WaveTrace: Motion Matching Input using Wrist-Worn Motion Sensors

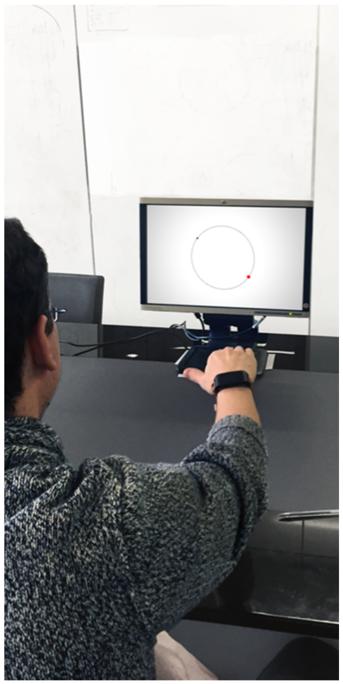

Figure 1: WaveTrace is an interaction technique based on motion matching between moving controls and pointing movements (recorded by an off-the-shelf smartwatch).

David Verweij
Eindhoven University of Technology
Eindhoven, the Netherlands
d.verweij@student.tue.nl

Vassilis Javed Khan
Eindhoven University of Technology
Eindhoven, the Netherlands
v.j.khan@tue.nl

Augusto Esteves
Edinburgh Napier University
Edinburgh, United Kingdom
a.esteves@napier.ac.uk

Saskia Bakker
Eindhoven University of Technology
Eindhoven, the Netherlands
s.bakker@tue.nl

CHI'17 Extended Abstracts, *May 06-11, 2017, Denver, CO, USA*
ACM 978-1-4503-4656-6/17/05.
http://dx.doi.org/10.1145/3027063.3053161

Abstract
We present WaveTrace, a novel interaction technique based on selection by motion matching. In motion matching systems, targets move continuously in a singular and pre-defined path – users interact with these by performing a synchronous bodily movement that matches the movement of one of the targets. Unlike previous work which tracks user input through optical systems, WaveTrace is arguably the first motion matching technique to rely on motion data from inertial measurement units readily available in many wrist-worn wearable devices such as smart watches. To evaluate the technique, we conducted a user study in which we varied: hand; degrees of visual angle; target speed; and number of concurrent targets. Preliminary results indicate that the technique supports up to eight concurrent targets; and that participants could select targets moving at speeds between 180 and 270°/s (mean acquisition time of 2237ms, and average success rate of 91%).

Author Keywords
Input technique; motion matching; wearables; smart watches; touchless interaction; gestural input.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

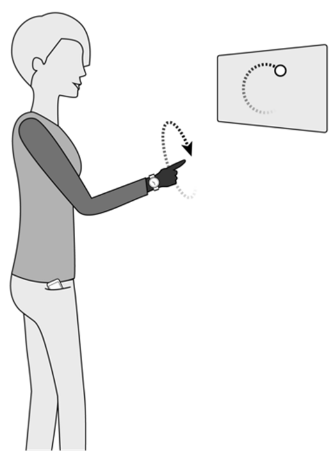

Figure 2: A target selection occurs when there is a high correlation (above 0.8) between the user and target motions.

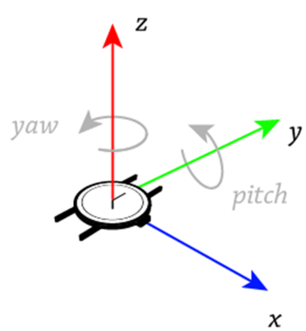

Figure 3: Readily available IMU's provide movement data. WaveTrace relies on *yaw* and *pitch* values provided by the IMU, which respectively represent rotation around the z and y axes.

Introduction

This paper introduces WaveTrace, a motion matching input technique that tracks user input through wrist-worn motion-sensors. Unlike traditional interfaces where targets are static (to facilitate pointing), motion matching interfaces display targets that move continuously in a specific trajectory (Figure 2) – with users selecting a target by simply mimicking that target's movement in real-time. The most common implementation of the technique works by computing the Pearson's correlation between the corresponding x- and y-coordinates of the user input and each target's positions within a certain time window. If the correlation in the x- and y-axis exceeds a certain threshold, the corresponding target is selected.

This idea was initially introduced to minimise pointing actions with a cursor [5,11], but has recently been explored in different scenarios. In Pursuits [10] (public displays), Orbits [4] (smart watches), and AmbiGaze [9] (smart rooms), an eye-tracker captures the input naturally provided by the user's eyes while these follow various moving targets. Similarly, in TraceMatch [3] and PathSync [2] (smart TVs), a simple optical system attached to a TV tracks the user's body for motion matching input.

Despite their promise, the systems still present several limitations that stops them from becoming ubiquitous. Pursuits, TraceMatch and PathSync use an optical tracking system that is attached to a single device (public display, TV). For example, PathSync is limited to six users, which need to stand in the Kinect's field of view to interact with one device. As such, not only is this approach hard to scale, optical systems are particularly susceptible to environmental lighting conditions and field of view issues. Furthermore, these input systems have also raised many privacy concerns in the past [8], especially when operating in the context of smart homes where they video record not only user input, but all activity in a particular region in space. On the other hand, wearable eye-trackers such as the ones used in Orbits and AmbiGaze are still quite costly ([7], ~€1400), and mostly are not positioned as consumer items (e.g., they require external computing power).

This paper explores, arguably for the first time, the feasibility of inertial measurement units (IMUs) in the support of motion matching input. IMUs are readily available and accessible in many of the devices we carry with us, including smart watches, fitness trackers, and jewellery (Figure 3). The goal of this work is to enable the use of such everyday devices as direct, uniform, and scalable input in the interaction with the many smart objects that start to surround us.

The remainder of this paper introduces WaveTrace in more detail, and reports on a preliminary user study that explores the robustness of the technique and the participants' ability to select targets of different characteristics. We conclude the paper with a description of the future work that needs to be undertaken to further this technique, its application(s), and to continue addressing the limitations of earlier motion matching input techniques.

WaveTrace

As with PathSync [2] and TraceMatch [3], WaveTrace is a motion matching technique that supports target acquisition through pointing gestures. Unlike these techniques, which use a computer vision system to

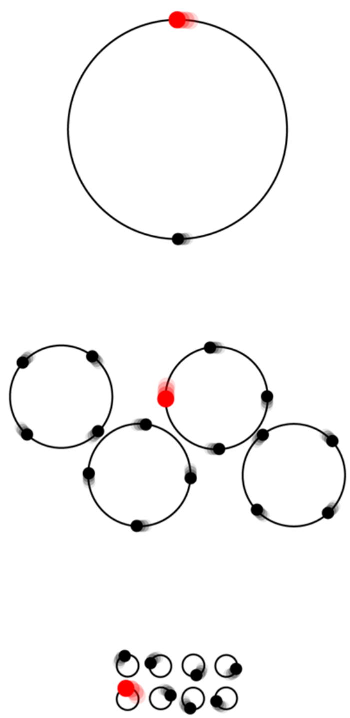

Figure 4: Three example trials for the user study presented in the paper. Participants were asked to point at the red target, which was slightly larger than the others to reduce visual search. The blur was added for representation purposes.

translate hand positions into x- and y-coordinates, WaveTrace was implemented using motion data from an off-the-shelf Android smart watch (Sony Smartwatch 3). This is achieved using the device's embedded IMU, which automatically translates arm and hand movements into, among other things, Euler angles (yaw, pitch, roll). The system logs and computes a Person's correlation between the x- and y-coordinates of the targets being displayed, and the corresponding *yaw* and *pitch* representations of the user's hand rotation (see Figures 2 and 3). Both *yaw* and *pitch* values are obtained through Android's fusion sensor 'Rotation Vector'. This is achieved at 195Hz. If any of the correlations between the x-yaw and y-pitch are higher than 0.8 (as suggested by [4,10]), the corresponding target is selected.

To calculate such correlations, a window size of 1500ms was used – 292 data points, respectively. This is based on previous work on gaze-based motion matching input [4,10], where 1000ms was used. An additional 500ms was required, as the user's hands are not as quick as the eye when catching up to moving stimulus [1].

Visually, WaveTrace is deeply inspired by Orbits [4], where each control is comprised of one or multiple targets that share the same circular trajectory (Figure 4). Target disambiguation is supported by giving each target a different phase. Furthermore, target speed and movement direction can also aid in their disambiguation. Finally, the technique supports both discrete and continuous control. To provide continuous input, the user simply needs to continuously match a target's movement.

The next section describes an on-going user study to test a range of target variables to explore the abilities of wrist-worn motion sensors as input for motion correlation selection. We report on success rates and acquisition times and effectiveness of different parametrical setups.

Pilot User Study

We conducted a user study to first, test the feasibility of our idea; and second, highlight performance differences between WaveTrace, which uses Euler angles to represent user input, and optical-based systems described in literature (where user input is represented as x and y in space). To achieve this, several of the target conditions found in PathSync [2] and TraceMatch [3] were replicated.

We recruited 13 participants, of which 2 were excluded from the study results as they did not follow the instructions. The resulting 11 participants (2F, 9M), aged between 23 and 42 (M = 32, SD = 5.99), were all right-handed. Participants rated their experience with computer technology and mobile devices with a 4.3 (SD=0.65) out of 5; and their experience with wearable technology with a 2.7 (SD=0.65) out of 5.

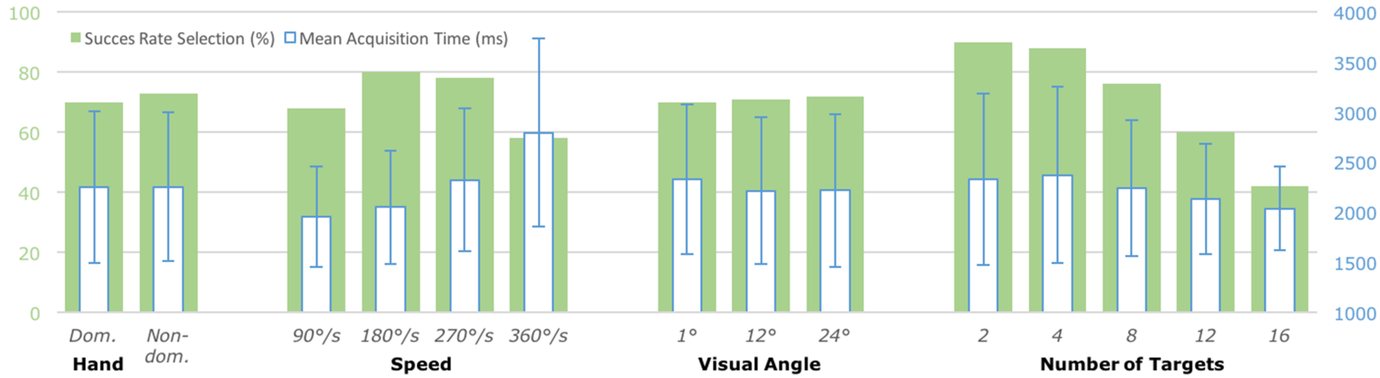

Figure 5: We plotted the acquisition time and success rate for each condition and its variables. White bars depict mean acquisition time. Error bars depict the corresponding standard deviation. Green bars depict the success rate selection.

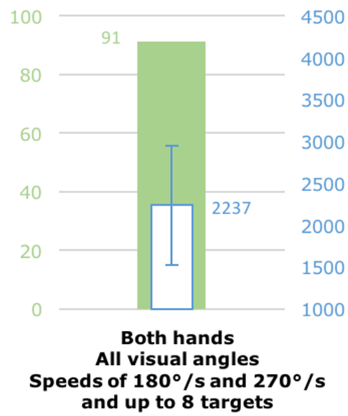

Figure 6: Taking out the least performing configurations for this technique (speeds 90°/s and 360°/s; number of targets 12 and 16) the technique has a performance of 91% success rate with a mean acquisition time of 2237 ms (*SD* = 711).

Each participant sat on a chair, 115 cm away from an external monitor (Figure 1). Participants' wore a Sony Smartwatch 3 on their wrist, which wirelessly transferred the rotation data to the system. For each presented trial, all participants were instructed to follow the red target as accurately as possible. The system presented a countdown starting at *5*. An instruction to 'start pointing at the screen' was shown from *2* onwards to ensure all participants started following a target approximately from the centre of the screen. At *0* the system presented a set of targets, one of them being red (which participants had to acquire in under six seconds). A countdown for the next trial was presented after any selection had occurred, or six seconds had passed (a time out). During each countdown participants were able to pause the system at their convenience.

To compare the performance of WaveTrace to previous motion matching techniques, trials varied the displayed targets in: degrees of visual angle (1°; 12°; 24°); speed (90°/s; 180°/s; 270°/s; 360°/s); and number of targets displayed in tandem (2; 3; 8; 12; 16) – Figure 4. In sum, there were 60 unique trials (3 x 4 x 5), which occurred four times: twice for the dominant hand (in a randomized order), and twice for the non-dominant hand (users switched hands halfway through the study) – 240 trials per participants in total. The hand condition was introduced to examine whether the hand's dominance significantly influences user performance with the technique. Each participant was asked to fill in the NASA-Task Load Index [6] after each hand condition. They were also given the chance to provide additional remarks or feedback at the end of the study. The study took no more than 60 min. per participant.

Results

Hand

A pair-wise t-test revealed no significant differences in user performance between dominant and non-dominant hands (Figure 5). This was true for both the success rate, $t(10) = -0.45$, $p = .66$; and acquisition times,

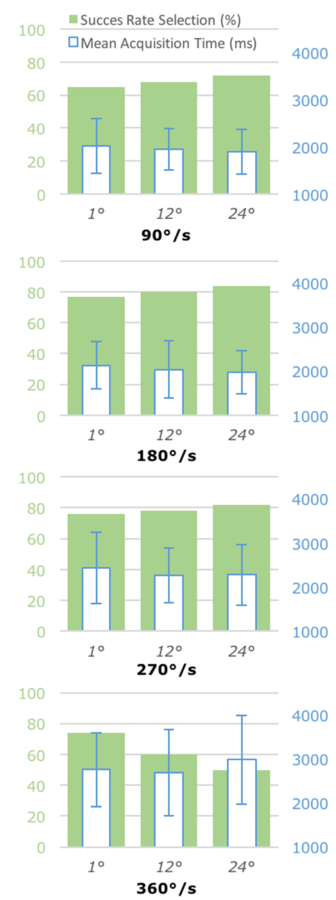

Figure 7: Comparing speed and visual angle indicates that bigger angles results in higher success rates and shorter acquisition times. The performance clearly decreases at the highest speed. The error bars depict the *SD*.

t(10) = -0.15, *p* = .88. However, the NASA TLX results suggest different perceived workloads for the two conditions. Seven of the eleven participants reported a higher workload when using the dominant hand, and several of these confirmed verbally after the study that the dominant hand was not their preferred choice for interaction with the trials presented.

Speed

A one-way repeated measures ANOVA revealed significant differences in the both the user success rate and acquisition times when acquiring targets that moved at different speeds: $F_{3,30} = 11.39$, $p < .01$; and, $F_{3,30} = 71.53$, $p < .01$, accordingly (Figure 5). Bonferroni-corrected post-hoc t-tests reveal that participants' success rate was significantly higher when faced with targets that moved at either 180°/s or 270°/s ($p = .01$). Participants acquisition times varied between these two speeds, being significantly lower for 180°/s ($p < .01$).

Size

A one-way repeated measures ANOVA revealed significant differences in the user acquisition times when acquiring targets displayed at different degrees of visual angle: $F_{2,20} = 5.82$, $p = .01$. This was not the case for the success rate: $F_{2,20} = 0.19$, $p = .83$ (Figure 5). Bonferroni-corrected post-hoc t-tests reveal that participants acquired targets moving in trajectories wider than 1° significantly faster ($p = .02$).

As both speed and visual angle influence the perceived change in position of the target, we highlighted the system's performance for each combination of speed and visual angle (Figure 7). With the exception of the fastest speed, the success rates and acquisition times seem to increase with larger visual angles. The opposite seems to happen at 360°/s. These results match our baseline (Clarke et al. [3]).

Number of Targets

A one-way repeated measures ANOVA revealed significant differences in both participants success rate and acquisition times when different number of targets were displayed: $F_{4,40} = 203.20$, $p < .01$; and, $F_{4,40} = 11.50$, $p < .01$, respectively (Figure 5). Bonferroni-corrected post-hoc t-tests reveal a success rate that significantly decreases when 8 or more targets are displayed in tandem ($p < .01$). No significant differences were found between 2 or 4 targets ($p = 1$). Furthermore, the only significant difference in acquisition times occurs when 16 targets are displayed in tandem ($p < .02$).

Discussion

The performance of WaveTrace seems to indicates that IMUs found in off-the-self devices can be used successfully for motion matching input. Furthermore, the results show similarity with other general motion matching input techniques, including similar success rates between WaveTrace, Orbits [4] and Pursuits [10] when up to 8 targets are displayed in tandem (Figure 6). The performance with these techniques decreases significantly when more targets are added on screen, even if these are invisible to the user [10].

More importantly, the results also show similarity with an optical baseline across study conditions. This includes how user performance decreased when targets moved at 360°/s [3]; or how the degrees of visual angle had little effect in user performance [2,3]. These are positive findings that highlight that an IMU-based approach not only works, but has a comparable performance with optical-based systems.

In addition to these findings, we also report on a novel upper threshold for target speed in manual motion matching: 270°/s. Participants' success rate with these targets was significantly better than with targets moving at the fastest speed of 360°/s, but not significantly worse than targets moving at 180°/s – the upper threshold previously described in [3].

Finally, no significant performance differences were found between participants interacting with their dominant and non-dominant hands. This is quite reassuring, as many people use watches and smart watches in their non-dominant hands. More so, more than half of participants reported a high subjective workload when using the dominant hand. We question if this is because participants were not used to having a wrist-worn device in their dominant hands. We plan on further exploring this topic in later work. Regardless, and due to the simplicity of the hardware requirements of our technique, it should be feasible to develop wristbands for input, which the user could wear on the preferred arm (in addition or as replacement of a smart watch).

Conclusion and Future work

In this paper we have explored the possibility of using wrist-worn sensors as input for motion matching selection. We have started to show the effectiveness of this technique using an off-the-shelf smart watch, and revealed optimal parameters to apply in implementations of this technique. The user study presented in this paper is ongoing and will produce additional data to be analyzed to provide more accurate conclusions.

Further work is planned to explore application areas for WaveTrace. By developing and (qualitatively) studying different prototype applications, we aim to provide insights into different application areas and the overall user experience with these. This will allow for the technique to further distance itself from the optical baselines described in the paper, which are particularly focused in supporting interaction with smart TVs and public displays. We aim to explore broader application areas such as: *smart phone* interaction using wrist movements; *shared displays* with layered information that can accessed through motion matching; and large *public displays* that support interaction from afar.

Finally, we also aim to develop a system architecture that allows for multiple input devices to seamlessly interact with different output devices. This would allow for an easier exploration in multiuser environments, such as public displays or smart homes.

Acknowledgements

This work was made possible by the Carnegie Trust Research Incentive Grant (70513). We would like to express our gratitude to Dzmitry Aliakseyeu for his support and to our participants, who made this research possible.

References

1. M. R. Burke and G. R. Barnes. 2006. Quantitative differences in smooth pursuit and saccadic eye movements. *Experimental Brain Research* 175, 4: 596–608. https://doi.org/10.1007/s00221-006-0576-6
2. Marcus Carter, Eduardo Velloso, John Downs, Abigail Sellen, Kenton O'Hara, and Frank Vetere. 2016. PathSync: Multi-User Gestural Interaction with Touchless Rhythmic Path Mimicry. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16), 3415–3427. https://doi.org/10.1145/2858036.2858284
3. Christopher Clarke, Alessio Bellino, Augusto Esteves>, Eduardo Velloso, and Hans Gellersen. 2016. TraceMatch: A Computer Vision Technique for User Input by Tracing of Animated Controls. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp '16), 298–303. https://doi.org/10.1145/2971648.2971714
4. Augusto Esteves, Eduardo Velloso, Andreas Bulling, and Hans Gellersen. 2015. Orbits: Gaze Interaction for Smart Watches Using Smooth Pursuit Eye Movements. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (UIST '15), 457–466. https://doi.org/10.1145/2807442.2807499
5. Jean-Daniel Fekete, Niklas Elmqvist, and Yves Guiard. 2009. Motion-pointing: Target Selection Using Elliptical Motions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '09), 289–298. https://doi.org/10.1145/1518701.1518748
6. Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9: 904–908. https://doi.org/10.1177/154193120605000909
7. Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction. 1151–1160. https://doi.org/10.1145/2638728.2641695
8. Franziska Roesner, Tadayoshi Kohno, and David Molnar. 2014. Security and Privacy for Augmented Reality Systems. *Commun. ACM* 57, 4: 88–96. https://doi.org/10.1145/2580723.2580730
9. Eduardo Velloso, Markus Wirth, Christian Weichel, Augusto Esteves, and Hans Gellersen. 2016. AmbiGaze: Direct Control of Ambient Devices by Gaze. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, 812–817. https://doi.org/10.1145/2901790.2901867
10. Mélodie Vidal, Andreas Bulling, and Hans Gellersen. 2013. Pursuits: Spontaneous Interaction with Displays Based on Smooth Pursuit Eye Movement and Moving Targets. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp '13), 439–448. https://doi.org/10.1145/2493432.2493477
11. John Williamson and Roderick Murray-Smith. 2004. Pointing Without a Pointer. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '04), 1407–1410. https://doi.org/10.1145/985921.986076