HulaMove: Using Commodity IMU for Waist Interaction

Xuhai Xu
xuhaixu@uw.edu
University of Washington
Seattle, WA

Jiahao Li
jhli1116@uw.edu
University of Washington
Seattle, WA

Tianyi Yuan
yuanty17@mails.tsinghua.edu.cn
Tsinghua University
Beijing, China

Liang He
lianghe@cs.uw.edu
University of Washington
Seattle, WA

Xin Liu
xliu0@cs.uw.edu
University of Washington
Seattle, WA

Yukang Yan
yyk@mail.tsinghua.edu.cn
Tsinghua University
Beijing, China

Yuntao Wang
yuntaowang@mail.tsinghua.edu.cn
Tsinghua University
Beijing, China

Yuanchun Shi
shiyc@mail.tsinghua.edu.cn
Tsinghua University
Beijing, China

Jennifer Mankoff
jmankoff@cs.uw.edu
University of Washington
Seattle, WA

Anind K. Dey
anind@uw.edu
University of Washington
Seattle, WA

ABSTRACT

We present *HulaMove*, a novel interaction technique that leverages the movement of the waist as a new eyes-free and hands-free input method for both the physical world and the virtual world. We first conducted a user study (N=12) to understand users' ability to control their waist. We found that users could easily discriminate *eight* shifting directions and *two* rotating orientations, and quickly confirm actions by returning to the original position (*quick return*). We developed a design space with eight gestures for waist interaction based on the results and implemented an IMU-based real-time system. Using a hierarchical machine learning model, our system could recognize waist gestures at an accuracy of 97.5%. Finally, we conducted a second user study (N=12) for usability testing in both real-world scenarios and virtual reality settings. Our usability study indicated that HulaMove significantly reduced interaction time by 41.8% compared to a touch screen method, and greatly improved users' sense of presence in the virtual world. This novel technique provides an additional input method when users' eyes or hands are busy, accelerates users' daily operations, and augments their immersive experience in the virtual world.

CHI '21, May 8–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8096-6/21/05…$15.00
https://doi.org/10.1145/3411764.3445182

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; *Interaction techniques*; *Ubiquitous and mobile computing systems and tools.*

KEYWORDS

Waist interaction; on-device sensing

ACM Reference Format:
Xuhai Xu, Jiahao Li, Tianyi Yuan, Liang He, Xin Liu, Yukang Yan, Yuntao Wang, Yuanchun Shi, Jennifer Mankoff, and Anind K. Dey. 2021. HulaMove: Using Commodity IMU for Waist Interaction. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3411764.3445182

1 INTRODUCTION

Modern ubiquitous interaction is increasingly being enriched by hands-free and eyes-free input from different parts of the human body, such as the face [19, 73], feet [48, 55] and mouth [9, 52]. In this paper, we contribute to that literature by introducing the waist as a novel interaction channel.

Although usually not used explicitly, core muscles are located around the waist and are commonly involved in a wide range of daily activities [42, 53]. Therefore, the movement of the waist can reflect a wide range of daily routines. Researchers often employ waist movement for full-body interaction [16, 40], activity recognition [7, 20] and health monitoring [77], *etc.* Most of this previous work leverages waist movement as a passive indicator, *i.e.*, achieving interaction, recognition, or monitoring goals via observing the waist movement that naturally happens during activities.

However, as the biggest joint of the human body, the waist can also be easily controlled in an active way. This observation gives rise to *HulaMove*, a novel technique that leverages the voluntary

movement of users' waists as a new eyes-free and hands-free input method, which can be useful for interactions in both the physical and virtual worlds. In real-world scenarios, it can serve as a convenient input channel in situations where a user's hands or eyes (or both) are busy. For example, users can easily answer a phone call when carrying heavy bags with their hands, or navigate between recipe pages in the kitchen where hands may be wet or busy holding cooking tools. In addition, in augmented and virtual reality (AR/VR) settings, using the waist for interaction can establish a stronger connection between a user's physical body and the virtual world, thus augmenting an immersive user experience.

Knowing how well users can control their waist movement is fundamental for waist interaction design. We conducted a user study (N=12) with a waist-controlled target acquisition task to evaluate users' waist control ability. We focused on two types of waist movement: 1) shifting, moving the waist in a certain direction while keeping the waist facing forward, and 2) rotating, spinning the waist clockwise or counterclockwise around the body's vertical axis (see Figure 1). We compared the use of a different number of directions (from 4 to 16 for shift, and from 2 to 8 for rotation) and three confirmation techniques (*Quick Return*, *Time Dwell* and *Button Click*). Our results show that users can easily discriminate 8 shifting directions and 2 rotation directions, and that *Quick Return* outperforms other confirmation techniques. Moreover, users indicated social concerns about shifting directly forward and backward, but not others.

We developed the design space of HulaMove to be easily performed and socially appropriate based on the results of the first user study, leading to eight waist gestures. Then, we implemented a real-time system for gesture detection and recognition. Novel interaction techniques involving different body parts typically require custom sensors (*e.g.*, fingertip cameras [62], ultrasonic wristbands [78], and capacitive fingernails [33]), limiting the scalability and generalizability to other applications. In contrast, HulaMove does not require any additional sensors other than the Inertial Measurement Unit (IMU) built in every smartphone. With minimal calibration, users can casually put a smartphone into a pocket near the waist (within 20 cm above or below the waist center, *e.g.*, a jacket pocket or a trouser pocket). The system can dynamically adapt to its position and robustly detect waist gestures. Using a hierarchical machine learning model, our system can successfully recognize eight waist gestures at an accuracy of 97.5%, with the false-positive rate as low as 0.1% in daily routines.

Finally, we ran a usability study (N=12) under two scenarios: a real-life scenario that simulated daily working conditions when users' hands and eyes were busy, and a VR scenario where users employed HulaMove to play an immersive game. Our results revealed that waist interaction only needed 58.2% of the interaction time compared to touchscreen interactions in real-life settings (*i.e.*, when phones need to be taken out of the pocket), and that using waist gestures in VR significantly increased the sense of presence and enhanced user experience. Participants provided positive feedback and stated a willingness to use HulaMove as a novel input technique during daily routines since it is fun and can be performed conveniently.

Our contributions of this paper are threefold:

- We created HulaMove, a novel eyes-free and hands-free input technique that leverages waist movement. We conducted a user study to understand users' ability to control their waist.
- We developed a design space for waist interaction and implement a real-time detection system using the IMU on a smartphone, without involving any customized hardware. Our best machine learning model achieved an accuracy of 97.5%.
- We evaluated HulaMove under both real-life and virtual reality scenarios and demonstrated good usability of waist interaction.

2 RELATED WORK

In this related work section, we first provide a general overview of interaction that utilizes different parts of the body. As HulaMove introduces substantial body movements, we pay special attention to the full-body interaction literature. Moreover, we used an IMU for interaction sensing, thus we also review research on activity recognition and monitoring with motion sensors.

2.1 Leveraging Different Body Parts for Interaction

Recent advances in Human-Computer Interaction have shown the capability of a wide range of human body parts for ubiquitous interaction. One of the major categories is on-body interaction, which refers to the use of body surfaces as an input or output channel [23, 67]. Examples include the palm [22, 68], arms [21, 32], fingers [29, 70], the face [57, 76], and ears [35, 73]. Past research also explored other parts beyond on-body surfaces to enable hands-free or eyes-free interaction, such as feet [49, 66], teeth [9, 74], and even hair [15]. We contribute to this area by introducing the waist as a novel interaction channel that can support both hands-free and eyes-free interaction.

Researchers have used various sensing techniques to support these interaction methods. For example, Stearns *et al.* [63] proposed TouchCam that employed two infrared sensors, an IMU and a camera installed on the finger to recognize touch on different on-body areas. Chan *et al.* [11] built Cyclops that used a fish-eye camera worn as a pendant or a badge to recognize limb actions. SkinMarks developed by Weigel *et al.* [70] used multiple forms of capacitive sensing and deformation sensing for on-skin gestures. Although supporting various input methods, these techniques require customized hardware, thus limiting their deployability. HulaMove strictly relies on the IMU on smartphones that are already commercially available. Therefore, our method has advantages over previous literature with high compatibility with common devices so that it can be easily adopted by users. The work that used the most similar sensing technique to HulaMove is from Jeremy *et al.* [55], who proposed to sense users' foot gestures with a smartphone placed in a pocket. However, their detection model relies on the phone to be placed at a strictly predetermined position, which introduces additional burden on users. In contrast, our algorithm allows users to casually put their phones anywhere close to the waist, such as a jacket pocket, or a front/back leg pocket. We leverage

a light-weight calibration stage to make the detection robust to phone positions.

2.2 Full-body Interaction

Since the waist is the biggest joint centered at the human body, using it for interaction naturally introduces substantial body movement, which has a similar paradigm as full-body interaction [16, 40]. Past research has explored full-body interaction in a variety of scenarios, such as art exhibition [17], education [5], therapy [54], and gaming [45, 60]. Under different contexts and situations, full-body movements are used in different ways [5, 36, 39]. For example, Kjölberg [36] proposed to embed full-body interaction into dancing as a form of bodily communication and artistic expression. Adachi *et al.* [5] built a full-body interaction system to help elementary school students study vegetation growth. Mora-Guiard *et al.* [43] proposed a design for children with Autism to explore and collaborate with neurotypical peers in a full-body interactive AR system. Gaming is the area that has significantly leveraged full-body interaction. Commercially available gaming devices such as the Xbox Kinect [4] and Nintendo Switch [2] enable whole-body gestures in video games. Gerling *et al.* [18] designed eight static or dynamic full-body gestures involving hands, arms, and legs (*e.g.*, pretend to fly, walk in-place) to motivate game experience for older adults. Xu *et al.* [71] proposed to use directional walking with Head Mounted Displays (HMDs) as gestures in VR/AR games. These full-body gestures and movement would inevitably involve some level of waist movement in an implicit way. Perhaps the closest work is Humantenna by Cohn *et al.* [14]. Although using a completely different sensing technique, their whole-body gesture set includes gestures similar to ours, such as torso rotation. However, none of the previous work explicitly introduces the waist as a major interaction channel, not to mention a detailed analysis of its interaction performance. We address this gap in our paper.

2.3 Motion Sensors for Activity Recognition and Monitoring

There has been extensive research on using motion sensors (including accelerometers, gyroscopes, magnetometers, and often combinations of these) for activity recognition [8, 38, 58]. Typical activities include standing, sitting, walking, running, climbing/descending stairs, cycling, and driving. Most researchers have focused on cases when the device is positioned at a specific location, such as front pant pockets [38], jacket pockets [30], or attached to the waist [26]/arm [58]. In a real-life scenario, however, sensors can be placed in multiple places, which leads to the on-body location problem, *e.g.*, sensor readings from the arm and pockets are very different. Two methods are commonly employed to tackle this problem: training one classifier per location separately, or training one classifier for all locations [59]. Another real-life problem is the undetermined device orientation; *e.g.*, the same device standing upright *v.s.* upside down will have reversed readings on one axis. Common methods include a) applying a transformation, b) using sign-invariant features, c) splitting horizontal and vertical (through gravity) gestures, and d) retraining per user [44]. We refer readers to a few good survey and tutorial papers for more details: [10, 44, 59]. In our paper, we develop a transformation-based method to address these two problems simultaneously using a natural, light-weighted calibration stage to enable robust waist interaction recognition.

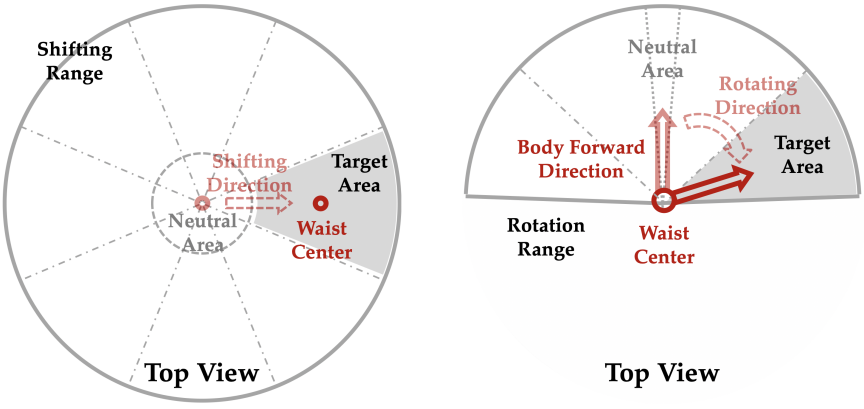
Beyond daily activity recognition, researchers also leverage waist motion as a tracker for physical health-related monitoring [77]. For example, Chen *et al.* [12] attached an accelerometer to a belt and built a system for reliable fall detection. Hjorth *et al.* [28] used a waist-worn accelerometer to measure and monitor children's sleep and physical activity overnight. Ahlrichs *et al.* [6] employed an accelerometer worn on the waist to detect the freezing of gait for Parkinson's disease diagnosis. Most of the previous work leverages waist movement passively and implicitly. In contrast, HulaMove leverages the waist in a proactive, voluntary approach to expand users' interaction bandwidth.

3 STUDY1: WAIST MOVEMENT CONTROL

Prior to developing the design space of HulaMove, we first need to understand users' capability of controlling their waist movement under different visual feedback conditions. Our first study aims to address this gap. The results can provide meaningful guidance on

(a) Shift (b) Rotate

Figure 1: Waist Gestures



(a) Shift (b) Rotate

Figure 2: Top View of Waist Gestures

waist interaction design, such as how many shifting or rotation directions they can easily distinguish. Moreover, confirmation is a fundamental aspect of any selection-based task. We also compare three common confirmation methods to find the best method for waist-based interaction.

3.1 Tasks and Variables

To investigate users' ability to control their waist movement, we designed a series of target acquisition and confirmation tasks [50, 74], and studied four important variables: waist gesture type, target acquisition region, confirmation technique, and visual feedback condition.

Waist gesture type: We focused on two main types of waist gestures in the design space: *shifting* (Figure 1a) and *rotating* (Figure 1b). We evaluated them independently in separate sessions. In the shifting session, a user's waist center was used to determine the selected area, *i.e.*, the cursor was in line with the waist center's position. They needed to horizontally move the cursor into a target area. In the rotating session, the cursor was vector centered at the user's waist center, pointing in their forward direction, and they needed to rotate their torso to face the target direction range.

Target acquisition region: In both the shifting and rotating sessions, the number of regions affects the width of the selection area (*i.e.*, the center angle of each sector). We compared different numbers of regions in each shifting/rotating direction, including 1, 2, 3, and 4. We also defined four directions in the shifting session (left, right, forward, backward) and two directions in the rotating session (clockwise and counter-clockwise). Therefore, the total numbers of regions were 4,8,12,16 in the shifting session and 2,4,6,8 in the rotating session. Figure 2 shows the examples of *two* regions in each direction in both sessions.

Confirmation technique: We investigated three confirmation techniques, which were used after users moved the waist cursor into the intended region. Button Click (*Click*): pressing a hand-held button; Dwell (*Dwell*): maintaining the cursor in the target region for 1 second; and Quick Return (*Quick Return*): quickly moving the waist cursor back to a neutral area (below the threshold, see Fig. 2). The threshold was adaptively set as 10% of their maximum shifting distance in the shifting session and 10% of the maximum rotating angle in the rotating session. For *Quick Return*, to avoid the noise introduced by subtle movements near the threshold, 300 ms was empirically chosen as the minimal time to recognize a "quick" return. The starting point of the returning procedure indicated the selected position.

Visual feedback conditions: Visual feedback is an important factor in interaction design. There were two visual feedback conditions in our study. In the *Full Feedback* condition (*Visual*), all regions were visible on screen outlined with borders, with the specified region for a trial in grey. Users' waist movements and the corresponding cursor were visualized in real-time. The correct waist location for shifting or direction for rotating was indicated by turning the specified region green. In the *No Feedback* condition (*No Visual*), only the neutral area was visible. There were no other regions for reference. The waist cursor would disappear once users moved across the threshold, and only appear again after returning into the neutral area. In real applications, *Visual* and *No Visual* could be suitable for different situations. Although HulaMove only had the *No Visual* condition (as shown in Section 4), we studied both conditions in Study 1 to provide a comprehensive understanding of users' performance.

3.2 Design and Procedure

We employed a within-subjects full factorial design with repeated measures. The four independent variables described in Section 3.1 were included: 1) visual feedback condition (*Visual*, *No Visual*), 2) waist gesture type (*shifting* and *rotating*), 3) selection method (*Click*, *Dwell*, *Quick Return*), and 4) the number of regions in each shifting/rotating direction (N = 1,2,3,4). There were four sessions in total (2 feedback × 2 gesture types). We pre-determined a counterbalanced order for the four sessions. In each session, we used a Latin square to balance the order of the selection method, and randomized the order of the number of regions N. Users repeated each condition three times.

3.2.1 Calibration. Since waist movement range varies across people, calibration is needed before users perform tasks. At the beginning of each session, users went through a calibration stage: they first demonstrated their maximum shifting distance and then moved in a circle (as if they are using a hula hoop), which captured the average maximum shifting distance in all directions. Then, they rotated their torso as far as possible, once clockwise and again counter-clockwise. This captured the average maximum rotating angle. After the calibration, the task regions then adapted to their waist movement ability and the direction they faced. In our study, the average maximum shifting distance was 17.1 cm (SD = 1.9 cm, Min = 14.1 cm, Max = 19.8 cm) and the average maximum rotating angle was 69.9° (SD = 5.9°, Min = 64.4°, Max = 84.0°).

3.2.2 Performance Metrics. The dependent variables included 1) *success rate*: the percentage of trials for a particular condition that resulted in successful target acquisitions; 2) *completion time*: the time from when the waist initially moved to leave the neutral area until the acquisition was confirmed; 3) *number of crossings*: the number of times the cursor crossed the edge of a target region once the cursor entered the target (subtracting 1 for all trials with *Quick Return* since the cursor inevitably crossed the edge when it returned). These measures complemented each other. *Success rate* and *completion time* indicated the overall completeness of the tasks, while *number of crossings* reflected their waist control performance.

3.2.3 Procedure. Users signed the consent form and began with a warm-up stage. After they indicated that they understood the procedure and the gestures, they went through the calibration and four sessions one after another. Each session took around five to ten minutes. A five-minute break was inserted after each sessions. Finally, the experimenter conducted a brief interview to obtain participants' feedback about waist interaction. The duration of the study was about fifty minutes.

3.3 Participants and Apparatus

After obtaining University IRB approval, we recruited 12 participants by snowball sampling (Female=5, Male=7, Age=25.2±1.8). All participants reported being healthy (height 171±6cm), weight 69±16kg, and waist size 77±8cm). To ensure the accurate measure

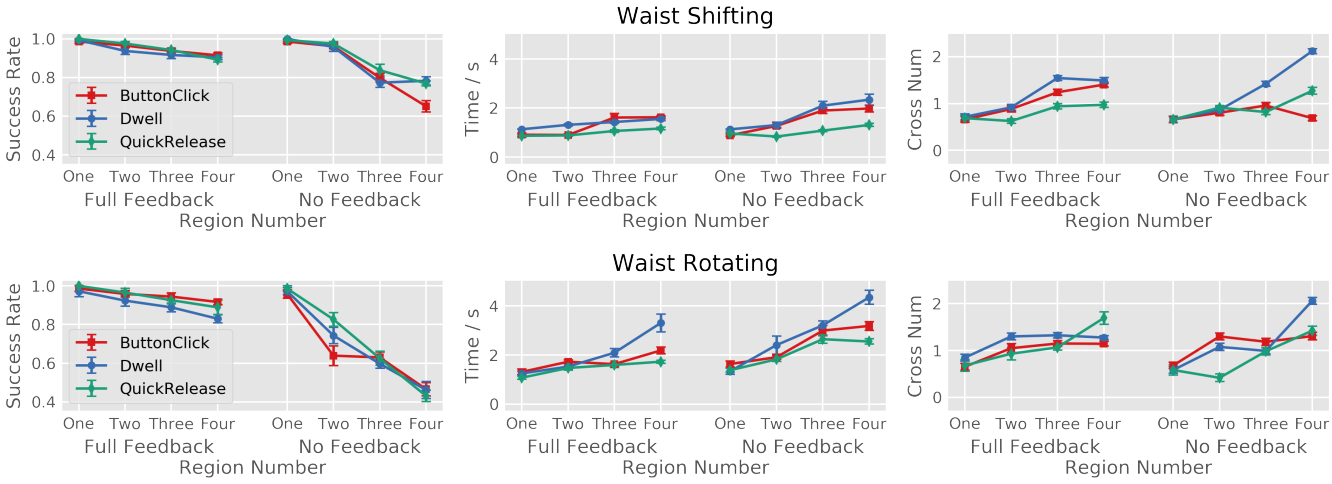


Figure 3: Results of Study 1. The figures on the top show the average success rate, completion time, and number of crossings in shifting sessions. The figures on the bottom show the results of the same three metrics in rotating sessions. In each figure, the left side shows the results with full visual feedback, while the right side shows the results without visual feedback. Error bars indicate the standard deviation.

of waist position and direction, we leveraged the tracking system embedded in an HTC VIVE Pro. We mounted one controller on the left and the other on the right side of participants' pants so that the center of the two controllers indicated the waist center and the forward direction was in line with participants' body facing direction. The head-mounted display (HMD) was not used in the study.

3.4 Results

Figure 3 shows the results of Study 1. Below, we summarize each factor separately.

3.4.1 Confirmation Techniques. One goal of our experiment was to find the best confirmation technique for waist interaction. We observed a consistent result: in both feedback conditions, *Quick Return* either outperformed the other confirmation techniques or had non-significant differences in success rate, speed, and waist movement control stability.

We confirmed our observation with statistical analysis. Since the data violated normality and homoscedasticity assumptions, we applied generalized linear mixed models (GLMMs) [41] using a Gamma link function, with confirmation techniques as the main factor. Table 1 summarizes the GLMMs' results, together with post hoc pairwise z-tests with Bonferroni adjustment. The comparison showed that in most cases, *Quick Return* had significantly better performance than other techniques ($p < 0.05$). The results are consistent across gesture types and visual feedback conditions. Therefore, we focused on the trials with *Quick Return* in the rest of the analysis.

3.4.2 Number of Regions. A primary goal of this first study was to determine how many regions in each shifting/rotating direction users can discriminate easily and control comfortably with decent performance. We applied GLMMs on the data using *Quick Return*, with the number of regions as the main factor. Similarly, z-tests with Bonferroni adjustment was used as the post hoc method. Table 2 summarizes all results.

There were a few observations. First, in general, performance decreased as the number of regions increased. The larger number of regions led to smaller ranges and increased the difficulties of waist movement control, especially when the number of regions was larger than 3. The success rate dropped greatly at $N = 3$ and $N = 4$ in both visual feedback conditions, and the drop was more significant in *No Visual*. *"When there were 12 or 16 targets, it was almost impossible for me to distinguish between two adjacent areas."* (P2).

Second, in waist shifting sessions, the difference between $N = 1$ and $N = 2$ was not significant. Among the 6 post hoc pairwise comparisons in the top part of Table 2, three comparisons did not indicate significance at level $p = 0.05$ and one showed that $N = 2$ had better performance than $N = 1$. In contrast, in waist rotating sessions, $N = 1$ had significantly better performance than $N = 2$ in most metrics. Based on these results, we chose $N = 2$ for waist shifting and $N = 1$ for waist rotating to maximize the number of gestures while having satisfactory performance, thus leading to 10 ($2 \times 4 + 1 \times 2$) waist gestures.

3.4.3 Visual Feedback. Unsurprisingly, removing the visual feedback increases the difficulty of the task, especially on the success rate metric. We observed a significant drop in success rate in both shifting ($\chi^2(2) = 10.1, p < 0.01$) and rotating sessions ($\chi^2(2) = 27.3, p < 0.001$). In the rotating session the completion time without feedback was also significantly longer ($\chi^2(2) = 39.0, p < 0.001$). Other results did not indicate significance. However, if we focus on $N = 2$ in the waist shifting session, and $N = 1$ in the waist rotating session, GLMMs did not indicate any significant difference between feedback types.

3.4.4 Subjective Feedback. During the experiment, we received some interesting comments from the participants. Three participants expressed their concerns about the social acceptance of forward-shifting gestures. Two participants showed similar concerns about backward-shifting gestures. To compare the 10 selected gestures more formally, we sent a short questionnaire to the 12 participants, asking three questions – physical demand, mental demand, and social appropriateness – for each gesture on a 7-point Likert scale.

Figure 4 showed the rating results of the 10 gestures, ranked from best to the worst. The results were consistent with the feedback during the study: forward shifting and backward shifting gestures received the most negative ratings, especially from a social perspective.

4 HULAMOVE SYSTEM DESIGN

The results of Study 1 are informative for the design of HulaMove: using *Quick Return* had the highest *Success* and lowest *Time* and *Number of Crossings* in general; Users could achieve similar performance in *No Visual* and *Visual* at $N = 2$ for shifting gestures and $N = 1$ for rotating gestures. We identified a few guidelines for waist interaction design:

(1) *Quick Return* is the best technique for confirmation
(2) Using $N = 2$ for waist shifting (eight directions) and $N = 1$ for rotating (two directions) balance the number of gestures and performance.
(3) Forward shifting and backward shifting gestures are considered socially inappropriate.

Therefore, our design space of HulaMove focuses on eight gestures with *Quick Return*: rotate left, rotate right, shift left, shift right, shift forward left, shift forward right, shift backward left, and shift backward right.

The majority of pockets where users put their phones were near the waist, such as jacket pockets, hoodie pockets, and trouser pockets. Leveraging a smartphone at these positions for waist gesture detection could obviate the need for any customized sensors, minimize user burden, and increase the scalability and generalizability. Therefore, we developed an algorithm and a real-time system using IMU data collected from a smartphone placed at any position near the waist (defined as 20 cm above or below the waist center) for waist gesture detection and recognition.

To detect a user's waist gestures from a phone that could be placed at multiple positions, we calibrated the IMU data by transforming the movement of the phone to the movement of the human body. After aligning the IMU signals with human body coordinates, a gesture detection module was applied to the aligned signals to identify any present waist gesture. Once a gesture motion was captured, it was fed into a gesture recognition module to get the output. Figure 5 visualizes the whole pipeline.

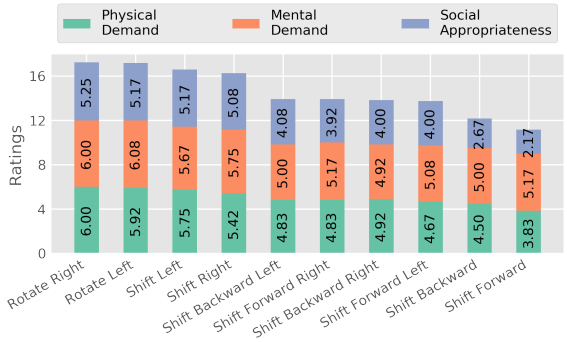


Figure 4: Ratings of the selected 10 waist gestures. Scores are pre-processed so that higher scores indicates better user experience.

4.1 Pre-processing and Calibration by Transformation

We collected IMU data (linear accelerometer ***acc***, and gyroscope ***gyro***, both represented as x, y, z 3-dimension vectors) at 30 Hz. The average duration of a waist gesture was 1.6 s (SD=0.3 s) according to two authors' manual annotations of over one hundred samples. Most waist movement's duration for gestures was between 1.0 and

Measures	Shifting & Full Feedback		Shifting & No Feedback	
	AOV ($\chi^2(2)$)	Post Hoc	AOV ($\chi^2(2)$)	Post Hoc
Success	$2.42, p = 0.290$	*Quick Return* ~ *Click* ~ *Dwell*	$2.74, p = 0.25$	*Quick Return* ~ *Click* ~ *Dwell*
Time	$34.81, p < 0.001^{***}$	*Quick Return* < *Click* ~ *Dwell*	$5735, p < 0.001^{***}$	*Quick Return* < *Click* < *Dwell*
NC	$28.44, p < 0.001^{***}$	*Quick Return* < *Click* ~ *Dwell*	$42.83, p < 0.001^{***}$	*Click* < *Quick Return* < *Dwell*
Measures	Rotating & Full Feedback		Rotating & No Feedback	
	AOV ($\chi^2(2)$)	Post Hoc	AOV ($\chi^2(2)$)	Post Hoc
Success	$9.71, p = 0.007^{**}$	*Click* ~ *Quick Return* > *Dwell*	$0.82, p = 0.662$	*Quick Return* ~ *Dwell* ~ *Click*
Time	$23.05, p < 0.001^{***}$	*Quick Return* < *Click* < *Dwell*	$13.69, p = 0.001^{**}$	*Quick Return* ~ *Dwell* ~ *Click*
NC	$3.66, p = 0.161$	*Quick Return* ~ *Click* ~ *Dwell*	$7.76, p = 0.02^{*}$	*Quick Return* ~ *Click* < *Dwell*

$^{*} < 0.05$, $^{**} < 0.01$, $^{***} < 0.001$, </> implies post hoc significance ($p < 0.05$). ~ means no significant difference. Same in Table 2

Table 1: Statistical analysis of confirmation techniques. *Success*: Success Rate, *Time*: Completion Time, *NC*: Number of Crossing. The relative positions of the selection method in post hoc columns indicate the mean value order. The more favorable techniques (higher *Success*, lower *Time* and *NC*) are put before the less favorable ones.

Measures	Shifting & Full Feedback		Shifting & No Feedback	
	AOV ($\chi^2(2)$)	Post Hoc	AOV ($\chi^2(2)$)	Post Hoc
Success	$113.2, p < 0.001^{***}$	$1 > 2 > 3 > 4$	$127.2, p < 0.001^{***}$	$1 \sim 2 > 3 > 4$
Time	$120.9, p < 0.001^{***}$	$1 \sim 2 < 3 < 4$	$159.0, p < 0.001^{***}$	$2 < 1 < 3 < 4$
NC	$46.2, p < 0.001^{***}$	$2 \sim 1 < 3 \sim 4$	$47.5, p < 0.001^{***}$	$1 \sim 3 \sim 2 < 4$
Measures	Rotating & Full Feedback		Rotating & No Feedback	
	AOV ($\chi^2(2)$)	Post Hoc	AOV ($\chi^2(2)$)	Post Hoc
Success	$15.2, p = 0.002^{**}$	$1 \sim 2 \sim 3 \sim 4$	$150.6, p < 0.001^{***}$	$1 > 2 > 3 > 4$
Time	$48.5, p < 0.001^{***}$	$1 < 2 \sim 3 \sim 4$	$162.8, p < 0.001^{***}$	$1 < 2 < 4 \sim 3$
NC	$17.0, p < 0.001^{***}$	$1 \sim 2 \sim 3 \sim 4$	$19.6, p < 0.001^{***}$	$2 \sim 1 \sim 3 \sim 4$

Table 2: Statistical analysis of region number. We focus on the data with *Quick Return* only. The relative positions of the region numbers in post hoc columns indicate the mean value order.

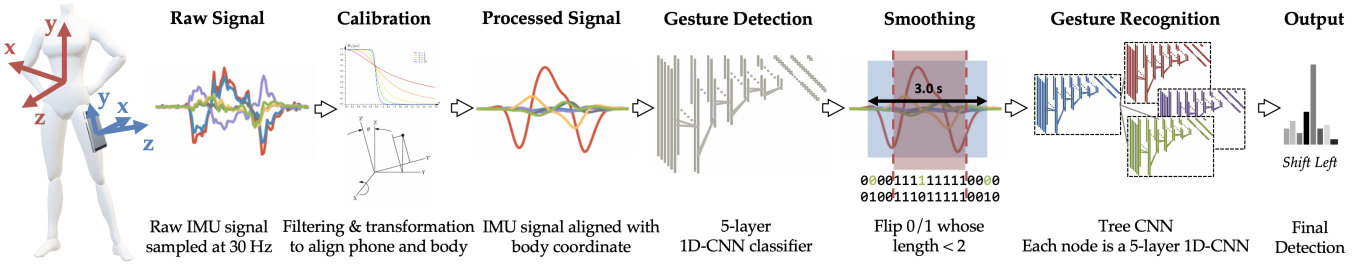


Figure 5: Overview of the system pipeline.

2.0 s (0.5-1 Hz). We applied a low-pass Butterworth filter [56] (cut-off at 2 Hz, twice the upper bound of waist gestures) on the raw IMU data to remove high-frequency noise and smooth the signal.

We then applied the calibration on the filtered data. The first step is to identify the 3×3 rotation matrix $\boldsymbol{A} = [\boldsymbol{A_X}; \boldsymbol{A_Y}; \boldsymbol{A_Z}]$ from the human body coordinate space to the phone coordinate space (see the leftmost of Figure 5), *i.e.*, $\boldsymbol{acc_p} = \boldsymbol{A} \cdot \boldsymbol{acc_h}$ and $\boldsymbol{gyro_p} = \boldsymbol{A} \cdot \boldsymbol{gyro_h}$, h/p indicates the human body/phone coordinate space. This matrix can transform the IMU signals originally collected in the phone coordinate space to the signals in the human body coordinate space. To obtain the matrix, we asked users to follow a count-down timer to perform four gestures after they put the phone in a pocket: a shift left, a shift right, a rotate left, and a rotate right. Then, we took a 3.0 s window starting from the timer to ensure that a gesture could be covered. We used the shift left and shift right data for calibrating the x-axis (*i.e.*, obtaining $\boldsymbol{A_X}$), since these two movements were in line with the x-axis in human body coordinate space. Similarly, the rotate left and rotate right data were used for calibrating the y-axis ($\boldsymbol{A_Y}$).

A shift right gesture with quick return can be divided into four stages: 1) speed up ($vel_{h_x} > 0, acc_{h_x} > 0$), 2) slow down until reaching the maximum shifting distance ($vel_{h_x} > 0, acc_{h_x} < 0$), 3) return back and speed up in the opposite direction ($vel_{h_x} < 0, acc_{h_x} < 0$), 4) slow down and stop at the original position ($vel_{h_x} < 0, acc_{h_x} > 0$). In the collected $\boldsymbol{acc}$ data (*i.e.*, $\boldsymbol{acc_p}$), the sequence of $|\boldsymbol{acc_p}|$ ($\sqrt{acc_{p_x}^2 + acc_{p_y}^2 + acc_{p_z}^2}$) has a small peak (during stage 1), followed by a wider peak (during stage 2 and 3), followed by another small peak (during stage 4). The upper part of Figure 6 illustrates this procedure.

Ideally, during a shift right gesture, the waist moves in $+x$ direction and returns in $-x$ direction in human body coordinates, *i.e.*, $\boldsymbol{acc_h} = [acc_{h_x}, 0, 0]^T$, $\boldsymbol{dir_{h_{sft}}} = [1, 0, 0]^T$, and $\boldsymbol{acc_p}$ have a consistent direction $\boldsymbol{dir_{p_{sft}}}$ (and its reverse) in the space. However, due to the human body's anatomic property and movement control error, the direction is not perfectly consistent. Thus, we select a few representative periods and use their average as the moving direction. Specifically, we choose the periods when $|\boldsymbol{acc_p}|$ ($= |\boldsymbol{acc_h}|$) is greater than its mean plus one standard deviation within the 3.0 s window, because small $\boldsymbol{acc_p}$ is more likely to be biased by noise. This identifies the three peaks that happen naturally during the four stages of a shifting gesture. As mentioned above, the direction of $\boldsymbol{acc_p}$ during stage 2 and 3 (the second peak) is opposite to stage 1 (the first peak) and stage 4 (the third peak). Therefore, after identifying the three peaks in $\boldsymbol{acc_p}$, we reverse the second peak and then calculate their average, $\boldsymbol{dir_{p_{sft\text{-}r}}}$, as the direction of the shift right gesture. To further reduce the error, we calculate the direction of the shift left gesture in the same way ($\boldsymbol{dir_{p_{sft\text{-}l}}}$), reverse its direction ($-\boldsymbol{dir_{p_{sft\text{-}l}}}$) so that it is also in $+x$ direction, and set the final shifting direction as the average of the two: $\boldsymbol{dir_{p_{sft}}} = (\boldsymbol{dir_{p_{sft\text{-}r}}} - \boldsymbol{dir_{p_{sft\text{-}l}}})/2$.

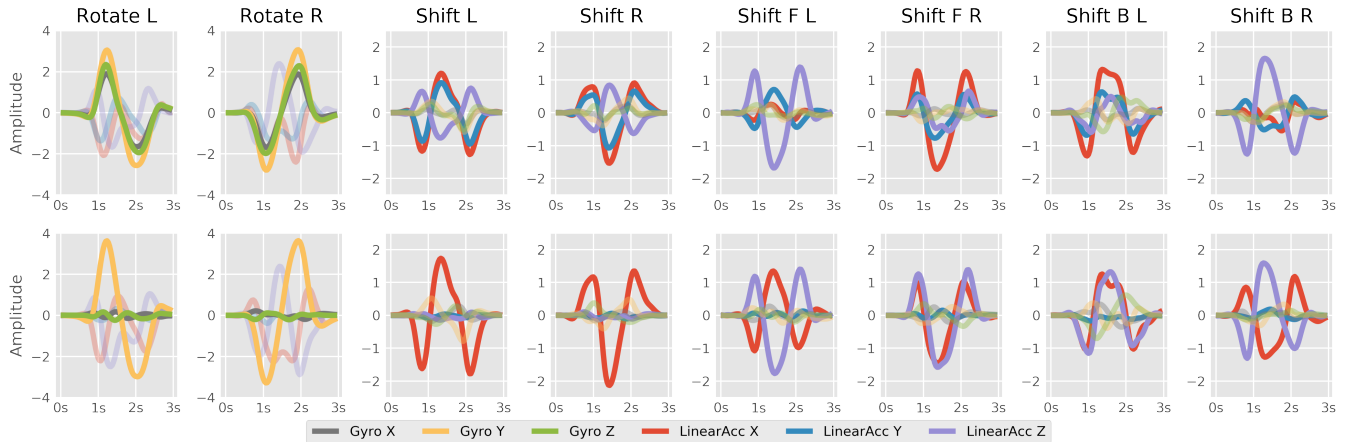


Figure 6: IMU signals of the eight waist gestures collected by a phone put in the right front trouser pocket. The top row shows the original data with a Butterworth low-pass filter at 2 Hz. The bottom row shows the data after transformation. Gyroscope/linear accelerometer signals are highlighted with higher opacity for rotating/shifting gestures respectively.

Then, we can easily calculate $\boldsymbol{A_X}$ by

$$\boldsymbol{dir_{p_{sft}}} = \boldsymbol{A} \cdot \boldsymbol{dir_{h_{sft}}}$$
$$\boldsymbol{dir_{p_{sft}}} = [\boldsymbol{A_X}; \boldsymbol{A_Y}; \boldsymbol{A_Z}] \cdot [1, 0, 0]^T$$
$$\boldsymbol{dir_{p_{sft}}} = \boldsymbol{A_X}$$

For the rotating gestures, the four stages are reflected by two peaks in $\boldsymbol{gyro_p}$, one for rotating to the maximum angle ($gyro_{h_y} > 0$) and one for turning back ($gyro_{h_y} < 0$). The same procedure can be applied on the rotate left and rotate right gestures to obtain $\boldsymbol{A_Y}$. Since a rotation matrix is an orthogonal matrix, if $\boldsymbol{A_X}$ and $\boldsymbol{A_Y}$ are not orthogonal, we adjust them slightly by rotating around their common vertical axis until they are orthogonal. Then, we have $\boldsymbol{A_Z} = \boldsymbol{A_Y} \times \boldsymbol{A_X}$.

Therefore, after a quick calibration stage with four gestures, we obtained the rotation matrix $\boldsymbol{A} = [\boldsymbol{A_X}; \boldsymbol{A_Y}; \boldsymbol{A_Z}]$ and transformed all incoming signals from the phone coordinate space to the human body coordinate space by $\boldsymbol{acc_h} = \boldsymbol{A}^{-1} \cdot \boldsymbol{acc_p}$ and $\boldsymbol{gyro_h} = \boldsymbol{A}^{-1} \cdot \boldsymbol{gyro_p}$. Figure 6 shows the eight gestures' typical IMU signals when a phone is put in the right front trouser pocket at a random orientation and its corresponding calibrated signals after transformation.

A phone placed in a pocket could change its position and orientation slightly when the user moves around. This would lead to the change of the rotation matrix $\boldsymbol{A}$. To mitigate this issue, our algorithm updated $\boldsymbol{A}$ dynamically once a shift left/right or rotate left/right was detected (as mentioned in Section 4.2 and Section 4.3). When one of these four gestures was detected, its corresponding $\boldsymbol{dir_p}$ was appended to a record list and we used a moving average to update $\boldsymbol{A_X}$ (for shifting) or $\boldsymbol{A_Y}$ (for rotating), followed by updating $\boldsymbol{A_Z}$ and $\boldsymbol{A}$.

4.2 Gesture Detection

After pre-processing and calibration, we applied a gesture detection classifier on the transformed signals using a sliding window. The size of the window was 3.0 s, with the step size as 0.2 s. The classifiers took both accelerometer and gyroscope data as the input ($30Hz \times 3\ s \times 6$) and used a 5-layer convolutional neural network (CNN) [37] for detection, with three 1-dimensional convolutional layers and two fully connected layers, all using the ReLu activation function [47]. A max-pooling layer [46], a batch normalization layer [31] and a dropout layer at a rate of 0.5 [61] were inserted between every two convolutional layers. The classifier output a 1 whenever the IMU signals belonged to a gesture and a 0 otherwise. Almost all waist gestures were longer than one second, so the presence of a gesture should lead the classifier to produce multiple 1's in succession; however, temporal shifts in the gesture signals and noise could make the classifier's serial output noisy. We remedied this issue by using a majority voting scheme to smooth the sequence, where adjacent sequences of consecutive 1's are merged if they were separated by one or two 0's and the same for consecutive 0's. A gesture was defined to be present whenever there were 3 or more consecutive 1's and followed by more than two 0's. Whenever a gesture occurred, the system took a 3.0 s window centered on the sequence of 1's and fed it into the gesture recognition module.

4.3 Gesture Recognition

The final step was to classify the captured gesture signals. The straightforward methods, including static and dynamic threshold, as well as signal processing and feature engineering, can be easily confused by many unexpected variations of users' waist movements, leading to biased rotation matrix and IMU signals. Another method is to train an eight-class classifier. However, such an approach would require a large amount of training data in each class. As we show in Section 4.4, these approaches had low performance on our dataset. Instead, we used a hierarchical tree-CNN [51] which divided the task into four easier complementary binary or ternary classification tasks. This significantly simplified recognition and improved the results. Specifically, the eight gestures could be split as follows:

(1) **Type Classifier**: the main signals for rotating gestures came from the gyroscope and the ones for shifting gestures came

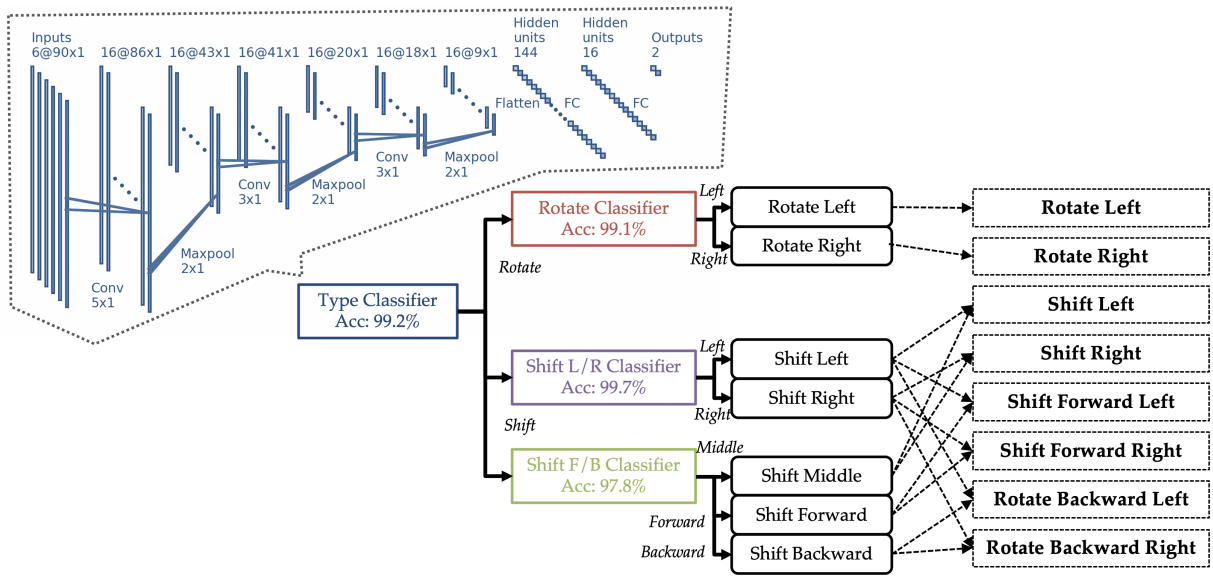


Figure 7: Tree-CNN framework of waist gesture recognition. The four rectangles indicate the four classifiers on tree nodes. We present one example of the classifier network structure. Others classifiers share the same structure other than the final output unit number. Their leave-one-out accuracy are in the box. The rounded rectangles indicate the direct output of the three second-layer classifiers. The dashed rectangles on the right are the final recognized gestures.

from the linear accelerometer. We first distinguish whether a gesture is rotating or shifting.

(2) **Rotation Classifier**: the two rotating gestures could be handled by a binary classifier.

(3) **Shift (Left/Right) Classifier**: among the six shifting gestures, three are shifting to the left side and three are to the right side. We used a binary classifier to extract this property.

(4) **Shift (Forward/Backward) Classifier**: two gestures are forward, two are backward, and two are neither of them. We used a ternary classifier to distinguish them. Combined with the *Shift(Left/Right) Classifier*, the system could uniquely identify a shifting gesture.

Figure 7 shows the classification tree. Each classifier had the same network structure as the gesture detection classifier besides the output unit number (2 for binary or 3 for ternary). Note that each classifier was trained on a different subset of the data, *e.g.*, the rotate classifier would be trained only on the data of rotating gestures.

4.4 Data Collection and Model Results

4.4.1 Data Collection. Three authors collected the data used to train our models using a custom Flask-based web app. The app sampled the IMU sensor on the phone at 30 Hz.

The authors accessed the web app using their own Android phone and followed the instructions to complete one data collection epoch: 1) placed the phone in a pocket near the waist, 2) performed the calibration stage with four gestures, 3) performed the eight gestures in a random order, in sync with the countdown timer presented on a laptop screen; 4) performed each gesture 10 times. Each author repeated this epoch multiple times. They changed the pocket and phone orientation every time to ensure variety in the data. Overall, 20 epochs were collected from the three authors in total, leading to 1600 gestures in total.

For each collected gesture, the timer counted down for 2 seconds, and then participants had another 4 seconds to complete the gesture. Data were recorded during all 6 seconds to capture IMU signals both with and without gestures. Moreover, three authors also kept the web app active and carried the phone in their pocket during their daily routines. Six hours of IMU data were collected and labeled as noise. Activities such as sitting, walking, lying down, standing still, climbing stairs were all included.

4.4.2 Data Cleaning and Augmentation. The data was organized into labeled windows. Recall that the timer started 2.0 s before the start of each gesture. We used a sliding 3.0 s window, with a 0.2 s step size to generate the training data. Since we sampled at 30Hz, each window included 90 samples. If more than 50% of them (>45) overlapped the gesture time (2.0-4.0 s), it was labeled as a positive example of that gesture. We additionally applied a non-overlapped 3.0 s sliding window to the longitudinally collected noise and labeled all samples as negative for gesture detection.

We then augmented the data set further by 1) flipping each signal [25], 2) adding random Gaussian noise [73]. This increased the size of the data by 3 times (each original sample s, + s flipped, + s with noise, + s flipped and with noise).

4.4.3 Results. Overall, more than 120 thousand noise samples and 14 thousand gesture samples were generated using this approach. Note that for each gesture recognition classifier, only a subset of the data was used, *e.g.*, *rotate classifier* only leveraged the two rotating gestures' data.

We labeled all data with an epoch number and used leave-one-epoch-out cross-validation to evaluate the performance of the gesture detection and recognition modules, *i.e.*, taking data from one epoch as the testing set and the rest of the epochs as the training set. For gesture recognition, we also added noise data to the training set in each loop to make the model more robust to noise. This does not affect the testing accuracy since the noise data was not added during the testing.

Gesture Detection. Combined with the smoothing technique, our gesture detection model captured all gestures successfully (100% true positive rate). When we tested the performance of the model without smoothing, the accuracy dropped to 90.1%. The misclassified samples were mainly near the 50% overlap area. This reflects the importance of our smoothing technique. Moreover, We also applied the model on a subset of noise data withheld from the training. The false positive rate was only 0.1%. When the model was tested with a sliding window mechanism with 0.2 s as the step size, the false positive rate further dropped to 0.03% (about 5.5 false positive per hour).

Gesture Recognition. The type, rotate, shift left/right, and shift forward/backward classifiers all achieved satisfactory average accuracy: 99.2%, 99.1%, 99.7%, and 97.8%, respectively. Combining all the four classifiers together, the final gesture recognition module achieved an accuracy of 97.5%. Figure 8 shows the confusion matrix. Compared with the moving-average threshold (Acc: 80.5%), a SVM (Acc: 82.1%), or a conventional multi-class CNN (eight class) (Acc: 92.0%), our tree-CNN structure significantly improved the performance of gesture recognition.

Combining the gesture detection and recognition modules, the overall performance of our model on the gesture dataset reached an accuracy of 97.5% and an F1-score of 97.2%. Moreover, as we show in the next usability study (Section 5), our model also had similar performance when applied on another 12 users' data, achieving an accuracy of 96.8% and an F1-score of 96.9%.

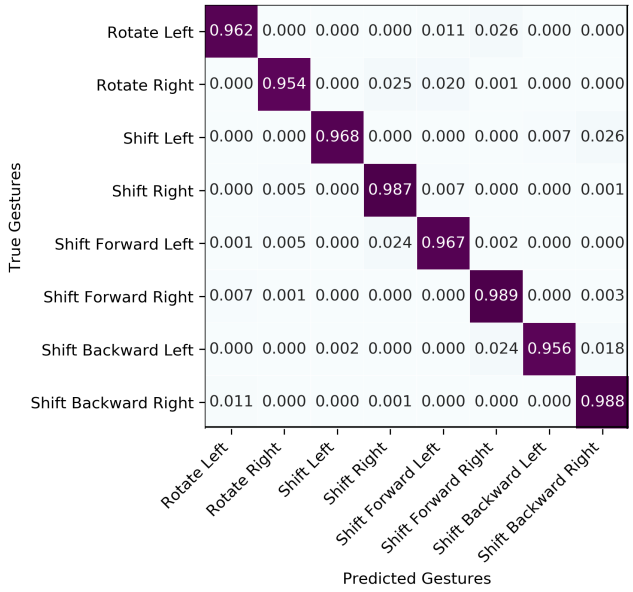


Figure 8: Confusion matrix of the eight waist gestures. The overall accuracy is 97.5%.

5 STUDY2: USABILITY EVALUATION

We conducted a second user study and evaluated a real-time implementation of HulaMove on its performance and usability.

5.1 Tasks and Variables

We envision HulaMove being used in a wide range of daily scenarios. We split them into two categories, interactions in the real physical world and in AR/VR, and chose typical tasks for the evaluation study.

5.1.1 Working in Physical World. As HulaMove has the advantage to provide eyes-free and hands-free interaction, it is suitable for many cases when users are standing and their hands are busy, such as cooking in the kitchen.

Task: We designed two common phone-based applications that often happen during these daily cases. Each of them involved a set of operations. 1) Music player: users controlled music with five actions, including play/pause, volume up, volume down, next song, and previous song; 2) Phone call: when a phone call came in, users could either answer, reject, or mute the call. The left side of Figure 9 illustrates the two tasks.

Setups: Two setups were involved in the study. One based on our system and the other based on touchscreen input (the standard input mechanism enabled by smartphones). In both setups, users held an empty paper box at hand, to simulate the daily cases where eyes and hands were busy. Moreover, the phone was put in their preferred trouser pocket (a common scenario in daily cases), so interactions were not visually available initially. 1) HulaMove: users used waist gestures to complete the task; 2) Pocket: users need to put down the box, remove the phone from the pocket, and then complete the task. The upper part of Table 3 shows the specific mapping of the two setups.

Note that HulaMove was not designed to replace the existing interaction techniques, nor was the purpose of this study to compare our method against the baseline. We picked the pocket setup as this is one of the most common and representative scenarios in daily life where HulaMove can be useful.

5.1.2 Games in VR. Another important use case of HulaMove is in AR/VR, where users can use waist gestures as another input channel. Such a connection between the virtual world and the physical body can potentially increase users' presence in the virtual world.

Task: We developed a VR game similar to BeatSaber [1]. Users stood still in the center of a virtual hallway, with a series of obstacles coming in front of them. Users could avoid obstacles by taking certain actions (slide left/right, jump, and squat). Some obstacles could also be destroyed with the hitting skill. Users started with three health points (HP) and lost one HP if they ran into an obstacle. The game ended when HP dropped to zero. The right side of Figure 9 shows a moment when a user was playing the game.

Setups: We have two setups for the VR game, one based on our system and the other based on a hand controller. 1) HulaMove: users put a phone into their preferred trouser pocket, and used waist gestures to play the game; 2) Hand controller: users pressed

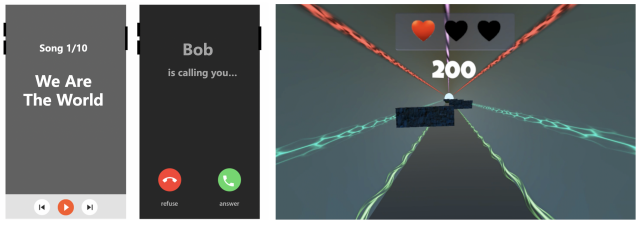


Figure 9: Task Setups of Study 2. Left) the UI of the two tasks in physical world. The two physical buttons on the left edge are used for volume adjustment, and the button on the right edge is used for muting a call. Right) a VR game similar to BeatSaber. Users can avoid or destroy the obstacles with different gestures. Gesture mappings are listed in Table 3.

buttons on the hand controller to play the game. The bottom part of Table 3 shows all operations available in the VR world.

5.2 Design and Procedure

We employed a within-subjects design for both tasks. In the physical tasks, the main independent variables were the task (music player and phone call) and the setup (waist *v.s.* pocket). In the VR game, the only independent variable was the setup (waist *v.s.* hand controller). This led to six sessions in total, four for the physical tasks (2 setups × 2 tasks) and two for the VR game (2 setups). The scenario order, the setup order and the inner task order, were all counterbalanced. For physical tasks, users repeated each operation five times. For the VR game, users played it once until they lost or for 10 minutes. If users lost the game within 3 minutes, they played it one more time.

5.2.1 Usability Metrics. For the two physical tasks, we measured the time that elapsed from when users began their interaction (following a count-down instruction) to when they completed each operation. We used a 7-point Likert scale NASA-TLX questionnaire [24] to measure the perceived workload and the effectiveness of the gestures in each setup [34, 72]. For the VR game, we used the same NASA-TLX questionnaire. Moreover, we also included four questions from the igroup presence questionnaire (IPQ) [65] to measure presence during the VR experience.

Task	Operation	Waist Gesture	Baseline
Music	Play/Pause	Rotate L/R	Virtual Btn
Music	Vol Up	Shift FL/FR	Physical Btn
Music	Vol Down	Shift BL/BR	Physical Btn
Music	Next	Shift R	Virtual Btn
Music	Previous	Shift L	Virtual Btn
Call	Answer	Shift R/FR/BR	Virtual Btn
Call	Reject	Shift L/FL/BL	Virtual Btn
Call	Mute	Rotate L/R	Physical Btn
VR	Slide Left	Shift L	TrackPad
VR	Slide Right	Shift R	TrackPad
VR	Jump	Shift FL/FR	TrackPad
VR	Squat	Shift BL/BR	TrackPad
VR	Destroy	Rotate L/R	Trigger

Table 3: The gesture mapping design of HulaMove gestures and baseline operations for the three applications. L/R/F/B are short for left/right/forward/backward.

5.2.2 Procedure. Users signed the consent form and began with a warm-up stage. Before the physical tasks and VR game sessions, users got themselves familiarized with the operations and interactions. They went through six sessions one by one and filled in the questionnaires. A 2-minute break was placed after each sessions. Note that at the beginning of each session with HulaMove, they went through a quick calibration stage with the four gestures to ensure our algorithm worked effectively. After they completed all tasks, the experimenter conducted a brief interview to collect their feedback. The duration of the study was about forty minutes.

5.3 Participants and Apparatus

Due to the difficulty of recruiting participants during a pandemic, we recruited 6 participants from Study 1 and 6 new participants (Female=4, Male=8, Age=25.0±2.0). The study was IRB-approved and all participants reported themselves as healthy. All participants were informed to wear pants with at least one pocket for the study.

We used a Samsung Galaxy S8 as the IMU data collector. The mobile phone sent data to a laptop (Windows 10 OS) via a Flask-based web application at 30 Hz. The recognition algorithm ran on the laptop in real-time. For the VR game, we used the same HTC VIVE Pro system as Study 1. The algorithm sent its recognition to the VIVE via a ZeroMQ-based local network.

5.4 Results

Our real-time system worked well with new users. Throughout the 12 participants' studies, there were only seven false-positive cases in total (less than 1% of the performed gestures) and the average number of crossing is also low (1.3±1.0). Among all detected gestures, each participant experienced less than two misclassified gestures on average (23 in total out of 717 gestures, leading to an accuracy of 96.8% and an F1 score of 96.9%). The errors happened separately and were distributed among participants (1.9 ± 0.9). Overall, the real-time performance of our system was consistent with our testing results in Section 4, indicating the robustness of the model. The average system delay time is 445 ms. Since our classifiers are all light-weight, the computation time is minimal. The main delay came from the gesture detection module as it needed to wait for two extra two sliding window steps (0.4 s) after a gesture was completed to produce two consecutive negative windows (0's) after a sequence of positive windows (1's) (see Section 4.2).

The study results indicated the advantages of our technique. We summarize all metrics in Figure 10. For the physical world tasks, we applied Wilcoxon signed-rank post hoc tests on each subjective question, respectively. The results did not indicate any significance. The overall subjective usability of HulaMove was similar to that of the most commonly-adopted touchscreen method. However, HulaMove greatly accelerated the interaction. A GLMM with a log link function on the time data, using the task and the setup as the main factors, showed the significance on *Setup* ($\chi^2(2) = 885.5, p <$

0.001^{***}), but not on *Task* ($\chi^2(2) = 0.6, p = 0.42$) nor the interaction between the two ($\chi^2(4) = 1.0, p = 0.31$). HulaMove significantly reduced interaction time by 41.8% (1.56 s *v.s.* 2.68 s) compared to the baseline.

In the VR game, although the new technique was perceived as more physically demanding ($W = 6.0, p = 0.04^{*}$, none of the other TLX questions showed significance), the difference was small (3.3 *v.s.* 2.3). More importantly, participants reported significantly higher scores on general presence ($W = 7.0, p = 0.01^{*}$), spatial presence ($W = 2.0, p = 0.01^{*}$), and engagement ($W = 7.5, p = 0.02^{*}$). The interview with participants also revealed positive feedback. P10 mentioned the improvement of virtual presence explicitly: *"I love to use my body to play the game* [in the VR session]. *I felt much more engaged!"* P7 wanted HulaMove to be used in his daily routines: *"It would be fantastic if my own phone can support this. Sometimes this can be really helpful."*. He often worked at a standing desk and thought having this technique would be convenient and make the whole experience more fun. Interestingly, we also received some inspiring feedback from participants. After playing the VR game, P5 asked *"Are you planning to use this as an exercise game?... This would be helpful when people have to stay at home because of the pandemic."* We discuss more potential applications of our method in Section 6.1. Overall, HulaMove provides a convenient and natural eyes-free and hands-free input method, which can significantly accelerate the interactions and improve the immersive experience in AR/VR.

6 DISCUSSION

We summarize the contributions and the meanings of HulaMove from various perspectives. From the research perspective, it provides the first systematic investigation of how well users can control the waist as an active input method, and the first design space of waist interaction. This adds an important piece to the existing body gesture-based interaction techniques. From the technological perspective, HulaMove can achieve high accuracy of recognition with only a commodity mobile phone without any customized sensor. This ensures the scalability and generalizability of our technique. From the design perspective, HulaMove is a new hands-free and eyes-free technique that can expand the existing input bandwidth. Moreover, using the body for interaction can establish an interesting body connection, which can lead to more engagement in the physical world and higher presence in the virtual world.

In this section, we first discuss the potential use cases of HulaMove in Section 6.1. We also discuss the relationship between our technique and other related interaction methods in Section 6.2. Finally, we reflect on the limitations and future work in Section 6.3.

6.1 Potential Applications

6.1.1 Eyes-free and Hand-free Interaction. Using the waist for interaction does not require visual attention. It can provide an auxiliary input channel that is both eyes-free and hands-free. This can be particularly useful when users are busy with their main work. For example, when users are working (*e.g.*, soldering circuits) and want to do some quick and simple operations on the phone such as changing the music, they need to put down the soldering iron, clean their hands, and pick up the phone to interact with. With HulaMove,

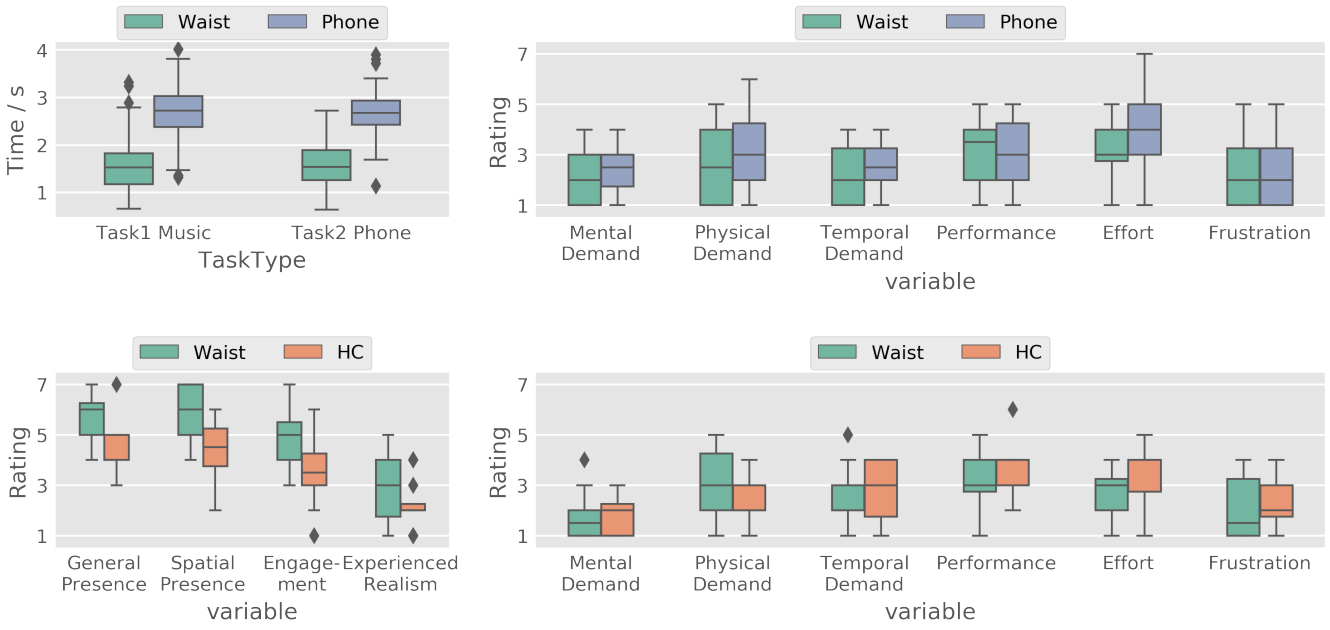


Figure 10: Results of Study 2. The top two figures show the completion time and NASA-TLX ratings of the physical world experience. The bottom two figures show the ratings of IPQ and NASA-TLX of the VR world experience.

their eyes and hands can still be kept on the circuits while performing the interaction. Similar cases can happen in daily routines when users are cooking in the kitchen, doing home chores, or carrying bags after shopping, *etc.* Moreover, when users are working on a computer at a standing desk, HulaMove can be leveraged as a convenient way for quick information retrieval. In addition to using finger gestures on a trackpad to switch pages, applications, and desktops, users can use HulaMove with a sense of 3D space around the body, which can potentially make the retrieval process more intuitive and interesting.

6.1.2 Immersive Experience. AR/VR provides another big application space where our technique can be useful. HulaMove requires users to move their waist for interaction, which can establish a tie between the physical body and the virtual environment. As found in Study 2, the connection can significantly increase users' sense of presence in the virtual world. Note that the mapping between waist gestures and effects needs to be carefully designed. We recommend the effect of waist gestures in the AR/VR should be linked to the virtual avatar body instead of some simple operations to increase immersion, such as controlling the virtual body to avoid obstacles, or fetching objects floating around the virtual body [75]. This is quite different from supporting quick operations in physical world applications mentioned above. In AR/VR, HulaMove should provide engaging effects to ensure a strong connection between the physical and virtual worlds.

6.1.3 Exercise Game. In Study 2, comments from P5 reveal the potential for HulaMove to be used for exercise. The NASA-TLX ratings in the bottom part of Figure 10 show that the only question that HulaMove had worse ratings is for physical demand. This "disadvantage" can actually be leveraged as one "feature" of HulaMove for it to be used as a low-cost, convenient input method for exercise games. Users can use waist gestures to interact with a video game or an immersive AR/VR game, similar to the Ring Fit Adventure [3]. The intensity of the game can be adapted to each individual's exercise plan or goal. In such a way, HulaMove can also provide health benefits beyond convenience and virtual presence.

6.2 Around-body Interaction and On-device Interaction

There are more dimensions to be explored in our design space. For instance, the shifting distance can potentially be split into multiple levels (*e.g.*, close to the center *v.s.* far from the center), which leads to concentric circles with different diameters and further divides the space around users' bodies into more regions. This is similar to the concept of around-body interaction proposed by Chen *et al.* [13] to expand the interaction space, switch between applications/modes, and increase devices' context-awareness by holding a mobile phone in multiple regions around a user's body. Instead of directly moving the phone, in HulaMove, users could move their waist to different regions (with a quick return) as different interaction gestures. We used touch input as the baseline in the usability study (Section 5). In addition to the standard interaction method, there are some techniques based on commercially available devices, such as capacitive screen-based interaction (*e.g.*, Pre-touch [27], EarTouch [69]), camera-based interaction (*e.g.*, Lip-interact [64], FrownOnError [76]) and microphone-based interaction (*e.g.*, Siri, EarBuddy [73]). HulaMove adds a new form of technique to this kind.

6.3 Limitation and Future Work

There are some limitations of our work. First, there are more aspects worth exploring in the design space. Other than the shifting distance mentioned in Section 6.2, the combination of shifting and rotating gestures, the speed and repetition of gestures, and the movement trajectory are all potential aspects for enriching HulaMove. We plan to investigate more of the design space in future work. Second, in the usability study, we only evaluated our system when participants were actively using waist gestures for interaction. Although we also tested noise robustness to some extent since non-gesture periods were inherently part of the study, participants did not verify detection errors/successes during daily routines. The testing results in Section 4.4 show that our system is robust to noise and the false positive rate is low. But the robustness in more realistic scenarios remains to be validated in the future. Moreover, comparing HulaMove against other eyes-free and hands-free interaction techniques – such as foot interaction [55] – is meaningful future work. It can help future researchers and designers identify the appropriate and efficient interaction methods in different situations. Third, our final goal is to deploy HulaMove on the phone. We plan to compile the machine learning models to the mobile version so that it can be run completely on the phone. The system delay due to the waiting for extra negative windows needs to be optimized by improving our algorithm to provide a smooth and quick system response, *e.g.*, by narrowing down the time window or shortening the step size. We also plan to reduce the sampling rate on the IMU to save battery life. Its impact on model accuracy will need to be further investigated.

7 CONCLUSION

We propose HulaMove, a novel interaction technique that leverages waist movements as a new hands- and eyes-free input method. Through a study involving a series of target acquisition and confirmation tasks, we finalized the design space of waist interaction. We found that users could easily and quickly distinguish eight shifting directions and two rotating directions, and that quick return was the optimal confirmation technique for HulaMove. We then developed an algorithm and a real-time system that can detect and recognize waist gestures from the IMU signals of a phone. With a quick calibration stage performing four gestures, the algorithm adapted to different phone positions and orientations, and transformed the IMU signals to align to human body coordinates. Our hierarchical tree-CNN model robustly detected waist gestures with an accuracy of 97.5%. Finally, we evaluated our system in both physical world and VR scenarios. Our results indicated that *HulaMove* significantly reduced the interaction time by 41.8% compared to a touchscreen method. Moreover, participants reported better user experience and higher presence in VR, and indicated a willingness to use waist interaction during their daily routines. Overall, HulaMove can provide an additional input channel when users' eyes and/or hands are busy, accelerate users' daily operations, and augment their immersive experience in AR/VR.

ACKNOWLEDGMENTS

This work is supported by the Natural Science Foundation of China under Grant No. 62002198. We thank all participants for their efforts under the pandemics.

REFERENCES

[1] . 2020. Beat Saber - A popular VR game. https://beatsaber.com/.
[2] . 2020. Nintendo Switch. https://www.nintendo.com/switch/.
[3] . 2020. Ring Fit Adventure. https://ringfitadventure.nintendo.com/.
[4] . 2020. Xbox One + Kinect. https://www.xbox.com/en-US/.
[5] Takayuki Adachi, Masafumi Goseki, Keita Muratsu, Hiroshi Mizoguchi, Miki Namatame, Masanori Sugimoto, Fusako Kusunoki, Etsuji Yamaguchi, Shigenori Inagaki, and Yoshiaki Takeda. 2013. Human SUGOROKU: Full-Body Interaction System for Students to Learn Vegetation Succession. In *Proceedings of the 12th International Conference on Interaction Design and Children* (New York, New York, USA) *(IDC '13)*. ACM, New York, NY, USA, 364–367. https://doi.org/10.1145/2485760.2485830
[6] Claas Ahlrichs, Albert Samà, Michael Lawo, Joan Cabestany, Daniel Rodríguez-Martín, Carlos Pérez-López, Dean Sweeney, Leo R Quinlan, Gearòid Ò Laighin, Timothy Counihan, et al. 2016. Detecting freezing of gait with a tri-axial accelerometer in Parkinson's disease patients. *Medical & biological engineering & computing* 54, 1 (2016), 223–233.
[7] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge Luis Reyes-Ortiz. 2013. A public domain dataset for human activity recognition using smartphones.. In *Esann*, Vol. 3. Esann, New York, NY, USA, 3.
[8] Alvina Anjum and Muhammad U Ilyas. 2013. Activity recognition using smartphone sensors. In *2013 ieee 10th consumer communications and networking conference (ccnc)*. IEEE, IEEE, New York, NY, USA, 914–919.
[9] Daniel Ashbrook, Carlos Tejada, Dhwanit Mehta, Anthony Jiminez, Goudam Muralitharam, Sangeeta Gajendra, and Ross Tallents. 2016. Bitey: An Exploration of Tooth Click Gestures for Hands-free User Interface Control. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Florence, Italy) *(MobileHCI '16)*. ACM, New York, NY, USA, 158–169. https://doi.org/10.1145/2935334.2935389
[10] Andreas Bulling, Ulf Blanke, and Bernt Schiele. 2014. A Tutorial on Human Activity Recognition Using Body-Worn Inertial Sensors. *ACM Comput. Surv.* 46, 3, Article 33 (Jan. 2014), 33 pages. https://doi.org/10.1145/2499621
[11] Liwei Chan, Chi-Hao Hsieh, Yi-Ling Chen, Shuo Yang, Da-Yuan Huang, Rong-Hao Liang, and Bing-Yu Chen. 2015. Cyclops: Wearable and Single-Piece Full-Body Gesture Input Devices. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. ACM, New York, NY, USA, 3001–3009. https://doi.org/10.1145/2702123.2702464
[12] Jay Chen, Karric Kwong, Dennis Chang, Jerry Luk, and Ruzena Bajcsy. 2006. Wearable sensors for reliable fall detection. In *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*. IEEE, IEEE, New York, NY, USA, 3551–3554.
[13] Xiang "Anthony" Chen, Julia Schwarz, Chris Harrison, Jennifer Mankoff, and Scott Hudson. 2014. Around-Body Interaction: Sensing & Interaction Techniques for Proprioception-Enhanced Input with Mobile Devices. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services* (Toronto, ON, Canada) *(MobileHCI '14)*. ACM, New York, NY, USA, 287–290. https://doi.org/10.1145/2628363.2628402
[14] Gabe Cohn, Daniel Morris, Shwetak Patel, and Desney Tan. 2012. Humantenna: Using the Body as an Antenna for Real-Time Whole-Body Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. ACM, New York, NY, USA, 1901–1910. https://doi.org/10.1145/2207676.2208330
[15] Christine Dierk, Sarah Sterman, Molly Jane Pearce Nicholas, and Eric Paulos. 2018. HäirIÖ: Human Hair as Interactive Material. In *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction* (Stockholm, Sweden) *(TEI '18)*. ACM, New York, NY, USA, 148–157. https://doi.org/10.1145/3173225.3173232
[16] Maiken Hillerup Fogtmann, Jonas Fritsch, and Karen Johanne Kortbek. 2008. Kinesthetic Interaction: Revealing the Bodily Potential in Interaction Design. In *Proceedings of the 20th Australasian Conference on Computer-Human Interaction: Designing for Habitus and Habitat* (Cairns, Australia) *(OZCHI '08)*. ACM, New York, NY, USA, 89–96. https://doi.org/10.1145/1517744.1517770
[17] Dustin Freeman, Nathan LaPierre, Fanny Chevalier, and Derek Reilly. 2013. Tweetris: A Study of Whole-Body Interaction during a Public Art Event. In *Proceedings of the 9th ACM Conference on Creativity & Cognition* (Sydney, Australia) *(C&C '13)*. ACM, New York, NY, USA, 224–233. https://doi.org/10.1145/2466627.2466650
[18] Kathrin Gerling, Ian Livingston, Lennart Nacke, and Regan Mandryk. 2012. Full-Body Motion-Based Game Interaction for Older Adults. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. ACM, New York, NY, USA, 1873–1882. https://doi.org/10.1145/2207676.2208324
[19] Mayank Goel, Chen Zhao, Ruth Vinisha, and Shwetak N. Patel. 2015. Tongue-in-Cheek: Using Wireless Signals to Enable Non-Intrusive and Flexible Facial Gestures Detection. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. ACM, New York, NY, USA, 255–258. https://doi.org/10.1145/2702123.2702591
[20] Piyush Gupta and Tim Dallas. 2014. Feature selection and activity recognition system using a single triaxial accelerometer. *IEEE Transactions on Biomedical Engineering* 61, 6 (2014), 1780–1786.
[21] 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* (Santa Barbara, California, USA) *(UIST '11)*. ACM, New York, NY, USA, 283–292. https://doi.org/10.1145/2047196.2047233
[22] 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* (Paris, France) *(CHI '13)*. ACM, New York, NY, USA, 889–898. https://doi.org/10.1145/2470654.2466114
[23] Chris Harrison, Shilpa Ramamurthy, and Scott E. Hudson. 2012. On-body Interaction: Armed and Dangerous. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction* (Kingston, Ontario, Canada) *(TEI '12)*. ACM, New York, NY, USA, 69–76. https://doi.org/10.1145/2148131.2148148
[24] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in Psychology*. Vol. 52. Elsevier, New York, NY, USA, 139–183.
[25] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, New York, NY, USA, 770–778.
[26] Apiwat Henpraserttae, Surapa Thiemjarus, and Sanparith Marukatat. 2011. Accurate activity recognition using a mobile phone regardless of device orientation and location. In *2011 International Conference on Body Sensor Networks*. IEEE, IEEE, New York, NY, USA, 41–46.
[27] Ken Hinckley, Seongkook Heo, Michel Pahud, Christian Holz, Hrvoje Benko, Abigail Sellen, Richard Banks, Kenton O'Hara, Gavin Smyth, and William Buxton. 2016. Pre-touch sensing for mobile interaction. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 2869–2881.
[28] Mads F Hjorth, Jean-Philippe Chaput, Camilla T Damsgaard, Stine-Mathilde Dalskov, Kim F Michaelsen, Inge Tetens, and Anders Sjödin. 2012. Measure of sleep and physical activity by a single accelerometer: Can a waist-worn Actigraph adequately measure sleep in children? *Sleep and Biological Rhythms* 10, 4 (2012), 328–335.
[29] Da-Yuan Huang, Liwei Chan, Shuo Yang, Fan Wang, Rong-Hao Liang, De-Nian Yang, Yi-Ping Hung, and Bing-Yu Chen. 2016. DigitSpace: Designing Thumb-to-Fingers Touch Interfaces for One-Handed and Eyes-Free Interactions. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. ACM, New York, NY, USA, 1526–1537. https://doi.org/10.1145/2858036.2858483
[30] Martin Hynes, Han Wang, and Liam Kilmartin. 2009. Off-the-shelf mobile handset environments for deploying accelerometer based gait and activity analysis algorithms. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, IEEE, New York, NY, USA, 5187–5190.
[31] Sergey Ioffe and Christian Szegedy. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint* 1, 1 (2015), 8.
[32] Yasha Iravantchi, Yang Zhang, Evi Bernitsas, Mayank Goel, and Chris Harrison. 2019. Interferi: Gesture Sensing Using On-Body Acoustic Interferometry. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, Article 276, 13 pages. https://doi.org/10.1145/3290605.3300506
[33] Hsin-Liu (Cindy) Kao, Artem Dementyev, Joseph A. Paradiso, and Chris Schmandt. 2015. NailO: Fingernails As an Input Surface. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. ACM, New York, NY, USA, 3015–3018. https://doi.org/10.1145/2702123.2702572
[34] Rushil Khurana, Nikola Banovic, and Kent Lyons. 2018. In only 3 minutes: perceived exertion limits of smartwatch use. In *Proceedings of the 2018 ACM International Symposium on Wearable Computers*. ACM, New York, NY, USA, 208–211.
[35] Takashi Kikuchi, Yuta Sugiura, Katsutoshi Masai, Maki Sugimoto, and Bruce H. Thomas. 2017. EarTouch: Turning the Ear into an Input Surface. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Vienna, Austria) *(MobileHCI '17)*. ACM, New York, NY, USA, Article 27, 6 pages. https://doi.org/10.1145/3098279.3098538
[36] Jin Kjölberg. 2004. Designing Full Body Movement Interaction Using Modern Dance as a Starting Point. In *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques* (Cambridge, MA, USA) *(DIS '04)*. ACM, New York, NY, USA, 353–356. https://doi.org/10.1145/

1013115.1013178
[37] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2017. ImageNet Classification with Deep Convolutional Neural Networks. In *Commun. ACM*. ACM, New York, NY, USA, 84–90. https://doi.org/10.1145/3065386
[38] Jennifer R. Kwapisz, Gary M. Weiss, and Samuel A. Moore. 2011. Activity Recognition Using Cell Phone Accelerometers. *SIGKDD Explor. Newsl.* 12, 2 (March 2011), 74–82. https://doi.org/10.1145/1964897.1964918
[39] Chen Liang, Chun Yu, Xiaoying Wei, Xuhai Xu, Yongquan Hu, Yuntao Wang, and Yuanchun Shi. 2021. Auth+Track: Enabling Authentication Free Interaction on Smartphone by Continuous User Tracking. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Tokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3411764.3445624
[40] Lian Loke and Toni Robertson. 2013. Moving and Making Strange: An Embodied Approach to Movement-Based Interaction Design. *ACM Trans. Comput.-Hum. Interact.* 20, 1, Article 7 (April 2013), 25 pages. https://doi.org/10.1145/2442106.2442113
[41] Charles E McCulloch and John M Neuhaus. 2005. Generalized linear mixed models. *Encyclopedia of Biostatistics* 4 (2005), 1.
[42] Stuart McGill. 2010. Core training: Evidence translating to better performance and injury prevention. *Strength & Conditioning Journal* 32, 3 (2010), 33–46.
[43] Joan Mora-Guiard, Ciera Crowell, Narcis Pares, and Pamela Heaton. 2017. Sparking social initiation behaviors in children with Autism through full-body Interaction. *International Journal of Child-Computer Interaction* 11 (2017), 62–71.
[44] Jafet Morales and David Akopian. 2017. Physical activity recognition by smartphones, a survey. *Biocybernetics and Biomedical Engineering* 37, 3 (2017), 388–400.
[45] Florian "Floyd" Mueller, Richard Byrne, Josh Andres, and Rakesh Patibanda. 2018. Experiencing the Body as Play. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. ACM, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173784
[46] Jawad Nagi, Frederick Ducatelle, Gianni A Di Caro, Dan Cireşan, Ueli Meier, Alessandro Giusti, Farrukh Nagi, Jürgen Schmidhuber, and Luca Maria Gambardella. 2011. Max-pooling convolutional neural networks for vision-based hand gesture recognition. In *2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*. IEEE, IEEE, New York, NY, USA, 342–347. https://doi.org/10.1109/ICSIPA.2011.6144164
[47] Vinod Nair and Geoffrey E. Hinton. 2010. Rectified Linear Units Improve Restricted Boltzmann Machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning* (Haifa, Israel) *(ICML'10)*. Omnipress, Madison, WI, USA, 807–814.
[48] Toni Pakkanen and Roope Raisamo. 2004. Appropriateness of Foot Interaction for Non-accurate Spatial Tasks. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems* (Vienna, Austria) *(CHI EA '04)*. ACM, New York, NY, USA, 1123–1126. https://doi.org/10.1145/985921.986004
[49] Toni Pakkanen and Roope Raisamo. 2004. Appropriateness of foot interaction for non-accurate spatial tasks. In *CHI'04 extended abstracts on Human factors in computing systems*. ACM, New York, NY, USA, 1123–1126.
[50] Gonzalo Ramos, Matthew Boulos, and Ravin Balakrishnan. 2004. Pressure Widgets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vienna, Austria) *(CHI '04)*. ACM, New York, NY, USA, 487–494. https://doi.org/10.1145/985692.985754
[51] Deboleena Roy, Priyadarshini Panda, and Kaushik Roy. 2020. Tree-CNN: a hierarchical deep convolutional neural network for incremental learning. *Neural Networks* 121 (2020), 148–160.
[52] Chris Salem and Shumin Zhai. 1997. An Isometric Tongue Pointing Device. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '97)*. ACM, New York, NY, USA, 538–539. https://doi.org/10.1145/258549.259021
[53] Kimitake Sato and Monique Mokha. 2009. Does core strength training influence running kinetics, lower-extremity stability, and 5000-M performance in runners? *The Journal of Strength & Conditioning Research* 23, 1 (2009), 133–140.
[54] Christian Schönauer, Thomas Pintaric, and Hannes Kaufmann. 2011. Full Body Interaction for Serious Games in Motor Rehabilitation. In *Proceedings of the 2nd Augmented Human International Conference* (Tokyo, Japan) *(AH '11)*. ACM, New York, NY, USA, Article 4, 8 pages. https://doi.org/10.1145/1959826.1959830
[55] Jeremy Scott, David Dearman, Koji Yatani, and Khai N. Truong. 2010. Sensing Foot Gestures from the Pocket. In *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology* (New York, New York, USA) *(UIST '10)*. ACM, New York, NY, USA, 199–208. https://doi.org/10.1145/1866029.1866063
[56] Ivan W Selesnick and C Sidney Burrus. 1998. Generalized digital Butterworth filter design. *IEEE Transactions on signal processing* 46, 6 (1998), 1688–1694.
[57] Marcos Serrano, Barrett M. Ens, and Pourang P. Irani. 2014. Exploring the Use of Hand-to-face Input for Interacting with Head-worn Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. ACM, New York, NY, USA, 3181–3190. https://doi.org/10.1145/2556288.2556984
[58] Muhammad Shoaib, Stephan Bosch, Ozlem Durmaz Incel, Hans Scholten, and Paul JM Havinga. 2014. Fusion of smartphone motion sensors for physical activity recognition. *Sensors* 14, 6 (2014), 10146–10176.
[59] Muhammad Shoaib, Stephan Bosch, Ozlem Durmaz Incel, Hans Scholten, and Paul JM Havinga. 2015. A survey of online activity recognition using mobile phones. *Sensors* 15, 1 (2015), 2059–2085.
[60] Misha Sra, Xuhai Xu, and Pattie Maes. 2018. BreathVR: Leveraging Breathing As a Directly Controlled Interface for Virtual Reality Games. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. ACM, New York, NY, USA, Article 340, 12 pages. https://doi.org/10.1145/3173574.3173914
[61] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research* 15, 1 (2014), 1929–1958.
[62] Lee Stearns, Uran Oh, Leah Findlater, and Jon E. Froehlich. 2018. TouchCam: Realtime Recognition of Location-Specific On-Body Gestures to Support Users with Visual Impairments. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4, Article 164 (Jan. 2018), 23 pages. https://doi.org/10.1145/3161416
[63] Lee Stearns, Uran Oh, Leah Findlater, and Jon E. Froehlich. 2018. TouchCam: Realtime Recognition of Location-Specific On-Body Gestures to Support Users with Visual Impairments. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4, Article 164 (Jan. 2018), 23 pages. https://doi.org/10.1145/3161416
[64] Ke Sun, Chun Yu, Weinan Shi, Lan Liu, and Yuanchun Shi. 2018. Lip-interact: Improving mobile device interaction with silent speech commands. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. ACM, New York, NY, USA, 581–593.
[65] Jacinto Vasconcelos-Raposo, Maximino Bessa, Miguel Melo, Luis Barbosa, Rui Rodrigues, Carla Maria Teixeira, Luciana Cabral, and António Augusto Sousa. 2016. Adaptation and validation of the Igroup presence questionnaire (IPQ) in a Portuguese sample. *Presence: Teleoperators and virtual environments* 25, 3 (2016), 191–203.
[66] Eduardo Velloso, Dominik Schmidt, Jason Alexander, Hans Gellersen, and Andreas Bulling. 2015. The feet in human–computer interaction: A survey of foot-based interaction. *ACM Computing Surveys (CSUR)* 48, 2 (2015), 1–35.
[67] 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* (Paris, France) *(CHI '13)*. ACM, New York, NY, USA, 1299–1308. https://doi.org/10.1145/2470654.2466170
[68] 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* (Copenhagen, Denmark) *(MobileHCI '15)*. ACM, New York, NY, USA, 217–226. https://doi.org/10.1145/2785830.2785885
[69] Ruolin Wang, Chun Yu, Xing-Dong Yang, Weijie He, and Yuanchun Shi. 2019. EarTouch: Facilitating Smartphone Use for Visually Impaired People in Mobile and Public Scenarios. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, Article 24, 13 pages. https://doi.org/10.1145/3290605.3300254
[70] 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* (Denver, Colorado, USA) *(CHI '17)*. ACM, New York, NY, USA, 3095–3105. https://doi.org/10.1145/3025453.3025704
[71] Wenge Xu, Hai-Ning Liang, Yuxuan Zhao, Difeng Yu, and Diego Monteiro. 2019. DMove: Directional Motion-Based Interaction for Augmented Reality Head-Mounted Displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300674
[72] Xuhai Xu, Alexandru Dancu, Pattie Maes, and Suranga Nanayakkara. 2018. Hand Range Interface: Information Always at Hand with a Body-centric Mid-air Input Surface. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Barcelona, Spain) *(MobileHCI '18)*. ACM, New York, NY, USA, Article 5, 12 pages. https://doi.org/10.1145/3229434.3229449
[73] Xuhai Xu, Haitian Shi, Xin Yi, WenJia Liu, Yukang Yan, Yuanchun Shi, Alex Mariakakis, Jennifer Mankoff, and Anind K. Dey. 2020. EarBuddy: Enabling On-Face Interaction via Wireless Earbuds. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. ACM, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376836
[74] Xuhai Xu, Chun Yu, Anind K. Dey, and Jennifer Mankoff. 2019. Clench Interface: Novel Biting Input Techniques. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, Article 275, 12 pages. https://doi.org/10.1145/3290605.3300505
[75] Yukang Yan, Chun Yu, Xiaojuan Ma, Shuai Huang, Hasan Iqbal, and Yuanchun Shi. 2018. Eyes-free target acquisition in interaction space around the body for virtual reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–13.

[76] Yukang Yan, Chun Yu, Wengrui Zheng, Ruining Tang, Xuhai Xu, and Yuanchun Shi. 2020. FrownOnError: Interrupting Responses from Smart Speakers by Facial Expressions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. ACM, New York, NY, USA, 14. https://doi.org/10.1145/3313831.3376810

[77] Che-Chang Yang and Yeh-Liang Hsu. 2010. A review of accelerometry-based wearable motion detectors for physical activity monitoring. *Sensors* 10, 8 (2010), 7772–7788.

[78] 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* (Montreal QC, Canada) *(CHI '18)*. ACM, New York, NY, USA, Article 437, 10 pages. https://doi.org/10.1145/3173574.3174011