SeeSaw - Rapid One-Handed Synchronous Gesture Interface for Smartwatches

Jason Wu, Cooper Colglazier, Adhithya Ravishankar, Yuyan Duan, Yuanbo Wang, Thomas Ploetz, Thad Starner
{jasonwu, ccolglazier3, adhiravishankar, yduan63, ywang936, thomas.ploetz, thad}@gatech.edu
School of Interactive Computing, College of Computing, Georgia Institute of Technology

ABSTRACT
We present SeeSaw, a synchronous gesture interface for commodity smartwatches to support watch-hand only input with no additional hardware. Our algorithm, which uses correlation to determine whether the user is rotating their wrist in synchrony with a tactile and visual prompt, minimizes false-trigger events while maintaining fast input during situational impairments. Results from a 12 person evaluation of the system, used to respond to notifications on the watch during walking and simulated driving, show interaction speeds of 4.0 s - 5.5 s, which is comparable to the swipe-based interface control condition. SeeSaw is also evaluated as an input interface for watches used in conjunction with a head-worn display. A six subject study showed a 95% success rate in dismissing notifications and a 3.57 s mean dismissal time.

ACM Classification Keywords
H.5.2. [Information Interfaces and Presentation]: User Interfaces — Input Devices and Strategies

Author Keywords
wearable computing; interaction; synchronous gestures; smartwatches; head-worn displays; one-handed input

INTRODUCTION
Wearable devices such as smartwatches and head worn displays (HWDs) provide convenient, readily available access to content. An obstacle to widespread adoption is the lack of appropriate on-the-go input techniques. Just as the keyboard and mouse are not suitable for smartphones, existing technologies such as the touchscreen and voice-based commands are not ideal for many wearable devices. Touch interfaces are hindered by finger occlusion on small form-factor devices, especially smartwatches [7]. Voice interactions can be effective but are not discreet. Interfaces for wearable devices should enable fast access "microinteractions" [1] while maintaining acceptable levels of privacy and expressiveness.

ISWC '18, October 8–12, 2018, Singapore, Singapore

ISBN 978-1-4503-5967-2/18/10. . . $15.00
DOI: https://doi.org/10.1145/3267242.3267251

Synchronous interfaces address some of the shortcomings of traditional gesture-based interactions, including their learnability and reliability. Instead of expressing user intent by performing multiple discrete gestures, a single motion is performed in synchrony with a visual or haptic target stimulus. Thus, minimal user learning is required [15]. This concept has been explored for smartwatch input and multi-user systems, using eye-gaze tracking, camera-based hand motion tracking, and magnetic ring motion correlation [3, 7, 10, 15]. However, these systems currently require additional external hardware, making it unsuitable for use with many smartwatches.

SeeSaw is a synchronous gesture interface that uses only a smartwatch's gyroscope. Gestures can be performed with the user's wrist and hand and require no additional hardware or software modification. The system is also designed to facilitate subtle interaction, minimizing attention from the user when interacting with the wearable computer. Alerts can be dismissed one-handed with speeds of 4 s - 5.5 s for smartwatch-only interactions and 3.6 s when a HWD is used for display. SeeSaw enables effective interaction with wearable computers in many common use cases where the watch hand may be occupied and provides a compelling alternate input modality to augment traditional touch interfaces.

RELATED WORK
Many wearable computers intend to facilitate quick microinteractions between the user and the system [1, 2]; yet the speech-to-text and touch-screen input systems for smartwatches and head-mounted computers pose challenges in social acceptability or in requiring two-handed input, respectively. Research efforts explore alternatives to address these challenges.

One-handed Input
One-handed input is preferable in many everyday use scenarios where either the user's hands are occupied, or a more discreet mode of interaction is needed. Serendipity is an example of such a finger gesture recognition system capable of recognizing five fine-motor gestures [14]. To expand the capabilities of gesture recognition systems, additional software and hardware modifications are often introduced to allow for a larger gesture set and higher detection accuracy. The ViBand input system uses a custom smartwatch software kernel to allow for increased sampling rate from sensors, which enables more accurate and expressive user input and the ability to sense external objects through touch [9]. Hardware modifications are also possible, as demonstrated by numerous projects such

as WristWhirl, WristFlex and Tomo, that allow for a larger gesture set and continuous input [5, 8, 16].

Synchronous Gesture Interfaces

Synchronous gestures are similar to rhythmic patterns in that they both allow the user to express intent over time, but in the case of synchronous gestures, the stimulus is presented to the user indicating the expected gesture or pattern [11]. Motion correlation has been implemented successfully in many camera-based systems, often allowing for robust, multi-user selection on large displays [3, 4, 11]. Synchronous gestures have also been explored using smooth pursuit tracking, as the eyes are able to closely follow the motion of the target. Previous work explored gaze-tracking using external trackers [7, 12, 13] and EOG glasses [6]. However, in all cases, gaze-tracking requires the use of specialized eye-tracking hardware and may interfere with the user's ability to comfortably view on-screen content. FingOrbits seeks to replace pursuit tracking with finger movement by using a specially-designed thumb ring [15]. The FingOrbits system is implemented using an IMU and contact microphone connected to a laptop running a FFT-based detection algorithm. Recent work by Reyes et al. has investigated removing additional external powered hardware for synchronous gesture interfaces by constructing a thumb ring with a passive rare-earth magnet, showing viable accuracy and speed for notification response applications on smartwatches [10]. While all of these input interactions show the effectiveness and advantages of synchronous gesture interfaces, they all require external hardware and can not be used for out-of-the-box operation with commodity smartwatches.

INTERFACE DESIGN

Gesture Interaction

The SeeSaw input interaction is a synchronous gesture interface that allows users to perform input through synchronous wrist rotations. The primary input interaction used by SeeSaw is a repetitive twisting of the wrist. When a stimulus signal is provided, it is rendered as a flashing target on the screen or as a haptic vibration. Syncing with a target is achieved by tilting the wrist away from the body when the target is illuminated and tilting the wrist back towards the body when the target is dimmed (Figure 1). When a haptic stimulus is used, the user times the tilting motion so that the vibration occurs when the wrist is away from the body. The user is able to provide a positive or negative response to the stimulus by gesturing in-sync or 180 degrees out-of-sync. In pilot testing with the haptic interface, users were able to respond to notifications quickly and with high accuracy while keeping their watch hand at their sides. Anecdotally, such interactions are not noticed by bystanders, whereas a swipe on the watch's touchscreen interrupts face to face conversation. SeeSaw also supports detecting repetitive twisting gestures without a stimulus. These gestures are performed by repeatedly tilting the wrist inward and outward at a constant frequency.

ALGORITHM DESIGN & IMPLEMENTATION

SeeSaw is a motion correlation algorithm designed to facilitate rapid synchronous gesture detection and to minimize false-triggering. The algorithm is implemented as a multi-stage

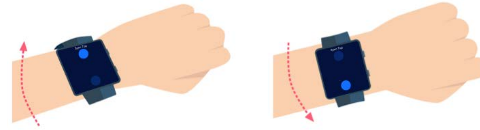

Figure 1: SeeSaw Wrist-Tilting Gesture Interaction

pipeline where sensor input from the smartwatch is processed in multiple steps: Signal Preprocessing, Synchrony Detection, Lag Adjustment, and Output Processing.

Algorithm 1: SeeSaw gesture detection algorithm

function SyncDetect (X_W, s, T);
Input : Window of sensor data X_W, stimuli s, and period T
Output : Sync detected in X_W
$X'_W \leftarrow$ PreProcess(X_W)
if $s = \emptyset$ **then**
    `/* No stimulus, use autocorrelation */`
    $\rho \leftarrow$ AutoCorrelation(X'_W, T)
else
    `/* Stimulus provided, use correlation */`
    $\phi \leftarrow$ SignalOffset(s)
    $\tau_{\text{delay}} \leftarrow$ LagAdjust(X'_W, s)
    $r \leftarrow$ GenerateReference($T, \phi + \tau_{\text{delay}}$)
    $\rho \leftarrow$ Correlation(X'_W, r)
end
$\rho' \leftarrow$ EWMA(ρ)
return $\rho' \geq \rho_{\text{thresh}}$

Signal Preprocessing consists of feature extraction (using the sensor axis with most variance), windowing, resampling, and detrending the incoming sensor data. If a stimulus signal is provided, the correlation between the sensor window and the stimulus is computed in the Synchrony Detection stage; otherwise, autocorrelation is used. Time-delay analysis using cross correlation is beneficial for improving the accuracy of with-stimulus synchrony detection and is used for Lag Adjustment. Finally, the detector output is smoothed using an exponentially weighted moving average (EWMA) to prevent false-triggering.

The parameters for these algorithms were selected using a combination of trial-and-error and using the results of a pilot study. The sensor window length is set to $t_{window} = 1.5$ s, using the x-component of the gyroscope sensor sampled at 10 Hz. The output is smoothed using a decay factor $\alpha = 0.35$.

EVALUATION

A commodity Sony Smartwatch 3 SWR50 is used for development and testing. The smartwatch features a quad-core 1.2 GHz ARM processor, 512 MB of RAM, and a 9 DoF IMU. Android 6.0.1 and Android Wear 1.5 were installed on the device. The SeeSaw synchronous gesture detector was implemented as an Android library which was used to construct Java applications that ran unmodified on the default operating system.

User Study - Gesture Comparison

A within-subjects study was conducted in a semi-controlled environment with a total of four experimental conditions (partially counterbalanced) to compare the synchronous gesture

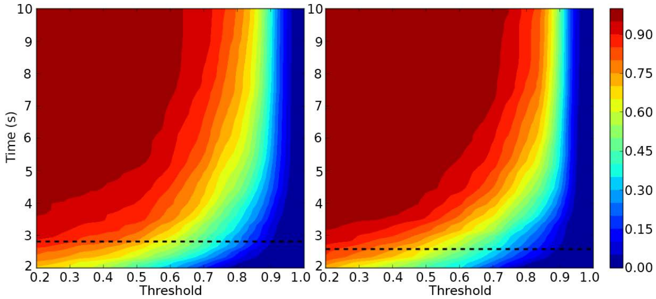

Figure 2: Precision vs Threshold vs Sync Time

(with visual and haptic stimuli) against a traditional swipe-to-unlock gesture across the watch's touchscreen. Participants were given two distraction tasks - 1) driving a car in a simulator and 2) walking while holding a filled cup of water. These tasks represent situation impairments where a watch-hand only interface might be desirable. Similar situations are common: riding a bicycle, holding an umbrella, holding a leash, etc. Participants were given a practice session of 25 notifications before each syncing session for them to become accustomed to a novel gestural input while performing the distraction task. The system was evaluated by its accuracy and detection speed in notification response scenarios with 12 participants (7M/5F, ages 20-26) in our institution's usability lab. The trigger threshold was set to a high value ($\rho_{thresh} = 0.85$) to enable post-hoc analysis of lower thresholds and to avoid early false-triggering. All participants were paid $10 to complete the study, which lasted approximately one hour. Following participation in the notification response study, participants were asked to complete a NASA TLX assessment for both the syncing gesture and the swiping gesture (Table 1).

To evaluate the SeeSaw interaction for controlling a HWD, a total of six participants were given the opportunity to participate in an additional 45-minute study for another $10. The HWD evaluation study tested only the driving condition, with two sessions (one practice and one evaluation) of 25 notifications triggered randomly around every 20 seconds (Table 2). The trigger threshold was set to $\rho_{thresh} = 0.8$.

Table 1: Gesture Comparison TLX

	Sync	Swipe	*p*-value
TLX Mental	8.9	3.2	0.06
TLX Physical	7.3	14.3	0.06
TLX Temporal	11.1	6.0	0.09
TLX Performance	7.7	7.8	0.95
TLX Effort	8.8	7.9	0.69
TLX Frustration	6.8	4.9	0.60
TLX Overall	50.7	44.2	0.46

Most participants found that notification dismissal with both gestures was more difficult during the sitting condition, due to the driving task. Swiping took 9.2% longer ($t_{sit} = 2.81$ s, $t_{walk} = 2.57$ s), and it took more time to reach the same accuracy for the syncing gesture. However, the detector's false positive rate was much lower for the sitting condition as the user's periodic motion while walking more easily matches the sync gesture. Using a threshold of $\rho_{sit} = 0.6$ with a false positive rate of 8 per hour, notifications can be dismissed in four seconds with 85% accuracy. While users found it easier to dismiss notifications while walking due to the less cognitively demanding distraction task and the availability of visual attention, the motion of walking and swinging arms introduced many more false positives. Using a threshold of $\rho_{walk} = 0.73$ with a false positive rate of 21 per hour, notifications can be dismissed in 5.5 seconds with 85% accuracy.

Table 2: Smartwatch & HWD Notification Dismissal

	HWD Dismissed	HWD Time
P1	93.3%	2.7 s
P2	90.0%	3.2 s
P3	93.3%	3.5 s
P4	100%	4.7 s
P5	100%	3.2 s
P6	93.3%	4.0 s
Overall	95.0%	3.6 s

Compared to the swiping gesture, the syncing gesture had a slightly higher workload on the NASA TLX assessment ($M_{sw} = 44.2$, $M_{sy} = 50.7$). While none of the differences of means reached statistical significance ($\alpha = 0.05$), trends suggest swiping was less mentally and temporally demanding, but more physically demanding. Based on our anecdotal observations, we hypothesize that the syncing gesture requires less overall concentration for users with more rhythmic ability (e.g., through playing musical instruments).

Discussion

We evaluate the SeeSaw interface in a best-case sitting activity with little movement and a worst-case walking activity with large amounts of periodic movement. Surprisingly, participants who became familiarized with the haptic stimulus were able to dismiss notifications with SeeSaw without needing to view the watch screen. Results show that, by using different correlation thresholds, the SeeSaw detector can be used for rapid, one-handed, and gaze-free notification dismissal for both. Detecting walking using current smartwatches is a common feature, which implies that the correlation threshold value could be raised during walking to help avoid the higher false positives during that activity.

Anecdotally, participants spilled water from the cup while walking more often when using the swipe gesture than with SeeSaw. When water was spilled on the touchscreen, participants noted that the accuracy of the swipe applications suffered dramatically. While the swipe gesture was faster for both tasks, these observations suggest potential advantages to SeeSaw: watch-hand only interaction, less movement artifacts from the interaction affecting a primary task (like holding a glass of water or umbrella), and more tolerance to wet conditions. Thus, SeeSaw can be used in conjunction with traditional swipe gestures to augment current smartwatch interactions for situations where the non-watch hand is occupied.

Compared to the smartwatch-only condition, the smartwatch and HWD system allowed participants to respond to notifications more quickly and accurately. All participants achieved over 90% accuracy, and the overall mean dismissal time is around one second less than the smartwatch-only system. Although the system was only evaluated for the sitting condition, data collected between stimulus events from both walking and sitting activities from the previous gesture comparison study can again be used to calculate the false positive rate for different thresholds. Using a threshold $\rho_{HWD} = 0.76$, with a false positive rate of one per hour for sitting and ten per hour for walking, notifications are dismissed in 3.5 s with 90% accuracy.

LIMITATIONS AND FUTURE WORK

During the user studies conducted, the detector was set to trigger at a constant threshold. Additional work can be done by building a pose-detector or activity recognizer that could dynamically adjust the triggering threshold. This effort would be aided by evaluating the detector across a wider range of everyday motion data from an in-the-wild data set. The study design can also be improved by conducting experiments on a larger number of participants.

Given the good performance of the smartwatch & HWD system, we intend to explore autocorrelation for the smartwatch-only system. Conversely, the HWD system might be extended to enable selecting between multiple simultaneous synchronization targets.

A primary advantage of the syncing gesture is its ability to be performed subtly with one-hand. In addition to evaluating the gesture's effectiveness in different scenarios by examining the false positive data, we intend to explore the social acceptability of the gestures in different scenarios using a user perception study or a social acceptability rating scale. Finally, we intend to explore the usefulness of SeeSaw as an activation gesture for eyes-free interactions. Perhaps tilting the wrist synchronously with rotating the head, without any visual or haptic stimulus, can be used as an activation or notification dismissal gesture. We hypothesize such interactions will require lower concentration and achieve higher accuracy.

CONCLUSION

We presented SeeSaw, a synchronous input interface for smartwatches that enables rapid, one-handed input for commodity smartwatches without any hardware or software modifications. In contrast to many gesture interfaces that detect user intent through gesture classification, the synchronous gesture interface uses flashing or vibrating stimuli to allow users to select UI elements or respond to notifications. We evaluated SeeSaw as a smartwatch-only interaction and found that users were able to dismiss incoming notifications from 4 s - 5.5 s using the one-handed, gaze-free interaction. During our evaluation of the smartwatch and HWD system, we found SeeSaw to provide excellent accuracy (95%) and dismissal speed (3.6 s) with a low false positive rate.

REFERENCES

1. D.L. Ashbrook. 2010. *Enabling mobile microinteractions*. Georgia Institute of Technology.
2. D.L. Ashbrook, J.R. Clawson, K. Lyons, T.E. Starner, and N. Patel. Quickdraw: the impact of mobility and on-body placement on device access time. In *Proc. CHI '08*. 219–222.
3. M. Carter, E. Velloso, J. Downs, A. Sellen, K. O'Hara, and F. Vetere. PathSync: Multi-user gestural interaction with touchless rhythmic path mimicry. In *Proc. CHI '16*. 3415–3427.
4. C. Clarke, A. Bellino, E. Velloso, H. Gellersen, and others. TraceMatch: A computer vision technique for user input by tracing of animated controls. In *Proc. Ubicomp '16*. 298–303.
5. A. Dementyev and J.A. Paradiso. WristFlex: low-power gesture input with wrist-worn pressure sensors. In *Proc. UIST '14*. 161–166.
6. M. Dhuliawala, J. Lee, J. Shimizu, A. Bulling, K. Kunze, T.E. Starner, and W. Woo. Smooth Eye Movement Interaction Using EOG Glasses. In *Proc. ICMI '16*. ACM, New York, NY, USA, 307–311.
7. A. Esteves, E. Velloso, A. Bulling, and H. Gellersen. Orbits: Gaze interaction for smart watches using smooth pursuit eye movements. In *Proc. UIST '15*. 457–466.
8. J. Gong, X. Yang, and P. Irani. WristWhirl: One-handed Continuous Smartwatch Input using Wrist Gestures. In *Proc. UIST '16*. 861–872.
9. G. Laput, R. Xiao, and C. Harrison. Viband: High-fidelity bio-acoustic sensing using commodity smartwatch accelerometers. In *Proc. UIST '16*. 321–333.
10. G. Reyes, J. Wu, N. Juneja, M. Goldshtein, K.W. Edwards, G.D. Abowd, and T.E. Starner. 2018. SynchroWatch: One-Handed Synchronous Smartwatch Gestures Using Correlation and Magnetic Sensing. *ACM IMWUT* 1, 4, Article 158 (Jan. 2018), 158:1–158:26 pages.
11. E. Velloso, M. Carter, J. Newn, A. Esteves, C. Clarke, and H. Gellersen. 2017. Motion Correlation: Selecting Objects by Matching Their Movement. *ACM TOCHI* 24, 3 (April 2017), 22.
12. E. Velloso, M. Wirth, C. Weichel, A. Esteves, and H. Gellersen. AmbiGaze: Direct control of ambient devices by gaze. In *Proc. DIS '16*. 812–817.
13. M. Vidal, K. Pfeuffer, A. Bulling, and H.W. Gellersen. Pursuits: Eye-based Interaction with Moving Targets. In *CHI EA '13*. ACM, New York, NY, USA, 3147–3150.
14. H. Wen, J. Ramos Rojas, and A.K. Dey. Serendipity: Finger gesture recognition using an off-the-shelf smartwatch. In *Proc. CHI '16*. 3847–3851.
15. C. Zhang, X. Wang, A. Waghmare, S. Jain, T. Ploetz, O.T. Inan, T.E. Starner, and G.D. Abowd. FingOrbits: interaction with wearables using synchronized thumb movements. In *Proc. ISWC '17*. 62–65.
16. Y Zhang and C. Harrison. Tomo: Wearable, low-cost electrical impedance tomography for hand gesture recognition. In *Proc. UIST '15*. 167–173.