Identifying Everyday Objects with a Smartphone Knock

Taesik Gong
School of Computing
KAIST
cathena913@kaist.ac.kr

Hyunsung Cho
School of Computing
KAIST
hyunsungcho@kaist.ac.kr

Bowon Lee
Dept. of Electronic Engineering
Inha University
bowon.lee@inha.ac.kr

Sung-Ju Lee
School of Computing
KAIST
profsj@kaist.ac.kr

CHI'18 Extended Abstracts, April 21–26, 2018, Montreal, QC, Canada
ACM 978-1-4503-5621-3/18/04.
https://doi.org/10.1145/3170427.3188514

Abstract
We use smartphones and their apps for almost every daily activity. For instance, to purchase a bottle of water online, a user has to unlock the smartphone, find the right e-commerce app, search the name of the water product, and finally place an order. This procedure requires manual, often cumbersome, input of a user, but could be significantly simplified if the smartphone can identify an object and automatically process this routine. We present Knocker, an object identification technique that only uses commercial off-the-shelf smartphones. The basic idea of Knocker is to leverage a unique set of responses that occur when a user knocks on an object with a smartphone, which consist of the generated sound from the knock and the changes in accelerometer and gyroscope values. Knocker employs a machine learning classifier to identify an object from the knock responses. A user study was conducted to evaluate the feasibility of Knocker with 14 objects in both quiet and noisy environments. The result shows that Knocker identifies objects with up to 99.7% accuracy.

Author Keywords
Object identification; interaction with objects; smartphone sensing; machine learning

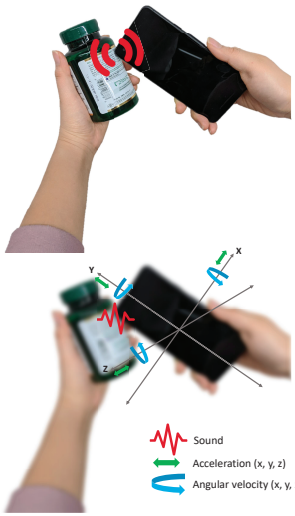


Figure 1: An example knock on a medicine bottle with a smartphone. Three types of responses can be sensed by the smartphone from the impulse.

ACM Classification Keywords

H.5.2. [Information interfaces and presentation (e.g., HCI)]: User Interface

Introduction

Smartphones play an essential role in our lives, enabling ubiquitous computing with a simple tap of the finger. Smartphones often provide us with the link between physical objects and online services. Despite the popular use of smartphones with physical objects, the process of linking physical objects to smartphone services is still cumbersome. When purchasing goods through e-commerce smartphone apps, for instance, a user has to follow a series of manual procedure, i.e., unlocking the phone, finding and launching the right app, locating the desired product inside the app, and placing an order. Had the smartphone known the object of interest and the following routine of the user's desired action involving the object, the procedure would be shortened and provide more seamless and efficient interaction between the physical objects and smartphone services.

There have been numerous approaches in object identification. Attaching tags to objects has been widely proposed, where tags are used to retrieve information of the objects. QR codes, RFID tags [8], near-field communication (NFC) [2], and acoustic barcodes [4] have been utilized to recognize and automatically select the target service from the mobile devices. These tag-based systems, however, require instrumentation of numerous objects with tags or custom sensors.

Vision-based solutions utilize computer vision and machine learning techniques to identify objects captured within the frame of the smartphone camera [1, 5]. However, vision-based systems are easily affected by the lighting condition of the environment and the misalignment of the smartphone with the target object, which reduces the usability.

The recent advance in speech recognition technologies has turned a wide array of systems into voice controllable systems (VCS) such as Apple Siri, Amazon Alexa, and Google Assistant. VCSs execute online services as commanded by the user in human language. Although the technology itself is promising, questions still remain on wider deployability due to the innate complexity of natural languages, e.g., numerous languages and dialects. Moreover, current VCSs rely on cloud services for high complexity tasks, which could cause delay in realtime interactions.

Another approach in recognizing and controlling appliances is sensing electromagnetic (EM) emissions [6, 9]. This approach exploits the uniqueness of EM signals emitted by electronic appliances for object recognition. The EM-based approach is, however, limited to electronic appliances since non-electronic objects do not emit EM signals. It also requires additional hardware to be attached to the smartphone for the EM sensing.

We argue that utilizing a knock is a viable alternative and present Knocker that identifies everyday objects when a user simply "knocks" on an object of interest with a smartphone (Figure 1). A knock generates a set of responses, i.e., the knock sound, linear acceleration, and rotational force. These are unique per object according to its characteristics, e.g., material, shape, size, etc. Knocker captures this multimodal response with a smartphone's built-in microphone, accelerometer, and gyroscope and feeds it into a support vector machine (SVM) classifier for object identification.

In contrast to previous approaches, Knocker requires no instrumentation on objects and is readily available with com-

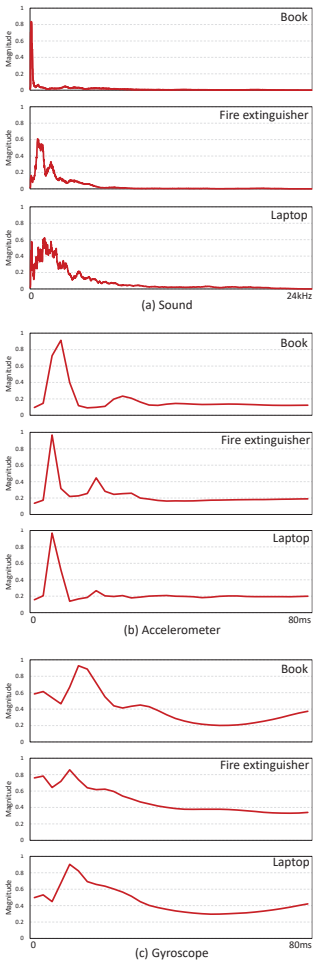


Figure 2: Example knock responses in (a) sound, (b) accelerometer, and (c) gyroscope, of three objects: a book, a fire extinguisher, and a laptop.

modity off-the-shelf smartphones. Moreover, it provides robustness against changes in lighting conditions and ambient noise. We also believe Knocker enables an efficient link between objects and desired services, with a single snap.

In the remainder, we detail the design of Knocker, present preliminary user study with 14 objects, and discuss novel applications possible with Knocker along with future directions.

Knocker

When a user knocks on a physical object with a smartphone, the knock generates a unique set of responses (according to the material, shape, size, etc). The basic concept of Knocker is analyzing the set of responses to identify each object.

The most prominent feature of the knock is the *knock sound* generated by the collision between the smartphone and the object, which can be captured by monitoring the microphone of the smartphone. The knock also exerts a force to the smartphone in proportion to the strength of the knock, as the form of linear acceleration and rotational force. Each object exhibits a different pattern of the force, and it can be captured by the rapid changes in the built-in accelerometer and gyroscope sensor values in the smartphone. In addition to the knock sound, we leverage the accelerometer and gyroscope values to identify objects, which are both distinctive per object and noise-tolerant. Figure 2 illustrates the examples of the different responses for three objects: a book, a fire extinguisher, and a laptop. The data is normalized and averaged over 50 knocks per object. Observe that each object has a distinctive set of responses in the form of sound, accelerometer and gyroscope values.

Knock Characteristics

We first analyze several characteristics of knocks with data collected in our pilot study with 14 objects (see Figure 3). We observe that when a user knocks an object with a smartphone, there is an abrupt peak in the amplitude of both sound and accelerometer. Knocker detects the knock when the peak is observed, and extracts the knock segment from the raw data.

We found that the duration of the knock responses ranges from 20 ms to 60 ms. Given 48 kHz is the common sampling rate for the built-in microphone of a smartphone and 400 Hz for both an accelerometer and a gyroscope, we use 4096 samples for the sound and 32 samples for the accelerometer and gyroscope signals. This configuration spans signals of about 85 ms for the microphone and 80 ms for accelerometer and gyroscope. This setting allows sufficient capture of the knock-related responses while minimizing computational overhead.

Classification

We apply three types of features for classification; the sound, accelerometer values, and gyroscope values acquired during a knock. We use the magnitude spectrum of the knock sound analyzed by the Fast Fourier Transform (FFT). For both the accelerometer and gyroscope data, we found that using the raw time series as feature gives better performance than the magnitude spectrum because of their relatively low sampling rate (400 Hz) and inability to capture high frequency responses.

We use a set of 2113 features as follows: the normalized magnitude spectrum of the sound from the DC to the Nyquist Frequency (2049), the normalized magnitude series of the accelerometer (32), and the gyroscope (32). We employ a sequential minimal optimization-based support vector machine (SVM) with polynomial-kernel as the classifier, pro-

Figure 3: Objects used for the experiment. The knock position and the desired condition of the object are specified.

vided by the Weka machine learning toolkit [3]. SVM is a widely used machine learning technique that constructs an optimal hyperplane for classification. We adopt SVM since it requires less training data and runtime complexity compared with deep learning techniques and outperforms other classifiers in our experiments.

User Study

We conducted an in-lab user study with 14 objects to evaluate the feasibility of Knocker. We implemented the Knocker prototype on Google Pixel 2 and collected knock data from 15 participants. The goal of this user study is to seek answers to the following questions:

- How does the accuracy change with different knock styles of different users?
- Is Knocker tolerant to a noisy environment?
- How does the accuracy differ across the 14 objects?

Participants and Procedure

We recruited 15 voluntary participants (aged 21-29, mean: 24.5; 5 females, 10 males). All are right-handed and familiar with using smartphones (usage period: 57-102 months, mean: 77.2 months). The experiment was carried out in a meeting room with all windows and doors closed. Each participant performed experiments in two settings: (i) a quiet environment and (ii) a noisy environment. For the noisy environment, we played the Billboard hot music playlist on another smartphone in full volume, which was measured 55-70 dBA at the distance of objects. Note that 60 dBA is comparable to the noise level of conversations in restaurants, and 70 dBA to that of a vacuum cleaner [7]. Half of the participants ran in a quiet setting first and then in a noisy setting, while the other half of the participants did reversely.

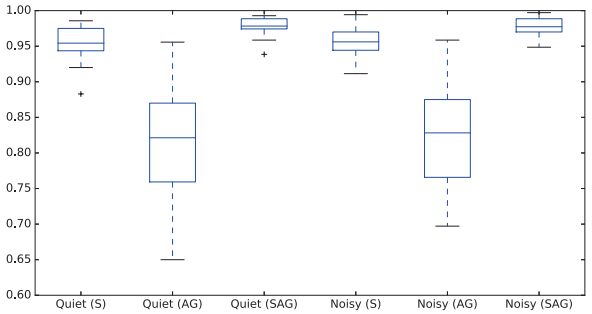
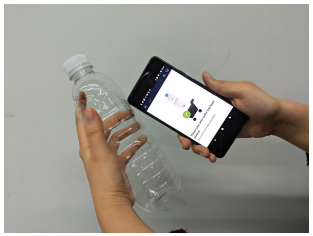


Figure 4: Overall accuracy from the user study.

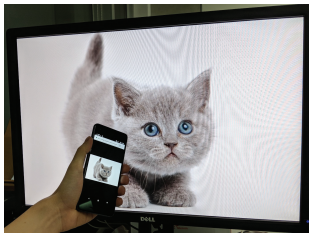
The objects used in the experiment are shown in Figure 3. Each participant was asked to knock 50 times on each object in each environment. We guided them to knock freely on any position of the objects as long as it was within the same surface of the object that we informed in advance (e.g., to knock on any “body part", not the “cap" of the water bottle). Participants were also instructed to knock on each object in its desired condition considering real use cases (e.g., beverage cans in hand, a laptop on the desk, and a guitar on the knees). We put no constraints on the strength of a knock, the hand and body postures, and the grip method. In total, 21,000 knocks (15 x 14 x 50 x 2) were collected and used for analysis.

Results and Discussion

We used 10-fold cross validation to evaluate the accuracy. Figure 4 plots the overall accuracy from the user study. In the x-axis, “Quiet” and “Noisy” represent music-off and music-on environment, respectively. “S" in the parenthesis refers to using only the sound as features, whereas “AG" refers to using the accelerometer and gyroscope. “SAG" is the accuracy when the sound, accelerometer, and gyroscope are used all together as features.



(a) Purchase water



(b) Transfer a photo

(c) Open a guitar tuner

Figure 6: Example applications of Knocker: (a) purchase water through an e-commerce app, (b) transfer a photo to a monitor, and (c) open a guitar tuner app.

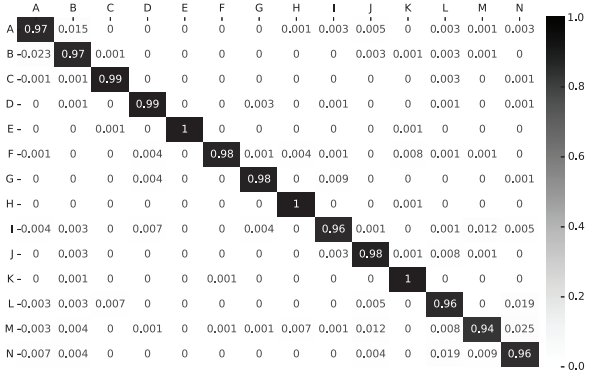


Figure 5: Confusion matrix of object classification. The ground truths are listed in the rows while the predicted classes are listed in the columns. The labels are same as in the Figure 3.

Each box plot shows a variation among users, mainly caused by different knock styles across users. It shows that accelerometer and gyroscope values are more dependent on individual knock styles than sound is. However, combining the three types of features (SAG) shows over 95% accuracy with a small variation among users, which is better than using only the knock sound or the accelerometer-gyroscope pair. Interestingly, we found that Knocker works in the noisy environment with little degradation on accuracy. We attribute this result to two reasons. First, the knock sound is dominant over the ambient noise because of the short distance from the knock spot to the built-in microphone of the smartphone. Second, music consists of a wide range of frequencies, while the knock sound has distinct peaks in certain frequencies per object.

Figure 5 shows the confusion matrix for 14 objects. The data used for this matrix is “Noisy (SAG)” with all users. A small degradation in accuracy is observed in the case between two similar objects, beverage cans (A, B) and water bottles (M, N), but the accuracy is not heavily biased towards a specific set of objects. We believe this result suggests the practical feasibility of Knocker.

Applications

With accurate object identification provided by Knocker, we envision users can perform object-specific, preregistered online services and actions by simply knocking on the object with the smartphone. Here we suggest a few examples.

Many people order goods online. Using Knocker, when a user finishes the last bottle of water in stock, for example, she can simply knock on the empty bottle to place an order of the same item in a preregistered e-commerce app (Figure 6(a)), instead of having to repeat the long tedious routine of making an online purchase every time. For electronic appliances, Knocker can facilitate the interaction between the devices by providing a one-knock photo transfer function to a monitor for a larger view (Figure 6(b)) or a document printing function for a printer, without the time and effort in connection setup. A user can easily find the instructions for use of a fire extinguisher on the smartphone with a knock, even in an emergency situation surrounded by smoke. In case of a guitar, the knock can trigger to display the guitar chord chart or open a guitar tuner app while holding the guitar on the knees (Figure 6(c)).

In addition to mapping an object to one action, Knocker has the potential for a variety set of applications. For example, in the connected world of IoT, a user can map a knock on the bedside table to a combined set of services available on the smartphone such as turning on/off the light, closing/opening the curtain, and setting the alarm on/off depending on the time of the knock. With this mapping, a user can get the services above and be ready to start the day in

the morning with one knock on the bedside table with her smartphone, in bed without even getting up.

Conclusion and Ongoing Work

We present the design of the knock-based object identification technology with a commodity off-the-shelf smartphone and the user study to evaluate its feasibility with various users and objects. Throughout this study, we show that Knocker is promising, with up to 99.7% accuracy and its potential to a wide range of applications using a simple, low latency setup. We also show the efficacy of the multimodal fusion of sound, accelerometer, and gyroscope for this task.

Our ongoing work includes implementing the introduced applications on a commodity smartphone. To investigate the availability of Knocker with diversified objects and environments, we are experimenting with a variety of scenarios, for example, a bicycle in an outdoor environment or a laptop in a cafe with babble noise. Another ongoing work is expanding the input space; one can map a different function to a different number of knocks in sequence that can be viewed as a similar concept to the single and double click of traditional mouse systems. The capability of expanding input space is the distinctive characteristic of Knocker compared with the existing methods such as tag-based systems.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No.2016R1A2B4014068).

REFERENCES

1. Adrian A. de Freitas, Michael Nebeling, Xiang 'Anthony' Chen, Junrui Yang, Akshaye Shreenithi Kirupa Karthikeyan Ranithangam, and Anind K. Dey. 2016. Snap-To-It: A User-Inspired Platform for Opportunistic Device Interactions. In *Proc. CHI '16*. ACM, 5909–5920.
2. Geven and others. 2007. Experiencing Real-world Interaction: Results from a NFC User Experience Field Trial. In *Proc. MobileHCI '07*. ACM, 234–237.
3. Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. 2009. The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter* 11, 1 (2009), 10–18.
4. Chris Harrison, Robert Xiao, and Scott Hudson. 2012. Acoustic Barcodes: Passive, Durable and Inexpensive Notched Identification Tags. In *Proc. UIST '12*. ACM, 563–568.
5. Hernisa Kacorri, Kris M. Kitani, Jeffrey P. Bigham, and Chieko Asakawa. 2017. People with Visual Impairment Training Personal Object Recognizers: Feasibility and Challenges. In *Proc. CHI '17*. ACM, 5839–5849.
6. Gierad Laput, Chouchang Yang, Robert Xiao, Alanson Sample, and Chris Harrison. 2015. EM-Sense: Touch Recognition of Uninstrumented, Electrical and Electromechanical Objects. In *Proc. UIST '15*. ACM, 157–166.
7. IAC Library. 2018. Comparitive Examples of Noise Levels. (2018). Retrieved January 6, 2018 from `http://www.industrialnoisecontrol.com/comparative-noise-examples.htm/`.
8. Roy Want and others. 1999. Bridging Physical and Virtual Worlds with Electronic Tags. In *Proc. CHI '99*. ACM, 370–377.
9. Robert Xiao, Gierad Laput, Yang Zhang, and Chris Harrison. 2017. Deus EM Machina: On-Touch Contextual Functionality for Smart IoT Appliances. In *Proc. CHI '17*. ACM, 4000–4008.