FingOrbits: Interaction with Wearables Using Synchronized Thumb Movements

Cheng Zhang, Xiaoxuan Wang, Anandghan Waghmare, Sumeet Jain, Thomas Ploetz
Omer T. Inan, Thad E. Starner, Gregory D. Abowd
chengzhang, wangxx, anandghan, sumeet, thomas.ploetz, inan, thad, abowd@gatech.edu
Georgia Institute of Technology

ABSTRACT
We present FingOrbits, a wearable interaction technique using synchronized thumb movements. A thumb-mounted ring with an inertial measurement unit and a contact microphone are used to capture thumb movements when rubbing against the other fingers. Spectral information of the movements are extracted and fed into a classification backend that facilitates gesture discrimination. FingOrbits enables up to 12 different gestures through detecting three rates of movement against each of the four fingers. Through a user study with 10 participants (7 novices, 3 experts), we demonstrate that FingOrbits can distinguish up to 12 thumb gestures with an accuracy of 89% to 99% rendering the approach applicable for practical applications.

Author Keywords
Wearables, Interaction, Gesture recognition, Ring

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation : User interfaces; Input devices and strategies

INTRODUCTION
Designing input capabilities for wearables is more challenging than for traditional computing devices like laptops because of the intentional smaller form factor. The substantially reduced interaction space limits the complexity and functionality of interactions. Furthermore, wearables rely mostly on touch-based input and a wearable interaction is typically done while the user is on the go. Consequently, wearable interaction techniques require minimal effort, which ideally translates to hands-free or at least one-handed input that is quickly accessible and socially acceptable.

In this paper, we present FingOrbits, a concept for thumb based interaction with wearables such as heads up displays or smart watches. Users wear a thumb ring that contains a contact microphone that communicates with the wearable. Through rubbing a thumb against the fingers of the same hand users perform specific input gestures. These gestures require little to no user training. Utilizing a signal processing and classification framework FingOrbits is able to recognize up to 12 different one-handed input gestures that are defined through three different movement ("rubbing") patterns that can be executed at each of the four fingers.

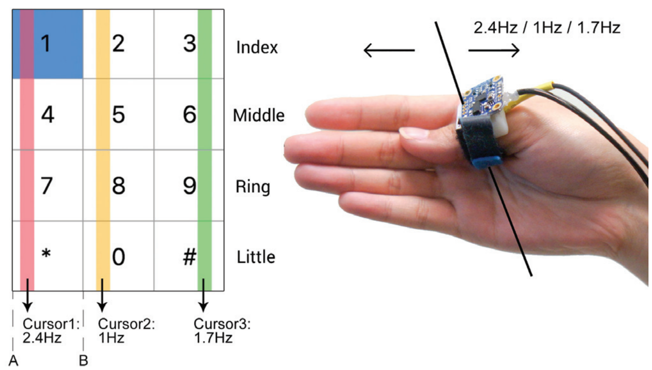


Figure 1. The FingOrbits system and experimental setting.

Our work is inspired by the recently presented Orbits system [4], which allows users to interact with displays through synchronizing their eye movements (captured using eye trackers) with moving dots on the screen. We adapt the concept of synchronized movements for one-handed input: *i)* Instead of an eye tracking system only a simple thumb ring with integrated contact microphone is required; *ii)* No exact trajectory match is required but rather only a frequency, which provides for a more flexible interaction. We evaluate our prototype system in a user study and demonstrate its potential.

RELATED WORK
The motivation for wearables comes from the desire to interact with computing devices at any time and in (almost) any environment. Such scenarios are very challenging as in many situations a user's hand(s) might be occupied, which renders interactions with the wearable device difficult. Combined with the small interaction space on wearables the overall range of interactions is rather constrained. In order to tackle such challenges a few interaction techniques have been proposed that facilitate single handed input for wearables.

Cameras have been used extensively to track user input through hand gestures. For example, the Digits system [6] tracks hand poses using a wrist mounted camera. PinchWatch

ISWC'17, September 11-15, 2017, Maui, HI, USA

ACM ISBN 978-1-4503-5188-1/17/09...$15.00
https://doi.org/10.1145/3123021.3123041

[7] operates similarly through utilizing a chest worn camera, while CyclopsRing [1] uses a thumb worn camera. Alternatively, inertial measurement units (IMUs)[8], magnetic sensing[2, 5] and acoustic sensing[10] have been used recognzing finger position or gestures.

We explore Orbits-like input mechanisms which involves following a moving element and syncing to one of the elements for making a selection. Orbits have previously been explored for eye tracking based interactions [4, 3]. We use similar hardware as has been employed for FingerSound [9], which was used for the recognition of Graffiti like input gestures.

SYSTEM DESCRIPTION

The FingOrbit system consists of a ring worn on the thumb (details below) and an accompanying visual interface which guides users to perform gestures. The visual interface is a standard number pad consisting of digits from $0 - 9$ and the symbols * and # presented in the tabular format of four rows and three columns. Each row starting from row 1 (topmost) to row 4 (bottommost) is mapped to the index, middle, ring, and little finger, respectively. Each column has a horizontally moving cursor which moves within the bounds of its column (Figure 1 – Left). The moving cursor frequencies are 2.4Hz, 1.0Hz, and 1.7Hz (from left to right). For instance, the leftmost cursor moves with a frequency of 2.4Hz in the first column between points A and B repeatedly as shown in figure 1. During testing, the gesture to be performed is marked by a highlighted cell. For instance, (Figure 1) depicts cell 1 (row 1, col 1) as highlighted. The study participant is expected to rub the against index finger (row 1) at a frequency of 2.4Hz matching the cursor frequency in column 1.

Hardware

To capture thumb movements we built a 3D printed ring (Figure 1 – Right) that houses a contact microphone (Knowles BU-21771) and an IMU (Bosch BNO055) that are connected to a pre-amplifier and Teensy 3.2 board, respectively. An USB audio board is used to connect the sensing platform to a laptop via USB. The focus of our current prototype is to show feasibility and thus all data processing is performed on the wired laptop. Future iterations of our system will focus on more autonomous operation as well as more comfortable to wear hardware. The microphone's recording rate was set to 44.1kHz whereas the IMU sampled at 200Hz.

Data Processing

Figure 2 gives an overview of the data processing pipeline that is used for FingOrbits. To detect the start and end of an intended gesture, we slide a $100ms$ window with no overlap on the received sound data. For each sliding window, we first mask silent portions by applying an empirically determined threshold. We then calculate the average energy of the audio signal for every frame extracted by the sliding window process. Once the maximum energy of a frame exceeds an empirically determined activation threshold, we mark this frame as the start of a gesture. If the maximum energy for any subsequent, five consecutive frames (i.e., for at least 0.5 seconds) falls below the threshold, we end the gesture.

If a gesture segment is longer than 1.3s, then we examine the dominant frequency (band) of the gyroscope data (i.e., the frequency with highest power in the whole frequency spectrogram). To estimate this frequency, we first find the axis of gyroscope data with highest energy and then calculate the FFT for the data of this axis.

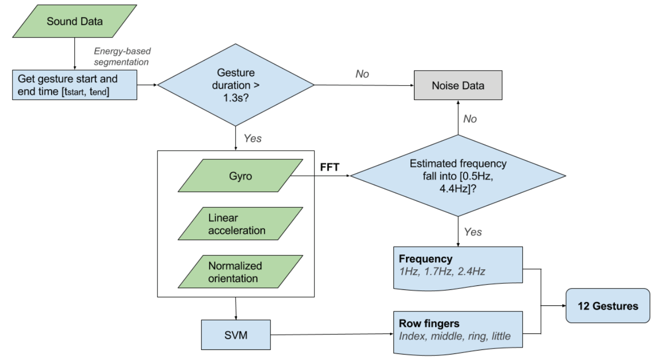


Figure 2. Flow chart of FingOrbit's data processing procedure.

If the estimated dominant falls in the range $[0.5\text{Hz}–4.4\text{Hz}]$, we label the extracted segment as a gesture and proceed to the next step. Otherwise, we mark it as noise do not advance. The range used was determined empirically with the prototype system. Once a gesture is confirmed, we use the following steps to estimate the frequency of the thumb movement and identify on which finger the thumb is moving. First, we compare the dominant frequency of the gyroscope data to a set of preset frequencies (1Hz, 1.7Hz, 2.4Hz) and find the closest match. For instance, if the estimated dominant frequency is 1.2Hz, it would suggest that the user is performing a gesture to match the frequency of 1Hz. These preset frequencies are chosen based on the results of a formative preliminary study that aimed at identifying thumb moving frequencies that are both comfortable to perform by users and discriminative w.r.t. three different states (slow, medium, fast).

To recognize which finger the thumb is rubbing against, we use gyroscope, linear acceleration, and orientation data from the IMU. To remove the influence of body orientation on the orientation data, we normalize the raw orientation data by subtracting the mean value on each axis before further processing. For each sensor, we extract statistical features for each axis: minimum, maximum, mean, energy, variance, standard deviation, zero crossing rate, and entropy. To represent the relationship between axes, we also extract the Pearson-correlation and energy ratio between each pair of the three axes on each sensor. In total we extract 90-dimensional feature vectors (per extracted gesture segment) and feed it to a support vector machine classification back-end to determine the finger on which the gesture was performed.

EVALUATION

In order to explore the practicalities of our FingOrbits system we conducted an experimental evaluation. We asked a total of ten participants (five males, average age: 27.8 ± 2.39, seven novices, three experienced users) to use the system and to perform thumb gestures. The general task was to match frequencies of movement patterns that were displayed on the laptop screen through rubbing the instrumented thumb against one of the other fingers (according to protocol; explained below).

Our study was performed in a controlled setting where participants faced a computer screen while resting their arms on the table. The system provided visual and auditory cues to assist the user with the study. During the study, randomly individual cells of the displayed grid were highlighted. Participants then had to match their thumb's rubbing frequency with the corresponding cursor frequency of the cell. Figure 1 illustrates the experimental apparatus.

Each participant finished three sessions – practice, training, and testing. Per practice session three repetitions of each gesture / cell were asked to be performed, whereas five iterations per gesture / cell were performed for both training and testing. The gesture to perform was selected randomly using a balanced distribution. Before the practice session started, a researcher explained the functionality of the system and demonstrated how to perform FingOrbits gestures. Participants were guided to perform the rubbing gestures until they saw a match between the frequency of the thumb movement and the corresponding cursor. Gestures recorded during the training sessions were used for training the analysis system (classification back-end). In both the practice and training sessions, real-time feedback of the recognized column was provided to the user by highlighting the recognized column on the screen . The three experienced users skipped the practice session and only performed the last two sessions. Both the real-time classification results and the segmented raw sensor data were recorded for further analysis.

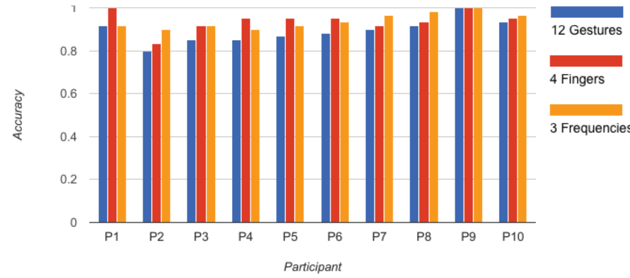


Figure 3. Accuracy for each participant

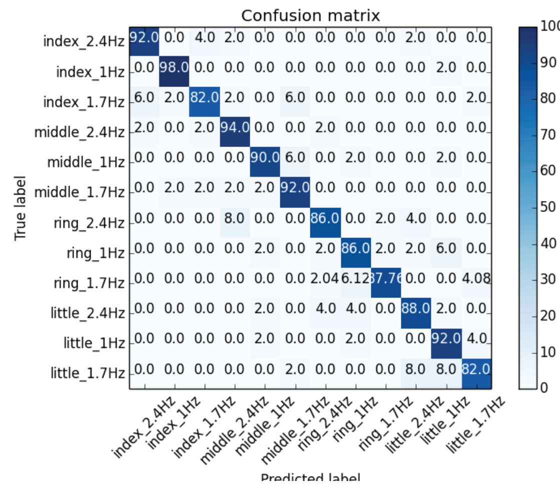


Figure 4. Confusion matrix for all 12 gestures

Results

The overall accuracy across all ten participants is 89% for recognizing all 12 gestures, where P8 to P10 are the expert users. Figure 3 presents the accuracy for each participant in different classification tasks. In the task of recognizing 12 gestures, P1 and P9 presents the highest accuracies of 91.7% and 100% , while P2 and P8 have the lowest accuracies of 80% and 91.5% across novice and expert users. Disaggregation of the recognition results gives insights into the performance of the two classification steps: *i)* which finger is the thumb rubbing against; and *ii)* which gesture (frequency) is the participant performing. On average, discriminating the four fingers the thumb rubs against works with 94% accuracy. The subsequent step of discriminating the three possible rubbing frequencies works with 94% accuracy.

As the confusion matrix illustrates (Figure 4), the gesture that has been recognized with lowest accuracy (82%) corresponds to rubbing the thumb against the index finger with the middle frequency (1.7Hz). The lowest accuracy (89.95%) for classifying three frequencies across 4 fingers is 1.7Hz as shown in Figure 5a. The ring finger was the most confused finger as shown in Figure 5b. In fact, most participants perceived rubbing their thumb against their ring and pinky fingers as rather uncomfortable, which explains the poor recognition performance and perhaps a greater variability of the performed gestures.

DISCUSSION

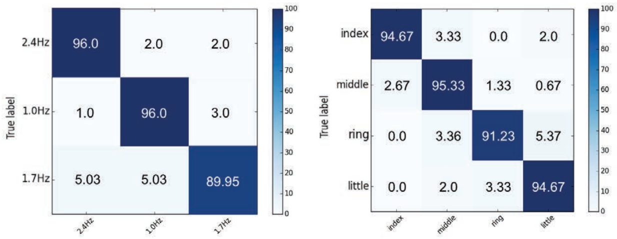
In this paper we presented FingOrbits, a concept for interacting with wearables through synchronized thumb movements. The developed concept and system adopts the Orbits idea of synchronizing user input with predefined motion patterns that serve as stimuli and as interaction triggers [4]. Unlike the original, eye tracking based Orbits system our approach tracks thumb movements using an instrumented ring. FingOrbits users can perform up to 12 different gestures through rubbing their thumb against other fingers using different movement frequencies. Our system provides high classification accuracies.

Design implications and Applications

Applications on wearables may not require all 12 input gestures that FingOrbits can recognize. Reasons for a limited recognition lexicon could be usability (e.g., rubbing the thumb against the pinkie is not very comfortable for many) or simplicity that leads to lower cognitive load during interacting. In what follows we will discuss exemplary applications of FingOrbits. We will provide recommendations for gesture sets to be used and –in light of this– link back to the results of our user study.

Number Input

Current interfaces for entering numbers (digits) into heads-up devices such as Google Glass require the use of a paired device (e.g., a smartphone), which slightly undermines the general idea of wearables. FingOrbits could be used to facilitate numerical input through utilizing 10 rubbing gestures while the glasses display the moving cursor. Based on our current apparatus it would be advisable to represent the digits 1–9 using the upper three fingers and the three different frequencies



a. Classification results of three frequencies b. Classification results of four fingers

Figure 5. Results for discriminating frequencies and fingers separately.

each, and to map the digit 0 to 1Hz movement on the pinkie. Such a setting would result in over 90% recognition accuracy.

Navigation, Shortcuts, and Music Player

Having four gestures can also be useful for a variety of applications. For instance, they can be used for shortcuts (matching each gesture to an application), navigation (using four directional slides), and controlling a music player (next song, previous song, pause, play). According to our user study using the index and middle fingers as interaction targets for FingOrbits results in high classification accuracy and maximal comfort. As such, the combination with the lowest and the highest frequencies would provide a reliably recognizable gesture set with average classification accuracies of the four functions to be performed at about 95%.

Quick response to notification

Many applications only require binary input functionality such as accepting or declining phone calls, or turning on/ off a notification. For such scenarios, FingOrbits should be used with two fingers (index and middle) and a different rubbing frequency for each finger (1Hz and 2.4Hz). According to our user study, such a configuration would lead to almost ideal classification results (99%) without collecting any training data from the user.

Eyes-free interaction

To investigate how FingOrbits works when the gestures are performed in an eyes-free fashion, we conducted an additional study with our three experienced users. We followed the same procedure as for the previous experiment, except that the arm of the participant was laying down naturally under the table (in the lap), and we did not provide the moving cursor from the testing session. All three participants performed eyes-free FingOrbits gestures for the first time.

The overall accuracy was 84.4%, i.e., a moderate drop of about 10% accuracy compared to the results of the first study. The accuracy of classifying the gestures with rubbing frequencies of 2.4Hz, 1Hz, 1.7Hz and were 96.7%, 98.3%, and 83.3% respectively. Consequently, if FingOrbits is used in an eyes-free fashion without displaying the moving cursor, one should stick to only two, rather distinct rubbing frequencies. A further analysis revealed that the accuracy of recognizing two frequencies on just the index finger and recognizing two frequencies on both index and middle finger are 100% and 91.7%, respectively. These preliminary results illustrate that FingOrbits has potential for real-world applications.

User Experience

During the study, we noticed FingOrbits gestures were not immediately intuitive to some participants in the practice session. Some users mistakenly thought they had to match the exact moving trajectory of the moving cursor, which was especially challenging for the highest speed. Some users were confused as to when a gesture was complete. These issues disappeared with training.

Some everyday activities, such as clapping, can produce false positives for FingOrbits. One potential solution is to adopt an activation gesture to start the whole recognition system. Sweeping the thumb on the index finger at 1Hz is a good candidate, based on its high recognition rate (98%) for this gesture across all 10 participants). However, we do not yest have empirical evidence of the false positive rate for that gesture.

REFERENCES

1. Chan, L., Chen, Y.-L., Hsieh, C.-H., Liang, R.-H., and Chen, B.-Y. Cyclopsring: Enabling whole-hand and context-aware interactions through a fisheye ring. In *Proc. UIST* (2015).
2. Chen, K.-Y., Lyons, K., White, S., and Patel, S. utrack: 3d input using two magnetic sensors. In *Proc. UIST* (2013).
3. Dhuliawala, M., Lee, J., Shimizu, J., Bulling, A., Kunze, K., Starner, T., and Woo, W. Smooth eye movement interaction using eog glasses. In *Proc. ICMI* (2016).
4. Esteves, A., Velloso, E., Bulling, A., and Gellersen, H. Orbits: Gaze interaction for smart watches using smooth pursuit eye movements. In *Proc. UIST* (2015).
5. Huang, D.-Y., Chan, L., Yang, S., Wang, F., Liang, R.-H., Yang, D.-N., Hung, Y.-P., and Chen, B.-Y. Digitspace: Designing thumb-to-fingers touch interfaces for one-handed and eyes-free interactions. In *Proc. CHI* (2016).
6. Kim, D., Hilliges, O., Izadi, S., Butler, A. D., Chen, J., Oikonomidis, I., and Olivier, P. Digits: freehand 3d interactions anywhere using a wrist-worn gloveless sensor. In *Proc. UIST* (2012).
7. Loclair, C., Gustafson, S., and Baudisch, P. Pinchwatch: a wearable device for one-handed microinteractions. In *Proc. MobileHCI* (2010).
8. Tsai, H.-R., Wu, C.-Y., Huang, L.-T., and Hung, Y.-P. Thumbring: private interactions using one-handed thumb motion input on finger segments. In *Adjunct Proc. MobileHCI* (2016).
9. Zhang, C., Waghmare, A., Kundra, P., Pu, Y., Gilliland, S., Ploetz, T., Starner, T., Inan, O., and Abowd, G. Fingersound: Recognizing unistroke thumb gestures using a ring. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* (September, 2017).
10. Zhang, C., Xue, Q., Waghmare, A., Jain, S., Pu, Y., Hersek, S., Lyons, K., Cunefare, K. A., Inan, O. T., and Abowd, G. D. Soundtrak: Continuous 3d tracking of a finger using active acoustics. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 1*, 2 (June 2017), 30:1–30:25.