Gait Gestures: Examining Stride and Foot Strike Variation as an Input Method While Walking

Ching-Yi Tsai
ching-yi@princeton.edu
Princeton University
United States

Ryan Yen
r4yen@uwaterloo.ca
University of Waterloo
Canada

Daekun Kim
d332kim@uwaterloo.ca
University of Waterloo
Canada

Daniel Vogel
dvogel@uwaterloo.ca
University of Waterloo
Canada

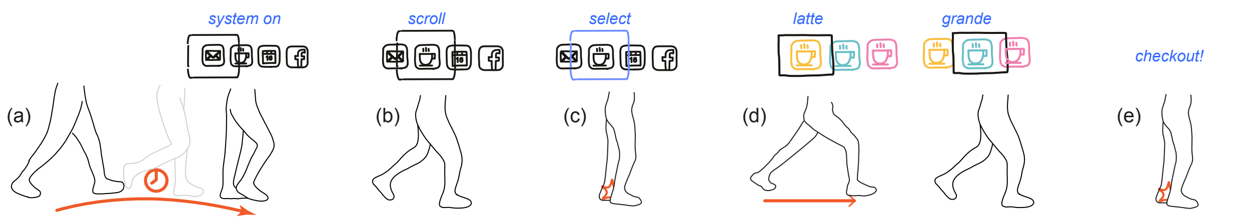

Figure 1: Example "walk-through" using a subset of Gait Gestures for AR interaction: a user wearing AR glasses is walking to work when they decide to order a coffee on the way; (a) they *slow-step* during a stride to 'activate' the system; (b) as they continue to walk, a 'list' of applications slowly scrolls by synchronized with each stride; (c) they 'select' the coffee app by *brushing one foot against the other* mid-stride; (d) following the instruction on the 'menu' buttons, they choose the serving size item with a right-footed BIG-STEP; (e) finally, they brush their feet together to 'confirm' the order. See section 6 for more examples of gait gesture interactions.

ABSTRACT

Walking is a cyclic pattern of alternating footstep strikes, with each pair of steps forming a stride, and a series of strides forming a gait. We conduct a systematic examination of different kinds of intentional variations from a normal gait that could be used as input actions without interrupting overall walking progress. A design space of 22 candidate Gait Gestures is generated by adapting previous standing foot input actions and identifying new actions possible in a walking context. A formative study (n=25) examines movement easiness, social acceptability, and walking compatibility with foot movement logging to calculate temporal and spatial characteristics. Using a categorization of these results, 7 gestures are selected for a wizard-of-oz prototype demonstrating an AR interface controlled by Gait Gestures for ordering food and audio playback while walking. As a technical proof-of-concept, a gait gesture recognizer is developed and tested using the formative study data.

CCS CONCEPTS

• **Human-centered computing** → **Interaction tech**.

KEYWORDS

interaction technique, walking, mixed reality, foot-based gesture

ACM Reference Format:
Ching-Yi Tsai, Ryan Yen, Daekun Kim, and Daniel Vogel. 2024. Gait Gestures: Examining Stride and Foot Strike Variation as an Input Method While Walking. In *The 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24), October 13–16, 2024, Pittsburgh, PA, USA.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3654777.3676342

UIST '24, October 13–16, 2024, Pittsburgh, PA, USA

ACM ISBN 979-8-4007-0628-8/24/10
https://doi.org/10.1145/3654777.3676342

1 INTRODUCTION

The average American adult takes around 5,000 steps each day [8], and active commuting by walking has obvious health benefits [46]. Many people walk throughout their day, such as going to a meeting in another building, getting coffee, meeting friends for dinner, enjoying trails in parks, running errands, and more. During these walks, there is often a need to accomplish small tasks like ordering food ahead, controlling music playback, and sending short text messages (e.g. 'be there soon"). Currently, most people use a smartphone, but touch input while walking is error prone [20] and looking down at a phone screen while walking is hazardous [2]. So, a typical strategy is to stop walking for a moment, retrieve the phone, perform the task, then put the phone away and continue walking . . . but we ask: *Is there a better way to accomplish small tasks without interrupting a walk?*

New developments in mobile augmented reality (AR) avoid the need to look down at a phone screen, but AR input while walking remains an open problem. Speech has issues with background noise, privacy, and social acceptability [51]. Eye-tracking requires overloading gaze direction for both cursor control and walking navigation, creating a safety issue and being prone to fatigue [15]. Using hand-based input may not be practical when hands are cold or covered in winter, or when carrying something. Additionally, common hand gestures, such as pinch, are slower and more difficult to perform when walking [59].

Instead, our approach is to leverage *the way you walk* to provide input to AR. A key observation is that if someone is already going to be walking for several minutes, the time to accomplish a small task is less critical than the ability to do so without interrupting the walk.

Walking has a natural rhythm of steps, called strides in a gait cycle, and our approach is to slightly alter these stride movements to interact with an AR interface over time. AR headsets are becoming more portable and we expect they could easily support 6 DOF foot tracking relative to the body (e.g. using an egocentric camera approach [16, 43, 52]).

Our work relates to previous research on foot input, such as using foot taps on targets [34], tilting the heel or ankle to trigger events [3], and kicking with variable direction or velocity [13]. However, it is important to note that almost all previous work assumes a stationary standing or seated posture. There has been some exploration of foot input when walking. Müller et al. [36] propose using lateral shifts when walking to follow a surface projected "line" to control modes and trigger events (e.g. stepping into a virtual sub-path would select a system option). However, this changes the trajectory of a walk, it assumes high quality environment mapping, and how the method is activated and deactivated was not discussed. Most similar to our work are Yamamoto et al. [57] and Smus and Kostakos [50]. Both propose two running gestures, a lateral side step and "skipping" (two consecutive steps with the same foot). However, a running gait is very different than walking and a gesture like skipping is not very compatible. Moreover, these short papers only report initial results with very small studies.

We conduct a systematic examination of intentional variations from a normal walking gait that could be used as input actions without interrupting overall walking progress. We refer to these as "Gait Gestures". To construct a candidate set of gestures, we identify foot-based gestures in the existing literature that appear compatible with forward body movement and walking ergonomics, and we add new time-based foot gestures specific to walking. With a candidate set of 22 Gait Gestures, we conduct a 25-person formative study to examine suitability in terms of perceived walking compatibility, movement easiness, and social acceptability. Based on the results, 7 Gait Gestures are incorporated into a prototype interaction design to demonstrate how they can control three common types of mobile applications for music, food ordering, and voice calls. Using Wizard-of-Oz recognition and a portable pass-through AR headset, we conduct a usability evaluation in which people use Gait Gestures with our interaction design as they walk an indoor circuit of hallways. Finally, as a technical proof-of-concept, we developed a gait gesture recognizer tested on the formative study data. For the 7 gestures used in our interaction design and a normal walking class to test false-positive performance, the recognizer achieves above 92% accuracy with high precision and recall.

In sum, we contribute a new foot gesture space to perform input actions while walking. We explore suitable gestures and their feasibility using a formative study, an interaction prototype, and a proof-of-concept recognizer. Data and code are available[1].

2 RELATED WORK

We focus our review on foot-based input followed by general work exploring challenges and approaches to enable input while walking.

2.1 Foot-based Input

The origins of using foot input for controlling an interactive system can be dated back to operating machines and instruments with pedals [7]. Early examples of foot input for computers focused on cursor positioning and selection methods, such as pioneering work by Pearson and Weiser [40, 41].

Related to these early approaches, many previous systems and studies use feet to tap on a target. Crossan et al. [9] investigate foot tapping for mobile phone input, for example, invoking menu commands through a single or double-foot tap when standing still. They show it can be as accurate as pulling out a phone from a pocket and performing touch interaction. Paelke et al. [39] demonstrate a technique to interact with a menu on a mobile device using standing kicks. Using a downward-facing camera on the device, the system detects the foot colliding with targets overlaid from a view of the floor. When standing and wearing an AR headset, Müller et al. [34] use foot taps on targets to prevent arm fatigue and avoid social disturbance with voice input. Their evaluation compares a direct mapping, where the foot taps on a floor target, with an indirect mapping, where the foot taps on floor areas mapped to a plane of targets floating in the air. Saunders and Vogel [44] also explore indirect target selection while standing, but with variations like heel taps, toe taps, and kicks. They later apply this technique to operate a desktop computer at a standing desk [45]. Felberbaum and Lanir [10] elicit mappings for various foot tapping and dragging movements to system commands. They include scenarios when standing or sitting in front of a computer screen and on an interactive floor.

2.1.1 Gestural Foot Input. Other work uses foot and body movements that could be considered simple gestures. Han et al. [13] explore kick direction and velocity as input for mobile devices, Kadobayashi et al. [19] use single directional steps as input for a large wall display, and Xu et al. [56] compare single directional steps with hand gestures as input for headset AR. Schmidt et al. [47] use foot-based tangibles to improve the affordance of foot-based gestures. Scott et al. [48] developed a pocket-based sensing system to detect more subtle foot movements like plantar flexion and heel rotation as input for mobile devices. Müller et al. [35] investigate 1-D selection tasks with toe movement, including flexion and extension. A study by Alexander et al. [3] elicits user-defined foot gestures and their mappings to various GUI commands.

In every work discussed so far, the user is essentially stationary when performing foot input. Even when taking small direction steps, as in Kadobayashi et al., the user returns to a neutral position. Perhaps a conceptual step closer to using foot input while walking, are walking-in-place (WIP) methods for locomotion in virtual environments. Methods include lifting only the heel off the ground [38], directional steps and foot-drags [24], side-steps [54], and using the amplitude of WIP movements to control locomotion speed [22]. While WIP techniques share some resemblance to a gait cycle, the explicit goal of WIP is not to actually walk through physical space but to restrict movement to a small fixed area.

2.1.2 Foot Input while Running. Yamamoto et al. [57] and Smus and Kostakos [50] both propose variations of a running gait cycle to provide input to a music player. They define two gestures, a lateral side step and "skipping" (two consecutive steps with the same foot). Each gesture can be performed with the left or right foot, enabling four different input actions. These are short papers published as adjunct proceedings, suggesting preliminary work, but small studies (n = 4 [57] and n = 7 [50]) compared this approach

[1] Study data and analysis code: https://github.com/exii-uw/gait-gestures

to a handheld touchscreen device and headphone buttons with promising initial results.

Compared to walking, running entails a significantly higher speed and more vigorous gait cycle. In a running gait cycle, only one foot touches the ground at the same time and there is a phase when both feet are in the air. This is very different to the speed and stability of a walking gait cycle which has a phase where both feet are on the ground at the same time and one foot is contacting the ground at all times. The increased stability of a walk cycle enables a wider range of foot gestures, and makes high-energy, heterogeneous gestures like skipping unsuitable.

2.2 Other Mobile Input While Walking

Mobile device input while walking has been an active research topic, with a frequent goal of reducing side effects in terms of input stability and safety. Kane et al. [20] show how increasing text and button sizes can make touchscreen phone input more stable when walking, and Goel et al. [12] show how footstep data can be used to correct touchscreen input displacement when typing while walking. Several works explore systems to avoid colliding with obstacles when focused on a phone screen while walking. For example, Hincapié-Ramos and Irani [14] add visual warnings in the periphery of the screen using a depth camera mounted to the back of a phone. Ahn and Kim [2] provide the same style of warning using an ultrasonic sensor. The NotifyEye [31] system routes notifications to a lightweight AR headset, reducing the need to look at a phone screen while walking. All these systems assume touchscreen input on a phone.

Alternative input methods can be used when walking, but these may also have stability issues. For instance, Zhou et al. [59] found that performing a pinch gesture while walking takes 570 ms longer than when stationary. Moreover, the hands may not even be available due to cold weather or when carrying objects. Researchers have proposed novel hands-free input methods that could be used when walking. For example, tongue and lip movements [18] or silent speech [25], but they require specialized tracking equipment such as electrode arrays or dental retainers with capacitive touch sensors. Using gaze input for precise eye-based cursor control when walking can distract the user from monitoring obstacles in their path, posing a safety risk not unlike looking down at a phone.

With the advancement of portable AR headsets, researchers have investigated AR interaction while walking. Lages and Bowman [28] explore different adaptation strategies for interface transitions and placement. Müller et al. [36] use lateral shifts when walking for discrete input selection, with different options rendered on the ground as parallel sub-paths. Kumar et al. [27] combine this technique with eye movement for unlocking AR headsets.

Unlike prior research, our approach eliminates the need for target-dependent interfaces, altering paths to trigger commands, highly specialized sensors, or utilizing other body parts for input.

3 GAIT GESTURE SPACE

A walking gait has different events and phases. There are variations in terminology [23, 37, 42], but they commonly refer to the same foot and body action. We base our terms on those summarized by Kharb et al. [23] (Figure 2).

A *strike* is when the heel touches the ground (Kharb et al. use the term "initial contact") and lift is when the toe or the foot leaves the ground (Kharb et al. use the term "toe off"). The *stance* is the phase when the foot is in contact with the ground after a strike, ending with a lift. The stance can be further divided into four stages: *loading*, when the foot first takes the weight; *mid stance*, from the opposite foot lift until the heel begins to rise; *terminal stance*, until the opposite foot makes contact; and finally *pre swing*, until the foot lift. The *swing* is the phase between the lift and before the next strike. During the swing, the foot is in the air. The swing can be further divided into three stages: the *initial swing* until the two feet are adjacent, followed by the *mid swing* until the tibia is vertical, and finally the *terminal swing* until the strike. A *stride* is the combination of stance and swing phases, starting and ending with a strike by the same foot. A key characteristic, other than stride duration, is the stride length. Finally, the *gait* is the cycle of strides alternating between feet.

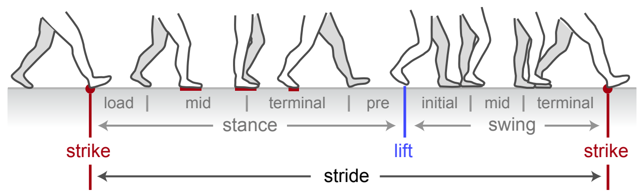

Figure 2: Gait cycle of right foot with strike event, stance phase, lift event, and swing phase that form a single stride. Stance and swing are further divided into stages.

3.1 Requirements and Sources

We focus on foot actions that are *target-independent*. Previous work proposing foot input while standing or sitting frequently adopts a 2D target selection paradigm [34]. This is typically achieved in two ways: direct selection of targets displayed on the floor (e.g. [5]) or by controlling a "foot cursor" for indirect selection of targets displayed somewhere other than the floor (e.g. on a desktop monitor [45]). We recognize direct input is possible while walking by stepping on targets, considering playing a slow and calm version of the children's "hop-scotch" sidewalk game. But, this assumes strict control of targets projected in the environment, something that may be technically feasible with modern AR headsets, but in practical terms, choosing where to position targets relative to the person as they walk in a dynamic street environment is very difficult. Consider street hazards to avoid, surface and colour variation to compensate for, and how to position a target to be simultaneously easy to step on but easy to avoid.

For these reasons, our gesture space is composed of *target-independent* lower-body actions. This removes technical requirements for accurate projection mapping and challenges with direct target selection in a street environment. Each gesture can be interpreted by a system as an event, which then triggers an action based on the current system state. Our intuition was also that focusing on gestural actions would make Gait Gestures more compatible with walking than using a target selection paradigm.

To generate a set of candidate foot and lower body gestural actions that are compatible with walking, we first examined prior

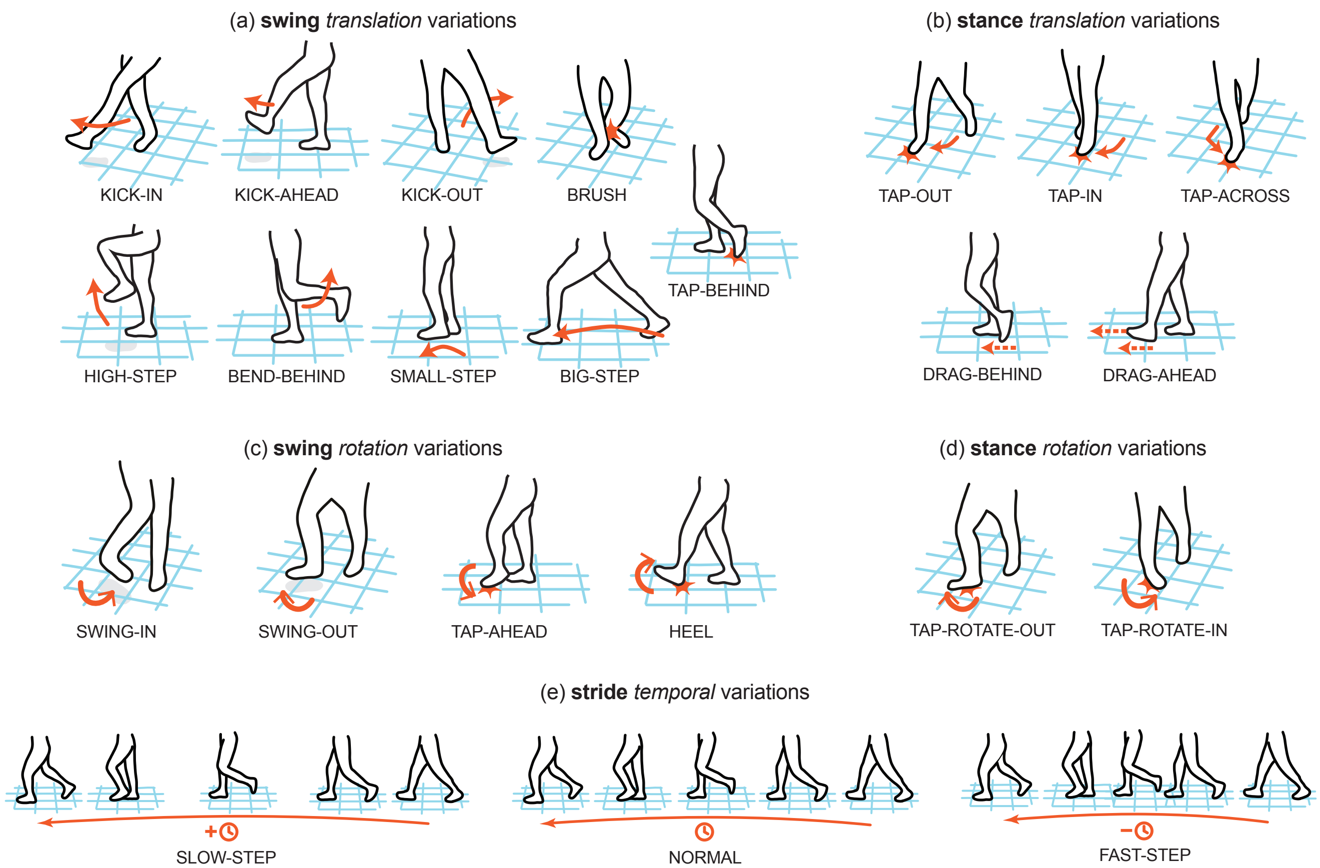

Figure 3: Gait gesture space grouped by variation characteristics.

work using foot gestures when seated or standing (see previous section). Foot actions were considered for inclusion in our gesture space if: (1) they are not dependent on a specific floor target location; (2) they do not require large shifts in body centre of gravity; (3) they do not introduce large deviations from a predominately forward direction; (4) they do not lift both feet of the air at once; and (5) they do not deviate far from how a foot normally contacts the ground when walking.

For example, the body weight shifting actions from [54, 56] do not satisfy requirement 2, and most backward tapping or step-back motions used in walk-in-place or swing-in-place virtual reality locomotion techniques [4, 24] do not satisfy requirement 3. However, many previous foot actions, such as tapping techniques [34] and kicking motions [13, 39], can be integrated into a gait cycle by eliminating returning the foot to its original position to stand in place. Besides selecting candidate gestures from previous sitting or standing foot input systems, the requirements also exclude some potential foot actions and tighten the design space. For example, a potential gait input can be utilizing foot eversion or inversion. When seated, there is little load on the feet so these kinds of actions are possible, but such changes to the loading of the foot during the stance would make walking difficult, and even cause injury.

In addition, we identified dimensions associated with higher-level aspects of walking that have not been explored before, but satisfy our requirements. We propose SLOW-STEP and FAST-STEP gestures to leverage temporal differences in stride and SMALL-STEP and BIG-STEP to explore variation in stride length. Neither of these dimensions would apply to standing foot gestures.

3.2 Candidate Gestures

The final set of 22 candidate gestures is visualized in Figure 3. This gesture space is partially derived from compatible foot gestures in Velloso et al. [53, page 21]. Four of our gesture categories cover spatial variations, separated by the swing and stance portions of a stride over two movement axes: lateral—medial, dorsiflexion—plantarflexion. Temporal variation is a new category we introduce.

Organized by movement type, we provide brief descriptions and establish a concise name for each gesture to enable efficient reference in later sections. Most gestures can be characterized by (1) whether its primary movement variation occurs during the swing or stance phases and (2) whether the movement it introduces is primary translation or rotation. Note that all gestures can be performed with either foot.

During Stance, with Translation.

- TAP-OUT: Position the strike at a lateral position compared to a *normal* stride, away from the opposite foot.
- TAP-ACROSS: Position the strike at a medial position compared to a *normal* stride, in front of the opposite foot.

- TAP-IN: Position the strike at a medial position compared to a *normal* stride, closer to the opposite foot.
- DRAG-AHEAD: In the mid-stance portion, slide the foot along the ground ahead of the body.
- DRAG-BEHIND: In the pre-lift portion, slide the toe or ball of the foot along the ground behind the body.

During Stance, with Rotation.

- TAP-ROTATE-OUT: Position the strike with more lateral rotation than a *normal* stride, pointing away from the opposite foot.
- TAP-ROTATE-IN: Position the strike with more medial rotation than a *normal* stride, pointing towards the opposite foot.

During Swing, with Translation.

- BIG-STEP: Increase the swing distance to make a longer stride length compared to a *normal* stride.
- SMALL-STEP: Decrease the swing distance to make a shorter stride length compared to a *normal* stride.
- BRUSH: In the mid-swing portion, slightly touch the opposite foot with the swinging foot.
- TAP-BEHIND: During the initial swing stage, briefly touch the floor with the tip of the toe and then continue with the mid swing stage.
- KICK-IN: Perform a kicking action in the medial direction during mid and terminal swing portion. A kicking action is a brief extension beyond the normal swing angle, then returning to a normal swing position and trajectory.
- KICK-OUT: Perform a kicking action in the lateral direction during the mid and terminal stages of the swing portion.
- KICK-AHEAD: Perform a kicking action in the forward direction during the mid and terminal swing portion.
- HIGH-STEP: Raise the foot higher than a NORMAL stride during the mid swing portion.
- BEND-BEHIND: Bend the knee to raise the foot higher than usual behind the body during the initial swing portion.

During Swing, with Rotation.

- TAP-AHEAD: During the terminal swing, rotate the foot (plantarflexion) so that the toe strikes the floor first instead of the heel.
- HEEL: During the terminal swing, rotate the entire foot, so that it incorporates more dorsiflexion on strike than a *normal* stride.
- SWING-OUT: During the mid and terminal swing, rotate the foot more laterally than a NORMAL stride during the swing phase.
- SWING-IN: During the mid and terminal swing, rotate the foot more medially than a NORMAL stride during the swing phase.

Temporal gesture.

- SLOW-STEP: Take a slower stride than a NORMAL stride.
- FAST-STEP: Take a faster stride than a NORMAL stride.

It is important to note that we group gestures into stance or swing categories to convey the main characteristics and the primary phase affected. In practice, a gesture is likely to impact the other phase as well, for instance, executing a TAP-IN might slightly affect the terminal stage of the preceding swing.

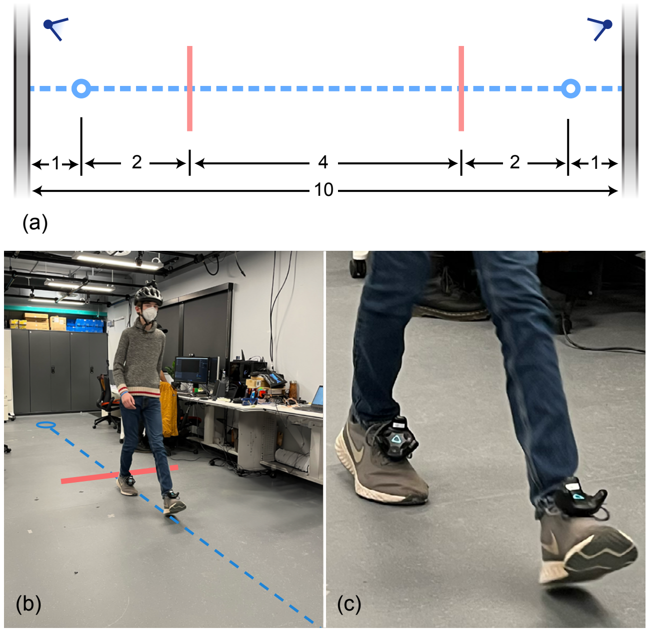

Figure 4: Study setting: (a) top-down diagram showing walking path in blue with red lines representing thresholds where a chime played instructing the participant to perform a gesture on next stride, two triangles are webcam positions; (b) view of participant (walking path and gesture threshold superimposed for illustration, no lines were visible to participants); (c) Lighthouse trackers attached to footwear.

4 STUDY

The goal of this study is to evaluate the suitability of the 22 Gait Gestures. The study is conducted in a lab for internal validity, high quality tracking, and logistical practicality. The participant walks along a straight path, and performs a gesture twice after a threshold distance signalled by a chime. They repeat this twice for each gesture and for each foot, and rate each gesture for perceived movement easiness, social acceptability, and walking compatibility. Logged data is used to calculate relative changes from a normal walk in terms of completion time, and positional and rotational changes.

4.1 Participants

We recruited 25 participants from mailing lists and word-of-mouth. Their ages were 19 to 61 (M=26.6, SD=8.5), 11 identified as women and 14 as men, and all but one said they were "right-footed". Regarding walking frequency, they reported an average score of 3.8 (on a numeric 1 to 5 scale, with 1 "walk very little" and 5 "walk very much"), and they estimated on average more than 5,800 steps per day. Nine participants said they engage in activities involving dexterous foot movements, such as soccer, taekwondo, and dancing. Remuneration was $15. Our Research Ethics Board approved the protocol and participants provided their written consent.

4.2 Apparatus

The study was conducted indoors using a 10-meter-long overground walking path with a width of 1.2 meters (Figure 4a). This created enough space to complete 5 to 6 strides (approximately a dozen steps) at a normal walking pace. We did not use a treadmill since it can introduce significant bio-mechanical differences compared to overground walking [49].

Movement logs were captured using three wireless VIVE 6DOF trackers (each 100ø × 42mm, 89g): two mounted on the participant's footwear using 3D printed mounts and a third attached to a provided bicycle helmet (Figure 4b,c). Four VIVE 2.0 lighthouse receivers covered the entire walking path volume. Additionally, an RGB camera was placed at each end of the walking path to record walk trials using a software-synchronized trigger. The study protocol and data collection were implemented with Unity (version 2021.3.16f1).

4.3 Task

The task is to walk along the 10m path and perform the required gesture in the next stride after a chime is played. Chimes are at two virtual checkpoints (triggered when the helmet tracker passes), creating two gesture repetitions per "walk". When the participant reached the end of the path, they turned around and, when ready, waited for another audio cue before walking back to the start position. As before, they performed the same gesture two more times as they crossed the checkpoints in the return direction. This task was then repeated with the other foot, creating 8 gesture stride data segments per gesture (4 with each foot).

After they reached the end of the path, the experimenter informed the participant to turn around with another system audio cue. The participant then turned and walked back to the starting point. During this returning walk, the same two checkpoints also cued the participant to perform the gestures with the same foot.

4.4 Procedure

After attaching the trackers and donning the helmet, the participant familiarized themselves with the task by practicing walks with and without a randomly selected gesture. At all times, they were instructed to use "their most comfortable walking speed". Before each sequence of walks with a new gesture, the experimenter described and demonstrated the gesture, and then the participant practiced fitting the required gesture into their stride in the "most comfortable manner". During this time, they practiced the gesture with both feet. When they were ready to do the recorded walk, the experimenter informed them of the foot to use (left or right) for the first out-and-back "walk". After completing the required gesture walks with both feet, the experimenter verbally collected subjective ratings in a post-gesture questionnaire, along with additional comments and feedback. This was repeated for all 22 Gait Gestures.

A mandatory 3-minute break was taken every 5 gestures, but the participant could rest between gesture walks whenever they wished. The total session took approximately 1.5 hours, and although walking at a comfortable pace was not fatiguing, the breaks provided an additional measure of safety in this regard.

4.5 Design

This is a within-subjects design with the primary independent variable GESTURE representing the 22 different walking gestures introduced in Section 3 (i.e. BIG-STEP, BRUSH, SLOW-STEP, etc.). In addition, FOOT is a secondary independent variable with two variations (LEFT vs. RIGHT). For each GESTURE and for each FOOT, the participant completed two "walks" along the 10m path, repeating the gesture two times per walk. The order for the GESTURE was shuffled for each user, and so was FOOT for each GESTURE. In summary: 22 GESTURES × 2 FOOT conditions × 4 repetitions = 176 gesture trials per participant. Each participant also completed 2 walks with out performing any gestures (referred to as NORMAL). Not there is no FOOT variation for NORMAL.

The post-gesture questionnaire captures 3 subjective measures using semantic differential anchoring in the form of continuous numeric ratings between 1 and 7 with labelled poles:

- *Movement Easiness*: "How easy was it to perform the gesture while walking?" (1 as "very hard", 7 as "very easy")
- *Social Acceptance*: "How would you feel performing this walking gesture in a public space?" (1 as "very uncomfortable", 7 as "very comfortable")
- *Walking Compatibility*: "How much did performing the gesture change your normal walking movement?" (1 as "very much", 7 as "very little")

We also calculate objective metrics to characterize overall patterns in the gestures. Using logs of the tracked feet and head, the metrics are differences in each GESTURE from the participant's normal walk in terms of *Duration*, *Position*, and *Rotation*. Relevant to this analysis, each participant completed two logged NORMAL "walks" at their most comfortable speeds without performing any gesture before the rest part of the study with other gestures.

5 RESULTS

We first present results for subjective ratings with a clustering analysis to identify promising gestures, and then we describe the data processing pipeline and computed objective characteristics. Since there are 22 levels of GESTURE, we do not perform a statistical analysis of differences (i.e. there are 231 pairs to compare). Instead, we use descriptive statistics, confidence intervals, clustering, and participant comments. Figure 5 provides a summary of both sets of quantitative results.

5.1 Subjective Ratings

Movement Easiness. Examining the results (Figure 5a), regardless of rotational or translational, gestures involving medial movement are generally rated lower (harder to perform) than those with lateral movement (*e.g.* TAP-ROTATE-OUT vs. TAP-ROTATE-IN and TAP-OUT vs. TAP-IN). One participant commented that these almost feel like *"tripping yourself"* [p8]. While FAST-STEP is generally considered socially acceptable and compatible with walking, many participants found it *"exhausting"* [p1, p12, p17, p25], resulting in a lower movement easiness score.

Rotational movements are typically rated as more challenging to perform, with the exception of TAP-BEHIND and HEEL which received high ratings across all three aspects. Regarding dragging

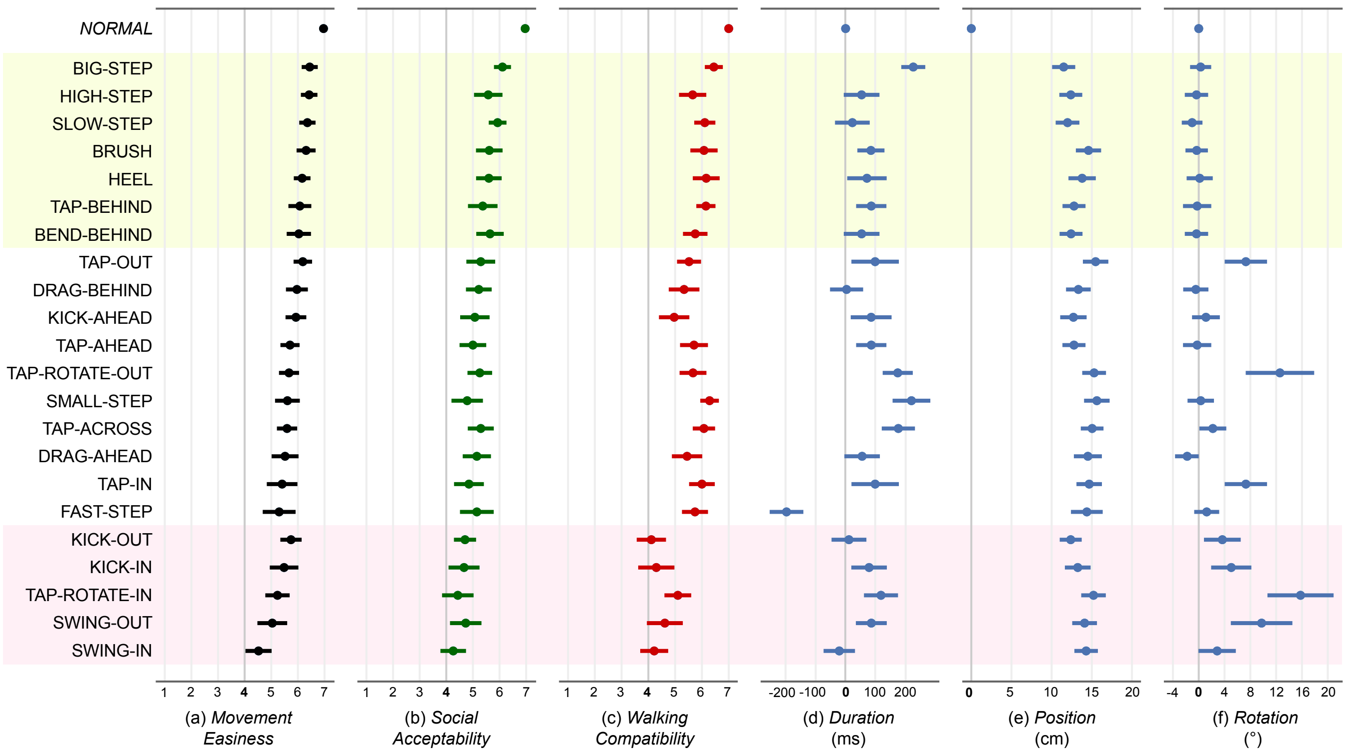

Figure 5: Study results for subjective ratings (a, b, c) and objective characteristics (d, e, f) by GESTURE, with NORMAL for comparison. Gestures are clustered by subjective ratings and ordered by *Movement Easiness* (green, white, and red colour bands for high, middle, and low groups of gestures). Error bars are 95% CI.

motions, DRAG-BEHIND scored better than DRAG-AHEAD. These results may reflect the difficulty of the action itself and how well each of these foot actions align with the corresponding gait cycle phase. Participants provide some insight, for example: *"Tapping behind my body was a bit confusing at first, but seems natural after I incorporate it in my walk."* [p12], *"compared to the behind one [DRAG-BEHIND], this [DRAG-AHEAD] seems to be weird"* [p6].

Social Acceptability. The trend in results (Figure 5b) are closely aligned with *Movement Easiness*. Notably, lateral rotations, SWING-OUT and TAP-ROTATE-OUT, were mentioned by several participants to be *"socially inappropriate"* [p4, p5, p6]. Moreover, for TAP-OUT, KICK-OUT, and KICK-IN, participants expressed concerns about its potential to disrupt pedestrians: *"a sudden side-stepping would block someone walking past you"* [p20], *"May look like tripping them intentionally."* [p15], or *"accidentally kick them in the body"* [P7]. While DRAG-BEHIND and DRAG-AHEAD were generally viewed as socially acceptable, participants noted they might *"hurt their shoes"* [p1, p12, p22], making them reluctant to use these gestures.

Walking Compatibility. Even though certain gestures might be challenging to execute based on *Movement Easiness*, those predominantly involving tapping or stepping movements received high ratings in *Walking Compatibility* (Figure 5c). As one participant noted, *"its movement [FAST-STEP] generally being the same as normal walking"* [p8]. In contrast, all three gestures based on kicking motions were rated lower because, as another participant pointed out, *"you need to pull your feet back after a kick to proceed with the next step"* [p14].

5.1.1 Clustering. As a high-level summary of subjective results, we use clustering to identify groups of gestures that appear most promising for regular use, those suited for infrequent commands, and those that should be used with caution or avoided. K-means clustering is used to find 4 groups (3 as explained above with one more for NORMAL) with feature vectors formed from all three subjective ratings: *Movement Easiness*, *Social Acceptability*, and *Walking Compatibility*. Clusters were ranked based on their average *Movement Easiness* ratings.

Figure 5 visualizes the results as three horizontal highlighted bands of colour. The highest-rated cluster is light green with 7 gestures: BIG-STEP, HIGH-STEP, SLOW-STEP, BRUSH, HEEL, TAP-BEHIND, and BEND-BEHIND. The lowest-rated cluster is light red with 5 gestures: KICK-OUT, KICK-IN, TAP-ROTATE-IN, SWING-OUT, SWING-IN. There are 10 gestures in the middle-rated cluster, which is visualized without any colour highlight. In general, the top cluster has gestures that received high average scores across all three subjective criteria and the bottom cluster has gestures rated low in all three aspects. The gestures in the mid-rated cluster are generally are rated well in one or two of the subjective categories.

5.2 Gait Cycle Detection

To compute objective characteristics related to gesture duration and movement, the logged tracker data was processed to identify gait cycles. The goal is to segment each "walk" in the study into a series of right foot and left foot strides, then label strides that include a gesture. After which, objective characteristics for *Duration*, *Position*, and *Rotation* may be calculated relative to normal walking.

Three 6DOF trackers log the position and orientation of the head and each foot each foot at ~80 Hz. The tracking volume uses standard HTC Lighthouse coordinates where the position is measured in metres and the Y axis is the floor plane normal. We manually aligned the tracker on each foot to be along the shoe centre line with the tracker's Z axis facing the shoes' forward.

Foot Strike Calibration. To account for variations in participant height and shape of head and shoes, we performed a one-time calibration to find the normal up-vector and translational offset from the floor of each foot when flat on the floor. The participant was instructed to face forward along the walking path (i.e. the Z-axis direction) and place each foot on a specifically marked point. The positional and angular offsets were recorded and used for calibration. We refer to components of the 3D position as P_x, P_y, and P_z and the components of the 3D rotation as Θ_x, Θ_y, and Θ_z.

Step Normalization. To compensate for participants with different step heights, we normalize their P_y (vertical) positional data (i.e. along the Y axis) using their individual maximum step height. The normalized P_y values are used to compute vertical acceleration, essential for estimating height changes during walking and for identifying specific events in the gait cycle, such as heel strikes. Note that for analysis, we set P_z to zero at the beginning of each stride and use relative increasing values during the stride.

Direction Normalization and Calibration. Recall the participant first walked along the path to the end position, then turned around and walked back along the same path. All data from the return walk is rotated 180° around the Y axis to normalize the overall walking direction. Since participants make small deviations from the ideal centre line of the walking path, we also adjust position data. For each trial, the mean direction of movement is calculated as the vector from the starting and ending head positions projected to 2D points on the floor. This overall direction vector is used to rotate the positional data around the Y axis to align with the walking path centre line (i.e. the tracking volume Z axis). P_x values perpendicular to the walking path are preprocessed relative to the head X-axis positional movement.

Noise Removal. When walking, unexpected body vibrations can cause higher-frequency noise in sensor readings. Previous studies have suggested that low-pass filters typically have cutoff frequencies between 0.9 and 3 Hz for the normal walking pace [17, 21]. However, we found 3 Hz blurred differences between gesture strides and normal strides, leading to incorrect stride segmentation. Through experimentation, we found an 8 Hz low-pass zero-lag Butterworth filter adequately removed high-frequency noise from the linear acceleration signal without removing characteristic movements of normal and gesture strides (Figure 6a).

Stride Detection. A stride is delineated by heel-to-ground strikes, these strikes can be identified by local minimums in the forward acceleration signal (P_z''), see Figure 6b. A sliding window algorithm

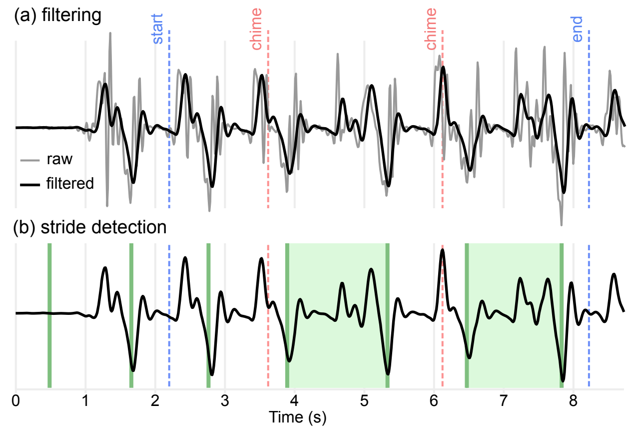

Figure 6: Gait cycle detection example using a LEFT foot BEND-BEHIND walking segment by p22: (a) raw and filtered P_z''; (b) foot strike and stride segmentation of P_z'' with green vertical lines delimiting each stride, strides shaded green are *gesture strides* manually labeled using synchronized video.

was employed to detect these events and define one step as the interval between two consecutive heel strikes. We specifically utilize Mueen's Algorithm for Similarity Search (MASS) to compute the distance profile [33]. This involves comparing the matrix profile of the NORMAL stride as the query pattern with the matrix profile for the sub-sequence of the acceleration of P_z to identify similar patterns. The matrix profile is a vector that stores the Z-normalized Euclidean distance between any sub-sequence within a time series and its nearest neighbour [1]. This approach facilitates the accurate detection of patterns in the time series data.

Labelling Gesture Stride. Using the time markers from the segmented gait cycle, we created a simple annotation tool to play corresponding video segments of the participant as captured by the two webcams. Each stride is manually classified as including a specific "gesture" or just "normal" walking.

To minimize ambiguity, if any stride is visually hard to determine, we analyze variances in gait dynamics. These are deviations in the signal preceding or succeeding the "M" shape (as shown in Figure 6 with a green background), alongside positional data and the logged timing of participants passing checkpoints. A total of 4,400 gesture strides were annotated (176 strides × 25 participants).

5.3 Objective Characteristics

With the segmented and labelled strides, the objective characteristics are computed. Each is a proportional ratio, for example, CLICK stride duration over NORMAL stride duration for the participant. Recall the NORMAL stride was recorded at the start of the session.

Examining the duration data, BIG-STEP and SMALL-STEP have the longest average duration, in contrast to FAST-STEP, which has the shortest. A participant commented on SMALL-STEP saying *"It demands additional focus and energy to ensure the step is taken with less distance"* [P13]. Unexpectedly, SLOW-STEP exhibited only a marginal increase in duration compared to NORMAL. However, participants confirmed that *"these indeed feel like longer time durations"* [P13].

Aside from BIG-STEP, gestures that received high ratings for *Movement Easiness* exhibited only slight differences in duration compared to NORMAL, including HIGH-STEP, BRUSH, and HEEL. Similarly, gestures that scored well in *Walking Compatibility* have minimal rotational deviation from NORMAL. Considering positional variations from the NORMAL stride are noticeable in nearly all gestures, these findings might suggest that participants find positional changes more acceptable in their subjective experience of these gestures compared to rotational adjustments. Overall, the top cluster of gestures also have small deviations from NORMAL in terms of objective characteristics.

6 INTERACTION DESIGN PROTOTYPE

To illustrate how Gait Gestures can be used as input, we created a proof-of-concept interaction design for an AR headset application. The interface covers simple mobile tasks while walking, like receiving a phone call, controlling a music player, and ordering food, along with general functions like activation and adjusting system settings like volume or brightness. We use a Wizard-of-Oz approach for gesture recognition to simplify development and focus only on interaction design. This also avoids recognition quality confounds in a usability evaluation we conduct with the prototype.

6.1 Gesture to Interaction Mapping

The design uses seven gestures that participants rated well in the previous study for walking compatibility, easiness, and social acceptability (*i.e.* all gestures in the highest cluster). Each is mapped to one or more functions that either trigger a global system command or, together with other gestures, define a higher-level interface widget with associated interactions. The chosen gesture-to-function mappings mainly consider the expected frequency of use (*e.g.* BEND-BEHIND mapped to "home" function) and semantic compatibility (*e.g.* left-foot gesture triggering the button on the left). We use a few simple "widgets" instead of more complex and specific task interfaces requiring a larger gesture set. This makes the interaction more consistent, simple, and scalable. These widgets prioritize compatibility with walking, even if task interactions take longer. Walking already consumes time, so this can be acceptable.

Global commands include a HIGH-STEP gesture to 'activate' the system, a SLOW-STEP to 'deactivate', a TAP-BEHIND to return to the previous screen in an app, a BEND-BEHIND to return the system home screen, and a BIG-STEP gesture for a global "option" to trigger the settings or options menu whenever available..

Several tasks use interactions with a *List* widget to select an item from a collection of items (Figure 7). Each item is represented as a small thumbnail image or icon, possibly with a 1 to 3-word description. Our prototype uses a List to open an app from the system home screen, choose an artist or playlist in the music app, and select a dish or beverage in the food ordering app. The List interaction exploits the pace of walking, with a new item side-scrolling into a central box every two strides. When the desired item is in the central box, a BRUSH gesture selects it. If needed, a HIGH-STEP pauses the scrolling to allow time to think about the item before committing, or to pause the interaction to focus on walking (*e.g.* to cross a street).

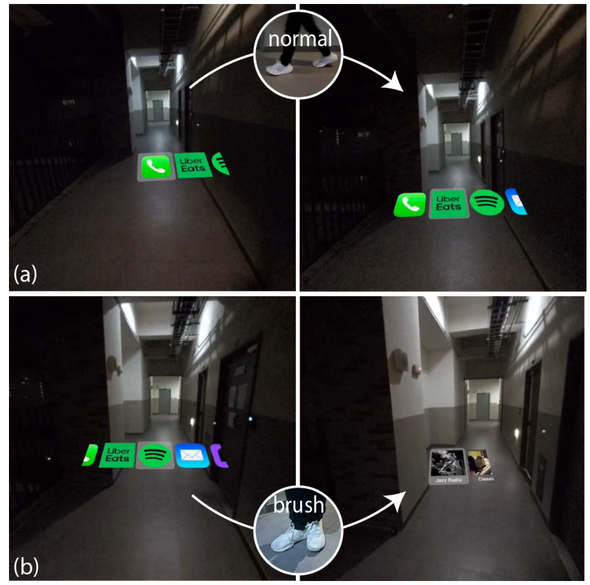

Figure 7: List widget example: (a) items scroll right-to-left as the user takes normal steps; (b) a BRUSH gesture selects the current central item, in this example the music app, then a new List widget appears with a list of playlists. To allow time to think or navigate a real world obstacle, a HIGH-STEP temporarily freezes scrolling.

A *Menu* widget is used to choose among a small number of options in a non-linear manner (Figure 8). After the Menu is invoked, all options are displayed as icons in a horizontal row. The display presents symmetrical left and right foot options around a central item that is most commonly accessed. Our prototype uses a 5-option Menu where the central option is selected with a BRUSH, the options immediately on either side of centre with a left HEEL and right HEEL, and the outermost options with a left BIG-STEP and right BIG-STEP.

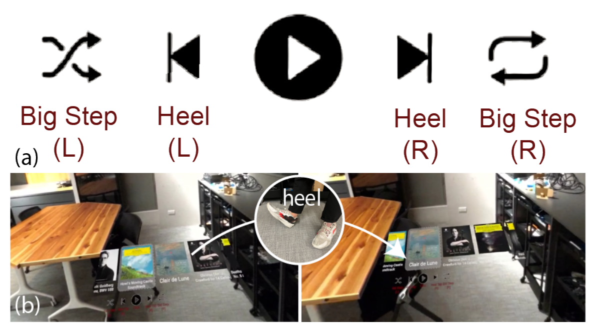

Figure 8: Menu widget example: (a) five options with a central play/pause option mapped to BRUSH; (b) a right HEEL picks the right-of-center option, skipping to the next track in this music app example.

A *Slider* widget (Figure 9) is used to control numeric parameters like brightness, volume, and quantity. Users activate an decrease or increase mode using a left or right BIG-STEP respectively. Once a mode is selected, subsequent normal steps adjust the value accordingly. The global TAP-BEHIND command can also be used to exit the adjustment mode.

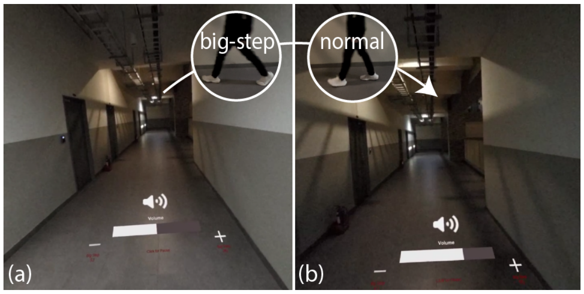

Figure 9: Slider widget example for volume adjustment: (a) a visualization of the current volume is shown with left and right options to "-" decrease or "+" increase; (b) using a right BIG-STEP activates increase mode, after which normal steps increases the volume.

6.2 "Walk"-through

To demonstrate how Gait Gesture input could enable AR interactions in context, we present a usage scenario of a user wearing AR glasses on their way to work:

Jay is on their usual 15-minute walk to work. As they head down the sidewalk, they use a SLOW-STEP to 'activate' the system. The system powers on, and presents a List of apps that slowly scrolls by with a new item entering a central box every two steps. Jay waits for the music app as they walk, then selects it with a BRUSH. This opens another List with the same scrolling interaction for Jay to choose a playlist. Once selected, Jay adjusts the volume using a Slider in which he uses a BIG-STEP with his right foot to switch to increase mode, and takes 3 normal steps to increase the volume by three levels. With music playing, Jay uses a SLOW-STEP to deactivate the system and continue their walk.

A bit later, Jay decides to order a coffee at a shop a few blocks ahead. They activate the system again with a SLOW-STEP and the last app (music) is shown by default. They use a BEND-BEHIND to return to the home screen, then use a series of List interactions to select the food ordering app and the coffee shop. As they use another List to select a beverage, they approach a street and use a HIGH-STEP to pause the system. After safely crossing, they use a HIGH-STEP to resume the List to select a latte. A Menu of size options is shown, and Jay uses a BIG-STEP with their right foot to choose extra-large (they feel tired from a late night). A BRUSH confirms the order and Jay looks forward to picking it up as they pass the coffee shop in a few minutes.

As they approach the office, there is an incoming call from a colleague at work. Rather than talk, Jay uses a TAP-BEHIND to decline, which also activates the system with a Menu of common text messages to send. Jay chooses "be there soon" with List interactions ended by a BRUSH, and the system deactivates again.

Note that these interactions unfold relatively slowly and in sync with Jay's walk. Some of them could happen over a whole block, like choosing a playlist and ordering coffee. However, in the context of Jay's long walk, these represent a relatively small amount of time. Importantly, they did not need to stop to perform these interactions, and they are already spending the time walking anyway.

6.3 System Implementation

The interactions described above are implemented in a prototype system in Unity that runs un-tethered on a Quest Pro headset in AR pass-through mode. Wizard-of-Oz gesture recognition is accomplished using socket communication with a Python application running on a laptop. A study facilitator walks alongside and manually sends events to the headset when they observe the user performing a gesture.

6.4 Usability Evaluation

To gain a general understanding of the usability of Gait Gestures in an application context, we conducted a small usability study with 8 people using the prototype interface and system. There was no overlap with participants in the main study. The study was conducted inside a large building along hallways that formed an approximately 170m path in a mostly rectangular circuit. The hallways were approximately 3 metres wide and free of obstructions with few encounters with other people.

6.4.1 Procedure. The experimenter introduced the seven Gait Gestures to the participant with diagrams and demonstrations. The relationship between gestures and system interactions was explained, and then the participant performed each gesture themselves. Afterward, they donned the headset and started the walk-through without practice. The gestures performed (including NORMAL steps) were recorded by the experimenter and logged by the system.

The specific tasks were: (1) play a song; (2) adjust the volume; (3) order a latte beverage; and (4) decline or accept an incoming call. This required using all seven Gait Gestures, either for global commands or to interact with the three widgets. Many commands and widgets had to be used multiple times.

The entire walking session, including demonstrations, took approximately 5 minutes. Afterward, the participant answered a questionnaire, with 7-point numeric continuous scale ratings of overall easiness, satisfaction, and gesture-command compatibility, along with standard NASA-TLX and SUS. After, the experimenter conducted a short semi-structured interview (see Appendix B.1 for questionnaire and interview questions).

6.4.2 Results. Most participants used the full 170m circuit to complete the tasks, starting and ending near the same place. On average, they spent 238s and executed 43.1 Gait Gestures. For comparison, walking 170m at a typical walking speed of 4.8km/h is 228s.

Overall, users responded positively to the system. On gesture easiness: "How easy was it to perform the gestures while walking?" (1 very hard -7 very easy) participants reported an average of 6.1 (SD = 0.1). On overall satisfaction: "I would like to use it for interaction while walking in the future" (1 very little - 7 very much) they reported an average 6.9 (SD = 0.2). On function compatibility: "I would like to use it for interaction while walking in the future" (1 very little - 7 very much) they reported an average of 5.8 (SD = 0.2).

Regarding NASA-TLX and SUS (Figure 10), TLX values suggest some temporal demand (which we believe is from learning to choose

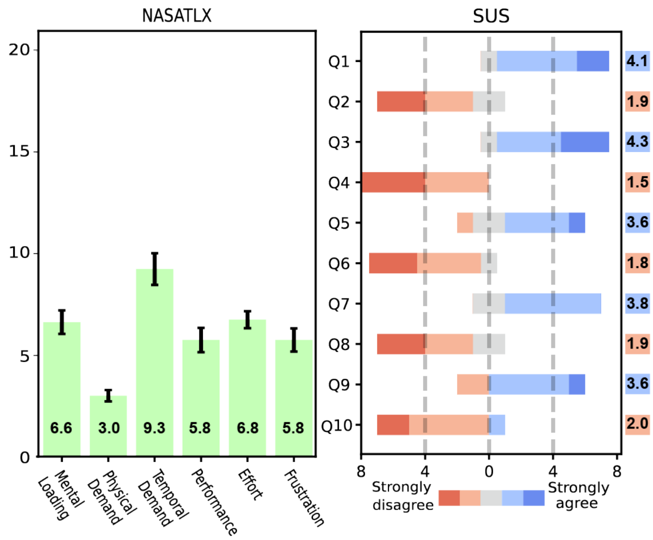

Figure 10: Usability test results for NASA-TLX and SUS

and timing the gestures), but overall scores are reasonable. The overall SUS score is 75.9, suggesting "good" [6].

During the experiment and in the interview, participant comments suggested the tasks were *"easy to perform and intuitive"* [P3], and *"seems most of the time it only requires normal walking, the input is not hard at all."* [p5]. However, participants also provide suggestions to improve further the interaction design: *"I think similar to the slider, the list that scrolls with my strides should have a return button, too"* [p4], and *"the scrolling speed could be even faster for a longer list, or let me switch to a fast-scrolling mode."* [p7] While these recommendations could improve the system further, the current result indicates the feasibility of using Gait Gestures in practice.

7 PROOF-OF-CONCEPT RECOGNIZER

The interaction prototype and usability evaluation suggest Gait Gesture input can be used to accomplish tasks while walking. To evaluate one aspect of technical feasibility, we built a recognizer and tested it offline using data from the main study. The approach builds on the gait cycle identification described earlier. Gait information like segmented strides are used to construct high-level features, and these features are used in a Random Forest Classifier.

7.1 Data Preparation and Feature Generation

The dataset is 6DOF (position and rotation) time series from the two foot trackers and the head logged at 80 Hz. It is segmented by "walk" and gesture type (or normal walking), then segmented into right and left foot strides using the gait cycle segmentation described earlier.

Some features are calculated using relative differences in the NORMAL stride and a gesture stride per participant. Dynamic Time Warping (DTW) distance is used to align the time series for each stride to enable this calculation. Standard low-level features such as mean, standard deviation, maximum, minimum, range, and median are calculated per stride, along with accumulated changes and differences using the aligned NORMAL stride segment. Additionally, higher-level features are calculated, for example: the portion of time above the ground; the portion of time where movement is above average; the relative segment time when the largest movement occurred; the time taken from the largest movement to the average; whether both feet touch each other; the length of the stance phase (the time the foot is fixed on the ground); the portion of time the foot halts in midair; and the rotation acceleration when the foot touching the ground. All low-level and higher-level features used in the Random Forest classifier are in Appendix Table 2.

7.2 Training and Results

This dataset of features is normalized using the Standard Scaler method to ensure uniformity in distribution, then divided into 70% training and 30% test sets (the random state used for splitting the data was 42). Optimal hyperparameters for the Random Forest classifier were determined using the GridSearchCV method with an exhaustive search over specified parameter values. The model was trained using the best hyperparameters and performance was evaluated on the 30% held out test set for overall accuracy, precision, and recall.

The overall accuracy of 92% is with precision and recall metrics both 0.92, indicating a well-performing model that balances false positives and false negatives effectively. The confusion matrix (Figure 11) shows that all gestures are recognized at similar levels, with no significant pattern of misclassification between gestures or with normal walking.

We also evaluate the model through a user-based cross-validation approach via LeavePGroupsOut, which leaves out data from seven participants in each fold. Across all participants, the average accuracy is 0.89 ± 0.13 (95% CI) with precision 0.91 ± 0.10 and recall 0.89 ± 0.13. This indicates a robust performance across different users. Detailed statistics are provided in Appendix A.1.

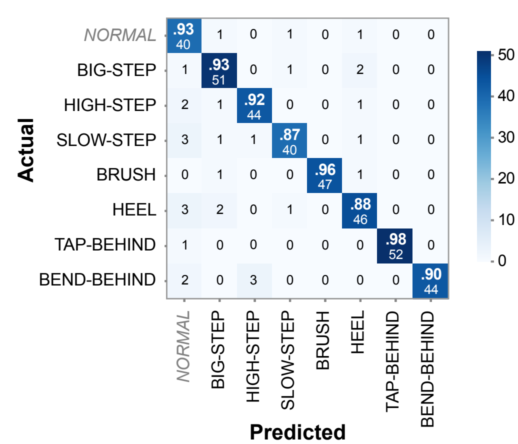

Figure 11: Recognizer confusion matrix

These are encouraging results with a relatively simple recognizer. However, it is important to note that while our model shows high accuracy, it is built on a balanced dataset and its performance in real-world scenarios may differ significantly if the distribution of gestures varies greatly from the training data.

8 DISCUSSION

In this section, we reflect on different aspects of Gait Gestures, suggest possible future work, and identify limitations.

Gait gestures are tied to the cadence of walking, so the cumulative time to complete a task with multiple gestures will be slower than equivalent interactions with conventional input like touchscreens. However, performing a single gait gesture does not appreciably increase stride time from normal walking. Our results suggest a Gait Gesture adds approximately 100 ms to normal stride time, which is less than the additional time to perform pinch gestures while walking [59]. Moreover, performing a single gait gesture is comparable to gaze dwell or saccade selection times, which take 700 ms and 300 ms respectively [26].

People adapt their walking gait according to their surroundings, consider reducing pace before crossing a street, stepping to the side to avoid an obstacle, and stepping over a puddle. We did not test these kinds of variations, but the inclusion of SLOW-STEP as a global command to turn the system on or off is an explicit way to avoid potential false positives. Another global command that would mitigate false positives after they occur is a global "undo" function mapped to another gesture. Perhaps using DRAG-BEHIND from the middle cluster given the semantic association with "erasing behind". Future work could gather data with more variations of walking and examine how these affect Gait Gesture recognition.

There are also natural variations in gait between different people. Our proof-of-concept recognizer performed well even when subsets of our participants were left out as a test set. However, we did not explicitly recruit participants with highly varied walking gaits, for example due to walking style, dexterity, or stride length, nor did we test with different footwear, like heeled dress shoes or industrial safety boots. An individual's previous experience with foot-related activities in particular may have an affect. For example, p17 has significant dance training and p20 is an avid soccer player. They both expressed preferences for specific gestures that seem to be influenced by their background. Notably, p17 expressed a strong preferences for BRUSH which is related to tap dancing, and p20 favoured KICK-IN. A future Gait Gesture system could include the ability to re-map gestures to match such individual preferences.

Additional information about how Gait Gestures are performed could be used for input. For example, gestures like BIG-STEP and SLOW-STEP could be expanded to include continuous variations like *how much bigger* or *how much slower* the step is. This could be leveraged as an advanced level of continuous input modality, such as setting the volume based on increased stride length with BIG-STEP, or using the relative increase in duration in SLOW-STEP for item selection. This kind of overloading of discrete gestural actions has parallels to Liao et al.'s dwell variation input technique [29]. Further research into these kinds of continuous variations is needed need to understand usability limits and suitability, as well as sensing and recognition approaches.

Our formative study used highly accurate VIVE 6DoF absolute position trackers. To collect comparable data in the wild and to conduct an expanded usability with a fully implemented recognition system, a fully portable foot-mounted tracking system is required. A straightforward solution is to adapt existing 6DoF trackers using inside-out or SLAM tracking. For example, we experimented with mounting a Quest Touch Pro controller on each show. Initial results are promising for the purpose of running a study or gathering data, but of course not practical for real deployment. The VIVE portable full-body Ultimate Tracker provides similar functionality and would likely be less cumbersome.

For practical deployment, we imagine two different approaches. Our stride detection relies on acceleration data, so IMU sensors mounted in shoes could be a feasible tracking solutions. This approach has been used in foot gesture sensing before [58]. To avoid instrumenting shoes, an approach using head-mounted or body-mounted cameras for ego-centric body tracking [16, 30, 43, 52, 55] may soon be compact and robust enough for commercial deployment of a Gait Gesture device.

8.1 Limitations

Our study primarily used a single short, straight, flat path. As discussed above, walking style can change based on surroundings. Particular variations, such as hills or curves, will influence the gait cycle. This, in turn, likely alters the perceived subjective experience of performing gestures. Relying on our prototype's global on-and-off gesture could simply restrict Gait Gesture use for straight, flat paths. However, some gestures may be more suited to different kinds of terrain. For example, SWING-OUT may be easier to perform and more natural to execute when turning a corner.

Regarding the gesture recognizer, our current dataset is based on gestures performed four times each in repeated sessions. Introducing a fully random order could add more variability. Additionally, since our P_x (lateral) preprocessing is relative to the head's x-axis movement, it may be affected by arbitrary head movements in practical applications. Despite these potential improvements that could be made, the recognizer performs well across diverse movements from different participants.

Regarding learnability, our initial usability study suggests people can effectively process and recall the 7 top-rated gestures implemented in our prototype. This aligns with Miller's "Magic Number 7 ± 2" principle [32]. However, challenges may arise if future interaction designers require more functions, such as discriminating global commands by foot, which could introduce additional cognitive resources to discover and remember. We acknowledge that overloading users with too much feedback or abrupt changes without a gradual introduction might lead to confusion. To counteract this, we displayed gesture names alongside Menu options in our prototype (Figure 8), in line with how some desktop applications present keyboard shortcuts. A more advanced learning strategy is to show in-situ gesture demonstrations next to the menu buttons [11].

9 CONCLUSION

We presented an investigation of deliberate deviations from a walking gait to serve as mobile AR input commands, all while maintaining the natural flow of walking. We believe the concept of Gait Gestures also demonstrates more generally how computer input can be integrated into an existing primary motor task, but with the goal of compatibility, not necessarily absolute efficiency.

ACKNOWLEDGMENTS

We thank Professor Mike Y. Chen for providing space and equipment at National Taiwan University for conducting the usability test. Additionally, we extend our gratitude to Professor Lung-Pan Cheng for allowing Ching-Yi to complete this work in his spare time during his research-assistantship. This work was made possible by NSERC Discovery Grant 2018-05187 and Canada Foundation for Innovation Infrastructure Fund 33151 "Facility for Fully Interactive Physio-digital Spaces."

REFERENCES

[1] Alireza Abdoli, Sara Alaee, Shima Imani, Amy Murillo, Alec Gerry, Leslie Hickle, and Eamonn Keogh. 2020. Fitbit for Chickens? Time Series Data Mining Can Increase the Productivity of Poultry Farms. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (Virtual Event, CA, USA) *(KDD '20)*. Association for Computing Machinery, New York, NY, USA, 3328–3336. https://doi.org/10.1145/3394486.3403385

[2] Euijai Ahn and Gerard J. Kim. 2013. Casual Video Watching during Sensor Guided Navigation. In *Proceedings of the 12th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry* (Hong Kong, Hong Kong) *(VRCAI '13)*. Association for Computing Machinery, New York, NY, USA, 275–278. https://doi.org/10.1145/2534329.2534378

[3] Jason Alexander, Teng Han, William Judd, Pourang Irani, and Sriram Subramanian. 2012. Putting Your Best Foot Forward: Investigating Real-World Mappings for Foot-Based Gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 1229–1238. https://doi.org/10.1145/2207676.2208575

[4] Yangyi Ang, Puteri Suhaiza Sulaiman, Rahmita Wirza O. K. Rahmat, and Noris Mohd Norowi. 2019. Swing-In-Place (SIP): A Less Fatigue Walking-in-Place Method With Side-Viewing Functionality for Mobile Virtual Reality. *IEEE Access* 7 (2019), 183985–183995. https://doi.org/10.1109/ACCESS.2019.2960409

[5] Thomas Augsten, Konstantin Kaefer, René Meusel, Caroline Fetzer, Dorian Kanitz, Thomas Stoff, Torsten Becker, Christian Holz, and Patrick Baudisch. 2010. Multitoe: high-precision interaction with back-projected floors based on high-resolution multi-touch input. In *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology* (New York, New York, USA) *(UIST '10)*. Association for Computing Machinery, New York, NY, USA, 209–218. https://doi.org/10.1145/1866029.1866064

[6] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies* 4, 3 (2009), 114–123.

[7] Ralph M Barnes, Henry Hardaway, and Odif Podolsky. 1942. Which pedal is best. *Factory Management and Maintenance* 100, 98 (1942), 267.

[8] David R Bassett Jr, Holly R Wyatt, Helen Thompson, John C Peters, and James O Hill. 2010. Pedometer-measured physical activity and health behaviors in United States adults. *Medicine and science in sports and exercise* 42, 10 (2010), 1819.

[9] Andrew Crossan, Stephen Brewster, and Alexander Ng. 2010. Foot Tapping for Mobile Interaction. In *Proceedings of the 24th BCS Interaction Specialist Group Conference* (Dundee, United Kingdom) *(BCS '10)*. BCS Learning & Development Ltd., Swindon, GBR, 418–422.

[10] Yasmin Felberbaum and Joel Lanir. 2018. Better Understanding of Foot Gestures: An Elicitation Study. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173908

[11] Dustin Freeman, Hrvoje Benko, Meredith Ringel Morris, and Daniel Wigdor. 2009. ShadowGuides: visualizations for in-situ learning of multi-touch and whole-hand gestures. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces* (Banff, Alberta, Canada) *(ITS '09)*. Association for Computing Machinery, New York, NY, USA, 165–172. https://doi.org/10.1145/1731903.1731935

[12] Mayank Goel, Leah Findlater, and Jacob Wobbrock. 2012. WalkType: Using Accelerometer Data to Accomodate Situational Impairments in Mobile Touch Screen Text Entry. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 2687–2696. https://doi.org/10.1145/2207676.2208662

[13] Teng Han, Jason Alexander, Abhijit Karnik, Pourang Irani, and Sriram Subramanian. 2011. Kick: Investigating the Use of Kick Gestures for Mobile Interactions. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services* (Stockholm, Sweden) *(MobileHCI '11)*. Association for Computing Machinery, New York, NY, USA, 29–32. https://doi.org/10.1145/2037373.2037379

[14] Juan David Hincapié-Ramos and Pourang Irani. 2013. CrashAlert: Enhancing Peripheral Alertness for Eyes-Busy Mobile Interaction While Walking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 3385–3388. https://doi.org/10.1145/2470654.2466463

[15] Wen-jun Hou, Kai-xiang Chen, Hao Li, and Hu Zhou. 2018. User Defined Eye Movement-Based Interaction for Virtual Reality. In *Cross-Cultural Design. Methods, Tools, and Users*, Pei-Luen Patrick Rau (Ed.). Springer International Publishing, Cham, 18–30.

[16] Dong-Hyun Hwang, Kohei Aso, Ye Yuan, Kris Kitani, and Hideki Koike. 2020. MonoEye: Multimodal Human Motion Capture System Using A Single Ultra-Wide Fisheye Camera. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '20)*. Association for Computing Machinery, New York, NY, USA, 98–111. https://doi.org/10.1145/3379337.3415856

[17] Sampath Jayalath and Nimsiri Abhayasinghe. 2013. A gyroscopic data based pedometer algorithm. In *2013 8th International Conference on Computer Science & Education*. IEEE, IEEE, New York, NY, USA, 551–555.

[18] Arata Jingu, Yudai Tanaka, and Pedro Lopes. 2023. LipIO: Enabling Lips as both Input and Output Surface. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 695, 14 pages. https://doi.org/10.1145/3544548.3580775

[19] R. Kadobayashi, K. Nishimoto, and K. Mase. 1998. Design and evaluation of gesture interface of an immersive walk-through application for exploring cyberspace. In *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*. IEEE, New York, NY, USA, 534–539. https://doi.org/10.1109/AFGR.1998.671003

[20] Shaun K. Kane, Jacob O. Wobbrock, and Ian E. Smith. 2008. Getting off the Treadmill: Evaluating Walking User Interfaces for Mobile Devices in Public Spaces. In *Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services* (Amsterdam, The Netherlands) *(MobileHCI '08)*. Association for Computing Machinery, New York, NY, USA, 109–118. https://doi.org/10.1145/1409240.1409253

[21] Xiaomin Kang, Baoqi Huang, and Guodong Qi. 2018. A novel walking detection and step counting algorithm using unconstrained smartphones. *Sensors* 18, 1 (2018), 297.

[22] Pinachuan Ke and Kenina Zhu. 2021. Larger Step Faster Speed: Investigating Gesture-Amplitude-based Locomotion in Place with Different Virtual Walking Speed in Virtual Reality. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE, New York, NY, USA, 438–447. https://doi.org/10.1109/VR50410.2021.00067

[23] Ashutosh Kharb, Vipin Saini, YK Jain, and Surender Dhiman. 2011. A review of gait cycle and its parameters. *IJCEM International Journal of Computational Engineering & Management* 13, 01 (2011).

[24] Woojoo Kim and Shuping Xiong. 2021. User-defined walking-in-place gestures for VR locomotion. *International Journal of Human-Computer Studies* 152 (2021), 102648. https://doi.org/10.1016/j.ijhcs.2021.102648

[25] Naoki Kimura, Tan Gemicioglu, Jonathan Womack, Richard Li, Yuhui Zhao, Abdelkareem Bedri, Zixiong Su, Alex Olwal, Jun Rekimoto, and Thad Starner. 2022. SilentSpeller: Towards mobile, hands-free, silent speech text entry using electropalatography. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 288, 19 pages. https://doi.org/10.1145/3491102.3502015

[26] Oleg V. Komogortsev, Young Sam Ryu, Do Hyong Koh, and Sandeep M. Gowda. 2009. Instantaneous saccade driven eye gaze interaction. In *Proceedings of the International Conference on Advances in Computer Entertainment Technology* (Athens, Greece) *(ACE '09)*. Association for Computing Machinery, New York, NY, USA, 140–147. https://doi.org/10.1145/1690388.1690412

[27] Abhishek Kumar, Lik-Hang Lee, Jagmohan Chauhan, Xiang Su, Mohammad A. Hoque, Susanna Pirttikangas, Sasu Tarkoma, and Pan Hui. 2022. PassWalk: Spatial Authentication Leveraging Lateral Shift and Gaze on Mobile Headsets. In *Proceedings of the 30th ACM International Conference on Multimedia* (Lisboa, Portugal) *(MM '22)*. Association for Computing Machinery, New York, NY, USA, 952–960. https://doi.org/10.1145/3503161.3548252

[28] Wallace S. Lages and Doug A. Bowman. 2019. Walking with Adaptive Augmented Reality Workspaces: Design and Usage Patterns. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (Marina del Ray, California) *(IUI '19)*. Association for Computing Machinery, New York, NY, USA, 356–366. https://doi.org/10.1145/3301275.3302278

[29] Yi-Chi Liao, Yen-Chiu Chen, Liwei Chan, and Bing-Yu Chen. 2017. Dwell+: Multi-Level Mode Selection Using Vibrotactile Cues. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (Québec City, QC, Canada) *(UIST '17)*. Association for Computing Machinery, New York, NY, USA, 5–16. https://doi.org/10.1145/3126594.3126627

[30] Yuxuan Liu, Jianxin Yang, Xiao Gu, Yijun Chen, Yao Guo, and Guang-Zhong Yang. 2023. EgoFish3D: Egocentric 3D Pose Estimation From a Fisheye Camera via Self-Supervised Learning. *IEEE Transactions on Multimedia* 25 (2023), 8880–8891.

https://doi.org/10.1109/TMM.2023.3242551
[31] Andrés Lucero and Akos Vetek. 2014. NotifEye: Using Interactive Glasses to Deal with Notifications While Walking in Public. In *Proceedings of the 11th Conference on Advances in Computer Entertainment Technology* (Funchal, Portugal) *(ACE '14)*. Association for Computing Machinery, New York, NY, USA, Article 17, 10 pages. https://doi.org/10.1145/2663806.2663824
[32] George A Miller. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review* 63, 2 (1956), 81.
[33] Abdullah Mueen, Sheng Zhong, Yan Zhu, Michael Yeh, Kaveh Kamgar, Krishnamurthy Viswanathan, Chetan Gupta, and Eamonn Keogh. 2022. The Fastest Similarity Search Algorithm for Time Series Subsequences under Euclidean Distance. http://www.cs.unm.edu/~mueen/FastestSimilaritySearch.html.
[34] Florian Müller, Joshua McManus, Sebastian Günther, Martin Schmitz, Max Mühlhäuser, and Markus Funk. 2019. Mind the Tap: Assessing Foot-Taps for Interacting with Head-Mounted Displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300707
[35] Florian Müller, Daniel Schmitt, Andrii Matviienko, Dominik Schön, Sebastian Günther, Thomas Kosch, and Martin Schmitz. 2023. TicTacToes: Assessing Toe Movements as an Input Modality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 520, 17 pages. https://doi.org/10.1145/3544548.3580954
[36] Florian Müller, Martin Schmitz, Daniel Schmitt, Sebastian Günther, Markus Funk, and Max Mühlhäuser. 2020. Walk The Line: Leveraging Lateral Shifts of the Walking Path as an Input Modality for Head-Mounted Displays. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376852
[37] Donald A Neumann et al. 2002. *Kinesiology of the musculoskeletal system.* elsevier, Amsterdam, Netherlands. 627 pages.
[38] Niels C. Nilsson, Stefania Serafin, Morten H. Laursen, Kasper S. Pedersen, Erik Sikström, and Rolf Nordahl. 2013. Tapping-In-Place: Increasing the naturalness of immersive walking-in-place locomotion through novel gestural input. In *2013 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, New York, NY, USA, 31–38. https://doi.org/10.1109/3DUI.2013.6550193
[39] Volker Paelke, Christian Reimann, and Dirk Stichling. 2004. Kick-up Menus. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems* (Vienna, Austria) *(CHI EA '04)*. Association for Computing Machinery, New York, NY, USA, 1552. https://doi.org/10.1145/985921.986128
[40] G. Pearson and M. Weiser. 1988. Exploratory Evaluation of a Planar Foot-Operated Cursor-Positioning Device. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Washington, D.C., USA) *(CHI '88)*. Association for Computing Machinery, New York, NY, USA, 13–18. https://doi.org/10.1145/57167.57169
[41] Glenn Pearson and Mark Weiser. 1988. Exploratory Evaluations of Two Versions of a Foot-Operated Cursor-Positioning Device in a Target-Selection Task. *SIGCHI Bull.* 19, 3 (jan 1988), 70–75. https://doi.org/10.1145/49108.1046356
[42] Chandra Prakash, Rajesh Kumar, and Namita Mittal. 2018. Recent developments in human gait research: parameters, approaches, applications, machine learning techniques, datasets and challenges. *Artificial Intelligence Review* 49 (2018), 1–40.
[43] Helge Rhodin, Christian Richardt, Dan Casas, Eldar Insafutdinov, Mohammad Shafiei, Hans-Peter Seidel, Bernt Schiele, and Christian Theobalt. 2016. EgoCap: egocentric marker-less motion capture with two fisheye cameras. *ACM Trans. Graph.* 35, 6, Article 162 (dec 2016), 11 pages. https://doi.org/10.1145/2980179.2980235
[44] William Saunders and Daniel Vogel. 2015. The performance of indirect foot pointing using discrete taps and kicks while standing. In *Proceedings of the 41st Graphics Interface Conference* (Halifax, Nova Scotia, Canada) *(GI '15)*. Canadian Information Processing Society, CAN, 265–272.
[45] William Saunders and Daniel Vogel. 2016. Tap-Kick-Click: Foot Interaction for a Standing Desk. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems* (Brisbane, QLD, Australia) *(DIS '16)*. Association for Computing Machinery, New York, NY, USA, 323–333. https://doi.org/10.1145/2901790.2901815
[46] Christine Schäfer, Barbara Mayr, Maria Dolores Fernandez La Puente de Battre, Bernhard Reich, Christian Schmied, Martin Loidl, David Niederseer, and Josef Niebauer. 2020. Health effects of active commuting to work: The available evidence before GISMO. *Scandinavian journal of medicine & science in sports* 30 (2020), 8–14.
[47] Dominik Schmidt, Raf Ramakers, Esben W. Pedersen, Johannes Jasper, Sven Köhler, Aileen Pohl, Hannes Rantzsch, Andreas Rau, Patrick Schmidt, Christoph Sterz, Yanina Yurchenko, and Patrick Baudisch. 2014. Kickables: Tangibles for Feet. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 3143–3152. https://doi.org/10.1145/2556288.2557016
[48] 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)*. Association for Computing Machinery, New York, NY, USA, 199–208. https://doi.org/10.1145/1866029.1866063
[49] Marie B. Semaan, Laura Wallard, Valentin Ruiz, Christophe Gillet, Sébastien Leteneur, and Emilie Simoneau-Buessinger. 2022. Is treadmill walking biomechanically comparable to overground walking? A systematic review. *Gait & Posture* 92 (2022), 249–257. https://doi.org/10.1016/j.gaitpost.2021.11.009
[50] Boris Smus and Vassilis Kostakos. 2010. Running gestures: hands-free interaction during physical activity. In *Proceedings of the 12th ACM International Conference Adjunct Papers on Ubiquitous Computing - Adjunct* (Copenhagen, Denmark) *(UbiComp '10 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 433–434. https://doi.org/10.1145/1864431.1864473
[51] T.E. Starner. 2002. The role of speech input in wearable computing. *IEEE Pervasive Computing* 1, 3 (2002), 89–93. https://doi.org/10.1109/MPRV.2002.1037727
[52] Denis Tome, Thiemo Alldieck, Patrick Peluse, Gerard Pons-Moll, Lourdes Agapito, Hernan Badino, and Fernando de la Torre. 2023. SelfPose: 3D Egocentric Pose Estimation From a Headset Mounted Camera. *IEEE Trans. Pattern Anal. Mach. Intell.* 45, 6 (jun 2023), 6794–6806. https://doi.org/10.1109/TPAMI.2020.3029700
[53] 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 Comput. Surv.* 48, 2, Article 21 (sep 2015), 35 pages. https://doi.org/10.1145/2816455
[54] Congzhi Wang, Oana A. Dogaru, Patrick L. Strandholt, Niels C. Nilsson, Rolf Nordahl, and Stefania Serafin. 2018. Step aside: An Initial Exploration of Gestural Input for Lateral Movement during Walking-in-Place Locomotion. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology* (Tokyo, Japan) *(VRST '18)*. Association for Computing Machinery, New York, NY, USA, Article 34, 5 pages. https://doi.org/10.1145/3281505.3281536
[55] Weipeng Xu, Avishek Chatterjee, Michael Zollhöfer, Helge Rhodin, Pascal Fua, Hans-Peter Seidel, and Christian Theobalt. 2019. Mo2Cap2: Real-time Mobile 3D Motion Capture with a Cap-mounted Fisheye Camera. *IEEE Transactions on Visualization and Computer Graphics* 25, 5 (2019), 2093–2101. https://doi.org/10.1109/TVCG.2019.2898650
[56] 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)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300674
[57] Tetsuya Yamamoto, Masahiko Tsukamoto, and Tomoki Yoshihisa. 2008. Foot-Step Input Method for Operating Information Devices While Jogging. In *2008 International Symposium on Applications and the Internet*. IEEE, New York, NY, USA, 173–176. https://doi.org/10.1109/SAINT.2008.97
[58] Zihan Yan, Jiayi Zhou, Yufei Wu, Guanhong Liu, Danli Luo, Zihong Zhou, Haipeng Mi, Lingyun Sun, Xiang 'Anthony' Chen, Ye Tao, Yang Zhang, and Guanyun Wang. 2022. Shoes++: A Smart Detachable Sole for Social Foot-to-foot Interaction. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 2, Article 85 (jul 2022), 29 pages. https://doi.org/10.1145/3534620
[59] Yun Zhou, Tao Xu, Bertrand David, and René Chalon. 2016. Interaction on-the-go: a fine-grained exploration on wearable PROCAM interfaces and gestures in mobile situations. *Universal Access in the Information Society* 15 (2016), 643–657. https://doi.org/10.1007/s10209-015-0448-6

A RECOGNIZER

A.1 Additional Recognizer Results

Table 1: Additional Recognizer Test Results

	Precision	Recall	F1-Score	Support
NORMAL	0.77	0.93	0.84	43
BIG-STEP	0.89	0.93	0.91	55
HIGH-STEP	0.92	0.92	0.92	48
SLOW-STEP	0.93	0.87	0.90	46
BRUSH	1.00	0.96	0.98	47
HEEL	0.88	0.88	0.88	52
TAP-BEHIND	1.00	0.98	0.99	53
BEND-BEHIND	1.00	0.90	0.95	49
Accuracy			0.92	393
Macro Avg.	0.92	0.92	0.92	393
Weighted Avg.	0.93	0.92	0.92	393

A.2 Features Selected

Table 2: Recognizer features

Description	Methods	Attributes
The portion of the time movement is above zero	Computes the mean of the times where the pos_z_a attribute is greater than zero.	pos_z_a
The portion of the time movement is above average	Computes the mean of the times where the pos_y attribute is greater than its average.	pos_y
Moving average of pos_z	Computes the mean of the rolling window (size: 10) of the pos_z attribute.	pos_z
The position of pos_z moves from start to end	Subtracts the first element of the pos_z attribute from the last element.	pos_z
The portion of the time when the first peak of ang_x shows	Finds the first peak of the ang_x attribute and divides it by the length of the data.	ang_x
Time from the highest peak of ang_x to next 0 after that peak	Identifies the highest peak of ang_x and finds the time from there to the next zero, divided by the total time length.	ang_x
Number of peaks in pos_z_a	Counts the number of peaks in the pos_z_a attribute.	pos_z_a
Number of times pos_z_a crosses zero	Counts the number of times pos_z_a changes its sign from positive to negative or vice versa.	pos_z_a
Distance from peak to valley in pos_z_a	Compute the difference between the maximum and minimum values of the pos_z_a attribute.	pos_z_a
The amount of two feet touch in pos_z_a	Compute the pos_x the find collision between right and left foot attribute.	pos_x
Rotation acceleration after touching ground	Compute the angular movement acceleration when pos_z is 0 (on the ground)	angular_movement

B USABILITY EVALUATION SUPPLEMENTARY

B.1 Questionnaire

Table 3: Questionnaire used for evaluating interaction design prototype

Category	Question
Gesture Easiness	How easy was it to perform the gestures while walking? (1 very hard - 7 very easy, continuous scale)
Overall Satisfaction	How compatible were gestures with the system functions and commands in general? (1 very weird - 7 very matched, continuous scale)
Function Compatibility	I would like to use it for interaction with HMD while walking in the future. (1 very little - 7 very much, continuous scale)
NASA-TLX	
Mental loading	How mentally demanding was the task? (1 very low - 21 very high, continuous scale)
Physical Demand	How physically demanding was the task? (1 very low - 21 very high, continuous scale)
Temporal Demand	How hurried or rushed was the pace of the task? (1 very low - 21 very high, continuous scale)
Performance	How successful were you in accomplishing what you were asked to do? (1 Perfect - 21 Failure, continuous scale)
Effort	How hard did you have to work to accomplish your level of performance? (1 very low - 21 very high, continuous scale)
Frustration	How insecure, discouraged, irritated, stressed, and annoyed were you? (1 very low - 21 very high, continuous scale)
SUS (5-point Likert scale)	
Q1	I think that I would like to use this system frequently while walking with HMD.
Q2	I found the system unnecessarily complex.
Q3	I thought the system was easy to use.
Q4	I think that I would need the support of a technical person to be able to use this system.
Q5	I found the various functions in the system were well integrated.
Q6	I thought there was too much inconsistency in this system.
Q7	I imagine that most people would learn to use this system very quickly.
Q8	I found the system very awkward to use.
Q9	I felt very confident using the system.
Q10	I needed to learn a lot of things before I could get going with this system.