Designing and Evaluating Hand-to-Hand Gestures with Dual Commodity Wrist-Worn Devices

YIQIN LU, BINGJIAN HUANG, CHUN YU*, GUAHONG LIU, and YUANCHUN SHI†, Tsinghua University, China

Hand gestures provide a natural and easy-to-use way to input commands. However, few works have studied the design space of bimanual hand gestures or attempted to infer gestures that involve devices on both hands. We explore the design space of hand-to-hand gestures, a group of gestures that are performed by touching one hand with the other hand. Hand-to-hand gestures are easy to perform and provide haptic feedback on both hands. Moreover, hand-to-hand gestures generate simultaneous vibration on two hands that can be sensed by dual off-the-shelf wrist-worn devices. In this work, we derive a hand-to-hand gesture vocabulary with subjective ratings from users and select gesture sets for real-life scenarios. We also take advantage of devices on both wrists to demonstrate their gesture-sensing capability. Our results show that the recognition accuracy for fourteen gestures is 94.6% when the user is stationary, and the accuracy for five gestures is 98.4% or 96.3% when the user is walking or running, respectively. This is significantly more accurate than a single device worn on either wrist. Our further evaluation also validates that users can easily remember hand-to-hand gestures and use our technique to invoke commands in real-life contexts.

CCS Concepts: • **Human-centered computing → Gestural input**; **Mobile devices**; *User centered design.*

Additional Key Words and Phrases: hand gesture, wearable device, motion correlation, locomotion

ACM Reference Format:
Yiqin Lu, Bingjian Huang, Chun Yu, Guahong Liu, and Yuanchun Shi. 2020. Designing and Evaluating Hand-to-Hand Gestures with Dual Commodity Wrist-Worn Devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 1, Article 20 (March 2020), 27 pages. https://doi.org/10.1145/3380984

1 INTRODUCTION

Demand for easy-to-use input technologies in mobile scenarios has risen along side the popularization of smart phones and wearables. Traditional input methods are inhibited by mobility constraints. Previous works show that using such methods in mobile scenarios can increase user task load and lead to potential injuries or input errors [35, 54]. For example, when running, constant movement prevents runners from tapping and swiping on smartphones and smartwatches accurately. Thus, researchers are actively seeking supplementary input methods that are compatible with a larger breadth of scenarios.

*This is the corresponding author
†Also with Institute for Accessibility Development, Tsinghua University

Authors' address: Yiqin Lu, lu-yq16@mails.tsinghua.edu.cn; Bingjian Huang, huangbj16@mails.tsinghua.edu.cn; Chun Yu, chunyu@tsinghua.edu.cn; Guahong Liu, lgh17@mails.tsinghua.edu.cn; Yuanchun Shi, shiyc@tsinghua.edu.cn, Key Laboratory of Pervasive Computing, Ministry of Education, Beijing Key Lab of Networked Multimedia, Beijing National Research Center for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University, Beijing, China.

2474-9567/2020/3-ART20 $15.00
https://doi.org/10.1145/3380984

Hand gestures, including unimanual gestures [9, 25, 34, 62, 69], mid-air bimanual gestures [17, 45, 48, 55, 58] and on-skin gestures [6, 19, 61], do not require tangible input devices and have low cognitive loads, which are considered as supplementary options for command input. The growing requirements in more ubiquitous scenarios continuously motivate researchers to design easy-to-use hand gestures and develop advanced sensing technologies.

Wrist-worn devices (e.g., smartwatches, smart bands) have become increasingly mainstream. The mount location of these devices provides instantaneous command input, and their continuous contact with the user's skin allows embedded sensors (e.g., IMU, microphone and pulse sensors, etc.) to record users' daily activities and monitor their health [50].

Previous works have investigated the feasibility of using these devices to sense hand gestures. The device is usually worn on one hand, either the dominant hand, for tracking hand movements [7, 31, 66, 74] or the non-dominant hand, for detecting touch events [70, 72, 73]. However, few works have investigated potential bimanual gestures or improvements in gesture inference that are created by combining the sensing capabilities from devices worn on the user's dominant and non-dominant hands.

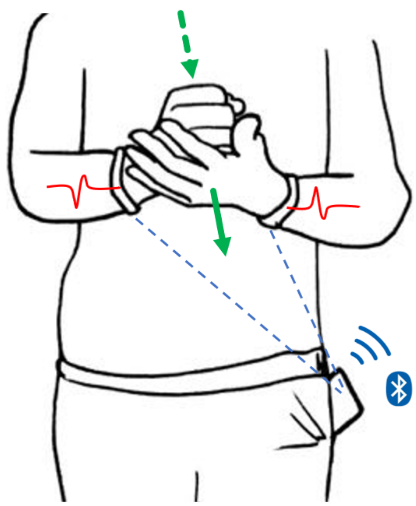


Fig. 1. Interaction using hand-to-hand gestures with two off-the-shelf wrist-worn devices. The user touches his left palm with his right fist, generating distinct signals from both hands which can be detected by the two smart bands on his wrists. The smart bands transfer the data to the user's smartphone via Bluetooth. The smartphone then recognizes the hand-to-hand gesture, and triggers a corresponding command.

We explore gestures that involve touching one hand with the other, which we refer to as "hand-to-hand gestures". These gestures vary by the point of contact on each hand (e.g., the palm, the back of the hand) and the contact type (e.g., tap, swipe). Hand-to-hand gestures have the following advantages for universal command input: 1) Unlike bimanual gestures in which the user's hands do not touch, hand-to-hand gestures have mutual haptic feedback on both hands. This feedback not only facilitates the self-awareness of gestures so that users will perceive the contact parts of hands and the quality of execution, but also helps users perform continuous gestures in scenarios where motor control is limited (e.g., running). 2) Hand-to-hand gestures are proprioceptive, meaning the user can execute them without visual interfaces. This is especially important when the user's hands are outside their field of view (e.g., VR/AR, walking on the street). Also, anatomical landmarks on hands (e.g., joints, gaps) can help users perform hand-to-hand gestures more easily [5, 64].

When the user performs hand-to-hand gestures, the vibration generated by the contact of their hands can be sensed by two sensors worn on the user's left and right wrists (Fig. 1). We also explore the feasibility of employing two off-the-shelf wrist-worn devices to sense hand-to-hand gestures. Our approach has the following advantages: 1) We use two off-the-shelf commercial smart wrist-worn devices without adding or modifying the hardware, which are not only lightweight and easy to implement, but have also been socially normalized. Moreover, both devices only work as data receivers connecting to the central smartphone, which consume little computing resources and battery power. 2) Dual wrist-worn devices can provide rich information on hand-to-hand gestures, including relative hand orientations and correlated motion directions. Sensing with two devices is more accurate than sensing with a single one and can remain effective even when the user is moving (e.g., running).

We conduct three user studies to explore and evaluate hand-to-hand gestures. In the first study, we explore a hand-to-hand gesture vocabulary according to users' preferences and derive 56 tap and swipe gestures. We also recruit users to rate gestures on four measures to provide guidance for gesture selection. In the second study, based on the data we collected from users in the first study, we derive gesture sets for different scenarios and collect sensor data in each scenario for recognition. Our results show that sensing with dual devices has a higher recognition performance than that with only a single one. In the third study, we evaluate hand-to-hand gestures and our gesture sensing system in different contexts and demonstrate their potential usability. Our major contributions are three-fold:

(1) We preliminarily explore the design space of hand-to-hand gestures and derive the gesture vocabulary with usable gesture sets in real scenarios.
(2) We study the sensing capability of dual wrist-worn devices and analyze cross-device features for more accurate gesture inference.
(3) We provide empirical results for gesture detection and recognition to demonstrate the superiority of using two devices in conjunction, as opposed to a single device.

2 RELATED WORK

We present related works in three aspects: First, we review previous work on bimanual hand gesture designs to show the variety and versatility of different types of hand gestures. We then summarize previous technical approaches on hand gesture sensing and recognition and discuss their pros and cons. Finally, we introduce works relevant to our sensing approach, which combine data from multiple sensors.

2.1 Bimanual Gesture Design

Gestures involving two hands include more degrees of freedom and are more efficient [8] than one-handed gestures in many interaction tasks. Although bimanual gestures can be derived from the combination of two simultaneous unimanual gestures, few of them are natural or ergonomic, because the motions of both hands tend to be constrained by the human body's kinematic system [16]. We summarize two categories of bimanual hand gestures and introduce related works focused on gesture design in each category.

2.1.1 Non-Touch Bimanual Gestures. Gestures that require both hands to cooperate with no contact are often customized for specific input vocabularies, such as graphical user interfaces (e.g., pan and zoom gestures [45, 55, 58, 67]), object manipulation (e.g., rotating and scaling in AR [48]) and mid-air drawing [17]. These gestures are usually axisymmetric or centrosymmetric, following humans' natural motor behavior and have been assigned conventional semantics in daily life (e.g., zoom in/out).

2.1.2 On-Skin Finger Gestures. In order to utilize the mutual control and bidirectional haptic feedback between two hands, many previous works have studied bimanual gestures which allow two hands to come into contact with one another. One idea is to treat the region on/around the non-dominant hand (e.g., palm [6, 18, 20, 61], back of the hand [56, 70, 73, 77, 78], forearm [20, 53, 68]) as a touchable surface and allow the dominant hand to perform on-skin gestures (e.g., tap, draw, pinch, etc.). Some researches have also studied touch accuracy and learnability on different locations of the hand [5, 19] and pointed out that landmarks on the hand (e.g., fingertips, metacarpal bones) are easy to touch and remember [19, 64].

Hand-to-hand gestures represent a kind of touch-based gestures that requires two hands to touch one another. Some of our hand-to-hand gestures have been mentioned in previous work, including in the context of on-skin finger gestures that we discussed above, clapping [26], forming an X-shape with two hands [6] and so on. To the best of our knowledge, the concept and design space of hand-to-hand gestures have not been explored independently and comprehensively. In this paper, we contribute to an understanding, design and evaluation of motion-based hand-to-hand gestures to enrich the vocabulary of bimanual input.

2.2 Hand Gesture Sensing and Recognition

We summarize related works by their sensing and recognition approaches. We also focus on the interaction techniques involving wrist-worn wearable devices in related works.

2.2.1 Inertial Sensing. The inertial sensor, also referred to as the inertial measurement unit (IMU), including the accelerometer (3-axes acceleration), gyroscope (3-axis rotation velocity) and magnetic sensor (3-axes orientation), are embedded in many consumer devices like smartphones and smartwatches. In previous work, wrist-worn inertial sensors have proven to be reliable in inferring hand postures using the orientation [28, 57] and hand motions using acceleration and rotation [7, 31, 59, 62, 66]. However, inertial signals from wrist-worn devices are better at detecting hand gestures with large-scale hand motions (e.g., hand waving [11, 59], rotating [7]) and general contact (e.g., clapping) [72] but weaker in sensing micro finger gestures (e.g., pinching, mid-air swiping) from ambient hand motions. In this paper, we rely on inertial sensing to recognize hand-to-hand gestures.

2.2.2 Acoustic Sensing. Sounds produced by hand movements [3, 20, 22, 62] or hand contact [70, 73] can be easily captured by wrist-worn microphones through bone conduction. Acoustic sensing has a high sample rate and provides rich signatures in both time and frequency domains to help recognize different hand gestures. Acoustic sensing can also be used for static hand posture inference by sweeping frequency signals between speakers and receivers [71] or by ultrasound imaging [40]. However, because the quality of acoustic signatures may be diminished by incidental ambient noise, it is often combined with other approaches (e.g., inertial sensing [62, 70, 73]) to compensate for its weakness.

2.2.3 Other Sensing Approaches. Computer vision technologies, including RGB cameras [17] or depth cameras (e.g., Kinect, Leap Motion) [18, 27, 34, 49, 56], can easily capture intuitive hand postures and motions as humans see them. In the past, researchers often mounted cameras on users' wrists to get a better field of view around the user's hands. While this might increase the system's accuracy in recognizing gestures, it also restricts the movement of the user's hands. Photoplethysmography (PPG) detects blood volume changes via a pulse oximeter and has been used for hand and finger motion detection [57, 74]. However, the signal is sensitive and noisy, which makes gesture recognition more challenging. A sensor array mounted on or around the wrist can infer hand gestures by sensing the deformation of the skin or the change of muscle distribution, including infrared proximity sensors [15, 39], EMG [21, 41], piezo/pressure sensors [12, 15, 41], capacitive sensors [51], tomography [75, 76] and electric fields [78]. However, these approaches require additional hardware.

2.3 Fusion from Multiple Sensors

In this work, we use two smartwatches to sense hand-to-hand gestures. After synchronization, the correlation of sensor signals between the two devices helps us better infer hand gestures. Some previous works have explored analyzing the correlation of data from multiple sensors in a hand motion event to enable detection in many scenarios. Device-to-device connection [24], authentication [37], or synchronization [23] can be achieved by shaking or bumping these devices and generating similar motion patterns in embedded sensors, which is robust enough to be separated from other concurrent motions [38]. A handshake [4, 65] or "High Five" gesture [29] between two people can be detected by the connected devices mounted on their wrists. The sensing of users' hands on many other devices can be augmented by wearing a smartwatch to allow for cross-device interaction with smartphones [10], identification of individual users on large multi-person touchscreens [63], and user authentication on smart devices [33].

3 TECHNICAL OVERVIEW: SENSING WITH DUAL OFF-THE-SHELF WRIST-WORN DEVICES

In this section, we introduce our hardware devices and sensing approaches. We introduce how we synchronize and align the two devices' signals and illustrate how pairing the signals enables several kinds of features that can be used for gesture inference, such as device orientation, motion correlation, and acoustic signals.

3.1 Apparatus

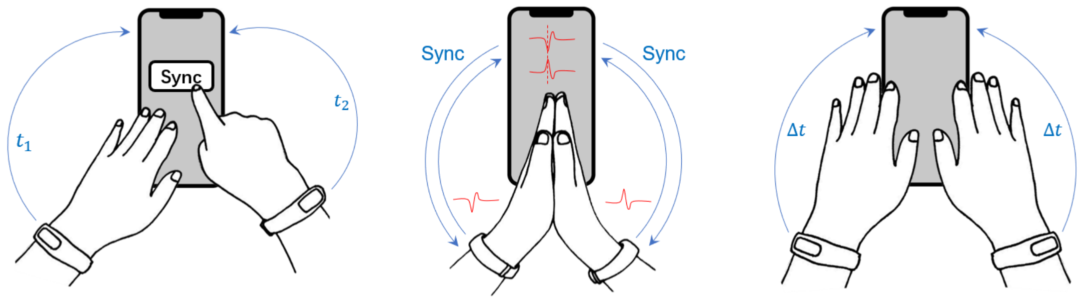
The scenario we envision is: the user wears a device on each wrist and invokes commands by performing hand-to-hand gestures. We use these wrist-worn devices as data receivers which consume little power for on-chip computation and a persistent Bluetooth connection with a smartphone for data analysis and command execution. In the following implementation, we use two off-the-shelf smartwatches as our hardware. We do not have specific requirements with regard to the computing power or interaction methods of smartwatches (smartwatches and smart bands can be used interchangeably).

We used two Apple Watch Series 4 devices as our hardware. On these devices, the sample rate of inertial data was 100Hz and the sampling interval was stable: the SD of the timestamp intervals was less than 1ms. The sample rate of the microphone was 44.1kHz. We developed a watch app using Objective-C and installed it on both Apple Watches (running WatchOS 5.1) for data collection. The two watch apps were connected to an iPhone X (running iOS 12.1) using Bluetooth. We obtained acceleration (3 axes), rotation velocity (3 axes), device orientation (Euler angle or quaternion) and microphone WAV data from the WatchKit interface in WatchOS. Other raw signals were inaccessible due to the lack of relevant WatchOS interfaces.

3.2 Synchronization and Aligning

The system timestamps for the sensor data collected from the smartwatches were inconsistent at first, and the delay of data transferred from the smartwatches to the smartphone via Bluetooth was unpredictable. Hence, it was necessary for the smartphone to synchronize two independent data streams from the smartwatches in the time domain. We designed a process to ensure synchronization (Fig. 2):

(1) When the app is launched, the smartphone is receiving independent data streams from each device with unaligned timestamps. To synchronize the streams, the user taps the "Sync." button on the iPhone app and keeps his or her hands still.
(2) The smartphone begins monitoring the inertial data and detects obvious signal perturbation. When the user performs a symmetric clapping gesture, the smartphone will observe significant motion peaks on in the acceleration data from both devices.
(3) The smartphone extracts the timestamp of the highest/lowest peak/valley as the reference timestamp for each wrist-worn device individually. Then, the smartphone issues a synchronization command to each wrist-worn device with a corresponding reference timestamp.
(4) Each wrist-worn device receives the synchronization command and starts to send data streams with the time relative to the reference timestamp (the original timestamp minus the reference timestamp). From this point on, the smartphone receives two data streams with consistent timestamps.



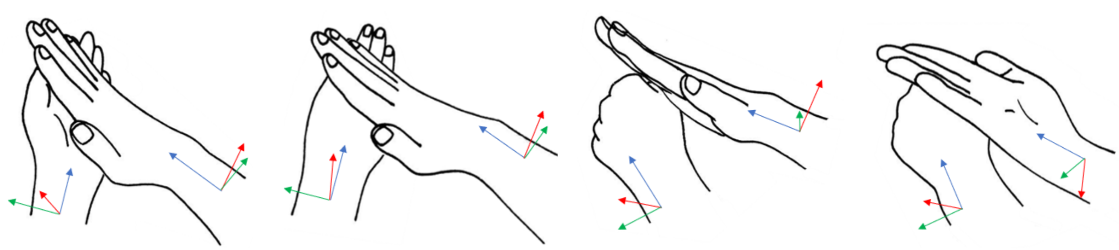
(a) (b) (c)

Fig. 2. Synchronizing the timestamps on both wrist-worn devices. (a) The timestamps t_1 and t_2 of two devices are not synchronized, and the user starts by press the "Sync." button. (b) The user claps, the smartphone synchronizes the timestamps of the two wrist-worn devices by aligning a single peak and valley interval in the signals and sends the information to the phone. (c) Both wrist-worn devices are synchronized with the aligned timestamp Δt relative to the clapping time.

After synchronization, however, the inertial data from the smartwatches is still not perfectly aligned, due to the instability of the interval between successive timestamps. Hence, we resample the inertial data using linear interpolation with the frequency of 100Hz. Two inertial data streams are synchronized and aligned with the pairwise discrete data points, which benefits the following analysis. The timestamp synchronization will persist until the user closes the app, which means he or she only needs to synchronize once in the beginning.

3.3 Device Orientation

Previous studies in the fields of ergonomics and biomechanics [46] show the limitations of human wrist rotation. For example, when fixing the forearm, the wrist can only rotate about 180° between palm-up and back-up hand postures (the back of the hand is referred to as "back" throughout the remainder of this paper). Therefore, the hand orientation and wrist orientation are relatively similar for most people when performing the same hand-to-hand gesture.



(a) Palm-to-Palm (b) Palm-to-Back (c) Palm-to-Fist (d) Back-to-Fist

Fig. 3. Two wrists show different orientations in different hand-to-hand gestures. Red arrows indicate the normal vector of the back. Green arrows indicate the direction of the outside of the body. Blue arrows indicate the direction of the fingers.

Wrist orientation differs in many different hand postures in terms of the different contact areas between the two hands. Fig. 3 shows the wrist's different orientations in different hand-to-hand gestures. Hence, two device orientations can be employed to infer hand postures.

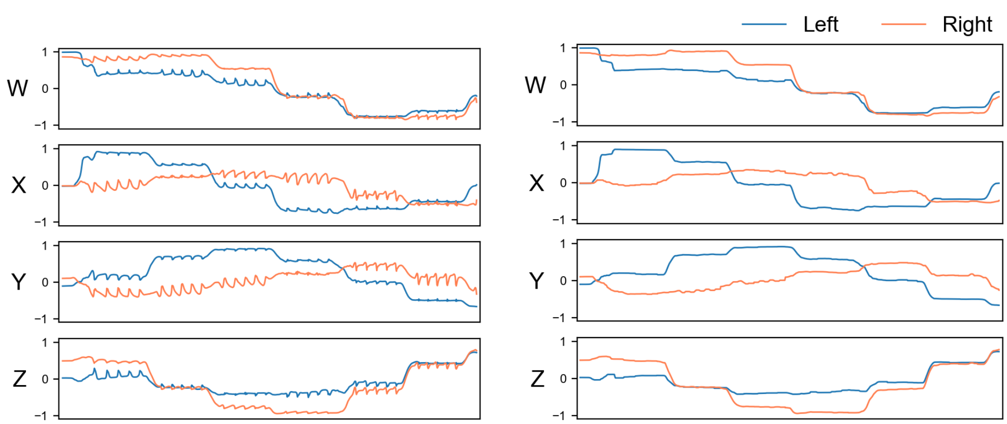
Wrist orientation is affected by the body's orientation. However, when performing the same gesture, the relative orientation between the two wrist-worn devices remains the same no matter what the orientation of hand or body is. That is, there will be a constant rotation between both devices. Using relative rotation to recognize gestures has also been employed in previous research [1] . We denote the orientations of left and right wrist-worn devices as q_L and q_R, respectively, in quaternion form. Then, we can derive the constant rotation Δq by observing two device orientations in real time as follows:

$$q_R = \Delta q * q_L \tag{1}$$

$$\Delta q = q_R q_L^{-1} \tag{2}$$

Equation (1) can also be expressed as: q_L rotates Δq then becomes q_R.

The orientation of both wrists will be affected by performing gestures. The wrists will move a little due to the motion produced by the gesture and generate turbulence in the orientation data (Fig. 4a), which makes the device orientations inconsistent during the gesture. To eliminate this turbulence's influence and make the devices' orientation data more stable for further analysis, we smooth the values of four quaternion components of the devices' orientation data respectively with a sliding window of 1000ms. Because the rotation angle between two successive frames is small due to the high sample rate, we simply use median filtering to smooth the signals, and the normalization is done after filtering for the quaternion computation. After smoothing, both devices' orientations are stable over time even when hand-to-hand gestures are performed (Fig. 4b).



(a) Before smoothing (b) After smoothing

Fig. 4. Smoothing the device orientations over four components of the quaternion to eliminate the turbulence brought about when the user performs hand-to-hand gestures. (a) Before smoothing, the gestures will generate turbulence in the orientation data. (b) After smoothing, the orientation data is stable, allowing for more accurate inference of hand postures.

Analyzing the relative rotation between the two devices' orientations is a powerful tool in recognizing hand-to-hand gestures. Nevertheless, the devices' orientations are not able to capture the full range of possible hand postures (e.g., some hand-to-hand gestures may have similar wrist orientations but different contact points on the hands).

3.4 Motion Direction and Correlation

When two hands come into contact, there is a pair of simultaneous and opposite motions on two hands. So, in theory, when the user performs hand-to-hand gestures, the wrist-worn devices on the left and right will sense a pair of motions that are equal in amplitude but opposite in direction at the same time. (Fig. 5).

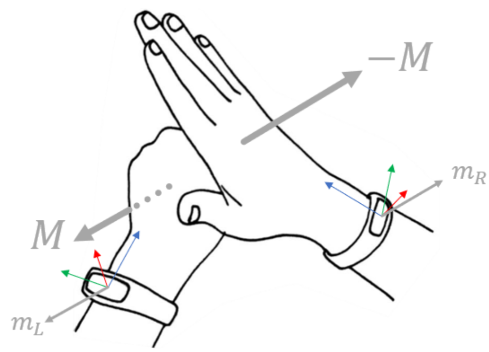


Fig. 5. A Palm-to-Fist gesture generates a pair of opposite motions M and $-M$ on the left and right hands, while the left and right wrist-worn devices sense a pair of motions m_L and m_R.

We denote the motions generated by the hand-to-hand gesture on the left and right hands as M and $-M$, respectively, in 3D vector form. We denote the observed motions from left and right wrist-worn devices as m_L and m_R, respectively, both in their own coordinate systems and in terms of their wrist orientation. The motion M can be related by the equation:

$$M = q_L^{-1} m_L q_L = -q_R^{-1} m_R q_R \tag{3}$$

where q_L and q_R denote orientations of the left and right wrists, respectively, in quaternion form as above. Equation (3) indicates that the motion can be recreated by inverse rotation, as shown in the following equation:

$$q_R q_L^{-1} m_L q_L q_R^{-1} = -m_R \tag{4}$$

According to Equation (2), Equation (4) can be rewritten as

$$\tilde{m}_L = (\Delta q) m_L (\Delta q)^{-1} = -m_R \tag{5}$$

where $\tilde{m}_L$ denotes the representation of the left wrist's motion in the coordinate system of the right wrist.

In practice, the hands' motions are not perfectly transferred to wrist-worn devices. This is due to multiple factors: 1) The hand is not a rigid body, so the motion will be buffered by the deformation of muscle. 2) The wrist has some flexural degrees of freedom, so the motion changes when it is conducted through the wrist. To quantitatively evaluate the correlation of observed motion pairs from wrist-worn devices, we define the dot product of $\tilde{m}_L$ and $-m_R$ to measure the matching level of two opposite motion vectors:

$$Cor(\tilde{m}_L, m_R) = \langle \tilde{m}_L, -m_R \rangle \tag{6}$$

$$= -(\tilde{m}_L^X m_R^X + \tilde{m}_L^Y m_R^Y + \tilde{m}_L^Z m_R^Z) \tag{7}$$

where $\langle a, b \rangle$ denotes the dot product of a and b, m_R^X, m_R^Y and m_R^Z denote the components of motion m_R on X-, Y-, and Z-axes in the coordinate system of right wrist, the same as $\tilde{m}_L$. The benefits of this dot product definition are two-fold: 1) It considers the angle between a pair of opposite motions: The closer the angle is to 180°, the higher the correlation between the pair of motions. 2) It merges the energy of the motion on three axes, which is equivalent to rotating the motions into one dimension and calculating the product.

In our work, we chose the acceleration as measured by the smartwatches to represent motion. The acceleration data from two wrists can not only be employed to help infer hand-to-hand gestures (as they are in recognizing unimanual gestures [7, 31, 59, 62, 66]), but can also distinguish hand-to-hand gestures from natural hand movements in daily activities. Previous works have also studied the feasibility of using dual origin signals and their correlation to detect the interaction between hands [33], devices [23], or between hand and device [63]. In our work, although hands may generate large motion in many activities, the special characteristics of motions between two hands help us to distinguish hand-to-hand gestures from natural hand movements.

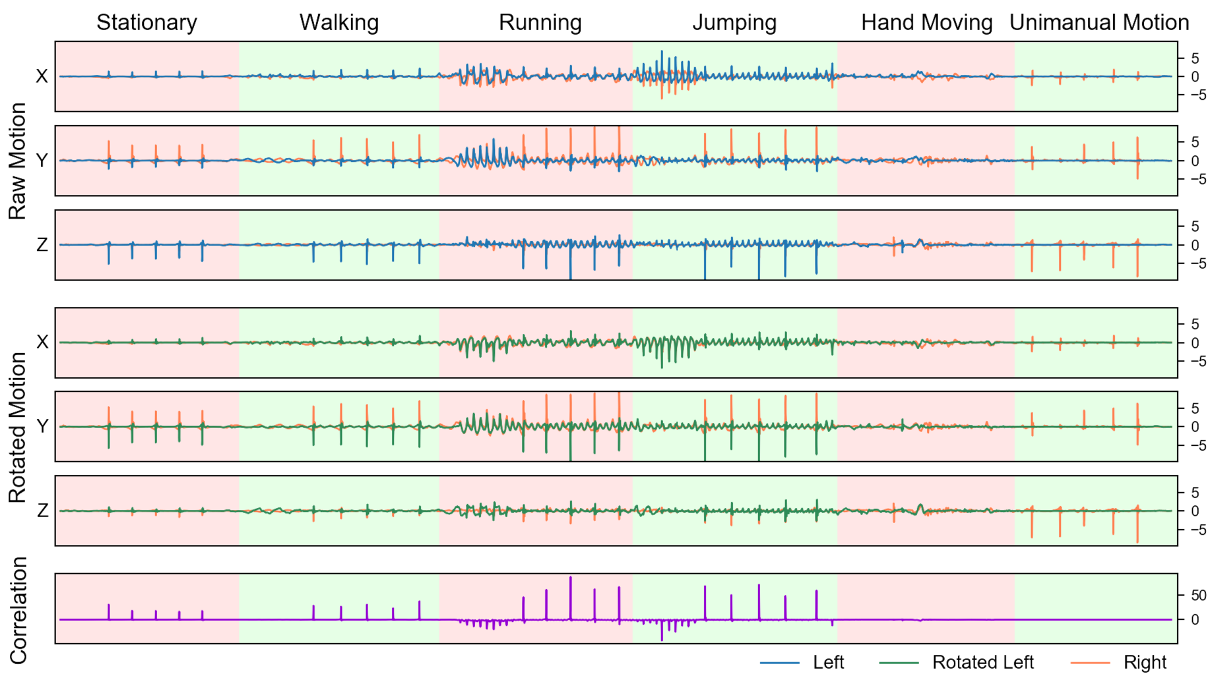


Fig. 6. An example of raw acceleration data (m_L and m_R, m/s^2), the accelerations rotated in the same coordinate system ($\tilde{m}_L$ and m_R, m/s^2) of two hands on three axes, and the correlation of the accelerations. In each of the stationary, walking, running, and jumping conditions, a hand-to-hand gesture (e.g., Bottom-to-Palm in this figure) is performed five times. In the hand moving condition, two hands are free to move without touching each other. In the unimanual motion condition, the right hand taps on the table five times, while the left is motionless. The correlation helps to distinguish hand-to-hand gestures from other natural hand movements: hand-to-hand gestures generate positive values in correlation, while other motions like running, jumping, waving hands, or unimanual touching generate zero or negative values.

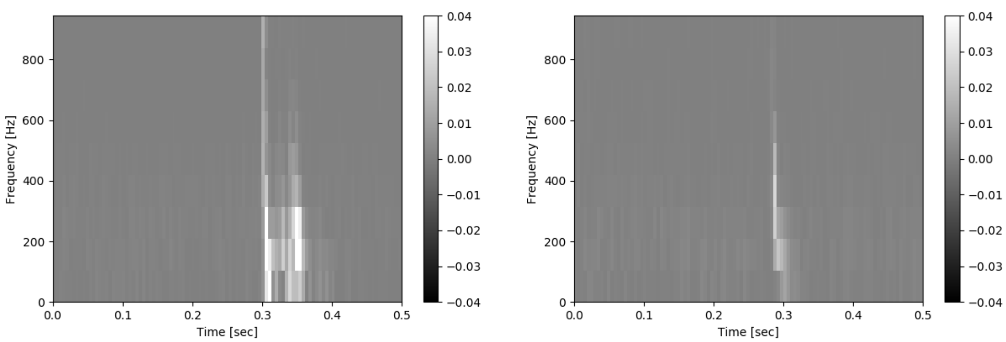
We show an example of how the correlation of two hands enhances the sensing of hand-to-hand gestures. Fig. 6 shows the raw motion data (m_L and m_R, m/s^2) and the motions rotated in the same coordinate system ($\tilde{m}_L$ and m_R, m/s^2) of two hands on three axes as well as the correlation of motions in different conditions. The same hand-to-hand gesture is performed five times in each of the stationary, walking, running, and jumping conditions, where the motion can be observed from the peaks on three axes, while no touch occurs in the hand moving condition and only the right hand touches the table five times in the unimanual motion condition. As we anticipated, walking and general hand motions generated small noise in the motion data, while running and jumping generated larger noise. When using acceleration signals, the noise generated by daily activities may interfere with the detection of gestures. However, when observing correlations, we can discover that hand-to-hand gestures will always produce significant positive correlation, while there will be no correlation or negative correlation (e.g., running or jumping) in non-hand-to-hand motion activities. Therefore, the correlation of motion from two smartwatches can be employed to distinguish hand-to-hand gestures from general hand movements, even when the user is running or jumping.

3.5 The Sound of Gestures

Device orientations and motion directions are effective in detecting and recognizing gestures. However, problems remain in distinguishing gestures with similar orientations and motions. For example, Palm-to-Back and IndexFinger-to-Back have similar motion characteristics because the relative orientations and contact areas of these two gestures are similar, which will generate similar signals in inertial data.

Thus, we apply microphone data and extract acoustic features to distinguish gestures with similar motions. Acoustic analysis is normally conducted in time and frequency domains, and features in both domains have been well explored. Since the sounds of gestures are relatively small and can be drowned out by ambient noises, we sought to recognize gesture patterns in the frequency domain.

Short-Time Fourier Transform (STFT) [2] is a typical algorithm to analyze non-stationary signals and extract both frequency and time features. By using STFT, we converted the audio time series of gestures into a 2-dimension time-frequency image, as shown in Fig. 7. The series reveals that different gestures may resemble each other in the time domain, but have distinguishable distributions in frequency domain. Therefore, we suggest that in addition to orientations and motions directions, deploying frequency features in classification can also help differentiate gestures.



(a) STFT of gesture Palm-to-Back

(b) STFT of gesture IndexFinger-to-Back

Fig. 7. An example of the Short-Time Fourier Transform (STFT) difference of gestures similar in IMU data.

We also implemented another widely used feature: Mel-Frequency Cepstral Coefficients (MFCCs) [44]. MFCCs are derived from a non-linear type of cepstral representation of microphone data, which are able to approximate human auditory features. The original process of MFCCs consists of five steps and the final result is in the time domain. In order to derive features in the frequency domain, we collect the generated sub-process features and combine them with inertial features to improve gesture detection and recognition.

3.6 Summary

In this section, we describe the technical implementation, including key features and the analysis of dual signals. We discuss using devices' orientations for gesture recognition, analyzing correlation of motions for hand-to-hand detection, and leveraging acoustics to improve gesture inference. More sensing capabilities are enabled by wearing two devices as opposed to a single device, and we validate the improved performance in Section 5. On the other hand, we found that it is hard to sense finger states using wrist-worn devices and must consider this limitation in our gesture design.

4 HAND-TO-HAND GESTURE DESIGN

In this section, we conducted interviews to collect users' comments on hand-to-hand gesture designs. Based on our findings, we present a representative gesture vocabulary in the hand-to-hand design space. We then collected users' subjective feedback for gestures in different measures. The goal of our study is to understand users' preferences with regard to hand-to-hand gestures and derive a gesture vocabulary.

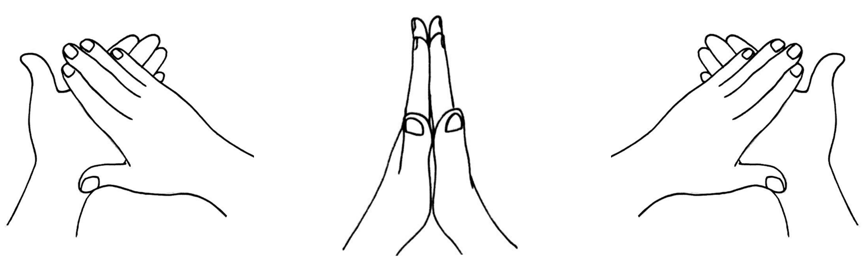
4.1 Understanding Users' Preferences

Before designing gestures, we first conducted a preliminary interview to collect users' opinions on hand-to-hand gestures.

4.1.1 Participants and Interview Procedure. We recruited ten participants (five male and five female), all right-handed with an average age of 25.8 (from 19 to 47, SD=2.77). In the interview, we first introduced the definition of hand-to-hand gestures, and then encouraged participants to propose as many gestures as possible and demonstrate gestures with their hands. We asked users to verbalize their thought process throughout the exercise.

We also considered gestures related to a set of predefined keywords, such as clap, swipe, finger tap, etc. When participants mentioned one of these keywords, we would ask them a question about their preferences with regard to the characteristics. To avoid bias from experimenters, we would not ask any questions unless participants explicitly mentioned the target keyword.

4.1.2 Findings. Nine participants proposed gestures including open hand, closed fist, and fist with one finger extended, and all participants agreed that these hand states should be differentiated for different gestures. With regard to the open hand posture, seven participants agreed that touching the palm and back of the hand should be regarded as two gestures, and two said it depended on the scenario. When we asked participants which finger or group of fingers they preferred in hand-to-hand gestures, all of them preferred the index finger.



(a) Right hand on top (b) Symmetric (c) Left hand on top

Fig. 8. Different forms of Palm-to-Palm gestures.

Gestures should not be distinguished in some conditions. In discussing gestures that share the same contact area, seven participants agreed that gestures only different in hand directions should be regarded as the same gesture. For example, similar Palm-to-Palm gestures, regardless of symmetry (Fig. 8a vs. Fig. 8b) and handedness (Fig. 8a vs. Fig. 8c), should be considered as one gesture. Moreover, eight participants preferred asymmetric gestures (Fig. 8a and Fig. 8c) because they felt more natural to perform. As for other factors, all participants unanimously agreed that the finger state should not be differentiated (i.e., flexion vs. relaxation). Eight participants agreed that tapping with different intensities should not be defined as different gestures because this could cause confusion.

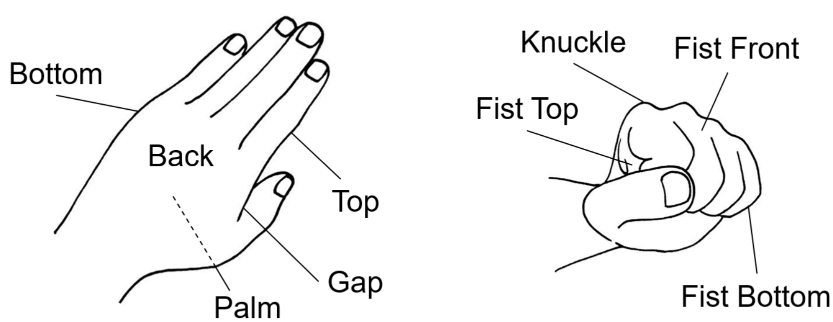
Participants also proposed uncommon touch areas that could complement gestures. Five participants suggested that the narrow sides of the hand could be possible contact areas. Moreover, five participants mentioned that the knuckles and gaps between fingers could be used as reference points [64] used to position a touch. Participants also provided gestures beyond simple taps. Nine participants mentioned swipe gestures between hands and fingers, as they noted that swipe gestures enable bidirectional (e.g., volume up and volume down) or quad-directional (e.g., grid menu selection) gestural input. Six participants considered double tap and long press as possible semantic complements for single tap gestures (e.g., single tap to turn on and double tap to turn off).

4.1.3 Discussion. Participants proposed a total of 35 different hand-to-hand gestures, which were limited for the following reasons: 1) Unlike elicitation experiments [9, 14, 30, 32, 67], we did not provide specific referents for participants to design gestures. 2) Participants' gesture proposals were constrained by their experience with prior interfaces and technologies due to legacy bias [43]. Nonetheless, these interviews still provided us with valuable insights. In the next subsection, we explore a gesture vocabulary based on these findings and attempt to simplify the design space.

4.2 Gesture Vocabulary

We are mainly concerned about three factors in our gesture design: the contact area of the dominant hand, the contact area of the non-dominant hand, and the motion type. To facilitate more concise illustrations, we set the right hand as the dominant by default.

Contact areas of the dominant (right) hand and the non-dominant (left) hand describe the location of motion and the relative orientation between hands. The contact areas we consider are shown in Fig. 9. Based on our interview findings, we only consider three hand states: open hand, closed fist, and single extended index finger. In open hand state (Fig. 9a), to avoid confusion, we selected five areas according to users' feedback: palm, back, top side, bottom side and gap. In closed fist state (Fig. 9b), we select four areas: top side, bottom side, front side and knuckle.



(a) Contact areas of open hand. (b) Contact areas of closed fist

Fig. 9. Contact areas we considered in the gesture design.

We consider two motion types in this work: tap and swipe. These gestures, mapped from graphical user interfaces (e.g., tap for clicking an icon and swipe for scrolling pages), were both widely mentioned by participants in our interviews. Tap gestures generally represent individual input commands, while swipe gestures represent multi-directional input commands. Double tap, long press or other tap-based gestures are not specially separated from tap gestures in terms of their similarity to corresponding single tap gestures. We will revisit this point later in our discussion.

The general design space of hand-to-hand gestures consists of thousands of gesture instances, which is too large to be studied. However, not every combination of contact and motion constitute a usable gesture. In this work, we derive a gesture vocabulary with 56 candidate gestures (41 tap gestures and 15 swipe gestures) based on our technical overview, user interviews, and fundamentals of ergonomic theory. Fig. 10 shows the gesture vocabulary arranged by relevance and named by design factors.

4.3 Subjective Evaluation

After deriving the gesture vocabulary, we sought to find out which gestures were most suited for real-life interactions. As there is a lack of quantitative metrics to measure hand gestures, we asked users to score each gesture as in previous works [6, 13, 14] to provide subjective evidence for the selection of usable gestures.

We recruited eighteen participants (eight male and ten female) for our subjective evaluation, with an average age of 23.7 (from 19 to 47, SD=6.2). All of participants were right-handed. We considered four measures in our evaluation:

- ***Usability*** [6, 13] measured ergonomics to reflect comfort of the gesture. The participants are required to consider the gesture not only in stationary conditions (e.g., sitting) but also under moving conditions (e.g., running). The higher the score, the easier the gesture is to perform.
- ***Memorability*** [6, 14, 30] measures the intuitiveness and learnability as a function of how long the gesture can be remembered. The higher the score, the easier the gesture is to memorize.
- ***Social Acceptance*** [6, 13, 32] measures if the user will feel uncomfortable or embarrassed, or if performing the gesture will disturb others in public settings. The higher the score, the more acceptable the gesture is in social environments.
- ***Disambiguity*** measures the difficulty of confusing the gesture with daily hand movements or with other gestures. The higher the score, the less ambiguous the gesture is.

At the beginning of the evaluation, we explained each measure to the participant. All gestures were shown on a web page with descriptions and example videos. The participants could rate gestures in any order. We divided gestures into groups, where similar or related gestures were put in the same group to facilitate comparison. For example, we placed Palm-to-Palm, Palm-to-Back, Back-to-Palm and Back-to-Back in the same group. We required the participants to view all gestures and practice them with their hands. Then, we asked participants to rate each gesture on a 7-point Likert scale for each of the four measures. After scoring the gestures, we also required the participants to revisit all gestures and encouraged them to adjust their scores so as to focus on the relative score among gestures. The whole subjective evaluation lasted for thirty to sixty minutes. Participants' scores are shown in Fig. 10.

The average scores of hand-to-hand gestures on *Usability*, *Memorability*, *Social Acceptance* and *Disambiguity* were 5.51 (4.28-6.83, SD=0.67), 5.31 (3.83-6.72, SD=0.71), 5.30 (4.11-6.72, SD=0.62) and 5.34 (4.16-6.00, SD=0.42), respectively. The results also showed some interesting correlations: 1) Some gestures with very high *Usability* were rated low on *Disambiguity*, such as Palm-to-Palm and FistFront-to-Palm, which were not suitable for real scenarios due to the high rate of false positives by the recognition system. 2) Gestures in which bones of the hands make contact with one another, such as FistFront-to-Back and FistFront-to-Top, were less preferred. Similarly, top/bottom gestures were rated lower than fist top/bottom gestures. 3) For swipe gestures, index finger on the dominant hand was more preferred than palm, back or bottom.

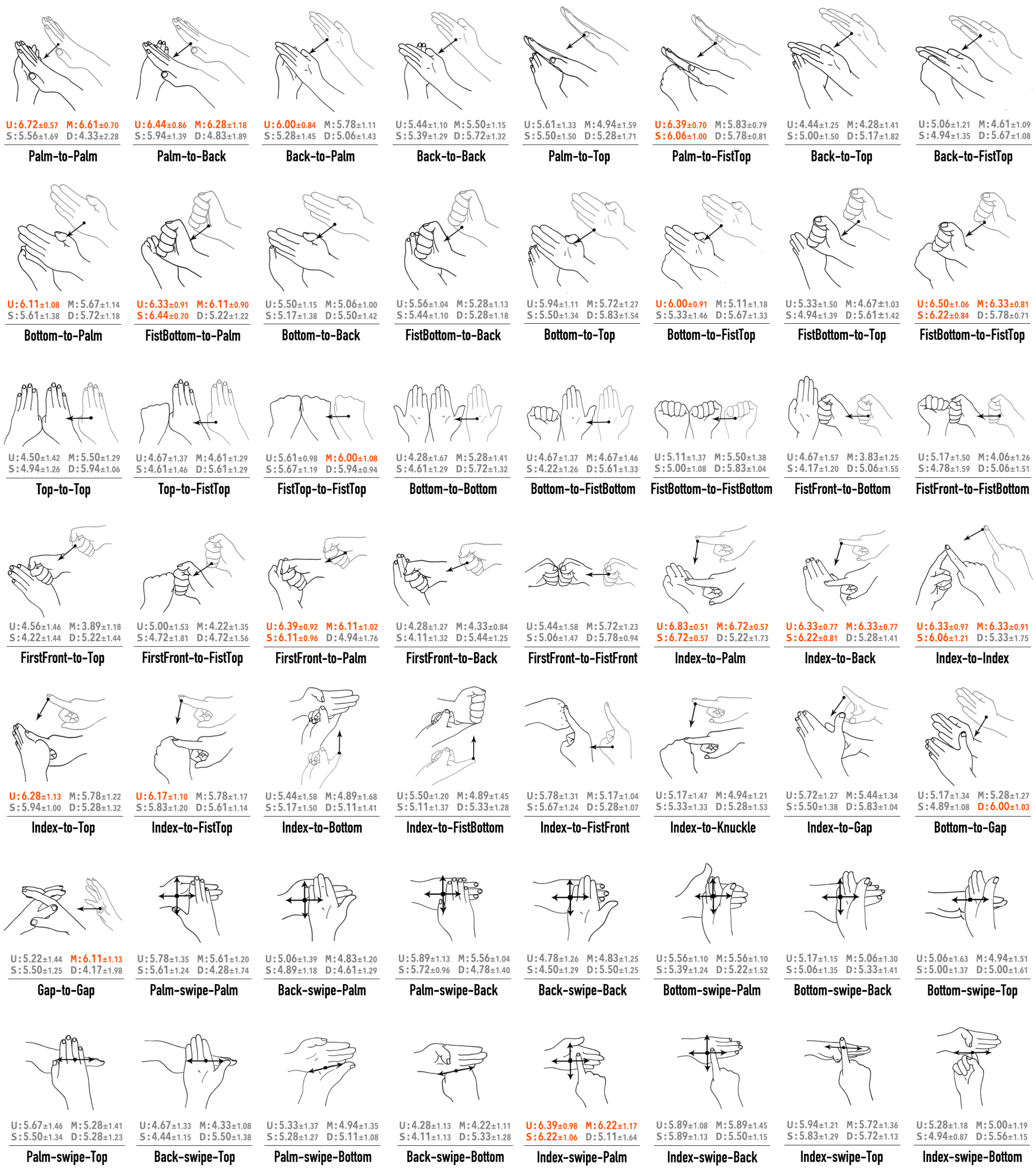


Fig. 10. Gesture vocabulary with 41 tap gestures and 15 swipe gestures arranged in terms of the relevance. Each gesture is named as [Contact Area of Right Hand]-[Motion Type: to/swipe]-[Contact Area of Left Hand]. The mean±sd. of users' subjective scores (1-7, the higher the better) on *Usability* (U), *Memorability* (M), *Social Acceptance* (S) and *Disambiguity* (D) is also shown. Scores are highlighted if they are no less than 6.

5 HAND-TO-HAND GESTURE RECOGNITION

The goal of this study is to verify the recognition capability of dual wrist-worn devices for hand-to-hand gestures. We consider the ergonomic constraints for a set of scenarios and select corresponding gesture sets. We then collect data from our recognition system and report the results for each gesture set.

5.1 Gesture Sets in Real Scenarios

Based on users' feedback in the subjective evaluation, we reviewed the pros and cons of each gesture on four measures. After the review, we sought to derive useful gesture sets for real-life scenarios. We consider two motor conditions in this study as informed by our users' feedback:

- **Stationary condition:** Typical scenarios are indoor and in-office environments, including remote control, VR/AR, etc. Under this condition, users sit or stand with their bodies relatively still and perform hand-to-hand gestures without motor restrictions.
- **Moving condition:** Typical scenarios include inputting a smartphone command input while walking or running. Under this condition, users' bodies are continuously moving, and the quality of performing hand gestures is limited by incidental arm swing.

From a technical perspective, the motion generated by hand contact under the stationary condition is high-quality and clear, while the motion data from the moving condition will contain more noises. From the user's perspective, the user's input ability is more limited under the moving condition as they have to be aware of their surroundings, and there will be fewer input requirements when walking or running, such as answering and hanging up the call or controlling music players. On the other hand, the input requirements (i.e., task complexity) for the number of gestures in the stationary condition is higher, such as controlling various electronic devices. For the sake of balancing the usability and memorability [42], we selected ten gestures for the stationary condition and five gestures for the moving condition from the gesture vocabulary with high scores on all measures (shown in Table. 1).

Table 1. Gesture sets selected for evaluation. The subjective scores for gestures on *Usability* (U), *Memorability* (M), *Social Acceptance* (S) and *Disambiguity* (D) are also shown.

(a) Gesture set for the stationary condition.

Gesture Name		U	M	S	D
Index-to-Palm		6.83	6.72	6.72	5.22
Index-to-Back		6.33	6.33	6.22	5.28
Index-to-Index		6.33	6.33	6.06	5.33
Index-to-FistTop		6.17	5.78	5.83	5.61
FistBottom-to-Palm		6.33	6.11	6.44	5.22
Palm-to-FistTop		6.39	5.83	6.06	5.78
FistBottom-to-FistTop		6.50	6.33	6.22	5.78
FistTop-to-FistTop		5.61	6.00	5.67	5.94
Index-swipe-Palm	↕↔	6.39	6.22	6.22	5.11
Index-swipe-Top	↔	5.94	5.72	5.83	5.72
Average		**6.28**	**6.14**	**6.13**	**5.50**

(b) Gesture set for the moving condition.

Gesture Name	U	M	S	D
Index-to-Palm	6.83	6.72	6.72	5.22
Index-to-Back	6.33	6.33	6.22	5.28
FistBottom-to-Palm	6.33	6.11	6.44	5.22
Palm-to-FistTop	6.39	5.83	6.06	5.78
FistBottom-to-FistTop	6.50	6.33	6.22	5.78
Average	**6.58**	**6.27**	**6.33**	**5.46**

Under the stationary condition, we chose eight tap gestures and two swipe gestures with high scores in all measures. The average scores of *Usability*, *Memorability*, *Social Acceptance* and *Disambiguity* of the gesture set are 6.28 (5.61-6.83, SD=0.33), 6.14 (5.72-6.72, SD=0.31), 6.13 (5.67-6.72, SD=0.31) and 5.50 (5.11-5.94, SD=0.30), respectively, which are 115%, 117%, 134% and 39% standard deviations higher than the mean scores of the gesture vocabulary, respectively. We involve one bi-directional swipe gesture (Index-swipe-Top, left/right) and one quad-directional swipe gesture (Index-swipe-Palm, left/right/up/down) with high scores.

Under the moving condition, we are more concerned about the usability than other measures due to the user's motor limitations. We selected five easy-to-perform tap gestures from the stationary condition's gesture set. The average scores of *Usability*, *Memorability*, *Social Acceptance* and *Disambiguity* of the gesture set are 6.48 (6.33-6.83, SD=0.21), 6.27 (5.83-6.72, SD=0.33), 6.33 (6.06-6.72, SD=0.26) and 5.46 (5.22-5.78, SD=0.30), respectively, which are 144%, 135%, 167% and 28% standard deviation higher than the mean scores, respectively. Fig. 11 shows the statistical differences of gesture sets.

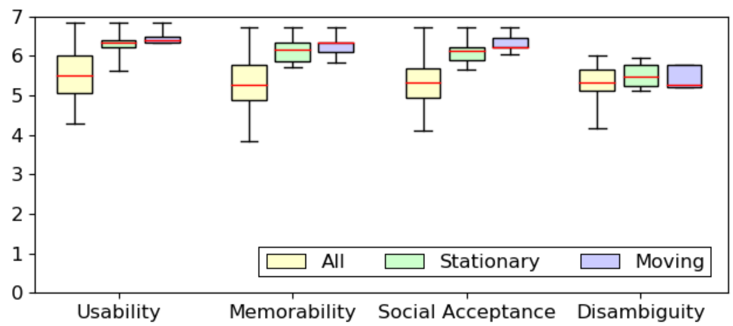


Fig. 11. The boxplot of scores of evaluated gesture sets on four measures.

5.2 Data Collection Procedure

We conducted two user experiments to collect sensor data from hand-to-hand gestures in real-life scenarios. We recruited nineteen participants, and all of them were right-handed.

5.2.1 Stationary Condition. Nine participants (six male and three female) were recruited in the data collection for the stationary condition, with an average age of 23.7 (from 19 to 28, SD=1.6). The tap gestures and swipe gestures were collected separately. We first collected the data for tap gestures. The experiment was divided into ten sessions. In each session, we required the participants to perform all eight tap gestures five times each, for a total of 50 samples for each tap gesture. We then collected the data from two swipe gestures, which consist of six swipe sub-gestures, including Index-swipe-Palm (left/right/up/down) and Index-swipe-Top (left/right). The swipe sub-gestures were collected using the same procedure as the tap gestures, with 50 samples for each swipe sub-gesture.

At the beginning of the experiment, we introduced the goal and the procedure. We required participants to sit still in front of a display in the lab office with the sound level of about 50dB. We required participants to wear two smartwatches at a relatively tight setting, (i.e., users did not report feeling uncomfortable and no slippage between watch straps and wrists). Next, we introduced eight tap gestures under the stationary condition and required the participants to practice these gestures to warm up. After the participants understood and were familiar with the gestures, the data-gathering sessions portion of the experiment began. At the beginning of this portion, the participants were first required to synchronize the two smartwatches using the "Sync." button on the app (for timestamp synchronization in data collection, as Fig. 2). Then, in each session, eight gestures were presented in a random order for counterbalancing. We showed a photo and description of each gesture to inform the participants about the current gesture that he or she should perform. The participants had a half to one minute break between every three sessions.

We also collected ambient data under the stationary condition. The participants were required to complete three tasks: 1) on-body interaction, including but not limited to touching different parts of the body, wearing and removing glasses, and adjusting their hair; 2) on-table interaction, including but not limited to operating devices like keyboards, mouses and phones, grabbing and dropping objects like cups, phones and books; 3) social interaction, including but not limited to waving and shaking hands, stand up and sit down, and hand moving during speech. The ambient data were collected for 2 minutes. The whole experiment lasted for about 45 minutes.

5.2.2 Moving Condition. The other ten participants (six male and four female) were recruited in the data collection for the moving condition, with an average age of 22.3 (from 19 to 25, SD=1.8). We collected data under two sub-conditions: walking and running. In each sub-condition, the experiment was also divided into ten sessions. In each session, the participants were required to perform all gestures five times each, with a total of 50 samples for each gesture in each sub-condition.

The experiments under the moving conditions were conducted in the hall corridor outside the lab. The walking and running sub-conditions were presented in a random order for counterbalancing. The procedure for each sub-condition was similar to the stationary condition. First, we required the participants to wear two smartwatches at a relatively tight setting, and we introduced the gestures. After the participants had practiced the gestures and warmed up, the experiment started and the participants were required to synchronize the watches. In each session, they started walking/running and performing gestures. The gestures were also presented in a random order, and we informed the participants about the current gesture using audio feedback. The participants had a one to two minute break between every three sessions in the walking sub-condition and every session in the running sub-condition.

We also collected ambient data under the moving condition. We required participants to walk or run with natural hand movements, including incidental arm swing, touching the body and clothes, and waving hands at bystanders. The whole experiment lasted for about 30 minutes.

5.3 Recognition Method and Results

We segmented sensor data from gestures using a 500ms window as positive samples. In total, we collected 14 (gestures) $\times$ 9 (participants) $\times$ 50 (times) = 6300 positive samples in the stationary condition and 2 (sub-conditions) $\times$ 5 (gestures) $\times$ 10 (participants) $\times$ 50 (times) = 5000 positive samples in moving conditions. We also randomly segmented from ambient data using a 500ms window and collected 63000, 25000, and 25000 negative samples under the stationary condition, walking sub-condition and running sub-condition, respectively. The window size of 500ms was used for all samples in feature extraction.

We report both detection and recognition results. With regards to detection, we used a Support Vector Machine (SVM) with Radial Based Function (RBF) kernel to classify hand-to-hand gestures and ambient hand movements. Because the ambient data has significantly more samples in real use, we balanced the training data of positive and negative samples to 1:10 to avoid negatively skewing our classification model in each of the stationary condition, walking sub-condition, and running sub-condition. The features we used in the classification model were derived from the IMU data of both devices, including statistics (min/max/mean/sd.) of the acceleration, rotation velocity and the correlation.

For recognition, we also used a SVM with an RBF kernel to classify the gestures. In the stationary condition, features could be divided into two categories: inertial features and acoustic features. Inertial features consisted of statistics (min/max/mean/sd.) of the acceleration, rotation velocity, and orientation (including relative rotation between two devices in quaternion, and pitch and roll angles. Yaw angle was not considered in order to exclude the effect of the body's direction). For acoustic features, we applied an MFCC feature extraction sub-process in frequency domain (26 features on each hand in total). In moving conditions, however, we only used inertial features, because we thought acoustic signals would become unreliable outdoors.

We uniformly used leave-one-out cross validation to report all classification results. That is, we tested the data from each participant with the model trained by the data from other participants, then averaged the results.

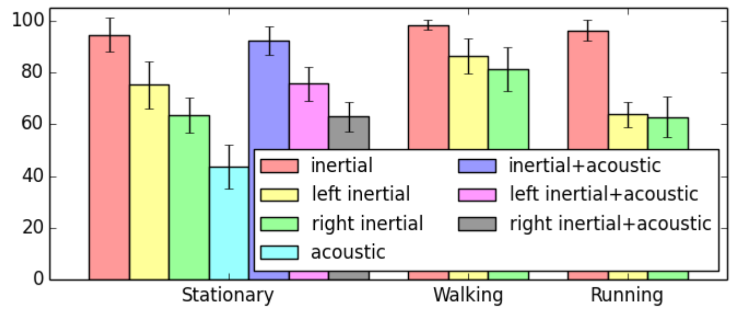


Fig. 12. Accuracy of gesture recognition using different configurations of features.

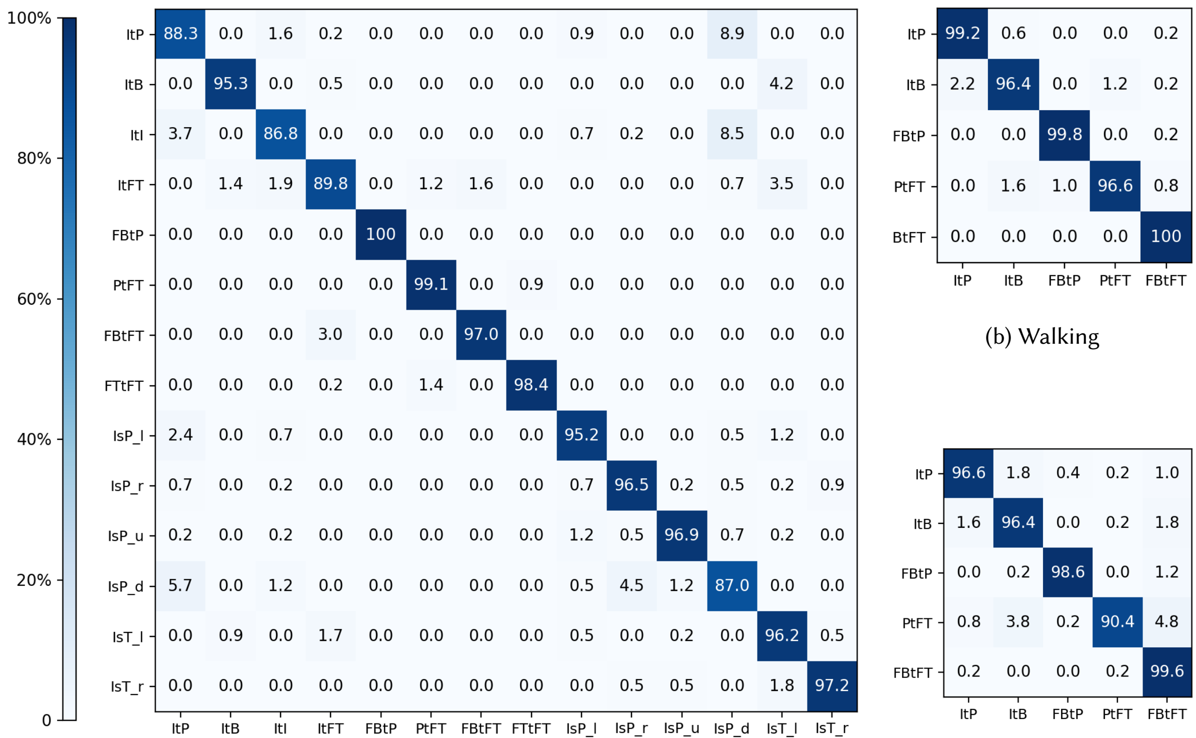
5.3.1 Stationary Condition. We first report the detection accuracy, which reflects the performance of extracting hand-to-hand gestures from ambient noise. During our analysis, we discovered that occasionally swipe gestures can be overwhelmed by ambient noise because the motions of swipe gestures are subtle and noises in the stationary condition are widely varied. So we believe it is necessary to apply a mode switch for the use of swipe gestures (i.e., perform a preceding tap gesture to notify the detector of the swipe input, then perform swipe gestures in a noiseless environment). Therefore, the detection reported here only includes tap gestures and ambient noises. The detection accuracy was 99.5%, with 96.6% precision and a 94.1% recall of positive samples.

We then reported the recognition performance of gestures in different configurations of sensors. The recognition accuracy for all fourteen gestures (eight tap gestures and six swipe sub-gestures) with dual wrist-worn devices using only inertial features was 94.6%. The confusion matrix of gestures is shown in Fig. 13a.

We also report the performance using only single wrist-worn device to illustrate the benefit of using two devices. The recognition accuracy with the left or right wrist-worn device using inertial features was 75.3% and 63.6%, respectively. RM-ANOVA showed the accuracy with two devices was significantly higher than the left ($F_{1,8} = 91.1$, $p < .001$) and right ($F_{1,8} = 428$, $p < .001$) device. This demonstrates that motion-based bimanual gesture detection can be improved by the use of two devices.

The recognition accuracy using only acoustic features of two devices was 43.7%, and RM-ANOVA showed that it was significantly lower than using inertial features ($F_{1,8} = 181$, $p < .001$). We also evaluated the performance using both inertial and acoustic features. However, the overall accuracy with both inertial and acoustic features was not higher than that with only inertial features (as shown in Fig. 12), and RM-ANOVA showed that the difference was not significant ($F_{1,8} = 2.54$, $p = .149$). We speculate this is because the gesture set did not contain multiple gestures with similar hand orientations and motion directions, and inertial features already had good performance for classification.

5.3.2 Moving Condition. The detection accuracy of hand-to-hand gestures in the walking and running sub-conditions was 99.8% and 99.6%, respectively. The precision and recall of positive samples were 99.8% and 99.9% under the walking sub-condition, while they were 100% and 99.6% under the running condition. The results show the performance of detecting hand-to-hand gestures from natural hand movements in moving conditions is very high (all greater than 99%), which indicates dual wrist-worn devices are reliable when the body is moving.



(a) Stationary

(b) Walking

(c) Running

Fig. 13. Confusion matrices for gesture recognition using dual inertial sensors in all conditions. The gesture is annotated as the abbreviation of the gesture name in Table 1.

The recognition accuracy of five gestures in walking and running sub-conditions was 98.4% and 96.3%, respectively. The confusion matrices of gestures are shown in Fig. 13b and Fig. 13c. Results showed hand-to-hand gestures could be also well used under moving conditions, which might enrich the interaction in outdoor and sports scenarios.

We also evaluated the performance of a single wrist-worn device. The recognition accuracy in the walking sub-condition using only the left or right wrist-worn devices was 86.3% and 63.8% (shown in Fig. 12), respectively, and in the running sub-condition was 81.2% and 62.7%, respectively. RM-ANOVA showed the accuracy with dual wrist-worn devices was significantly higher than the left ($F_{1,9} = 34.5$, $p < .001$) or right ($F_{1,9} = 151$, $p < .001$) device in the walking sub-condition, and also higher than the left ($F_{1,9} = 129$, $p < .001$) or right ($F_{1,9} = 185$, $p < .001$) device in the running sub-condition. This result was in line with that of the stationary condition, and demonstrates the superiority of sensing with two wrist-worn devices as opposed to one.

6 USING HAND-TO-HAND GESTURES IN REAL CONTEXTS

In this section, we validate the usability of hand-to-hand gestures and our gesture inference system in real-life contexts with smartphone applications and tasks.

6.1 Apparatus and Participants

We used the same hardware platform as in the previous studies. We introduced the configuration of the hardware and how we built the real-time sensing system in Section 3. We used the classification models (i.e., detection and recognition models) trained in Section 5 for real-time gesture inference. In summary, we used two smartwatches to record sensor data and transmit the data to the smartphone. The smartphone synchronizes two data streams and segments a 500ms window at 4Hz. The smartphone uses the detection model to check if a hand-to-hand gesture occurs in the window, if so, the smartphone then uses the recognition model to obtain the gesture type and invoke the corresponding command.

We recruited ten participants (seven male and three female) with an average age of 24.1 (from 21 to 28, SD=1.5), all right-handed.

6.2 Tasks and Procedure

In this study, we evaluated our system in a phone-in-pocket scenario. Users triggered a smartphone function by performing a hand-to-hand gesture without taking the phone out of their pocket. We chose *Phone Call*, *Music Player* and *Message* as typical applications and mapped each smartphone command with a gesture selected from gesture sets in Table 1 in Section 5. The mapping between gestures and commands, which was informed by the results of our pilot study, is shown in Table 2 below.

Table 2. Mapping between commands and gestures

Application	**Command**	**Gesture**
Phone Call	Answer Call	Index-to-Back
	Reject Call	Index-to-Palm
Music Player	Play/Pause	FistBottom-to-FistTop
	Next Music	FistBottom-to-Palm
	Previous Music	Palm-to-FistTop
Message	Read Message	Index-to-FistTop
	Delete Message	Index-to-Index
	Reply Message	FistTop-to-FistTop

We evaluated hand-to-hand gestures under stationary, walking, and running conditions. In the stationary condition, we required the participants to invoke all eight commands. We did not involve swipe gestures in this study because of the difficulty of distinguishing the swipe motion from natural hand movements according to our results in Section 5. In the walking and running conditions, the participants were only required to invoke commands of *Phone Call* and *Music Player* applications.

In this experiment, the participants wore one smartwatch on each wrist and wore earphones to receive audio notifications. When the experimenter issued a command, the participants would receive an audio stimulus telling them to perform the corresponding gesture. For example, if the participants heard "Message received from Alice", he or she would then perform an Index-to-FistTop gesture to read the message. After the system recognized a gesture, the participants would receive audio feedback. If the system did not detect the gesture, the participants were allowed to repeat the gesture, otherwise, the system moved on to the next command.

We compared hand-to-hand gesture interaction with the traditional method, in which a user took the phone out of their pocket and tapped a button on the screen to invoke the command, after receiving audio stimulus.

Before the experiment, we introduced hand-to-hand gestures to the participants and required them to remember the mapping. Next, the participant had 5-10 minutes to get familiar with the gestures and commands as well as the experiment procedure. Then, the experiment started, and the participant was required to finish twenty commands under each of the stationary, walking, and running conditions. The presented order of conditions was counterbalanced. The requirements of the conditions were the same as those in Section 5, and the participants were allowed to move their hands freely during idle time between the two commands. After the gesture session, the participants were required to complete the same tasks by taking the phone out of their pocket. We also asked participants for comments on hand-to-hand gestures in real-life use cases.

6.3 Results

We reported four types of errors: 1) *User error*: participant performed the wrong gesture. 2) *Recognition error*: system mistakenly recognizes a gesture as another. When the participant invoked an incorrect command, the experimenter would confirm whether the participant did a correct gesture (yes: recognition error; no: user error) and recorded the type. 3) *False positive*: system detected a gesture when the participant did not perform it. The system could automatically log these errors. 4) *False negative*: participant performed a gesture but the system failed to detect it. The experimenter manually recorded these errors. Task success rates and error rates in all conditions are shown in Fig. 14.

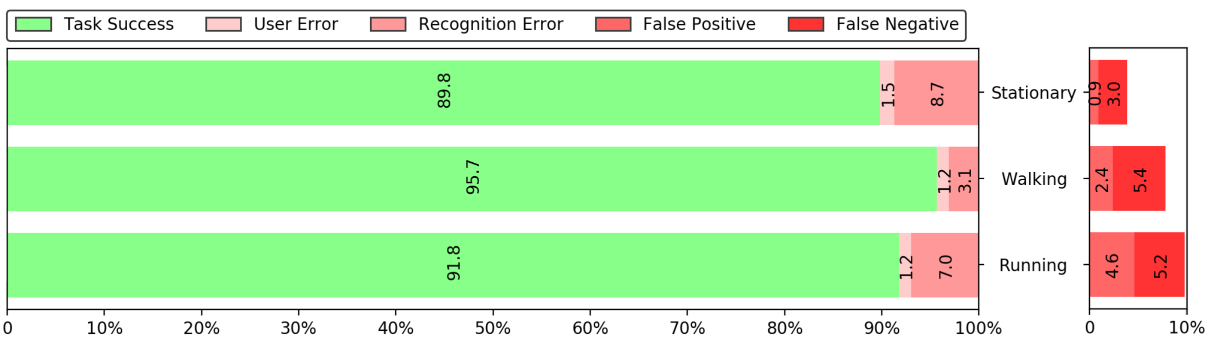


Fig. 14. Task success rates and error rates in all conditions.

Task success rates were over 89% under real-life use scenarios, which suggests that this interaction method is viable outside of the lab. User error rates were low, which indicated that the mapping between hand-to-hand gestures and commands was easy to remember in 5-10 minutes. The recognition error, false positive, and false negative rates were a little higher than the results reported in Section 5. Based on our observations, we speculated more errors were caused by users' incidental movements of their wrists when lifting their hands or putting hands down. Sometimes the participants performed gestures while lifting their hands or started putting their hands down before completing the gestures. These cases were not emphasized in our classification model training, which is the limitation of our system and can be improved by collecting more targeted training data.

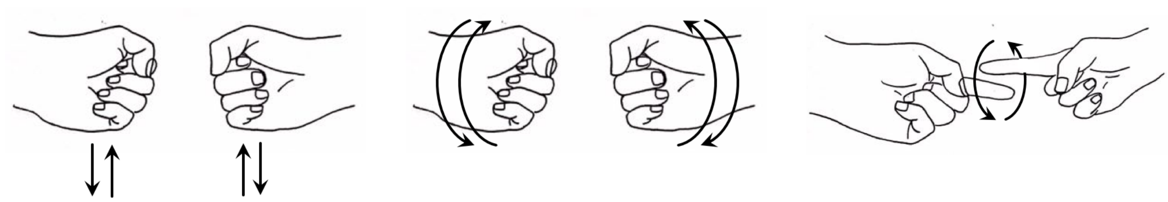
We also examined the task completion time, which we computed as the interval from a command being issued to the user to the system successfully recognizing the input. The average task completion time was 3.76s (SD=1.57), 4.27s (SD=1.77) and 3.90s (SD=2.07) under the stationary, walking, and running conditions, respectively, while that of taking the phone out of the pocket was 6.68s (SD=2.32). This result showed that using hand-to-hand gestures was 36% faster than the out-of-the-pocket interaction under the walking condition.

We asked the participants about the advantages and disadvantages of hand-to-hand gesture input. Participants generally provided positive comments with regard to this form of interaction: "When running, my hands are exactly in that place for gestures", "I can control my smartphone without looking at the screen and my hands", "It is acceptable for me to try again if a gesture does not respond". Participants also raised some concerns of gestures: "Palm-to-FistTop is sometimes loud, which makes me feel embarrassed", "FistTop-to-FistTop is sometimes painful".

7 DISCUSSION

7.1 Expanding Gesture Vocabulary for Dual Wrist-Worn Devices

We leverage the correlation of the motion generated by the hand contact on both left and right wrists to design and recognize hand-to-hand gestures. We show several gestures without hand contact in Fig. 15 that generate special and correlative motion. For example, bimanual simultaneous wrist twist (Fig. 15b) generates a pair of homodromous motions with large rotation velocity, which can be easily sensed by dual wrist-worn gyroscopes. This gesture is distinct from natural hand movements and has been used as a delimiter in previous work [20]. Bearing this in mind, gesture design for dual wrist-worn devices is not limited by hand-to-hand contact. Any hand gesture with correlative motion can also be considered.



(a) Simultaneous fist swing (b) Simultaneous wrists twist (c) Finger circles

Fig. 15. Other bimanual gestures which can also be sensed by the correlative motions using dual wrist-worn devices.

We could also enrich the gesture set by introducing a temporal element. In our interviews, some participants mentioned that double tap or long press could be variants of the single tap gesture. These variants can be used to supplement multi-level commands, for example, single tap to turn on and double tap to turn off. In our work, we did not evaluate these form factors in subjective scoring, but we think single tap gestures and their variants may have similar scores in all measures, especially in *Usability* and *Social Acceptance*.

The contact area of bimanual gestures can be also extended beyond the hand. Previous works have explored touch interactions on the forearm [53] or on the wrist-worn device itself [47]. To leverage the correlation between the inertial data collected by two wrist-worn devices, unimanual on-body gestures [36, 51, 52, 60] can be also extended as symmetric touch gestures with two hands. For example, left and right hands tap on the chest at the same time.

7.2 Sensing Capability & Hand-to-Hand Gesture Design Space

We do not explore the full design space of hand-to-hand gestures in this paper due to the limited sensing capability. As we mentioned in Section 3, we use off-the-shelf smartwatches as sensing devices, and the wrist-worn sensors cannot directly sense finger states, such as fist clenching, finger pinch, etc. In addition, current built-in sensors on commercial wrist-worn devices are not so reliable in all conditions: features derived by microphone sensors may be influenced by ambient noise; PPG sensors (measuring heart rate) are error prone during hand movements, etc.

We coupled our gesture vocabulary with a sensing capability. According to our results in Section 4, the users generally agreed that states of open hand and fist should be considered as two different gestures. If sensing hand states is possible, for example, adding EMG [41, 41], pressure [12, 41, 47] or infrared sensors [39] to wrist-worn devices, we can involve this factor in the design space and consider more finger gestures.

We also do not subdivide the contact areas in one gesture. For example, some of the participants we interviewed could distinguish several locations when touching fingers on the palm with the help of hand edges or reference points, even while running. In the stationary condition, users' accuracy can be much higher. Previous works [70, 73] have explored recognizing touch locations of a dial layout on the back of the hand with wrist-worn devices. However, the resolution in real-life contexts has not been investigated.

7.3 Limitations

First, our gesture design procedure is preliminary. We only considered limited design factors and motion types in the design space. Proposed gestures were only evaluated in the stationary condition, and require further investigation under mobile conditions. Second, we did not consider different preferences for tightness of the smartwatch. In our work, we required the participants to wear the watches in a relatively tight way. However, the tightness might affect the sensitivity of sensors and influence the system's recognition performance [74]. Also, we did not study the effect of gloves. Third, we did not study gesture-command mapping. We did not follow Wobbrock et al's work [67] to conduct user elicitation studies to design gestures in particular scenarios [9, 14, 30, 32], instead, the mapping of real tasks was pre-defined by the experimenters. How to design easy-to-remember mappings between commands and gestures needs to be further investigated in future works.

8 CONCLUSION

In this work, we preliminarily explore the design space of hand-to-hand gestures and the ability to recognize them using dual off-the-shelf wrist-worn devices. We come up with three design factors and design 56 candidate hand-to-hand gestures based on user interviews. We selected gesture sets for a stationary condition with fourteen gestures and moving conditions with five gestures based on subjective feedback from participants. We also present a sensing technique with dual consumer wrist-worn devices, including useful features like relative device orientation and motion correlation, and examine the performance of gesture inference. The detection accuracy of hand-to-hand gestures in the context of users' natural hand movements is over 99% under both stationary and moving conditions. The recognition accuracy of fourteen gestures in the stationary condition is 94.6%, and the accuracy of five gestures in walking and running conditions is 98.4% and 96.3%, respectively. Our evaluation in real-life task-driven contexts shows that our system is practical and yields a high success rate. We conclude that dual off-the-shelf wrist-worn devices are effective tools in sensing touch-based hand gestures. Our results suggest that hand-to-hand gestures have the potential to be used in people's daily lives to enrich interactions in many scenarios.

ACKNOWLEDGMENTS

We would like to thank all reviewers for their helpful comments. We would also like to thank Matt Andryc for his help in writing this paper. This work is supported by the National Key Research and Development Plan under Grant No. 2016YFB1001200, the Natural Science Foundation of China under Grant No. 61521002, No. 61672314, and also by Beijing Key Lab of Networked Multimedia.

REFERENCES

[1] S Alavi, D Arsenault, and A Whitehead. 2016. Quaternion-Based Gesture Recognition Using Wireless Wearable Motion Capture Sensors. *Sensors* 16, 5 (2016), 605.

[2] J. Allen. 1977. Short term spectral analysis, synthesis, and modification by discrete Fourier transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 25, 3 (June 1977), 235–238. https://doi.org/10.1109/TASSP.1977.1162950

[3] Brian Amento, Will Hill, and Loren Terveen. 2002. The Sound of One Hand: A Wrist-mounted Bio-acoustic Fingertip Gesture Interface. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems (CHI EA '02)*. ACM, New York, NY, USA, 724–725. https://doi.org/10.1145/506443.506566

[4] Antonio Augimeri, Giancarlo Fortino, Manoj R Rege, Vlado Handziski, and Adam Wolisz. 2010. A cooperative approach for handshake detection based on body sensor networks.. In *SMC*. 281–288.

[5] Joanna Bergstrom-Lehtovirta, Sebastian Boring, and Kasper Hornbæk. 2017. Placing and Recalling Virtual Items on the Skin. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 1497–1507. https://doi.org/10.1145/3025453.3026030

[6] Idil Bostan, Oğuz Turan Buruk, Mert Canat, Mustafa Ozan Tezcan, Celalettin Yurdakul, Tilbe Göksun, and Oğuzhan Özcan. 2017. Hands As a Controller: User Preferences for Hand Specific On-Skin Gestures. In *Proceedings of the 2017 Conference on Designing Interactive Systems (DIS '17)*. ACM, New York, NY, USA, 1123–1134. https://doi.org/10.1145/3064663.3064766

[7] Daniel Buschek, Bianka Roppelt, and Florian Alt. 2018. Extending Keyboard Shortcuts with Arm and Wrist Rotation Gestures. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 21, 12 pages. https://doi.org/10.1145/3173574.3173595

[8] W Buxton and B Myers. 1986. A study in two-handed input. *Proc Chi* 17, 4 (1986), 321–326.

[9] Edwin Chan, Teddy Seyed, Wolfgang Stuerzlinger, Xing-Dong Yang, and Frank Maurer. 2016. User Elicitation on Single-hand Microgestures. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 3403–3414. https://doi.org/10.1145/2858036.2858589

[10] Xiang 'Anthony' Chen, Tovi Grossman, Daniel J. Wigdor, and George Fitzmaurice. 2014. Duet: Exploring Joint Interactions on a Smart Phone and a Smart Watch. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 159–168. https://doi.org/10.1145/2556288.2556955

[11] Xiang 'Anthony' Chen and Yang Li. 2016. Bootstrapping User-Defined Body Tapping Recognition with Offline-Learned Probabilistic Representation. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. ACM, New York, NY, USA, 359–364. https://doi.org/10.1145/2984511.2984541

[12] Artem Dementyev and Joseph A. Paradiso. 2014. WristFlex: Low-power Gesture Input with Wrist-worn Pressure Sensors. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST '14)*. ACM, New York, NY, USA, 161–166. https://doi.org/10.1145/2642918.2647396

[13] Tilman Dingler, Rufat Rzayev, Alireza Sahami Shirazi, and Niels Henze. 2018. Designing Consistent Gestures Across Device Types: Eliciting RSVP Controls for Phone, Watch, and Glasses. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 419, 12 pages. https://doi.org/10.1145/3173574.3173993

[14] H. Dong, A. Danesh, N. Figueroa, and A. E. Saddik. 2015. An Elicitation Study on Gesture Preferences and Memorability Toward a Practical Hand-Gesture Vocabulary for Smart Televisions. *IEEE Access* 3 (2015), 543–555. https://doi.org/10.1109/ACCESS.2015.2432679

[15] Jun Gong, Xing-Dong Yang, and Pourang Irani. 2016. WristWhirl: One-handed Continuous Smartwatch Input Using Wrist Gestures. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. ACM, New York, NY, USA, 861–872. https://doi.org/10.1145/2984511.2984563

[16] Yves Guiard. 1987. Asymmetric Division of Labor in Human Skilled Bimanual Action. *Journal of Motor Behavior* 19, 4 (1987), 486–517. https://doi.org/10.1080/00222895.1987.10735426 arXiv:https://doi.org/10.1080/00222895.1987.10735426 PMID: 15136274.

[17] Sean Gustafson, Daniel Bierwirth, and Patrick Baudisch. 2010. Imaginary Interfaces: Spatial Interaction with Empty Hands and Without Visual Feedback. In *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology (UIST '10)*. ACM, New York, NY, USA, 3–12. https://doi.org/10.1145/1866029.1866033

[18] Sean Gustafson, Christian Holz, and Patrick Baudisch. 2011. Imaginary Phone: Learning Imaginary Interfaces by Transferring Spatial Memory from a Familiar Device. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (UIST '11)*. ACM, New York, NY, USA, 283–292. https://doi.org/10.1145/2047196.2047233

[19] Sean G. Gustafson, Bernhard Rabe, and Patrick M. Baudisch. 2013. Understanding Palm-based Imaginary Interfaces: The Role of Visual and Tactile Cues when Browsing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 889–898. https://doi.org/10.1145/2470654.2466114

[20] Teng Han, Khalad Hasan, Keisuke Nakamura, Randy Gomez, and Pourang Irani. 2017. SoundCraft: Enabling Spatial Interactions on Smartwatches Using Hand Generated Acoustics. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST '17)*. ACM, New York, NY, USA, 579–591. https://doi.org/10.1145/3126594.3126612

[21] Faizan Haque, Mathieu Nancel, and Daniel Vogel. 2015. Myopoint: Pointing and Clicking Using Forearm Mounted Electromyography and Inertial Motion Sensors. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 3653–3656. https://doi.org/10.1145/2702123.2702133

[22] Chris Harrison, Desney Tan, and Dan Morris. 2010. Skinput: Appropriating the Body As an Input Surface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 453–462. https://doi.org/10.1145/1753326.1753394

[23] Ken Hinckley. 2003. Synchronous Gestures for Multiple Persons and Computers. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology (UIST '03)*. ACM, New York, NY, USA, 149–158. https://doi.org/10.1145/964696.964713

[24] Lars Erik Holmquist, Friedemann Mattern, Bernt Schiele, Petteri Alahuhta, Michael Beigl, and Hans-W Gellersen. 2001. Smart-its friends: A technique for users to easily establish connections between smart artefacts. In *international conference on Ubiquitous Computing*. Springer, 116–122.

[25] Simon Ismair, Julie Wagner, Ted Selker, and Andreas Butz. 2015. MIME: Teaching Mid-Air Pose-Command Mappings. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '15)*. ACM, New York, NY, USA, 199–206. https://doi.org/10.1145/2785830.2785854

[26] Oskar Juhlin and Elin Önnevall. 2013. On the Relation of Ordinary Gestures to TV Screens: General Lessons for the Design of Collaborative Interactive Techniques. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 919–930. https://doi.org/10.1145/2470654.2466118

[27] David Kim, Otmar Hilliges, Shahram Izadi, Alex D. Butler, Jiawen Chen, Iason Oikonomidis, and Patrick Olivier. 2012. Digits: Freehand 3D Interactions Anywhere Using a Wrist-worn Gloveless Sensor. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology (UIST '12)*. ACM, New York, NY, USA, 167–176. https://doi.org/10.1145/2380116.2380139

[28] Han-Jong Kim, Seijin Cha, Richard C. Park, Tek-Jin Nam, Woohun Lee, and Geehyuk Lee. 2016. Mo-Bi: Contextual Mobile Interfaces Through Bimanual Posture Sensing with Wrist-Worn Devices. In *Proceedings of HCI Korea (HCIK '16)*. Hanbit Media, Inc., South Korea, 94–99. https://doi.org/10.17210/hcik.2016.01.94

[29] Yuhwan Kim, Seungchul Lee, Inseok Hwang, Hyunho Ro, Youngki Lee, Miri Moon, and Junehwa Song. 2014. High5: Promoting Interpersonal Hand-to-hand Touch for Vibrant Workplace with Electrodermal Sensor Watches. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14)*. ACM, New York, NY, USA, 15–19. https://doi.org/10.1145/2632048.2632072

[30] Christine Kühnel, Tilo Westermann, Fabian Hemmert, Sven Kratz, Alexander Müller, and Sebastian Möller. 2011. I'm home: Defining and evaluating a gesture set for smart-home control. *International Journal of Human-Computer Studies* 69, 11 (2011), 693 – 704. https://doi.org/10.1016/j.ijhcs.2011.04.005

[31] Gierad Laput, Robert Xiao, and Chris Harrison. 2016. ViBand: High-Fidelity Bio-Acoustic Sensing Using Commodity Smartwatch Accelerometers. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. ACM, New York, NY, USA, 321–333. https://doi.org/10.1145/2984511.2984582

[32] DoYoung Lee, Youryang Lee, Yonghwan Shin, and Ian Oakley. 2018. Designing Socially Acceptable Hand-to-Face Input. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology (UIST '18)*. ACM, New York, NY, USA, 711–723. https://doi.org/10.1145/3242587.3242642

[33] Ho-Man Colman Leung, Chi-Wing Fu, and Pheng-Ann Heng. 2018. TwistIn: Tangible Authentication of Smart Devices via Motion Co-analysis with a Smartwatch. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 2, Article 72 (July 2018), 24 pages. https://doi.org/10.1145/3214275

[34] Mingyu Liu, Mathieu Nancel, and Daniel Vogel. 2015. Gunslinger: Subtle Arms-down Mid-air Interaction. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology (UIST '15)*. ACM, New York, NY, USA, 63–71. https://doi.org/10.1145/2807442.2807489

[35] Joe Marshall, Alexandru Dancu, and Florian "Floyd" Mueller. 2016. Interaction in Motion: Designing Truly Mobile Interaction. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS '16)*. ACM, New York, NY, USA, 215–228. https://doi.org/10.1145/2901790.2901844

[36] Denys J. C. Matthies, Simon T. Perrault, Bodo Urban, and Shengdong Zhao. 2015. Botential: Localizing On-Body Gestures by Measuring Electrical Signatures on the Human Skin. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '15)*. ACM, New York, NY, USA, 207–216. https://doi.org/10.1145/2785830.2785859

[37] Rene Mayrhofer and Hans Gellersen. 2007. Shake well before use: Authentication based on accelerometer data. In *International Conference on Pervasive Computing*. Springer, 144–161.

[38] Rene Mayrhofer and Hans Gellersen. 2009. Shake well before use: Intuitive and secure pairing of mobile devices. *IEEE Transactions on Mobile Computing* 8, 6 (2009), 792–806.

[39] Jess McIntosh, Asier Marzo, and Mike Fraser. 2017. SensIR: Detecting Hand Gestures with a Wearable Bracelet Using Infrared Transmission and Reflection. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST '17)*. ACM, New York, NY, USA, 593–597. https://doi.org/10.1145/3126594.3126604

[40] Jess McIntosh, Asier Marzo, Mike Fraser, and Carol Phillips. 2017. EchoFlex: Hand Gesture Recognition Using Ultrasound Imaging. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 1923–1934. https://doi.org/10.1145/3025453.3025807

[41] Jess McIntosh, Charlie McNeill, Mike Fraser, Frederic Kerber, Markus Löchtefeld, and Antonio Krüger. 2016. EMPress: Practical Hand Gesture Classification with Wrist-Mounted EMG and Pressure Sensing. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 2332–2342. https://doi.org/10.1145/2858036.2858093

[42] George A. Miller. 1994. The Magical Number Seven, Plus or Minus Two: Some Limits on Out Capacity for Processing Information. *Psychological Review* 101, 2 (1994), 343–352.

[43] Meredith Ringel Morris, Andreea Danielescu, Steven Drucker, Danyel Fisher, Bongshin Lee, m. c. schraefel, and Jacob O. Wobbrock. 2014. Reducing Legacy Bias in Gesture Elicitation Studies. *interactions* 21, 3 (May 2014), 40–45. https://doi.org/10.1145/2591689

[44] K. S. R. Murty and B. Yegnanarayana. 2006. Combining evidence from residual phase and MFCC features for speaker recognition. *IEEE Signal Processing Letters* 13, 1 (Jan 2006), 52–55. https://doi.org/10.1109/LSP.2005.860538

[45] Mathieu Nancel, Julie Wagner, Emmanuel Pietriga, Olivier Chapuis, and Wendy Mackay. 2011. Mid-air Pan-and-zoom on Wall-sized Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 177–186. https://doi.org/10.1145/1978942.1978969

[46] Michael Nielsen, Moritz Störring, Thomas B. Moeslund, and Erik Granum. 2004. A Procedure for Developing Intuitive and Ergonomic Gesture Interfaces for HCI. In *Gesture-Based Communication in Human-Computer Interaction*, Antonio Camurri and Gualtiero Volpe (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 409–420.

[47] Simon T. Perrault, Eric Lecolinet, James Eagan, and Yves Guiard. 2013. Watchit: Simple Gestures and Eyes-free Interaction for Wristwatches and Bracelets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 1451–1460. https://doi.org/10.1145/2470654.2466192

[48] Thammathip Piumsomboon, Adrian Clark, Mark Billinghurst, and Andy Cockburn. 2013. User-Defined Gestures for Augmented Reality. In *Human-Computer Interaction – INTERACT 2013*, Paula Kotzé, Gary Marsden, Gitte Lindgaard, Janet Wesson, and Marco Winckler (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 282–299.

[49] Manuel Prätorius, Dimitar Valkov, Ulrich Burgbacher, and Klaus Hinrichs. 2014. DigiTap: An Eyes-free VR/AR Symbolic Input Device. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology (VRST '14)*. ACM, New York, NY, USA, 9–18. https://doi.org/10.1145/2671015.2671029

[50] Reza Rawassizadeh, Blaine A. Price, and Marian Petre. 2014. Wearables: Has the Age of Smartwatches Finally Arrived? *Commun. ACM* 58, 1 (Dec. 2014), 45–47. https://doi.org/10.1145/2629633

[51] Jun Rekimoto. 2001. GestureWrist and GesturePad: Unobtrusive Wearable Interaction Devices. In *Proceedings of the 5th IEEE International Symposium on Wearable Computers (ISWC '01)*. IEEE Computer Society, Washington, DC, USA, 21–. http://dl.acm.org/citation.cfm?id=580581.856565

[52] Munehiko Sato, Ivan Poupyrev, and Chris Harrison. 2012. Touché: Enhancing Touch Interaction on Humans, Screens, Liquids, and Everyday Objects. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 483–492. https://doi.org/10.1145/2207676.2207743

[53] Stefan Schneegass and Alexandra Voit. 2016. GestureSleeve: Using Touch Sensitive Fabrics for Gestural Input on the Forearm for Controlling Smartwatches. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers (ISWC '16)*. ACM, New York, NY, USA, 108–115. https://doi.org/10.1145/2971763.2971797

[54] Matthias Seuter, Max Pfeiffer, Gernot Bauer, Karen Zentgraf, and Christian Kray. 2017. Running with Technology: Evaluating the Impact of Interacting with Wearable Devices on Running Movement. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 101 (Sept. 2017), 17 pages. https://doi.org/10.1145/3130966

[55] Teddy Seyed, Chris Burns, Mario Costa Sousa, Frank Maurer, and Anthony Tang. 2012. Eliciting Usable Gestures for Multi-display Environments. In *Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces (ITS '12)*. ACM, New York, NY, USA, 41–50. https://doi.org/10.1145/2396636.2396643

[56] Srinath Sridhar, Anders Markussen, Antti Oulasvirta, Christian Theobalt, and Sebastian Boring. 2017. WatchSense: On- and Above-Skin Input Sensing Through a Wearable Depth Sensor. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 3891–3902. https://doi.org/10.1145/3025453.3026005

[57] Ke Sun, Yuntao Wang, Chun Yu, Yukang Yan, Hongyi Wen, and Yuanchun Shi. 2017. Float: One-Handed and Touch-Free Target Selection on Smartwatches. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 692–704. https://doi.org/10.1145/3025453.3026027

[58] Radu-Daniel Vatavu. 2012. User-defined Gestures for Free-hand TV Control. In *Proceedings of the 10th European Conference on Interactive TV and Video (EuroITV '12)*. ACM, New York, NY, USA, 45–48. https://doi.org/10.1145/2325616.2325626

[59] Tran Huy Vu, Archan Misra, Quentin Roy, Kenny Choo Tsu Wei, and Youngki Lee. 2018. Smartwatch-based Early Gesture Detection & Trajectory Tracking for Interactive Gesture-Driven Applications. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1, Article 39 (March 2018), 27 pages. https://doi.org/10.1145/3191771

[60] Julie Wagner, Mathieu Nancel, Sean G. Gustafson, Stephane Huot, and Wendy E. Mackay. 2013. Body-centric Design Space for Multi-surface Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 1299–1308. https://doi.org/10.1145/2470654.2466170

[61] Cheng-Yao Wang, Min-Chieh Hsiu, Po-Tsung Chiu, Chiao-Hui Chang, Liwei Chan, Bing-Yu Chen, and Mike Y. Chen. 2015. PalmGesture: Using Palms As Gesture Interfaces for Eyes-free Input. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '15)*. ACM, New York, NY, USA, 217–226. https://doi.org/10.1145/2785830.2785885

[62] Jamie A. Ward, Paul Lukowicz, and Gerhard Tröster. 2005. Gesture Spotting Using Wrist Worn Microphone and 3-axis Accelerometer. In *Proceedings of the 2005 Joint Conference on Smart Objects and Ambient Intelligence: Innovative Context-aware Services: Usages and*

Technologies (sOc-EUSAI '05). ACM, New York, NY, USA, 99–104. https://doi.org/10.1145/1107548.1107578
[63] Andrew M. Webb, Michel Pahud, Ken Hinckley, and Bill Buxton. 2016. Wearables As Context for Guiard-abiding Bimanual Touch. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. ACM, New York, NY, USA, 287–300. https://doi.org/10.1145/2984511.2984564
[64] Martin Weigel, Aditya Shekhar Nittala, Alex Olwal, and Jürgen Steimle. 2017. SkinMarks: Enabling Interactions on Body Landmarks Using Conformal Skin Electronics. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 3095–3105. https://doi.org/10.1145/3025453.3025704
[65] Tim Weißker, Erdan Genc, Andreas Berst, Frederik David Schreiber, and Florian Echtler. 2017. ShakeCast: using handshake detection for automated, setup-free exchange of contact data. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, 88.
[66] Hongyi Wen, Julian Ramos Rojas, and Anind K. Dey. 2016. Serendipity: Finger Gesture Recognition Using an Off-the-Shelf Smartwatch. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 3847–3851. https://doi.org/10.1145/2858036.2858466
[67] Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. 2009. User-defined Gestures for Surface Computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 1083–1092. https://doi.org/10.1145/1518701.1518866
[68] Robert Xiao, Teng Cao, Ning Guo, Jun Zhuo, Yang Zhang, and Chris Harrison. 2018. LumiWatch: On-Arm Projected Graphics and Touch Input. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 95, 11 pages. https://doi.org/10.1145/3173574.3173669
[69] Yukang Yan, Chun Yu, Xiaojuan Ma, Xin Yi, Ke Sun, and Yuanchun Shi. 2018. VirtualGrasp: Leveraging Experience of Interacting with Physical Objects to Facilitate Digital Object Retrieval. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 78, 13 pages. https://doi.org/10.1145/3173574.3173652
[70] Cheng Zhang, AbdelKareem Bedri, Gabriel Reyes, Bailey Bercik, Omer T. Inan, Thad E. Starner, and Gregory D. Abowd. 2016. TapSkin: Recognizing On-Skin Input for Smartwatches. In *Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces (ISS '16)*. ACM, New York, NY, USA, 13–22. https://doi.org/10.1145/2992154.2992187
[71] Cheng Zhang, Qiuyue Xue, Anandghan Waghmare, Ruichen Meng, Sumeet Jain, Yizeng Han, Xinyu Li, Kenneth Cunefare, Thomas Ploetz, Thad Starner, Omer Inan, and Gregory D. Abowd. 2018. FingerPing: Recognizing Fine-grained Hand Poses Using Active Acoustic On-body Sensing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 437, 10 pages. https://doi.org/10.1145/3173574.3174011
[72] Cheng Zhang, Junrui Yang, Caleb Southern, Thad E. Starner, and Gregory D. Abowd. 2016. WatchOut: Extending Interactions on a Smartwatch with Inertial Sensing. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers (ISWC '16)*. ACM, New York, NY, USA, 136–143. https://doi.org/10.1145/2971763.2971775
[73] Maotian Zhang, Qian Dai, Panlong Yang, Jie Xiong, Chang Tian, and Chaocan Xiang. 2018. iDial: Enabling a Virtual Dial Plate on the Hand Back for Around-Device Interaction. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1, Article 55 (March 2018), 20 pages. https://doi.org/10.1145/3191787
[74] Yu Zhang, Tao Gu, Chu Luo, Vassilis Kostakos, and Aruna Seneviratne. 2018. FinDroidHR: Smartwatch Gesture Input with Optical Heartrate Monitor. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1, Article 56 (March 2018), 42 pages. https://doi.org/10.1145/3191788
[75] Yang Zhang and Chris Harrison. 2015. Tomo: Wearable, Low-Cost Electrical Impedance Tomography for Hand Gesture Recognition. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology (UIST '15)*. ACM, New York, NY, USA, 167–173. https://doi.org/10.1145/2807442.2807480
[76] Yang Zhang, Robert Xiao, and Chris Harrison. 2016. Advancing Hand Gesture Recognition with High Resolution Electrical Impedance Tomography. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. ACM, New York, NY, USA, 843–850. https://doi.org/10.1145/2984511.2984574
[77] Yang Zhang, Junhan Zhou, Gierad Laput, and Chris Harrison. 2016. SkinTrack: Using the Body As an Electrical Waveguide for Continuous Finger Tracking on the Skin. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 1491–1503. https://doi.org/10.1145/2858036.2858082
[78] Junhan Zhou, Yang Zhang, Gierad Laput, and Chris Harrison. 2016. AuraSense: Enabling Expressive Around-Smartwatch Interactions with Electric Field Sensing. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. ACM, New York, NY, USA, 81–86. https://doi.org/10.1145/2984511.2984568