Poster: Sonicnect: Accurate hands-free gesture input system with smart acoustic sensing

Maotian CHANG

Ping LI

Panlong YANG

Jie XIONG
Singapore Management University, jxiong@smu.edu.sg

Chang TIAN

Poster – Sonicnect: Accurate Hands-Free Gesture Input System with Smart Acoustic Sensing

Maotian Zhang[1], Ping Li[1], Panlong Yang[1], Jie Xiong[2], Chang Tian[1]
[1]PLA University of Science and Technology, China
[2]Singapore Management University, Singapore

ABSTRACT

This work presents *Sonicnect*, an acoustic sensing system with smartphone that enables accurate *hands-free* gesture input. *Sonicnect* leverages the embedded microphone in the smartphone to capture the subtle audio signals generated with fingers touching on the table. It supports 9 commonly used gestures (click, flip, scroll and zoom, *etc.*) with above 92% recognition accuracy, and the minimum gesture movement could be 2cm. Distinguishable features are then extracted by exploiting spatio-temporal and frequency properties of the subtle audio signals. We conduct extensive real environment experiments to evaluate its performance. The results validate the effectiveness and robustness of *Sonicnect*.

1. SONICNECT

With the rapid increase of smart display devices (*e.g.*, smart TVs and game consoles), people need to control the intelligent services more easily, and interact with them more naturally. However, Kinect and most existing solutions on gesture recognition require additional hardware or customized devices, limiting their applicability to the wide public. Inspiringly, acoustic sensing techniques have been utilized for accurate gesture or text input and recognition. However, they may not designed for gesture input when continuous finger movements are involved, or not work in *hands-free* mode. Thus a natural question arises that, *can we design an accurate gesture input system in hands-free mode?*

In this work, we introduce *Sonicnect*, a working system for this vision, which enables accurate gesture input in *hands-free* mode with smart acoustic sensing. The microphone captures audio signals generated from the user's moving fingers, where gestures are then extracted and distinguished from the received audio signals. *Sonicnect* can support typical gestures, such as *click*, *scroll*, *flip* and *zoom*. Two challenges need to be formally addressed before realizing this system. The first challenge is that, *how to efficiently extract gesture information from subtle acoustic signals?* In dealing with this challenge, we leverage the extended amplitude spectrum density (eASD) to extract the basic features of gestures. The second challenge is *how to distinguish gesture direction accurately?* In response to this challenge, we leverage the Doppler effect to identify the *near and far* features to the microphone, such that the gesture direction can be further accurately distinguished. The design of *Sonicnect* is as follows.

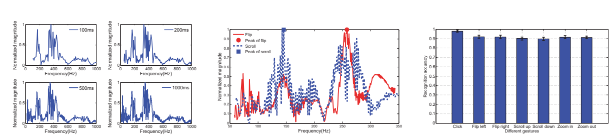
Gesture filter. The audio signals produced by touching and sliding the surfaces form a cluster of energy burst. In the beginning, the energy burst is large enough to be detected as the start of input event. *Sonicnect* leverages such unique profiles to detect the gesture input, and further segments audio signals.

Gesture recognition. We employ a two-stage approach to extract the gesture information from the recorded audio signals. At the first stage, *Sonicnect* leverages frequency information to extract the feasible feature to distinguish different gestures. We first investigate the ASDs for different time length in Figure 1a and set the audio signals length to $1000ms$ for each gesture. We plot eASD of two different gestures (flip right and scroll down) between $50Hz$ and $350Hz$, as Figure 1b shows. It is concluded that the eASD of two different gestures exhibit distinct values across frequencies, and their peaks are at different frequency.

At the second stage, we exploit the Doppler effect to identify the directions of different gestures. For example, gesture flip includes flip left and flip right, which can be categorized as far from and close to the microphone of smartphone, respectively. We conduct real environment experiments and observe that, with the increase of the distance between the sender (gesture input via the user's finger) and the receiver (the microphone), the frequency is drifting, showing the feasibility of using the Doppler effect. *Sonicnect* utilizes this observation to identify the direction of a gesture.

Recognition accuracy of *Sonicnect*. In a default setting scenario, a HUAWEI U9508 smartphone is placed on the top of a coated-wood table in an office room, and the user inputs the gestures near to it. The microphone at the bottom of the smartphone captures and buffers the audio signals which will be processed by the proposed approach.

Figure 1c illustrates the accuracy of each gesture achieved by *Sonicnect* (Error bars show standard deviation across 10 experimental runs). The overall recognition accuracy is 92.1%. The accuracy of gesture click, flip, scroll, and zoom are about 98%, 92%, 90% and 92%, respectively.



(a) (b) (c)

Figure 1: Frequency domain features of gestures ((a): ASD for different time segment, (b): eASD of two different gestures.), and gesture recognition accuracy of *Sonicnect* (c).

MobiSys'16 Companion June 25-30, 2016, Singapore, Singapore

ACM ISBN 978-1-4503-4416-6/16/06.
DOI: http://dx.doi.org/10.1145/2938559.2948830