PathSync: Multi-User Gestural Interaction with Touchless Rhythmic Path Mimicry

Marcus Carter[1], **Eduardo Velloso**[1], **John Downs**[1], **Abigail Sellen**[2], **Kenton O'Hara**[2] **Frank Vetere**[1]
[1]Microsoft Research Centre for Social NUI
The University of Melbourne
[marcusc][evelloso][jpdowns][f.vetere]@unimelb.edu.au
[2]Microsoft Research
Cambridge, UK
[asellen][keohara]@microsoft.com

ABSTRACT
In this paper, we present *PathSync*, a novel, distal and multi-user mid-air gestural technique based on the principle of rhythmic path mimicry; by replicating the movement of a screen-represented pattern with their hand, users can intuitively interact with digital objects quickly, and with a high level of accuracy. We present three studies that each contribute (1) improvements to how correlation is calculated in path-mimicry techniques necessary for touchless interaction, (2) a validation of its efficiency in comparison to existing techniques, and (3) a demonstration of its intuitiveness and multi-user capacity 'in the wild'. Our studies consequently demonstrate *PathSync's* potential as an immediately legitimate alternative to existing techniques, with key advantages for public display and multi-user applications.

Author Keywords
Touchless Interaction, PathSync, Kinect.

ACM Classification Keywords
H.5.2 User Interfaces (e.g., HCI): Interaction Styles

INTRODUCTION
In this paper we present and validate a novel touchless interaction technique with many advantages over existing techniques. *PathSync* is a form of distal touchless interaction based on the principle of rhythmic path mimicry [8, 9, 14, 44-46, 50]. By replicating the movement of a screen-represented pattern with their hand, users can distally interact with digital objects quickly and at a high level of accuracy (see Fig 1). In comparison to existing techniques, *PathSync* has several distinct and important advantages: it enables distal interaction with large screens, is suited to multiple users, avoids the Midas Touch problem without requiring users to memorize gestures, requires a small range of motion to interact, is intuitive, and has high discoverability.

ACM ISBN 978-1-4503-3362-7/16/05…$15.00.
DOI: http://dx.doi.org/10.1145/2858036.2858284

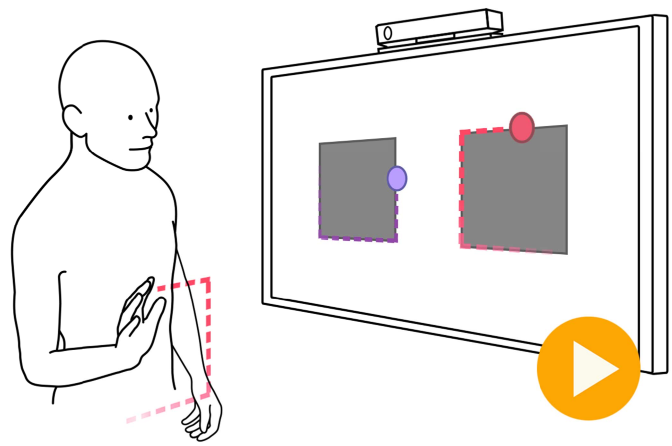

Figure 1. *PathSync* correlates the user's hand movement to a screen-represented pattern to identify interactions. This animated figure is best viewed in Adobe Reader

We begin with a brief overview of prior work on hand-based touchless interaction techniques and mimicry-based interaction. We then describe *PathSync* before overviewing the method and results of our three studies that evaluated human capacity to mimic moving targets, compared *PathSync* to the default Kinect interface technique, and deployed *PathSync* 'in the wild' on two public screens where it was successfully used by over 1000 users over a 28 day period. In the discussion, we overview the opportunities for this new genre of interaction technique - the synchronous-replication of a system represented pattern by the user – and its applicability to other modalities.

This paper therefore contributes the extension of rhythmic path mimicry to the modality of touchless hand gestures, demonstrating its advantages over existing techniques and opportunities for future use. Our three studies each contribute (1) necessary improvements to how correlation is calculated in path-mimicry techniques, (2) a validation of its efficiency in comparison to existing techniques, and (3) the verification of its intuitiveness and multi-user capacity 'in the wild'. Through reporting the development, verification and field-trial of this new interaction technique, we demonstrate its potential as an immediately legitimate alternative to existing techniques.

RELATED WORK
A wide range of research in HCI is concerned with 'touchless', 'mid-air' or 'remote' interaction; the ability to

interact with technology at a distance, without physically touching or holding hardware. The interest in touchless interactions reflects an expectation that – as technology becomes increasingly ubiquitous – computer use will become increasingly embedded within all areas of everyday life, requiring new forms of more 'natural' interactions [49].

While numerous recent research projects have explored other forms of touchless bodily interaction, in this paper, we focus on prior work involving hand-based touchless, gestural interaction, which has been implemented in a wide range of applications. The most commonly cited advantages for the touchless modality include:

1. improved hygiene, from a lack of touch
2. support for larger displays, as users can be further from the screen
3. multiple users, as everyone has a "controller"
4. alleviating the 'burden' of physical contact or requiring a device.

When supporting hand-based touchless interaction, there are two dominant styles of technique; *cursor-based* and those that employ a *gesture-library*.

Cursor Metaphor for Touchless Interactions

Cursor-based touchless hand interaction techniques are those that use the metaphor of a cursor to afford direct manipulation, such as by representing a user's hand in 3D space as a cursor on a 2D screen. This is the primary form of interaction on the Xbox 360 and Xbox One interfaces (although both interfaces use a combination of *cursor-based* and *gesture-library* interactions), and is therefore the most prevalent style of touchless hand-based interaction.

The 2010 Xbox 360 touchless interface utilized a *hover-to-select* technique, where the user would have to hold the cursor over an object for a defined amount of time to select it [26; evaluated in 38]. However, this technique has been noted for exacerbating 'Gorilla-Arm' fatigue issues [16,18] and slowing interactions [39] as the arm has to be held out in front of the user. In the more recent Xbox One console (released in 2013) the metro style user interface – which resembles the Windows 8/10 metro user interface of large, square tiles - is navigated by a *Press-to-Select* technique that leverages the familiar metaphor of a 'button press'. The user holds their hand in front of their body, maneuvers it so that the on-screen cursor (a large hand) is located above the selected object, and then 'pushes' their hand directly towards the screen to 'press' the tile. Yoo et al. [51] recently conducted a direct comparison of these two techniques, finding that 8 out of 10 participants preferred *Press-to-Select* despite it being less accurate than *hover-to-select* techniques, a finding they attribute to the perceived delay of the hover.

While it is common that these cursors 'snap-to' the center of buttons to attempt to make up for the lack of haptic feedback, this style of interaction has been shown to lack accuracy and mastery [38, 47], a key measure for a successful user interface [41]. 'Pressing' when fully extended is also difficult, causing errors when the 'press' motion moves the hand (and cursor) towards the center of the screen. Further, that the user has to use their full range of movement to interact with the screen is frequently attributed to causing meaningful fatigue after prolonged use [13, 16, 38], another significant limitation.

The Xbox interface utilizes a 1:1 mapping of the user's hand and cursor, which limits the size of the screen to the user's reach, but we note that alternative techniques for cursor-mapping overcome these limitations. Allowing the user to define the gesture space, creating a custom ratio, has been shown to be more comfortable without sacrificing accuracy on small displays [20]. Other methods like ray-casting and relative cursor pointing are often more suited to larger screens. Relative cursor pointing utilizes a clutch gesture (such as open hand/closed hand) to allow the user to choose when to move the cursor, permitting relative movements to control its location, a configuration well suited towards very large displays [22]. Ray-casting, or pointing, is similarly suitable for large displays, particularly when with semantic pointing where the size of the targets are adapted based on cursor distance [3].

An inherent limitation of these techniques is that they are proximal; despite being touchless, in order not to exacerbate these issues the user must stand centered on the screen, limiting the area available for interaction and the number of collocated users. Displaying a cursor also has the potential to be confusing with multiple users, where they also act to reveal the intent of a user to others, disallowing hidden or unattributed interactions. However, this can also be an advantage in collaborative contexts, where cursors can contextualize surrounding discussions [34], acting as an 'avatar' for the user.

While this method – drawing on a familiar metaphor for computer interaction - is the most common in commercial applications [see 4 for a recent review], the focus in research has overwhelmingly been on *gesture-library* techniques.

Gesture Libraries for Touchless Interactions

Gesture-library techniques involve the remote detection of static or moving hand gestures that correspond to a predefined set of movements, such as iconic gestures like pinching to delimit interactions [24], or symbolic gestures such as tracing letters of the alphabet in mid-air for text entry [29]. The effective use of gestures from gesture libraries requires that they be known and recalled to users before interactions, often resulting in high false-positive recall [43].

A prominent research context for touchless interaction is within the [critiqued, see 32] paradigm of 'natural' user interface (NUI) research. In the context of *gesture-library* work, this has often been interpreted as meaning that a natural hand-gesture interface is one where users can "interact with technology by employing the same gestures they employ to interact with objects in everyday life" [23, p. 36]. That is, the gestures used must either be implicit or have some symbolic relationship to the interaction task [11] (e.g. swipe to delete) or mimicry of conventions (e.g tracing a letter). While this interpretation of the NUI program has been more recently critiqued [see O'Hara et al. [33] who argue naturalness emerges through the interface's support of embodied practices], this perspective remains prevalent in *gesture-library* based research, where the design of the prescribed set of gesture is based upon conventions and emergent properties of the task.

This approach to creating gesture libraries has led to the considerable volume of work that principally assesses touchless interaction in terms of intuitiveness or ease of use [15, 42], or discoverability in contexts like walk-up displays [17, 36, 48]. Other perspectives create solutions based on user-defined [2, 35, 37] or context aware hand-gestures [6], or by improving the detection of hand gestures [12, 52]. These projects have demonstrated that it is possible to improve the user experience, accuracy and opportunities for *gesture-library* based research, making it a suitable modality for many applications. In addition, like *PathSync*, mid-air gestures are distal, supporting multiple users and interactions with very large displays.

However, *gesture-library* based techniques have two key limitations that render them unsuitable for numerous application contexts. Firstly, the user has to memorize multiple gestures, and recall what gesture does what in what context. This severely impacts the number of interactions possible at any one moment. Further, there is a limited number of different gestures that can be made by a single user and recognized as different gestures by a system, as they have to be sufficiently different in shape

In comparison to *PathSync*, *gesture-library* based techniques also require high activation windows to circumvent the 'Midas touch' problem [40]. That is, in order to disambiguate between a non-interactive hand movement and those that are intentional interactions, a gesture has to be sufficiently 'long', or be delimited somehow, to reduce the likelihood of someone unintentionally causing the interaction. Consequently, systems that rely on *gesture-libraries* have an extremely low level of discoverability, rendering them inappropriate for walk-up public screens, and are therefore unlikely to prompt the initial propagation of touchless, gestural interaction.

Rhythmic Mimicry as an Interaction Technique

The principle of rhythmic mimicry has been employed in a small number of discrete research projects, though never as a form of touchless gestural input. This style of interaction exploits the natural human ability to mimic external rhythms, be they spatial or audio based [10].

The earliest example of rhythmic mimicry that we are aware of is Williamson et al. [50], who developed a system with randomly moving circles to 'point without a pointer'. Users could select one of these circles by mimicking the movement of the circle on a laptop trackpad as a form of continuous interaction. The relative variance between the trackpad-pattern and the system represented pattern was calculated using a basic correlation algorithm, finding users could select these objects within 4-10 seconds. More thoroughly, Fekete et al. [10] extended this principle (which they refer to as "motion pointing") using recurring elliptical motions and an optical mouse input. As they found that the correlation between the elliptical motions and user input was not sufficiently accurate to reliably distinguish between multiple different shaped and phased ellipses, Fekete et al. contribute a 'move and stroke' technique, where the four closest matches are initially selected, requiring a second gesture (a stroke in one of four directions) is used to make the selection out of these four. The improved correlation method we contribute in Study 1 makes such extensions of the interaction unnecessary with gestural input.

More recently, Vidal et al. [44-46] developed *Pursuits*, an eye-gaze calibration and interaction technique that leverages *smooth pursuit* eye movements. When our eyes follow a moving target, they perform a smooth movement, impossible to otherwise replicate. The attributes of this smooth movement can be correlated with the location of the on-screen moving target, and used to calibrate gaze detection. Esteves et al. [8, 9] implemented this principle as a gaze-based interaction technique, where users could interact with *Orbits* – a target following a regular, circular path - by following its movement (such as a 'read notification' icon on a smart watch) for 1 second. Esteves et al. demonstrated that it is possible to distinguish between multiple simultaneous *Orbits* by alternating the speed, shape, direction and phase of the targets, as a particular smooth-eye movement can only be possible if the user is looking at an icon moving in the same direction and speed at that moment.

Each of these systems represent a form of rhythmic *path* mimicry, where both spatial and temporal properties are used to activate an interaction. In these examples, all possible interactions are simultaneously visualized along a depicted path. Using an alternative method, Ghomi et al. [14] evaluated the use of audio & visual rythmic 'beat' mimicry, and attempt to extend it to a system that requires users recall different rhythmic beats to interact (as multiple audio beats cannot be simultaneously displayed).

PATHSYNC

PathSync represents the extension of the principle of rhythmic path mimicry to touchless, hand-based interaction; the synchronous mimicry of a screen-based moving target with the user's hand. Similar to Orbits [8, 9], this technique works by depicting a shape (e.g. a square) with a target that follows its perimeter at a constant speed. The user interacts by moving their hand in-sync with the moving target (see Fig 1) for a specified length of time. By varying the direction, speed, shape and phase of the moving target, we can have a large number of interactions simultaneously displayed.

While prior studies have demonstrated the principle of rhythmic interaction using trackpad or optical mouse input, we believe that its key opportunities are as a form of *touchless* interaction, where our natural capacity to mimic external rhythms may make up for the lack of tangible feedback and require only relative accuracy (rather than absolute accuracy, which limits cursor-based techniques). In our subsequent studies, we contribute (1) the necessary improvements to how correlation is calculated in path-mimicry techniques, (2) a validation of its efficiency, and (3) its intuitiveness and multi-user capacity 'in the wild'.

RESEARCH DESIGN & RESULTS

Our research set out to identify the parameters of this novel interaction technique, understand its efficiency and discoverability, and interrogate and demonstrate its value in comparison to other touchless gestural interaction techniques currently available. We designed three sequential user studies that accomplish these aims and build the necessary foundations for the further use of *PathSync.*

STUDY 1 – IDENTIFYNG PARAMETERS

A major advantage of eye-based path-mimicry interaction techniques such as *Pursuits* [44-46] and *Orbits* [8, 9] is that once the eyes lock onto a moving object, their relative motion will closely match that of the object due to the physiological nature of eye-movements. However, users have substantially more voluntary control of their body, making it unclear whether the rhythm and trajectory of such gestures would be similar enough to the stimuli's to rapidly trigger interface commands with high precision.

Consequently, our first study was principally concerned with three questions that related to enabling hand-based rhythmic path mimicry gestures. First, can users match on-screen patterns with their hands close enough to trigger different controls? Second, if so; up to how many offset targets on the same trajectory can the technique distinguish? Third, do users' capabilities depend on the shape, size or speed of the pattern?

To answer these questions, we conducted a user study with 20 right-handed participants (14M/6F), aged between 20 and 63 years (mean=32). Only 2 participants had used the Kinect gesture system to any real extent (one more than 20 times and the other more than 5 times). Eight participants had used it fewer than 5 times, and 8 had never used it before. Upon arrival, participants signed a consent form and completed a demographics questionnaire. The study took place in a quiet lab setting, with participants standing in front of a 42" screen with a Kinect v2 mounted above it

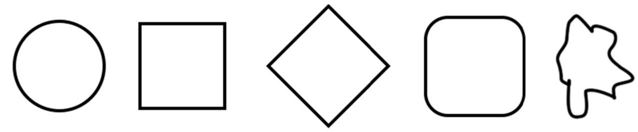

Figure 2. The 5 shapes trailed in Study 1.

which recorded participants' 3D joint positions at 30Hz, synchronized with the position of on-screen targets. Using *Open Broadcaster Software* (OBS), we captured the Kinect's video stream and skeletal view, as well as the TV screen in a single video file. Participants were shown a short abstract animation that illustrated the technique, asking them to match the movement of the target displayed on the screen with their hand. Before the trials, we ensured that they correctly understood the technique.

In each trial, participants were presented with a small circle that smoothly moved along different clockwise trajectories. In each trial, we varied the trajectory SHAPE (Circle, Square, Rounded Square, Diamond, Squiggle) (Fig. 2), the SIZE, i.e. the perimeter of the trajectory (200, 300, and 400px), and the SPEED of the moving target (4, 5, and 6s/cycle), for a total of 5x3x3=45 trials per participant. The order of the trials and the position on the screen where the shape was displayed were randomized across participants. After completing all trials, participants were asked to rank their preference regarding the comfort of mimicking the shapes, speeds and sizes. In this step, a sample of each level was displayed on the screen for their reference.

Study 1 Results

In our analysis, we used the positions of the right wrist (rather than hand), as the percentage of missing frames in the recording (4.2%) was smaller than for the hands (6.2%) and there were no substantial differences in the trajectories of these joints. Based on this dataset we made three significant improvements on the original algorithm used by Vidal et al. and Esteves et al. [8, 9, 44-46]. In their algorithm, in every window the calculated for each axis the individual correlation between the raw coordinates of the gaze point and the target. However, this approach fails when one of the coordinates remains constant (e.g. at the edge of a square) ,because in this case, the standard deviation of the coordinate values in the denominator is zero.

To address this problem, before computing the correlation we must rotate the data in order to maximize the variation along both axes. We achieve this by conducting a Principal Component Analysis (PCA) on the target trajectory data, which outputs a rotation matrix that maximizes the variation along one of the axes. For example, whereas the edge of a square would be unchanged after this rotation, the edge of a diamond would be rotated so that it is horizontal (or vertical). To distribute this variation across both axes, we rotate the data 45 degrees further (see Figure 3) and use the same rotation matrices computed with the target data to rotate the hand data. We then compute the

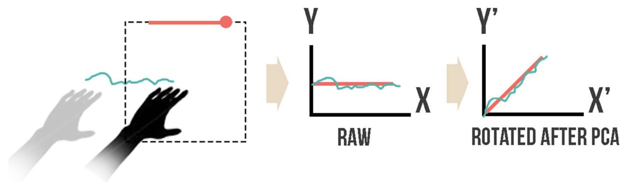

Figure 3. Before calculating the correlation, we rotate the data to distribute the variance across both axes.

Pearson's product-moment correlation coefficients within the window for each axis the following equation, and discard the largest, as proposed by Vidal et al.:

$$r = \frac{E[(Hand - \overline{Hand}) \times (Target - \overline{Target})]}{\sigma(Hand) \times \sigma(Target)}$$

The second modification we made is the use of a bi-level threshold [30, 31]. Instead of activating a control immediately when the correlation crosses a given threshold (e.g. .8 [8]), we first require a higher threshold for activating the window (.9), but allow a larger tolerance range by only deactivating it when it crosses a lower threshold (.6). The third improvement is in only activating the control after one full second of activated windows, which helps to remove spurious false activations. We consider a trial successful if there was at least one activation that satisfied all of these conditions. Figure 4 illustrates these steps. Similarly to Esteves et al., we used 1s (30 frames) windows for computing correlations.

Our first question regarded whether users' hand movements would match the movement of the stimuli closely enough to successfully select an object. Therefore, we computed the mean number of successful trials across all users for each of our study's conditions. Due to the complexity of the *Squiggle*'s shape, users failed to match its movement in most trials, with only 24.4% of successful trials. We therefore excluded this shape from our subsequent analyses. The other shapes, however, yielded very high success rates, with a mean of 93.8% successful trials.

These results, however, only reflect the performance of the system with a single target on the screen. To evaluate the robustness of our approach in distinguishing different trajectory shapes and phases, we simulated other moving targets and compared the activations against the user data. To understand how closely offset the targets can be, we simulated objects on the same trajectory, but an angular offset ranging from -180 to 180 degrees in 5-degree increments. The smaller the offset angle, the higher is the probability of a false positive activation, so we computed the minimum offset angle on either direction (discarding the smallest) that yielded at least one false positive activation in each trial. We then tested the effect of the SHAPE, the SIZE, and the SPEED on these angles with a three-way repeated-measures ANOVA (Greenhouse-Geisser-corrected where Mauchly's test revealed a violation

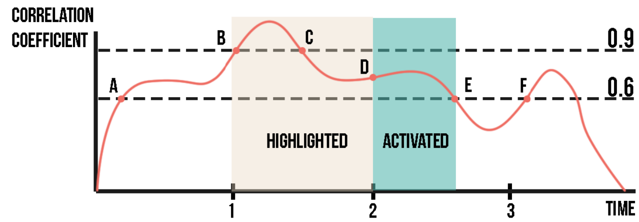

Figure 4. Hypothetical correlation curve: We highlight the object when it crosses the upper threshold (B), but not the lower (A). If the correlation remains above the lower threshold for one second, we activate the object (D), even if it goes below the upper threshold (C). We deactivate the object when it crosses the lower threshold (E).

of sphericity) and post-hoc pairwise t-tests (Bonferroni-corrected). We found a medium significant effect of the SHAPE of the trajectory ($F_{1.6,30.7} = 1.75, p < .05, ges = .06$), but not of the SPEED, SIZE or the interactions between variables. The *Circle* and the *Rounded Square* required a minimum offset of 37°, and the *Square* and the *Diamond* required a minimum of 28°. These results indicate that with a trajectory in the shape of a *Circle* or a *Rounded Square*, our algorithm is able to distinguish up to 9 simultaneous targets, and one with a *Square* or a *Diamond*, up to 12 clockwise targets. If we also consider targets moving counter-clockwise, the technique can support up to 18 and 24 targets, respectively.

An interesting behavior we observed in these data recording sessions was that, even though only the relative movement of the users' hands mattered, users still tended to perform the gesture roughly in the same direction where the shape was displayed on the screen. To test this hypothesis, we computed the Pearson's correlation coefficient between the mean positions of the shape on the screen to the mean 2D position of the users' wrists on the plane parallel to the screen. We found a high correlation (.58) between horizontal positions of the stimulus and the gesture, and a moderate correlation between their vertical positions (.36). These results show that even if position in which a gesture is performed does not depend on the position of the stimulus, users have a natural tendency of reaching out towards them. Further improvements could take this correlation in to account.

Study 1 Summary

In this first study, we interrogated the fundamental human capacity to mimic the movement of an on-screen target with their hand. We found that users are highly capable of matching on-screen movement, allowing for 18+ simultaneous targets of the same shape and speed without false-activation, sufficient for UI design [27]. We contributed three improvements to existing path mimicry algorithms. This therefore demonstrates the feasibility of *PathSync* as an interaction technique, but in a somewhat artificial task. Consequently, our subsequent study set out to evaluate *PathSync* in a 'real' task in comparison to an existing and widely adopted technique.

STUDY 2 – EFFICIENCY OF PATHSYNC

Our second study set out to compare users' performance with *PathSync* when selecting on-screen targets against *Press-to-Select* (*PtS*), the most common gestural interaction technique currently available. *PtS* works by representing (1:1) the user's hand on the screen with a cursor, initiating selection by extending their hand towards the screen to 'press' the button the cursor is over. While *PtS* also supports more complex transactions, we felt a direct comparison was suitable due to *PtS* commercial ubiquity.

For this study, we recruited 40 participants (21F/19M), aged between 19 and 35 years (mean=24) using posters on campus and internal mailing lists, none of whom had participated in the previous study. Participants were rewarded with a $5 coffee voucher. Five participants were left-handed and used the interface with this hand. Participants had little experience with Kinect interfaces: 24 had never used it before, 10 fewer than 5 times, 3 fewer than 20 times, and 3 more than 20 times. The recording setup was the same as in the previous study. Upon arrival, participants signed an informed consent form and completed a demographics questionnaire. Participants then completed a tutorial session for each technique (labelled A and B for the participants). For the *PtS* technique, we used the interactive tutorial that ships with the Kinect SDK. For *PathSync*, we designed a similar version that followed the same visual style (see Figures 4-6).

After the tutorials, participants completed a series of trials in which they were asked to select a particular object (e.g. a Carrot) among 5-9 other objects of the same category (e.g. Broccoli, Corn, etc., see Figure 6). We designed our interface in the Metro style, to match the ones found in the *Xbox One* and *Windows 8/10* interfaces. In the *PathSync* condition, a small protrusion in the same color of the button's background, moved around the button. The button was highlighted when correlation went above the upper threshold (.9), and if a high correlation (>.6) was maintained for one second, the button was selected. These thresholds were determined via simulation using study 1 data.

A trial was completed after a correct or incorrect button was selected, or after a 30s time-out. Participants completed a total of 48 trials, in 8 blocks of 6 trials, with alternating techniques for each trial, in a counter-balanced order. Random-order-practice was used because it has been shown to benefit motor learning more than block-practice [16, 19, 21]. In-between each trial, participants were asked to lower their hands as a resetting step. After completing all trials, participants were asked to fill in a questionnaire regarding their impressions of the techniques: their ease-of-use, frustration, and overall preference.

Study 2 Results

Overall the results indicated that *PathSync* is a comparable touchless-interaction technique to *Press-to-Select*, not considering the inherent discrete, distal and multi-user advantages of *PathSync* we document and explore in Study 3. We found no significant differences between the intuitiveness, efficiency and learnability of the techniques, and participants' subjective opinions were evenly distributed; 19 participants found *PathSync* easier, 20 found it faster, and 21 found it more frustrating.

Intuitiveness – Initial Proficiency

At the conclusion of both the tutorials, participants were asked if they understood the techniques. No participant said they did not understand either technique, and only one participant was uncertain about *PathSync*. That is, 39/40 were confident that they understood *PathSync* after the tutorial. However, every participant (N=21, only 2 of whom had used the Kinect more than 5 times previously) who received *Press-to-Select* first was uncertain about whether they understood that method, a confusion not present in those that received *PathSync* first. The confusion around *Press-to-Select* appeared to be that users were unsure if they had to pull their hand back after pushing a button. However, the relationship between the order of the techniques and this confusion is not clear.

In the last step of each tutorial, participants were asked to select a sequence of four targets in a specific order. We used the time to complete this task as a metric for the initial proficiency with the techniques, as users had the minimum experience necessary to complete it. We identified a significant ($p<.001$) but small (in absolute terms) difference between the two techniques, where the mean *PathSync* completion time was 33.1 seconds for these 4 targets, while *PtS* took 33.6 seconds for the same 4 targets. We examined the total tutorial completion time, and completion time for these specific steps based on age, gender, prior experience with the Kinect, and order of the tutorials, and found no significant differences. These results indicate that the two techniques have comparable intuitiveness.

Efficiency – Overall Proficiency

We tested the effects of the NUMBER OF OBJECTS and the INTERACTION TECHNIQUE on the SELECTION TIME for each technique with a two-way repeated-measures ANOVA, excluding the 7 trials (2 *PtS*, 5 *PS*) that timed out at our artificial limit of 30 seconds. Out of the 1920 trials, there were 7 timeouts and 71 incorrect selections. There were no significant ($p > .05$) differences between the error rates on each technique (1.9% for *PathSync*, vs 2.1% for *PtS*). We note that while 17 participants made at least one error with *PtS*, only 7 users made any errors with *PathSync* (with 18 of the 29 incorrect *PathSync* selections made by just two participants).

We found a slightly lower mean completion time in the *PathSync* condition (5.86s per target) than in the *PtS* condition (6.13s), but this difference was not statistically significant at the p=.05 level. We did not find any significant effect of the NUMBER OF OBJECTS or interaction effects.

We also note that prior experience with the Kinect did not affect a participant's average time-to-completion. There was no significant difference based on prior experience, order of the tutorials, or on the user's *perception* of which technique was faster. In summary, our results suggest that both techniques yield similar performance in selection tasks.

User Preference - Qualitative Results

At the conclusion of the study, participants were asked which technique was easier, faster, or more frustrating. 22 participants strongly preferred *PathSync*, rating it easier, faster and less frustrating, while only 6 participants rated *Press-to-Select* as positively. The remaining 12 participants were more ambiguous, liking or finding issues with both.

Our participants whose responses clearly indicated that they preferred *PathSync* (n=22) explained their preference by referring to *PathSync* as more responsive [P38], comfortable [P40], effective [P7], developed [P31] and more natural [P39], while comments about *Press-to-Select* considered it too sensitive [P32, P22, P18], hard to control [P25, P39, P19], or requiring too much concentration [P12]. Four participants explicitly commented that they felt *PtS* was more fatiguing than *PathSync*; *"PtS strained my muscles a bit*" [P28], "*PtS won't be as comfortable to the users as they will become tired*" [P31], and P29 and P33 felt strongly that *PathSync* was a better technique because it didn't matter where they mimicked the target; they could do the interaction in the most comfortable position for their hand.

Those participants that indicated that they preferred *Press-to-Select* (n=6) on each rating were less negative about *PathSync.* P15 felt "*it took ages to make PathSync work*", while P26 felt it was "*a little bit confusing*", but P3 noted she preferred *PathSync* initially, and P5 prefaced their ratings by commenting that they chose *Press-to-Select* "*because it's easier and you know exactly where you have to move your hand. PathSync takes time to realise how it works*".

Indeed, several participants [P10, P31, P31, P39, P40, P12] who rated *PathSync* positively speculated that - while they preferred *PathSync,* others might find it less intuitive because *"it is straight forward*" [P39], but, as P10 put it, "*its different, its weird, but once you get used to it, you know this is what you want*" [P10]. The remaining participants (n=12), whose preference was more ambiguous, generally felt that "*both are pretty good, I'm really happy to use either"* [P29]. This group highlighted issues such as the inconsistency of how long it took *PathSync* to register; "*sometimes you move just half a circle, later a quarter of a circle, then later you do 4 circles and it still doesn't recognize?"* [P21], or the frustrations of *Press-to-Select; "it is a little difficult because it is more sensitive, I don't like the push"* [P27]. P17 simply concluded that which technique they preferred would depend on what task they were doing.

We had hypothesised that the moving targets necessary for path-mimicry would meaningfully distract the users, and consequently randomised the number of tiles in each trial, finding it had no effect on either technique. 36 participants were explicitly asked following the trials if they felt the dots in the *PathSync* type were confusing or distracting, and 26 felt that it was not distracting at all. P13 even felt that the targets - which we made the same color as the tile to minimize how distracting they were - should more obvious to make it easier to find them. Of those remaining, some only found it "*a little distracting*" [P25, P32, P37, P40], while only 4 felt strongly about it being negatively distracting [P26, P35, P21, P30]. Two acknowledged it was a little distracting [P28 & P29], but felt that the time it took to locate the cursor once they had located the target was equivalent to the distractions of the numerous moving targets.

Finally, several users made comments against the 'push' gesture in *Press-to-Select,* suggesting that it slows the interaction down too much [P9, P8], and other clutch configurations such as "thumbs up" [P18], pointing [P24, P39] or grabbing [P6] could be better for target selection. We chose *Press-to-Select* for our comparison in this study because it is the most widely available, being at the core of the Xbox One user interface (with over 8 million consoles). Future research should explore user preference between different configurations of *cursor-based* interaction, as there are few comparative studies between touchless gestures.

Study 2 - Discussion

The purpose of this study was to compare *PathSync* to *Press-to-Select* in the specific context where *PtS* can perform optimally; with a single user, standing centered, ~3m from a large television. Our results indicate that *PathSync* is a comparable technique to *Press-to-Select* encouraging further research and implementation.

In particular, this study presented an extremely promising validation of the comparable intuitiveness of *PathSync*, both in terms of initial proficiency (based on tutorial completion speed) and how easy it was to understand the technique (based on participant's confidence in their understanding of the technique following the tutorials). These results are a positive but surprising result considering that *PathSync* does not rely on a commonplace interaction metaphor.

We also note a significant difference between the nature of the errors with these two techniques. With *PtS*, 17 participants made at least one error, while no participant made more than 4 errors. In contrast, only 7 participants made an error with *PathSync*, but 2 of these made up for almost two thirds of the total errors with that technique. As

Figure 8. ***Attract Mode*****, featuring an active silhouette and the only instructions given on how to interact with *PathSync*.**

our participants did not receive feedback if they made an error, we believe that these participants did not realize they were not interacting correctly; a situation that may not replicate in a real application. As our study took under 20 minutes, we did not identify any impacts of fatigue.

STUDY 3 – EVALUATING PATHSYNC IN THE WILD

As indicated in our review of prior work, gestural interfaces are commonly being employed for interaction with large public displays [1]. However, existing techniques are problematic where discoverability (how easy it is to understand how to interact) and multiple users are a key requirement. *PathSync,* we suggest, is therefore a highly suitable touchless interaction technique for this context. Consequently, our third study set out to implement *PathSync* in a social, public display application, and to demonstrate the opportunities for *PathSync* where current touchless interaction techniques fall short.

We developed *Social NUIz*, a multi-user quiz game, which was deployed on two indoor, public displays at our University campus for 4 weeks. The first location (*hallway*) was in a busy corridor between buildings, with thousands of passersby each day. Potential users were thus always going somewhere, such as to class, consequently meaning the system has a very low threshold for interaction (like most wall-mounted displays). If it didn't work, or wasn't intuitive, users would likely just continue on their journey. The second (the *library*) was next to a dozen large tables outside a campus library, a space typically used by individuals and small groups to study or eat lunch. The threshold for interaction is much lower in this location, but it is much harder to attract potential users.

We chose to develop a game-based application as previous work has shown that it is a popular application among public displays users and a suitable method for evaluating an interaction technique [5], and a quiz-game was chosen as we anticipated all potential users understanding what was expected of them (answering questions). A quiz also demonstrates the capacity for *PathSync* to work with multiple simultaneous users while hiding their answers from one another, as it is hard to identify an opponent's input.

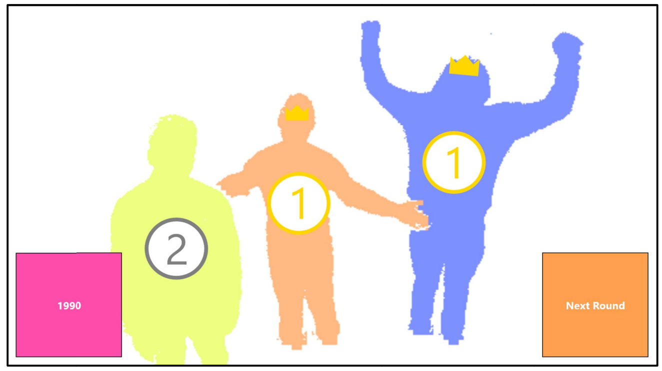

Figure 9. ***PathSync*** **attributes inputs to specific users, allowing us to track an individual's score.**

Social NUIz features an *Attract Mode*, a *Game Mode*, and a *High-Score* functionality. The *Attract Mode* presents two possible answers to a question, with an explanation of how to input an answer, see Figure 8. As is demonstrated in prior work, for these public displays to be highly effective in attracting users [48], we displayed a live silhouette of people who walked by in the background of the image. When an answer is registered, users were shown their live silhouette (a crown superimposed if correct), and their current score over their torso (Fig 9). During *Game Mode*, users were given trivia questions with 4 possible answers, and were given 25 seconds to input an answer. If all users detectable by the Kinect had provided an answer, we skipped the remaining time.

Up to 6 people can play *Social NUIz* at once—the number of bodies that the Kinect can simultaneously track. Each answer selection was attributed to the body of the corresponding player, allowing them to increase their score as long as they stayed within the field of view of the Kinect. If a player achieved a high-score, the quiz asked to take a photo of them (which they could easily avoid) that was then displayed in attract mode, challenging new players to beat their high score.

We conducted 6 hours of passive observation of use on each screen with the purpose of identifying how users initially approached the screen and how frequently someone attempted to interact with the game but failed to initiate the first start-orbit, and what misconceptions caused these failures. We also wanted to identify if people could learn how to interact with the system simply from observing another person interacting, without having seen the tutorial screen. These observations occurred at different hours over the week during semester.

Study 3 - Results

Over a 28 day period, 1065 people successfully input an answer to the *Social NUIz* game, where the longest streak was 50 questions (equivalent to 100 interactions, as users have to select 'next round'). We had 851 single users 176 pairs and 38 groups of three or more.

We passively observed 26 interactions with the screen over twelve 30 minute periods. 13 of these interactions resulted in successfully answering the attract mode question, while 13 were not successful. Reflecting the transient nature of the space, in 5 of these 'failed' interactions we did not observe any discernable attempt to select an answer; users were either pulled away from the system by their companions (n=2), only interacted to play with their silhouette (n=2) or read the question while on the phone (n=1). Of the remaining 8, the principal issues were assuming that the screen was a touch screen (n=4) or standing too close to the screen to be detected by the Kinect sensor (n=3). In one case, a user was correctly mimicking the path of the orbit but only a few inches from the screen, while another's touch interaction tried following the moving target. We note that 4 of the observed users that successfully interacted initially tried to touch the screen, before reading the instructions and matching the path touchlessly. As a result of the low-threshold for interaction that interactive public-screens have, none of these 8 users attempted to interact for longer than 10 seconds, almost immediately giving up and continuing to their destination.

As expected, usage data varied considerably across the two spaces. The *hallway* location had considerably more sessions (962 vs. 103), but the average length was shorter (47 seconds (SD=86) vs 170 seconds (SD=354)) and with fewer users at once in comparison to the *library* location (19.5% vs 26.2% having 2 or more users). 61.2% of the interactions in the *hallway* were only one round long; users answered the attract mode question, and then continued their journey. Observations indicated that the short period of interaction is primarily due to the transient nature of the space; users were intrigued by the attract mode question, and after finding out the answer (and whether they were right), they moved on rather than selecting 'next round'.

A key advantage of *PathSync* is its capacity to support multiple users. Sessions with more than 2 users were 141 seconds long (SD=267), or 3.9 questions (SD=6.1) on average, while sessions with 1 user were 38 seconds long (SD=72), or 2 questions on average. That is, while 70.1% of sessions with 2 users or more lasted more than 2 rounds, only 32.4% of single users interacted for more than the attract mode question. Our passive observations corroborated this advantage, finding that the interaction modality and design of *SocialNUIz* encouraged bystanders and spectators to become players as the game waited for their response if they were visible, and the hidden nature of an interaction supported the 'reveal' moment of the quiz game. We only observed one instance of gestures interfering with other users, which was due to the users believing their gestures had to physically line up with the on-screen targets, requiring both participants to attempt to interact in the same area.

Study 3 – Discussion

Based on our observations and the large volume of use (n=1065 sessions), we argue that *PathSync* is a sufficiently intuitive and robust interaction technique for public displays. The primary issues that we observed were the assumption that the screen was a touch screen, and standing too close for the Kinect sensor, easily solved by additional signage. There were over 426 sessions with 3+ successful *PathSync* interactions, emphasizing *PathSync's* potential as an immediately legitimate alternative to existing techniques. We also contend that the large number of sessions with a single user interacting for a single round (571 out of 851) does not challenge intuitiveness of the display, as users of public displays like these have a very low threshold for interaction [25, 48]. These users validate the intuitiveness of the technique, as they still input an answer despite having no interest in a prolonged interaction. Further, the support for social use was particularly evident, as multiple users played for longer and answered more questions in the game.

DISCUSSION

In this paper we have presented *PathSync*, a novel form of touchless hand-based gestural interaction based on the principal of rhythmic path mimicry. We demonstrated and validated the advantages and efficiency of this new technique through three studies; in Study 1, we contributed three improvements to the correlation algorithm used to better respond to the nature of gestural path mimicry; in Study 2, we demonstrated that *PathSync* is a comparable technique to the widespread *Press-to-Select* technique on each relevant measure; and in Study 3 we verified the multi-user capability, as well as the discoverability and learnability of this highly novel technique out of the lab.

We will now reiterate the advantages of *PathSync* over existing gestural interaction techniques before considering the alternative configurations of *PathSync* that extend opportunities for its use and overview how this new genre of interaction technique has further applicability.

Opportunities for PathSync

As we noted in our review of prior work, touchless gestures are cited as having four key advantages; (1) improved hygiene, (2) supporting larger displays, (3) multiple users and social use and (4) alleviating users the 'burden' of physical contact with remote detection. Like existing techniques, *PathSync* does not require physical contact, alleviating the 'burden' of physical contact and having hygiene advantages.

Firstly, our studies found that it is surprisingly natural and intuitive to replicate the movement of a target with a known path, such as in the case of a dot moving around the perimeter of a square. In study 1, we demonstrate that users are able to do this with a high level of accuracy. This means that the correlation thresholds can be extremely high (0.9), avoiding the *midas touch* issues that limit *gesture-library* techniques, and allowing short (>1 second) activation window. The comparable error rates and time-to-complete of *PathSync* in comparison to the Xbox One's *Press-to-Select* further demonstrate the validity of this as an alternative touchless interaction technique. As in our configurations of *PathSync* each target is associated with an icon that the user can select, all interactions are immediately contextualized requiring no memorization by the user, and permitting a large number of possible interactions at once.

Coupled with these results, a key advantage of *PathSync* is that it is a *distal* interaction technique; the location of the user's hand movement is disassociated from the location of the system represented pattern. As long as the user can see the pattern, and their movements are within the field of view of the Kinect sensor, they are able to interact. This means that of *PathSync* is well suited to both small screens (where tile and cursor size preclude *Press-to-Select* style *cursor-based* techniques) and very, very large screens (where 1:1 mapping of hand movements limits screen size), both foci of prior gestural interaction research.

This meaningfully opens up the opportunities for multi-user touchless interaction as users can more comfortably arrange themselves around larger displays and simultaneously interact without physical or virtual interruption. This is advantageous on very large public displays, but is also suited to the typical configuration of a TV lounge, where some couches are often perpendicular to the television. In the context of prior work that has commented on the capacity of gestural interfaces to blur the lines between spectator and player [7], opportunities for *PathSync* may include multi-user games that further explore this capacity. While cursors – as representations of users - can be useful for collaborative use cases [34], as demonstrated in *SocialNUIz* it is difficult to identify what another user's hand movements correspond with, allowing secretive interaction such as voting.

An additional opportunity presented by *PathSync* is that – by not requiring the representation of a cursor, or active feedback – it could be configured without a screen; for example, a target that followed the perimeter of buttons in an elevator, allowing users to select their destination without communicating touch-based diseases – a highly desirable advantage in contexts such as Hospitals or Nursing Homes. We note, however, that this lack of a familiar metaphor or active feedback is one of the primary limitations of *PathSync*, particularly in terms of discoverability that is key to such an application area.

Limitations and Future Work

While this paper has demonstrated and validated the advantages and efficiency of *PathSync*, several questions remain due to the limitations of the studies in this paper.

While Study 2 showed that our method of always displaying the paths is non-invasive and is able to not reduce available space on the screen, it would not be applicable for all UI designs. Other opportunities may emerge through exploring other methods for representing patterns that user's replicate (such as pulsating objects, flat paths rather than shapes, symbolic shapes, ecetera), and the capacity to overlay paths to increase density of selected objects. Alternate means of representing patterns on the screen could also improve the intuitiveness of the technique on public displays. Similarly, we noted in study 2 that while *PathSync* is comparable to *Press-to-Select* in a basic target selection task, *PtS* supports more complex transactions (such as grab and drag) which is necessary for a holistic touchless user interface. Future work should explore how *PathSync* might by integrated with these other techniques.

We also speculate that *PathSync* has fatigue advantages over other techniques, a claim which future work should interrogate. In the context of recent research [16, 20] that has investigated *rested* touchless interaction as a solution to the 'gorilla arm problem', we note that *PathSync* works comfortably when the elbow is rested, for example on a desk or couch arm. We believe that the distal nature of *PathSync* also allows users to do the necessary movements where it is most comfortable, and that the smaller ranges of motion required for *PathSync* similarly may reduce fatigue. Further, following Montero et al.'s [28] study of the social acceptance of gestural interfaces, this smaller range of motion required for *PathSync* may result in improved social acceptance. Finally, less-fatiguing finger movements may be suitable for *PathSync* interaction.

The concept of rhythmic path mimicry also has the potential in other modalities that have not yet been explored, though further extensions to how user-enacted patterns are correlated to system patterns would be necessary; a target moving along a sinusoidal wave cold be mimicked with a user's whistle, changing pitch relative to the target's path; a user could snap their fingers in beat with a pulsating light; choosing either interaction technique depending on context. Further research should explore the usability and opportunities that these new interaction techniques present.

CONCLUSION

In this paper, we have contributed new methods for correlating rhythmic mimicry for touchless hand-based interaction. We have shown that it is a comparable technique to the well-known *Press-to-Select* method found on the Xbox One interface, and we have demonstrated that it is sufficiently robust, intuitive and responsive for over 1000 users who used *PathSync* as the interaction technique for a quiz game 'in the wild'.

While the lack of active feedback and unfamiliarity of the interaction metaphor in *PathSync* mean other techniques may be more suitable for some applications, we have demonstrated that *PathSync* is an immediately legitimate alternative to existing techniques, with key advantages for public display and multi-user applications.

REFERENCES

1. Carmelo Ardito, Paolo Buono, Maria Francesca Costabile, and Giuseppe Desolda. 2015. Interaction with Large Displays: A Survey. *ACM Comput. Surv.* 47, 3, Article 46 (February 2015), 38 pages. DOI=10.1145/2682623
2. Ali Bigdelou, Loren Schwarz, and Nassir Navab. 2012. An adaptive solution for intra-operative gesture-based human-machine interaction. In *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces* (IUI '12), USA, 75-84. DOI=10.1145/2166966.2166981
3. Renaud Blanch, Yves Guiard, and Michel Beaudouin-Lafon. 2004. Semantic pointing: improving target acquisition with control-display ratio adaptation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '04), 519-526. DOI=http://dx.doi.org/10.1145/985692.985758
4. Arthur Theil Cabreira and Faustina Hwang. 2015. An analysis of mid-air gestures used across three platforms. In *Proceedings of the 2015 British HCI Conference* (British HCI '15), USA, 257-258. DOI=10.1145/2783446.2783599
5. Marcus Carter, John Downs, Bjorn Nansen, Mitchell Harrop, and Martin Gibbs. 2014. Paradigms of games research in HCI: a review of 10 years of research at CHI. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play* (CHI PLAY '14), 27-36. DOI=10.1145/2658537.2658708
6. Debaleena Chattopadhyay and Davide Bolchini. 2014. Touchless circular menus: toward an intuitive UI for touchless interactions with large displays. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces* (AVI '14), 33-40. DOI=10.1145/2598153.2598181
7. John Downs, Frank Vetere, and Wally Smith. 2015. Differentiated Participation in Social Videogaming. In *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction* (OzCHI '15), 92-100. DOI=http://dx.doi.org/10.1145/2838739.2838777
8. Augusto Esteves, Eduardo Velloso, Andreas Bulling and Hans Gellersen. 2015. Orbits: Enabling Gaze Interaction in Smart Watches Using Moving Targets. In *Proceedings of the 2015 ACM Conference on Pervasive and Ubiquitous Computing* (UbiComp '15).
9. 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 ACM User Interface Software and Technology Symposium* (UIST'15).
10. 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. DOI=10.1145/1518701.1518748
11. Lyndsey Fisk, Marcus Carter, Behnaz Rostami Yeganeh, Frank Vetere, and Bernd Ploderer. 2014. Implicit and explicit interactions in video mediated collaboration. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: the Future of Design* (OzCHI '14), 250-259. DOI=10.1145/2686612.2686650
12. Simon Fothergill, Helena Mentis, Pushmeet Kohli, and Sebastian Nowozin. 2012. Instructing people for training gestural interactive systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '12), 1737-1746. DOI=10.1145/2207676.2208303
13. Franca Garzotto and Matteo Valoriani. 2013. Touchless gestural interaction with small displays: a case study. In *Proceedings of the Biannual Conference of the Italian Chapter of SIGCHI* (CHItaly '13), USA, , Article 26 , 10 pages. DOI=10.1145/2499149.2499154
14. Emilien Ghomi, Guillaume Faure, Stéphane Huot, Olivier Chapuis, and Michel Beaudouin-Lafon. 2012. Using rhythmic patterns as an input method. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '12, 1253-1262. DOI=10.1145/2207676.2208579
15. Sukeshini A. Grandhi, Gina Joue, and Irene Mittelberg. 2011. Understanding naturalness and intuitiveness in gesture production: insights for touchless gestural interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '11), USA, 821-824. DOI=10.1145/1978942.1979061
16. Darren Guinness, Alvin Jude, G. Michael Poor, and Ashley Dover. 2015. Models for Rested Touchless Gestural Interaction. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction* (SUI '15), 34-43. DOI=10.1145/2788940.2788948
17. John Hardy, Enrico Rukzio, and Nigel Davies. 2011. Real world responses to interactive gesture based public displays. In *Proceedings of the 10th International Conference on Mobile and Ubiquitous Multimedia* (MUM '11), 33-39. DOI=10.1145/2107596.2107600

18. Juan David Hincapié-Ramos, Xiang Guo, and Pourang Irani. 2014. The consumed endurance workbench: a tool to assess arm fatigue during mid-air interactions. In *Proceedings of the 2014 companion publication on Designing interactive systems* (DIS Companion '14), 109-112. DOI=10.1145/2598784.2602795
19. Chien-Ho Janice Lin, Katherine J. Sullivan, Allan D. Wu, Shailesh Kantak, and Carolee J. Winstein. 2007. Effect of task practice order on motor skill learning in adults with Parkinson disease: a pilot study. *Physical therapy* 87, 9: 1120-1131.
20. Alvin Jude, G. Michael Poor, and Darren Guinness. 2014. Personal space: user defined gesture space for GUI interaction. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '14), USA, 1615-1620. DOI=10.1145/2559206.2581242
21. Timothy Lee and Elizabeth Genovese. 1988. Distribution of practice in motor skill acquisition: Learning and performance effects reconsidered. *Research Quarterly for Exercise and Sport* 59, 4: 277-287.
22. Ville Mäkelä, Tomi Heimonen, and Markku Turunen. 2014. Magnetic Cursor: Improving Target Selection in Freehand Pointing Interfaces. In *Proceedings of The International Symposium on Pervasive Displays* (PerDis '14), 112 -118, DOI=http://dx.doi.org/10.1145/2611009.2611025
23. Alessio Malizia and Andrea Bellucci. 2012. The artificiality of natural user interfaces. *Commun. ACM* 55, 3 (March 2012), 36-38. DOI=10.1145/2093548.2093563
24. Anders Markussen, Mikkel Rønne Jakobsen, and Kasper Hornbæk. 2014. Vulture: a mid-air word-gesture keyboard. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (CHI '14), 1073-1082. DOI=10.1145/2556288.2556964
25. Paul Marshall, Richard Morris, Yvonne Rogers, Stefan Kreitmayer, and Matt Davies. 2011. Rethinking 'multi-user': an in-the-wild study of how groups approach a walk-up-and-use tabletop interface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '11), 3033-3042. DOI=10.1145/1978942.1979392
26. Microsoft. Xbox 360 Kinect Gestures. Retrieved September 3, 2015 from http://support.xbox.com/en-AU/xbox-360/kinect/body-controller.
27. George Miller. 1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychology Review* 63, 2: 81.
28. Calkin S. Montero, Jason Alexander, Mark T. Marshall, and Sriram Subramanian. 2010. Would you do that?: understanding social acceptance of gestural interfaces. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*(MobileHCI '10), 275-278. DOI=10.1145/1851600.1851647
29. Tomoya Murata and Jungpil Shin. 2014. Hand Gesture and Character Recognition Based on Kinect Sensor. *International Journal of Distributed Sensor Networks,* Article ID 278460: 6 pages
30. Matei. Negulescu, Jaime Ruiz and Edward Lank. 2010. Exploring usability and learnability of mode inferencing in pen/tablet interfaces. In *Proceedings of the Seventh Sketch-Based Interfaces and Modeling Symposium* (SBIM '10), 87-94.
31. Matei Negulescu, Jaime Ruiz, and Edward Lank. 2012. A recognition safety net: bi-level threshold recognition for mobile motion gestures. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services* (MobileHCI '12), 147-150. DOI=10.1145/2371574.2371598
32. Donald A. Norman and Jakob Nielsen. 2010. Gestural interfaces: a step backward in usability. *interactions* 17, 5 (September 2010), 46-49. DOI=10.1145/1836216.1836228
33. Kenton O'Hara, Richard Harper, Helena Mentis, Abigail Sellen, and Alex Taylor. 2013. On the naturalness of touchless: Putting the "interaction" back into NUI. *ACM Trans. Comput.-Hum. Interact.* 20, 1, Article 5 (April 2013), 25 pages. DOI=10.1145/2442106.2442111
34. Kenton O'Hara, Gerardo Gonzalez, Abigail Sellen, Graeme Penney, Andreas Varnavas, Helena Mentis, Antonio Criminisi, Robert Corish, Mark Rouncefield, Neville Dastur, and Tom Carrell. 2014. Touchless interaction in surgery. *Commun. ACM* 57, 1 (January 2014), 70-77. DOI=10.1145/2541883.2541899
35. Thammathip Piumsomboon, Adrian Clark, Mark Billinghurst, and Andy Cockburn. 2013. User-defined gestures for augmented reality. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '13), 955-960. DOI=10.1145/2468356.2468527
36. Gustavo Rovelo, Donald Degraen, Davy Vanacken, Kris Luyten and Karin Connix. 2015. Gestu-Wan - An Intelligible Mid-Air Gesture Guidance System for Walk-up-and-Use Displays. In *Proceedings of the International Conference on Human-Computer Interaction* (INTERACT'15), 368-386. DOI= 10.1007/978-3-319-22668-2_28

37. Hassan Saidinejad, Mahsa Teimourikia, Sara Comai, and Fabio Salice. 2014. Static hand poses for gestural interaction: a study. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces* (AVI '14), USA, 379-380. DOI=10.1145/2598153.2600049

38. Lawrence Sambrooks and Brett Wilkinson. 2013. Comparison of gestural, touch, and mouse interaction with Fitts' law. In *Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration* (OzCHI '13), 119-122. DOI=10.1145/2541016.2541066

39. Matthias Schwaller and Denis Lalanne. 2013. Pointing in the air: Measuring the effect of hand selection strategies on performance and effort." *Human Factors in Computing and Informatics*. Springer Berlin Heidelberg: 732-747.

40. Julia Schwarz, Charles Claudius Marais, Tommer Leyvand, Scott E. Hudson, and Jennifer Mankoff. 2014. Combining body pose, gaze, and gesture to determine intention to interact in vision-based interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14), 3443-3452.

41. Ben Shneiderman. 1997. Direct manipulation for comprehensible, predictable and controllable user interfaces. In *Proceedings of the 2nd international conference on Intelligent user interfaces* (IUI '97), 33-39. DOI=10.1145/238218.238281

42. Helman Stern, Juan Wachs and Yael Edan. 2008. Optimal Consensus Intuitive Hand Gesture Vocabulary Design. In *2008 IEEE Int. Conference on Semantic Computing*, 96-103. DOI= 10.1109/ICSC.2008.29.

43. Radu-Daniel Vatavu and Ionut-Alexandru Zaiti. 2014. Leap gestures for TV: insights from an elicitation study. In *Proceedings of the 2014 ACM international conference on Interactive experiences for TV and online video* (TVX '14), 131-138. DOI=10.1145/2602299.2602316

44. 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. DOI=10.1145/2493432.2493477

45. Mélodie Vidal, Andreas Bulling, and Hans Gellersen. 2015. Pursuits: Spontaneous Eye-Based Interaction for Dynamic Interfaces. *GetMobile: Mobile Comp. and Comm.* 18, 4 (January 2015), 8-10. DOI=10.1145/2721914.2721917

46. Mélodie Vidal, Ken Pfeuffer, Andreas Bulling, and Hans W. Gellersen. 2013. Pursuits: eye-based interaction with moving targets. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '13), 3147-3150. DOI=10.1145/2468356.2479632

47. Daniel Vogel and Ravin Balakrishnan. 2005. Distant freehand pointing and clicking on very large, high resolution displays. In *Proceedings of the 18th annual ACM symposium on User interface software and technology* (UIST '05), 33-42. DOI=10.1145/1095034.1095041

48. Robert Walter, Gilles Bailly, and Jörg Müller. 2013. StrikeAPose: revealing mid-air gestures on public displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '13), 841-850. DOI=10.1145/2470654.2470774

49. Daniel Wigdor and Dennis Wixon. 2011. *Brave NUI World: Designing Natural User Interfaces for Touch and Gesture*. Morgan Kaufman: London.

50. 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. DOI=10.1145/985921.986076

51. Soojeong Yoo, Callum Parker, Judy Kay, and Martin Tomitsch. 2015. To Dwell or Not to Dwell: An Evaluation of Mid-Air Gestures for Large Information Displays. In *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction* (OzCHI '15), 187-191. DOI=http://dx.doi.org/10.1145/2838739.2838819

52. Bruno Zamborlin, Frederic Bevilacqua, Marco Gillies, and Mark D'inverno. 2014. Fluid gesture interaction design: Applications of continuous recognition for the design of modern gestural interfaces. *ACM Trans. Interact. Intell. Syst.* 3, 4, Article 22 (January 2014), 30 pages.