed system, we need not only to improve its accuracy, but also to adapt the system to various real environments, which is our future work.

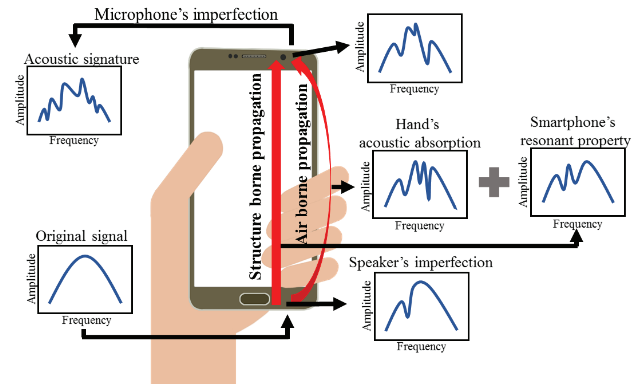


Figure 2: Sound delivery from the speaker to the microphone

Grip Sensing System

Grip Sensing Principle

In order to detect different hand postures using acoustic signature, we should understand how a sound signal is delivered from the speaker to the microphone, especially focusing on the change of the sound signal according to different hand postures, shown in Figure 2.

First, when the speaker emits sound signal, it is a little bit different from the sound the smartphone originally issues due to hardware imperfection. Then, the emitted sound signal is propagated into two ways, called structure- and air-borne propagations. The structure-borne sound is the sound that propagates through the body of the smartphone. While propagating, the structure-borne sound changes the frequency spectrum according to the resonant property of the smartphone, which is related to the boundary conditions and modal damping [4, 9]. The boundary condition of the smartphone changes according to how the user grasps the smartphone. On the other hand, the air-borne sound

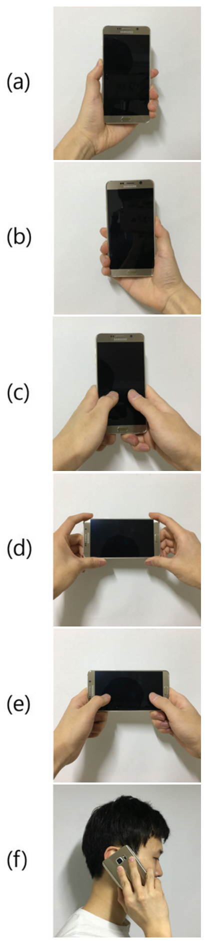


Figure 3: Various grips according to different smartphone usage functions

is the sound that propagates through the air. Because the human hand has different acoustic absorption coefficients for each frequency [1], the sound that goes directly to the microphone from the speaker through air is attenuated in certain frequency ranges, which depends on grips.

Through the two propagations, different grips entail different recorded sound signatures. For example, Figure 1 shows the recorded sound signatures of two different grips, which are made by the feature extraction process to be described in the next subsection.

Sound Feature Extraction Process

While we know that different grips yield different sound signatures, we need to extract features to distinguish the different grips. To this end, we design a sound signal shown in Figure 4. We determine the frequency range from 16kHz to 24kHz to prevent the user from noise. Also, the sound signal is a linear chirp signal sweeping from 16kHz to 18kHz, from 18kHz to 20kHz, from 20kHz to 22kHz, and from 22kHz to 24kHz. The length of each chirp signal is 200 samples and the interval between each chirp is 500 samples. As a result, the total time taken to sense a single grip is about (200*4 + 500*3)/48000=47.9ms. Also, we set volume as 50% of maximum volume.

These sound signal structure reduces the effect of the reflection sound. Because reflection sound changes differently depending on the location [8], it makes it difficult to identify the target grip. By using four short chirp signals instead of one long chirp signal or one short chirp signal, our system eliminates the influence of reflection sound coming far from while getting more features. Also, setting the volume as 50% of maximum volume allows the air-borne sound to ignore the effect of a remaining reflection sound. Note that one may think that a 50% sound signal is too noisy; however, according to our experiments, users usually do not recognize the sound signal because of the short duration and high frequency.

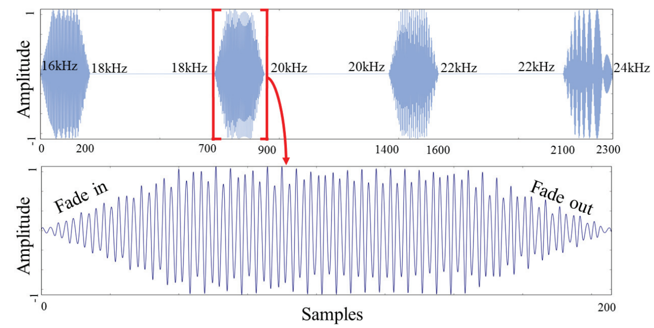
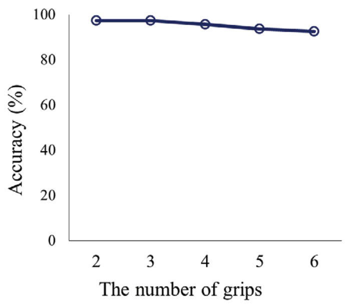


Figure 4: Structure of a sound signal: The upper figure shows the structure of the generated sound signal consisting of four chirps. The lower figure magnifies the second chirp.

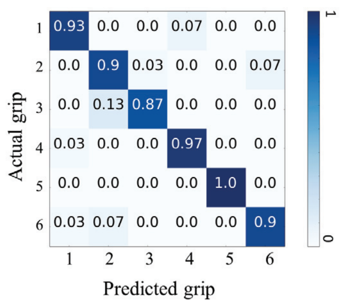
The recorded sound signal can extract the amplitude of each frequency through Fast Fourier Transform (FFT). We employ FFT using hamming widow with size 1024. A total of 172 features were used after excluding unnecessary parts. We classify each sound signature by implementing SVM with linear kernel.

Evaluation

We implemented the proposed system in SamSung Galaxy Note 5. We have ten participants with eight males and two females, and their ages range from 20 to 32 years. They are required to see six different grips shown in Figure 3, and asked to grasp the smartphone with the six grips. If individual participants grasp the smartphone with each grip in the same position, the accuracy is very high because of the constant reflection sound. To address the effect of the reflection sound, we let participants repeat the following



(a) accuracy



(b) confusion matrix

Figure 5: Classification results for different grips

process 30 times for each grip: each participant grasps the smartphone with the corresponding grip, puts down the smartphone, and changes the position. Then, we collect 180 sound signatures from each participant. We applied 3-fold cross validation to measure the accuracy of our system; here we use a user's data for the validation of the user only.

The accuracy according to the number of grips to be classified is shown in Figure 5(a), in which the numerical value stands for an average accuracy of the ten participants. We achieved an average accuracy of 93.0% to distinguish six grips. Also, Figure 5(b) shows a confusion matrix of the six grips of a participant whose accuracy is the closest to the average accuracy. The y-axis represents the actual grip taken by the participant, while the x-axis indicates the predicted grip of our system. The confusion matrix demonstrates that how accurately each grip are classified.

Conclusion and future work

In this paper, we presented a novel grip sensing system, which detects various hand postures on smartphones without any additional hardware and any screen touch event. Our evaluation results showed that our system classifies six different grips with 93.0% average accuracy for 10 different users. Beyond the feasibility, our future work is to improve the accuracy of the system as well as to adapt the system to various real environments.

Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2016R1D1A1B03930580) and the Ministry of Science, ICT & Future Planning (2017R1A2B2002458), and the program for fostering next-generation researchers in engineering of National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT & Future Planning (2017H1D8A2031628). This research was also supported by the MISP(Ministry of Science, ICT & Future Planning), Korea, under the National Program for Excellence in SW(2015-0-00914) supervised by the IITP(Institute for Information & communications Technology Promotion). Jinkyu Lee is the corresponding author.

REFERENCES

1. Eugene Ackerman and Fujio Oda. 1962. *Acoustic absorption coefficients of human body surfaces*. Technical Report. DTIC Document.
2. Lung-Pan Cheng, Fang-I Hsiao, Yen-Ting Liu, and Mike Y. Chen. 2013a. iRotateGrasp: automatic screen rotation based on grasp of mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 3051–3054.
3. Lung-Pan Cheng, Hsiang-Sheng Liang, Che-Yang Wu, and Mike Y Chen. 2013b. iGrasp: grasp-based adaptive keyboard for mobile devices. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 3037–3046.
4. D.J. Ewins. 2000. *Modal Testing: Theory, Practice, and Application*. Research Studies Press.
5. Mayank Goel, Alex Jansen, Travis Mandel, Shwetak N Patel, and Jacob O Wobbrock. 2013. ContextType: using hand posture information to improve mobile touch screen text entry. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2795–2798.
6. Mayank Goel, Jacob Wobbrock, and Shwetak Patel. 2012. GripSense: using built-in sensors to detect hand posture and pressure on commodity mobile phones. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*. ACM, 545–554.
7. Makoto Ono, Buntarou Shizuki, and Jiro Tanaka. 2013. Touch & activate: adding interactivity to existing objects using active acoustic sensing. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*. ACM, 31–40.
8. Yu-Chih Tung and Kang G Shin. 2015. EchoTag: accurate infrastructure-free indoor location tagging with smartphones. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. ACM, 525–536.
9. Yu-Chih Tung and Kang G Shin. 2016. Expansion of human-phone interface by sensing structure-borne sound propagation. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 277–289.