Demo: Device-Free Gesture Tracking Using Acoustic Signals

Wei Wang[†], Alex X. Liu[†‡], Ke Sun[†]
[†]*State Key Laboratory for Novel Software Technology, Nanjing University, China*
[‡]*Dept. of Computer Science and Engineering, Michigan State University, USA*
ww@nju.edu.cn, alexliu@cse.msu.edu, samsonsunke@gmail.com

Abstract

In this demo, we present LLAP, a hand tracking system that uses ultrasound to localize the hand of the user to enable device-free gesture inputs. LLAP utilizes speakers and microphones on Commercial-Off-The-Shelf (COTS) mobile devices to play and record sound waves that are inaudible to humans. By measuring the phase of the sound signal reflected by the hands or fingers of the user, we can accurately measure the gesture movements. With a single pair of speaker/microphone, LLAP can track hand movement with accuracy of 3.5 mm. For devices with two microphones, LLAP enables drawing-in-the air capability with tracking accuracy of 4.6 mm. Moreover, the latency for LLAP is smaller than 15 ms for both the Android and the iOS platforms so that LLAP can be used for real-time applications.

1. INTRODUCTION

As the size of mobile devices shrinks, it becomes harder for users to interact with small wearable devices, such as smart watches and smart glasses. Device-free gesture tracking allows users to interact with the device by moving their hands and fingers in the air. Therefore, device-free tracking technology removes the restriction on the input area and provides a user-friendly interaction mechanism for small mobile devices. Moreover, device-free tracking can also be a complementary input method for mobile phones, because it allows the user to interact with the device without blocking the screen.

In this demonstration, we present our acoustic signal based tracking system that uses the Low-Latency Acoustic Phase (LLAP) tracking algorithm in [1]. LLAP measures the phase and amplitude distortion in the sound wave reflected by the hand to achieve highly accurate gesture tracking. The key advantage of LLAP is that it reuses speakers and microphones that are already on the mobile devices to act as the gesture sensor. Therefore, LLAP can be directly deployed on existing mobile phones as software applications without incurring any hardware cost, while systems using the specialized 60 GHz radar chips need hardware upgrades [2, 3].

Compared to existing sound based ranging systems that use the Time-Of-Arrival/Time-Difference-Of-Arrival (TOA/TDOA) measurements [4, 5] or the Doppler shift measurements [6], LLAP has the following advantages. First, LLAP uses Continuous Wave (CW) signals in the frequency range of 17∼23 kHz that are inaudible to human. Existing ranging systems often periodically emit Chirp or impulse signals that generate audible sounds on COTS speakers [4]. Second, LLAP provides high tracking accuracy of several millimeters. Our system uses novel coherent detection algorithms to measure the phase of the reflected ultrasound. The phase of reflected signal changes by 2π when the object moves by a distance of half of the sound wavelength, which is around 2 cm in our system. Thus, phase resolution of $\pi/4$ converts to distance resolution of 1.25 mm and the tracking error of LLAP is merely 3.5 mm on average in our experiments. Third, LLAP can response to users with latency smaller than 15 ms. Existing Doppler based systems have to accumulate a data segment of long enough size, *e.g.*, 2048 audio samples (translated to 42.27 ms), to perform Fast Fourier Transform (FFT). This requirement greatly limits the responsiveness of such systems. In contrast, LLAP extracts the phase change from data segments as small as 16 samples, and can respond within 15 ms on COTS mobile devices. Therefore, LLAP can catch up with refresh rate of 60 frames per second that is commonly required by gaming applications. Fourth, LLAP has low computational requirements and can be easily implemented on COTS mobile phones. On our Android implementation, a Samsung Galaxy S5 only takes 4.32 ms to process a data segment with length of 512 samples (translated to 10.7 ms). Therefore, commercial devices have enough computational power to handle the real-time processing for LLAP. For the iOS platform, we can further reduce the computational cost by using the Digital Signal Process (DSP) interface. On an iPhone 6s, it only takes 0.05 ms to process data segment of 10.7 ms, which incurs CPU cost of less than 3% and "low" energy impact as indicated by Xcode debugging tools.

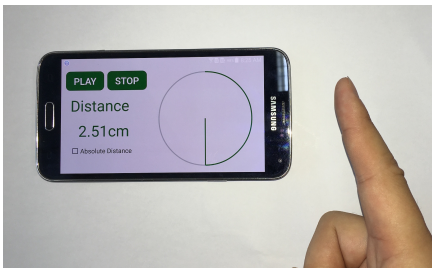

Figure 1: Demonstration using Android platform

Our demonstration will show both the one-dimensional tracking and two-dimensional drawing-in-the air applications. For one-dimensional tracking, LLAP achieves 3.5 mm tracking error and less than 15 ms latency. Figure 1 shows an example of demonstration scenario, which shows how our system can accurately detect small movement of a single finger. For two-dimensional, LLAP allows the user to draw characters and words in the air and achieves word recognition accuracy of more than 91%.

MobiCom'16 October 03-07, 2016, New York City, NY, USA

ACM ISBN 978-1-4503-4226-1/16/10.
DOI: http://dx.doi.org/10.1145/2973750.2987385

2. GESTURE TRACKING SYSTEM

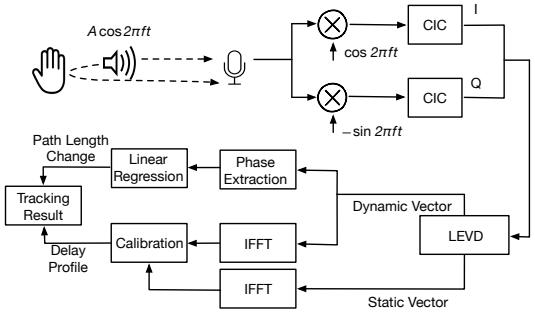

Figure 2: System Design

Figure 2 shows the overall system design of LLAP. LLAP first transmits CW signals in the form of $\cos(2\pi f_i t)$ at different frequencies of f_i from the speaker of the mobile device. We use 16 different frequencies with frequency interval of 350 Hz in the range of 17$\sim$ 23 kHz. The sampling rate of the audio signal is 48 kHz. We record the sound signal at the same time using the microphones of the same device on the same sampling rate of 48 kHz. When the human hand moves, the sound reflected by the hand will be distorted in phase and amplitude. LLAP uses the amplitude and phase distortions caused by the moving hand to perform gesture tracking.

As the first step, we use a digital down converter to convert the recorded audio into a complex baseband signal. The complex baseband signal of the recorded sound contains both a static vector and a dynamic vector as shown in [7]. The static vector represents the sound signal travelled through the Line-Of-Sight (LOS) path and reflected by static objects, such as walls and tables. As these objects do not move, the static vector only varies slowly with the time. The dynamic vector represents the sound signal reflected by the moving hand. When the hand moves, the phase of the dynamic vector changes according to the hand movement distance. Every time the hand moves by half of the sound wavelength, the phase of the dynamic vector changes by 2π. Thus, we can calculate the hand movement distance by measuring the phase of the dynamic vector.

To measure the phase of the dynamic vector, we design a novel algorithm called Local Extreme Value Detection (LEVD) to estimate the static vector form the mixed baseband signal. We then subtract the static vector from the baseband signal to obtain the dynamic vector. After getting the dynamic vector on different sound frequencies, we use the phase change of the dynamic vectors to estimate the movement distance. The movement distance at a single sound frequency can be calculated using the phase change and the sound wavelength at the given frequency. However, such single frequency estimations are susceptible to multipath interferences. To mitigate the multipath effect, we use a linear regression model to combine distance estimations at multiple sound frequencies.

The movement distance obtained from the previous algorithms are relative movement distances, but not the absolute hand locations. To obtain the absolute hand locations, we use the phase/amplitude change of the baseband signal over different frequencies to obtain a delay profile of the reflected sound. We apply Inverse FFT (IFFT) on the dynamic vector of 16 different frequencies to obtain the delay profile, which gives the path length estimation with accuracy of 4 cm in our system.

LLAP combines the fine-grained phase based movement distance measurement (with accuracy of 3.5 mm) and the coarse-grained delay profile based absolute path length estimation (with accuracy of 4 cm) to achieve highly accurate gesture tracking. The coarse-grained path length estimation provides initial position estimation for the tracking system and prevents error accumulation during tracking. After combined the distance measurements, LLAP obtains the location of the hand using triangulation. As many mobile phones have multiple microphones, by measuring the distance between the hand to two different microphones, we are able to locate the hand in a two-dimensional space. In this way, LLAP allows the user to draw in the air use a single mobile phone.

We have implemented LLAP on both the Android and the iOS platforms. For the Android platform, we implement signal processing algorithms as C functions using Android NDK. The rest of the application is implemented in Java. We provide both one-dimensional demo and two-dimensional demon on Android platforms using Samsung Galaxy S5. The one-dimensional demo shows how a user can control a simple UI widget, such as a scroll bar, using hands and fingers. The two-dimensional demo shows the drawing-in-the air capability. We will demonstrate how users can draw words and characters using LLAP. For iOS platform, we implement the system in Objective C using the vDSP interfaces. Unfortunately, iOS only provides mono recording capability. Therefore, we can only demonstrate the one-dimensional tracking capability on iOS. We have implemented a simple shooting game using LLAP, which can be installed on recent iPhones. We encourage users to install the application on their phone and get a better understanding of how device-free gesture tracking performs in various environments.

3. DEMO DESCRIPTION

In this demo, we will show the LLAP application on both the Android and iOS platforms. The attendance will get hands-on experience of our LLAP system on smart phones, such as Samsung Galaxy S5 and iPhone 6s. For attendance with recent models of iPhones, *e.g.*, iPhone 6/6s, we can provide iOS programs to be installed on their own mobile phones.

Equipment: Mobile phones, including Samsung Galaxy S5 and iPhone 6s, will be provided by us. As we use ultrasounds, we wish that there are no interferences in the ultrasound band, *i.e.*, sounds in 17$\sim$23 kHz frequency range. The demo can be setup within 10 minutes.

4. CONCLUSION

We will demonstrate the LLAP system, which brings high accuracy and low-latency tracking capability to COTS mobile devices. We hope that we can share our experiences on this system with the research community and inspire new types of device-free gesture input in the near future.

5. REFERENCES

[1] Wei Wang, Alex X. Liu, and Ke Sun. Device-free gesture tracking using acoustic signals. In *Proceedings of ACM MobiCom*, 2016.

[2] Jaime Lien, Nicholas Gillian, M Emre Karagozler, Patrick Amihood, Carsten Schwesig, Erik Olson, Hakim Raja, and Ivan Poupyrev. Soli: ubiquitous gesture sensing with millimeter wave radar. *ACM Transactions on Graphics*, 35(4):142, 2016.

[3] Teng Wei and Xinyu Zhang. mTrack: High-precision passive tracking using millimeter wave radios. In *Proceedings of ACM MobiCom*, 2015.

[4] Chunyi Peng, Guobin Shen, Yongguang Zhang, Yanlin Li, and Kun Tan. Beepbeep: a high accuracy acoustic ranging system using COTS mobile devices. In *Proceedings of ACM SenSys*, 2007.

[5] Rajalakshmi Nandakumar, Vikram Iyer, Desney Tan, and Shyamnath Gollakota. FingerIO: Using active sonar for fine-grained finger tracking. In *Proceedings of ACM CHI*, 2016.

[6] Sangki Yun, Yi-Chao Chen, and Lili Qiu. Turning a mobile device into a mouse in the air. In *Proceedings of ACM MobiSys*, 2015.

[7] Wei Wang, Alex X. Liu, Muhammad Shahzad, Kang Ling, and Sanglu Lu. Understanding and modeling of WiFi signal based human activity recognition. In *Proceedings of ACM MobiCom*, 2015.