Do That, There: An Interaction Technique for Addressing In-Air Gesture Systems

Euan Freeman, Stephen Brewster
Glasgow Interactive Systems Group
University of Glasgow
Glasgow, Scotland
{first.last}@glasgow.ac.uk

Vuokko Lantz
Sensor Systems
Nokia Technologies
Espoo, Finland
vuokko.lantz@nokia.com

ABSTRACT
When users want to interact with an in-air gesture system, they must first *address* it. This involves finding where to gesture so that their actions can be sensed, and how to direct their input towards that system so that they do not also affect others or cause unwanted effects. This is an important problem [6] which lacks a practical solution. We present an interaction technique which uses multimodal feedback to help users address in-air gesture systems. The feedback tells them how ("do that") and where ("there") to gesture, using light, audio and tactile displays. By doing that there, users can direct their input to the system they wish to interact with, in a place where their gestures can be sensed. We discuss the design of our technique and three experiments investigating its use, finding that users can "do that" well (93.2%–99.9%) while accurately (51mm–80mm) and quickly (3.7s) finding "there".

Author Keywords
Address; Gestures; Interactive Light; Rhythmic Input.

ACM Classification Keywords
H.5.2. User Interfaces: Interaction styles.

INTRODUCTION
In *Making Sense of Sensing-Systems*, Bellotti *et al.* [6] identified five usability problems which should be considered when designing sensing systems, such as for in-air gestures. The first of these was the way in which users *address* the system, which involves directing input towards it; this happens *before* control, when users first initiate interaction. When interacting with physical input devices, like keyboards or touchscreens, users direct their input explicitly, by reaching out and touching the controls. However, it is less clear how input should be directed when using in-air gestures, which are sensed in the environment and may be seen by many systems at once (as in Figure 1). Users must know where to gesture, so that their input can be sensed, and how to direct their input, so that they do not affect other gesture systems unintentionally.

This is the author's copy. The definitive version will appear in the proceedings of CHI '16 in the ACM Digital Library.

DOI: http://dx.doi.org/10.1145/2858036.2858308

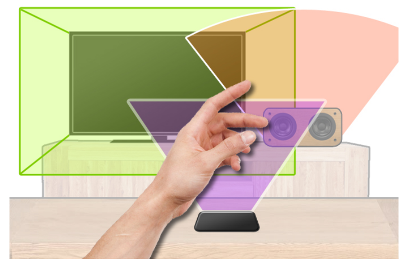

Figure 1. Users need to be able to *address* gesture systems, otherwise their actions may produce unwanted (or no) effects. Here, many devices are sensing input from the same space; without directing input towards one in particular, gestures may have unintended effects on other devices.

Without knowing where to perform in-air gestures, users may interact in a place where their hand movements cannot be detected by sensors. Even if they can initially be sensed, their hand movements during a gesture may take them out of the sensor space; for example, when waving from side to side, the hand may briefly leave the sensor range. Users also need to gesture where sensors can detect sufficient detail. Input too far from, or too close to, the sensor may mean gestures cannot be sensed accurately. Helping users find where to gesture, when they are addressing a system, can avoid these issues.

Users must also be able to direct gestures towards a specific system. Without doing so, gestures may be unintentionally sensed and acted upon by others; this is the *Midas Touch* problem [26]. While others have proposed ways of directing input towards gesture systems, these have limitations which mean they may be impractical in real use. Some are only intended for use with a single system and others may suffer from ambiguity over which system users want to address.

In this paper, we investigate a solution to the address problem, presenting a technique called "do that, there". It is a way to address gesture systems, not control them, and can be used by many systems at once. It helps users direct input towards a position-based gesture system by showing them how ("do that") and where ("there") to address it. Once users have addressed a system, they can control it with gestures. We do not look at specific applications or gestures; instead, our work focuses on address. We designed our technique so that it can be used by a wide variety of devices, including those without a screen for displaying visual feedback. Use of alternative outputs (audio, tactile and interactive light) means that devices of all shapes and sizes can help users do that, there.

We present three experiments which evaluate our interaction technique, first studying "that" and "there" on their own, as they could also be used in other interactions, then finally combining them. We found that our technique can help users find "there" with good accuracy. They also "do that" well, directing their input successfully with minimal feedback. When these are combined, they remain successful, with users completing almost all gestures successfully.

RELATED WORK

Users need unified ways of interacting with the increasing number of 'smart' devices in their environments. *Proxemic-Aware Controls* [28] used mobile devices as a "universal control" for such systems, with proximity and orientation giving varying amounts of control. In-air gestures complement this style of control, although (in terms of *gradual engagement* [31]) they provide more convenient, but limited, distal control when users do not wish to approach and engage more. For in-air gestures to provide effective control, however, users must be able to address the device they want to interact with.

The way users address an interface is "*so fundamental that it is often taken for granted in UI design*" [6]. Yet, such a fundamental part of interaction can cause a variety of challenges for users. Bellotti *et al.* [6] discuss how an unwanted response, and no response when expected, are potential problems when addressing a sensing system. Unwanted responses come from unintentional gesture input (the Midas Touch problem [26]). This may occur when ordinary movements, like reaching for something, are detected and treated as meaningful input. Gestures intended for one system may also be misinterpreted as input by another. Solutions have been proposed for overcoming these issues, by directing input, and will be discussed below. These can help avoid unintentional input, but have limitations which make them impractical when there is more than one gesture system in the environment; this scenario becomes more likely as in-air gesture-sensing technology continues to improve, in terms of resource demands and size, and gets introduced to more of our devices, like it already has with smartphones and televisions, for example.

Another aspect of addressing a system is knowing where to interact. Users must gesture within range of the sensors for their movements to be sensed; otherwise, they may get no response when one is expected. Gesture systems may have ambiguous input areas and it is not always clear where sensors are located and what they can and cannot see. One approach to help users find where to gesture is to give feedback when they can be sensed. Large surfaces often show users what the sensors see; for example, *StrikeAPose* [49] displayed a stylised version of depth-camera data, allowing users to see their position in the field of view. Xbox Kinect games often use similar visualisations. *Proxemic Flow* [48] gave feedback from a floor display; "halos" around users' feet used a traffic light metaphor to show tracking quality. Users can use this information to adapt how they interact, for example moving closer to the sensor or towards the centre of its field of view.

Not all systems have these display capabilities, so other outputs are required. Morrison *et al.* [33] looked at how audio feedback could help users understand how well they can be seen. They found that telling users which body parts could not be seen offered no benefit over simply saying that the body was not in full view; it was also more complex. Although feedback saying the body could not be fully seen was more ambiguous than naming unseen body parts, such feedback could encourage users to explore the input space and develop their own understanding about how they need to gesture. Gaver *et al.* [17] argued that ambiguity can be a useful design resource as it encourages users to engage more and conceptually understand how interfaces work. This could be useful in gesture systems as it may help users gain a better understanding of what they need to do to be sensed properly.

Directing Gesture Input

Many have explored pointing at things to direct input towards them. *GesturePen* [46] was a stylus which allowed users to select between objects by pointing at them in mid-air. Users pointed at objects ("that one there") and then confirmed their selection by pressing a button on the stylus. *DopLink* [5] also let users direct input by pointing at a device, using flick-gestures rather than button presses for selection. *PICOntrol* [40] allowed users to interact with distant objects by aiming a pico projector at them. Projectors were also used to show interface controls in the space surrounding objects. These interactions gave users no way to overcome ambiguous selections. If users pointed towards two or more objects which were close to each other, they had no way of resolving selection ambiguity. With *Point & Control* [8], users pointed a smartphone towards an object to address it and, if selection was ambiguous, a list of possible targets was displayed on the touchscreen. Delamare *et al.* [12, 13] used wrist-rotation to choose between objects. Users pointed in the general area of the device they wanted to address; then, they rotated their wrist to select between objects in that area.

An alternative to pointing at a device to address it is to use an *activation gesture*. These direct input towards one system by showing intention to interact with it. Activation gestures must be uncommon so that they are not performed accidentally [26, 18]. Once gesture input is active, sensing continues until users perform *closure* gestures [18] or leave the sensor space. Activation gestures include dynamic hand movements, like the Xbox 360 "Wave to Kinect" gesture, static hand poses, like pointing, and full-body poses, like the *Teapot* gesture (placing hand on hip) from *StrikeAPose* [49].

When using activation gestures, it is important to select a good pose or gesture. O'Hara *et al.* [37] found that their users often activated theirs unintentionally, even though it was an uncommon movement. Activation gestures must be unique to every system if users are to specify *which* system they want to address. This would require many gestures and a way of revealing them to users, which they may be unable to do. The target could also be inferred using other inputs, like body pose and gaze [41]. However, such techniques could also lead to ambiguity, like with pointing gestures. Maglio *et al.* [29] found that users did not always look at devices before addressing them and there were also occasions when users did not look at them at all during interaction. If sensors are embedded within devices, rather than external like a Kinect, then users may not know where to look in the first place, adding further ambiguity about what they are addressing. We build on this work by investigating an alternative way of directing input towards gesture systems, with minimal feedback requirements.

Giving Gesture Feedback

Our research investigates ways of addressing in-air gesture systems, which are becoming increasingly diverse as novel sensing capabilities are added to more devices. For a solution to be effective and widely used, it needs to be usable by these diverse devices, from small devices with no screen to those with large displays. We therefore focus on three types of output which can be used in the absence of a screen: audio, tactile and interactive light displays. By focusing on these novel display types, our interaction technique can be used by a variety of devices, with or without screens.

Audio displays are widely available to gesture systems; speakers are commonly used for delivering feedback and users often wear earphones when using some devices, like phones and entertainment systems. Others have investigated how audio displays can be used by gesture systems. For example: Morrison *et al.* [33] used sound to tell users which of their body parts a gesture-sensing system could and could not "see"; Charbonneau *et al.* [11] used sound to help teach body gestures in games; and Morrison-Smith and Ruiz [34] used audio feedback to teach motion gestures.

Giving tactile feedback about in-air gestures is more challenging than audio, as users cannot feel feedback from devices they do not touch during interaction. However, recent technologies have made tactile feedback about in-air gestures possible. Ultrasound haptic displays [24, 10] create tactile sensations in mid-air and some wearables, like activity trackers and smartwatches, have tactile displays for delivering alerts. Freeman *et al.* [15] explored this area further, comparing ultrasound haptics to wearable tactile displays. They found that tactile feedback about in-air gestures was successful and could enhance gesture interaction.

Finally, we consider the use of interactive light displays for giving feedback about gestures (as in [16]). These displays use light to illuminate the space around a device. For example, Qin *et al.* [38] and Müller *et al.* [35] used LEDs embedded around mobile devices to illuminate surrounding spaces for display. We extend this idea and use interactive light displays for gesture feedback, allowing devices with no screens, or small screens, to present feedback in their surrounding area.

Rhythmic Input

Our interaction technique uses rhythmic input, as will be discussed in the following section. Rhythm has been used in interaction techniques before (e.g. [14, 30, 7]), allowing users to control input by repeating actions in time. We build on these by looking at rhythmic in-air gestures for the first time. We draw on neuroscience, where rhythmic hand movements are better understood (e.g. [4, 23, 22, 39, 44, 45, 47]).

DO THAT, THERE

We now describe two hardware prototypes created to allow us to look at the use of our interaction technique. We chose two devices so we could investigate "do that, there" with different form factors: mobile phones and small household appliances. These types of device are similar in that both have small screens, although users might gesture at them in different ways, using smaller movements near a mobile phone and larger movements towards a device opposite the room.

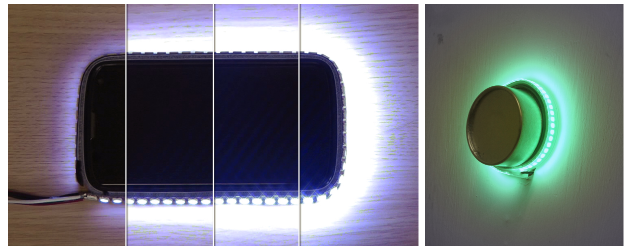

Figure 2. *Left:* smartphone prototype, displaying varying brightness levels. *Right:* dial, illuminating the surrounding wall with green light.

Some phones now have in-air gesture interfaces (e.g. Samsung's *Galaxy S4*) and many have explored gestures near phones (e.g. [9, 27, 51, 42]). Gestures allow input without first lifting the phone, even from across the room [51], provide an input space larger than the screen [25, 20, 21] and may let users interact more expressively [43]. These phones may benefit from our interaction and the displays we use; their small screens limit the amount of feedback which can be given and on-screen feedback further restricts the space available for content. Multimodal feedback from other outputs keeps the screen free and gives a larger output space.

We also consider small household devices, like 'smart' thermostats [36]. An increasing number of household devices are being enriched with novel input methods and gestures could give users another way of interacting with them. However, such devices typically have small screens, or no screen at all, limiting the amount of feedback that can be given. We explore if other outputs, as well as our interaction technique, can help users address and interact with them using in-air gestures. This also lets us evaluate our interaction in a different context, gesturing across the room rather than over a device.

Our phone prototype (Figure 2, left) has 60 *NeoPixel* LEDs [3] around the device edge to illuminate surrounding surfaces, similar to Qin's light display phone prototype [38]. These LEDs can change hue and brightness independently and are controlled by an Arduino microcontroller. A Leap Motion [1] centred over the phone tracks finger movements above it. Our household appliance prototype (Figure 2, right) has a dial form factor, inspired by modern 'smart home' music systems and thermostats. The dial (⌀10cm) has 46 *NeoPixel* LEDs [3] around its edge for interactive light feedback. The dial uses an Xbox Kinect [32] for sensing gestures. Both prototypes deliver audio feedback using laptop speakers. Tactile feedback is given from an on-wrist tactile display, as this was successful in similar work [15]. We use a Precision Microdrives [2] C10-100 actuator (used by [15]) attached to a fabric watch strap. This is driven using a laptop audio output, to synchronise audio and tactile cues.

We now describe "do that, there", our interaction technique for addressing in-air gesture systems. This technique builds on the research discussed earlier; it overcomes limitations with existing interactions for addressing gesture systems and it uses the novel displays discussed before for feedback.

There: Showing Users Where to Gesture

When addressing a gesture system, users must first identify where to perform gestures, otherwise their gestures may not be sensed properly, or at all. We give multimodal feedback to help guide users, showing them where they should provide input ("there"). Rather than explicitly guide them, we give feedback telling them how well they can be seen by sensors (similar to *Proxemic Flow* [48]); this encourages them to explore the input space and form their own understanding of how it works, an idea discussed earlier. We do this by estimating how well users can be seen, as a function of the distance between their hand and a "sweet spot" within sensor view; in this case, this point is the centre of the field of view, at a distance where sufficient detail can be sensed.

Interactive light displays use brightness of white light to present this information (as in Figure 2), with brighter light when users are easily sensed; we mapped brightness inversely (100%–0%) to distance from the "sweet spot". In this case, hue was not used because it is used in the "do that" feedback. With audio and tactile displays, we give 'Geiger counter' feedback, with pulsing tones and vibrations being given more frequently when users can be better sensed. These cues were 35ms long and we mapped intervals between them (70ms–370ms) to distance from the "sweet spot". Tones were 370 Hz (F#4) sine waves and vibrations were 150 Hz (best actuator frequency), both generated in *Pure Data*. All three types of feedback are more salient when users are gesturing in a good location and are less noticeable when the position is bad.

Do That: Showing Users How to Direct Input

Users also have to be able to specify that gestures are intentional and directed towards the system they wish to interact with. We discussed ways of doing this earlier, including pointing gestures, activation gestures and with gaze. However, these were limited by potential ambiguity and being unsuitable for use when there are multiple gesture systems. Instead, we propose a new input technique called *rhythmic gestures* for directing input, for use by many systems at once.

Rhythmic gestures are movements repeated in time with a rhythm, which is shown using interactive light animations. We use light animations as these can convey spatial and temporal information, and research has found good performance synchronising hand movements with a visual rhythm [23, 22, 4]. Like activation gestures, rhythmic gestures allow users to show intent to interact; however, they also allow them to direct input (through the movement and its tempo). An example is waving a hand from side to side, once per second. When using this technique, each gesture system would have its own rhythmic gesture and would only accept input if users had first performed the appropriate rhythmic gesture to address it.

We selected five hand movements for rhythmic gestures, shown in Figure 3. Gesture systems reveal their rhythmic gesture using light animations (or on-screen ones, for larger displays), with the movement of light showing the gesture motion and its tempo. The following list describes the animations for our five gestures and Figure 4 shows an example. These were designed to show how to move relative to the dial LED display and may be inappropriate for different display setups.

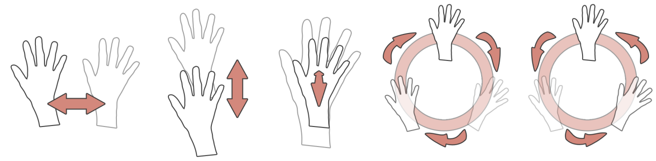

Figure 3. Our five rhythmic gesture movements, from left to right: Side-to-Side (SS), Up-and-Down (UD), Forwards-and-Backwards (FB), Clockwise (C), Anticlockwise (AC).

Figure 4. Movement of light for the Side-to-Side gesture animation.

Gesture animations use white light to show movement, turning green when users are gesturing in time with them. Users *match* a rhythmic gesture if they have performed three correct movements in time with its rhythm; this ensures that input is intentional and users are not making other movements (like reaching forward and picking something up).

- **Side-to-Side:** A group of lights moves from one side to the other and back again, continuously; see Figure 4;
- **Up-and-Down:** Like Side-to-Side, but light moves up and down the interactive light display, instead;
- **Forwards-and-Backwards:** Lights on at 12 o'clock, 3 o'clock, 6 o'clock and 9 o'clock; more lights fade in, gradually filling the gaps until all are on; then reversed;
- **(Anti-)Clockwise:** A group of lights continuously moves around the light display in the appropriate direction.

Audio and tactile feedback are given about rhythmic gesture movements. At the end of each sensed movement (for example, after users complete a circular movement or when their side-to-side movement changes direction), a 200ms tone (523 Hz, C5) and vibration (150 Hz) are presented using the audio and tactile display. This feedback confirms that movements have been sensed and can let users determine if they are performing movements in time with the visual animation.

Do That, There: Combining the Interactions

Our "there" and "do that" interactions were designed to be used together, so that gesture systems could give users feedback telling them how well they can be sensed, while also showing them which gesture they need to perform to select it. There are many ways of combining feedback for the "do that" and "there" interactions, across the three output types we use. We created five designs which allow us to investigate aspects of feedback design. First, we considered when rhythmic gesture feedback should be given: should this be given while users are locating where to gesture, or only after they have started a gesture? Second, we created designs to investigate if feedback for "there" is still useful after users have started a rhythmic gesture. Finally, we chose designs which explore modality combinations: should "do that" and "there" use the same output channels, or be presented separately? We investigate these questions in Experiment 3 of this paper. Our five feedback designs are:

- **All:** "there" and "do that" at the same time using all output channels together, from the very start of an interaction;
- **All-Short:** Same as *All*, except feedback for "there" stops once users begin matching a rhythmic gesture;
- **Split:** "there" using audio and tactile, "do that" using interactive light. Each from the very start of interaction;
- **Split-Short:** Same as *Split*, except feedback for "there" stops once users begin matching a rhythmic gesture;
- **Split-Swap:** Same as *Split-Short*, except audio and tactile channels are used for rhythmic gesture feedback once feedback for "there" stops.

Light was used for "do that" in all designs, because it is needed for gesture animations. *All* vs *Split* explores modality use, with *Split* using each modality for different information. *-Short* investigates if "there" cues should stop once users begin gesturing. The *Split-Swap* design shares the use of non-visual feedback between "there" and "do that"; there is no *All-Swap* because other designs give the same outcome.

When performing a rhythmic gesture, hand movement means that the calculation of "there" feedback will change continuously. This is undesirable, as users could be gesturing in a good location but may receive constantly changing feedback guiding their movements elsewhere. To overcome this problem, we use a weighted average hand position when estimating how well users can be seen. As gesture movements are repeated, the mean hand position will not change much unless users are searching the input space. Another issue is that many devices giving audio and tactile feedback at once could be noisy and obtrusive, especially for "there". We do not investigate this in this work, however gesture systems could make an informed choice about which one should give feedback; for example, the system users are addressing, or the system whose "sweet spot" they are closest to.

EXPERIMENT 1: "THERE"

Our first experiment investigated the "there" part of the interaction, to see if the feedback was effective for guiding hand movements, as this is necessary for helping users find where to gesture. We used the mobile phone prototype (from Figure 2) for this study, with the Leap Motion tracking hand position. Participants sat at a table with the mobile phone in front of them and were asked to locate target points over it, using their dominant hand's extended index finger. Feedback was given about the fingertip position relative to the target. Points were randomly positioned within a volume from (-100mm, 200mm, -100mm) to (100mm, 300mm, 100m), relative to the sensor. This range was used so that hands would remain fully visible to the sensor during tasks. Tasks started when users moved their hand over the sensor and ended when they pressed the spacebar on a keyboard in front of them with their other hand; this meant final finger position could be accurately recorded.

There were seven conditions, representing each unimodal and multimodal combination of the three outputs: Light Only (L), Audio Only (A), Tactile Only (T), Light & Tactile (LT), Light & Audio (LA), Audio & Tactile (AT), All (LAT). Participants completed a block of 15 tasks for each condition; condition order was balanced using a Latin square. At the beginning of the experiment sessions, we gave a short tutorial about the interaction, giving participants a chance to try the tasks and experience the feedback. During tasks, we measured *Time*, the total task time, and *Distance*, the Euclidean between target point and final finger position ($\sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}$). Sixteen people participated (six female, two left-handed); mean age was 26.1 years (sd 3.4 years). All were paid £6.

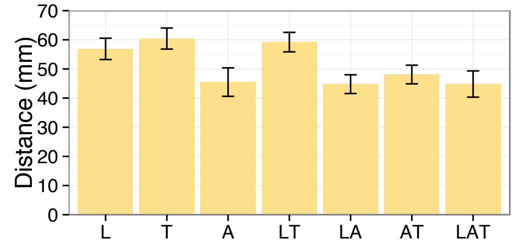

Figure 5. Mean *Distance* for each condition.

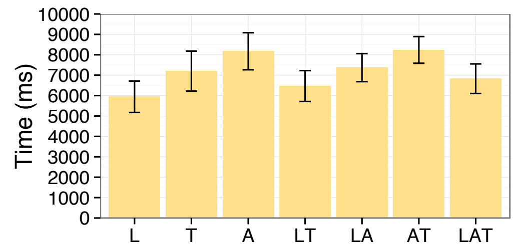

Figure 6. Mean *Time* for each condition.

Results

Mean *Distance* was 51mm (sd 15mm); see Figure 5 (all error bars show 95% CIs). A repeated-measures ANOVA found a significant effect of condition: $F(6, 90) = 15.83$, $p < 0.001$. *Post hoc* pairwise t-tests found all conditions containing audio (A, LA, AT, LAT) had significantly lower *Distance* than those without (L, T, LT): all $t \geq 3.47$, $p \leq 0.01$ (all p-values in this paper were Bonferroni-corrected for multiple comparisons).

Mean *Time* was 7174ms (sd 2819ms); Figure 6 shows *Time* per condition. *Time* was not normally-distributed so data were transformed using the Aligned-Rank Transform [50], allowing use of parametric methods. A repeated-measures ANOVA found a significant effect of condition: $F(6, 90) = 6.20$, $p < 0.001$. *Post hoc* t-tests found L and LT were both significantly faster than A and AT: all $t \leq -3.8$, $p \leq 0.005$.

Discussion

Users located targets more accurately when given audio feedback; tactile did not perform as well, despite using the same temporal design as audio. Some participants reported that tactile feedback often felt continuous so the minimum interval for the Geiger counter feedback may have been too low for it to be accurately perceived. This was not a problem for audio feedback, possibly because the auditory modality has a higher temporal resolution than tactile.

Accuracy with audio feedback often came at the expense of task time, however; *Time* was significantly higher for conditions with light and without audio (L and LT) than for those without light and with audio (A and AT). Participants liked interactive light feedback because it responded quickly to their hand movements, allowing them to see when they were in the "right area". Some participants described using light to initially position their hand and then using other feedback to more accurately locate targets. Light with audio could be the most effective as users would benefit from faster localisation (from light) and better accuracy (from audio).

This experiment focused on accurate localisation to fully investigate the effectiveness of the feedback for "there". This resulted in longer interaction times (seven seconds) than may be necessary for locating where to gesture, as users focused on accuracy. It is also unknown if the findings are the same for guiding hand movements, as finger movements were used here. Experiment 3 addresses these points: it combines "do that" and "there", which will show if users need as much time to find where to gesture when also giving input; and it uses hand movements from a greater distance away.

EXPERIMENT 2: "DO THAT"

Our second experiment investigated the "do that" part of the interaction, to see how well users could perform rhythmic gestures and to investigate the design space. These gestures were studied in isolation here so that they could be more fully understood than if they were studied as part of the combined interaction. We used the dial prototype (Figure 2) for this experiment; it was mounted on a wall and users sat 2.5m opposite, as though controlling a device from across the room. We chose the dial for this study because rhythmic gestures could be more reliably sensed from a distance using the Kinect.

Each task required users to 'match' a rhythmic gesture which was shown to them using the dial's interactive light display. The earlier description of rhythmic gestures explains criteria for matching a rhythmic gesture. There was a twelve second limit in which to match a gesture; this was chosen to reduce fatigue from trying to match more difficult gestures.

The five gesture movements discussed earlier were used: Side-to-Side (SS), Up-and-Down (UD), Forwards-and-Backwards (FB), Clockwise (C) and Anticlockwise (AC). These were combined with a variety of gesture intervals: 500ms, 700ms, 900ms and 1100ms. These were chosen because they exceed the minimum interval for synchronising hand movements with a repeating visual stimulus (460ms [39]) yet are not too long that interaction is unnecessarily time consuming. Research also suggests that rhythmic movements become less accurate as the movement interval increases beyond 800ms [47], so our longest interval of 1100ms was still close to this value.

Users were given four types of feedback about rhythmic gestures during the experiment, as described earlier: additional audio feedback (Audio), additional tactile feedback (Tactile), audio and tactile together (Both) and no additional feedback (None). These feedback designs were used in addition to the interactive light animations about the rhythmic gestures.

	C	**AC**	**FB**	**SS**	**UD**
500ms	73%	69%	89%	97%	97%
700ms	91%	84%	92%	100%	98%
900ms	98%	94%	95%	100%	100%
1100ms	98%	97%	91%	100%	100%

Table 1. Success for each gesture. [Anti-]Clockwise (AC, C), Forwards-and-Backwards (FB), Side-to-Side (SS), and Up-and-Down (UD).

We investigated different feedback designs to see how extra feedback about gesture movements affected performance. Research suggests that rhythmic movements are performed more accurately in time when additional discrete feedback is given about movements [44, 45], so we wanted to investigate if this was the case for rhythmic gestures.

There were four conditions in this experiment, one for each type of feedback. For each, participants completed a block of twenty tasks (one for each gesture and interval combination). This meant just one trial per combination, but we wanted to cover a broad design space using a within-subjects design without too many trials. We measured *Success* (if gesture matched or not) and *Time-Match*, the time taken to match a rhythmic gesture. After each task, we asked participants to rate the difficulty of matching the gesture rhythm (*Difficulty-Match*), using a ten-point scale from 1 (easiest) to 10 (most difficult); ratings were given verbally and noted by experimenter. Sixteen people took part (four female); their mean age was 28.9 years (sd 4.5 years). All were paid £6.

Results

Success rates

Users successfully matched 1193 of 1280 rhythmic gestures (93.2%); Table 1 shows the success rate of each gesture and interval combination. Logistic regression was used to analyse the effect of gesture, interval and feedback on *Success*. A repeated-measures ANOVA on the regression model found that: **Feedback** had no significant effect on *Success* ($\chi^2(3) = 3.76$, $p = 0.29$); **Gesture** ($\chi^2(4) = 65.71$, $p < 0.001$) and **Interval** ($\chi^2(3) = 55.48$, $p < 0.001$) did; and no interactions between factors were significant (all $p \geq 0.06$).

Post hoc Wilcoxon's tests for **Gesture** found that *Success* was higher for SS and UD than all other gestures: all $z \geq 3.50$, $p \leq 0.004$. *Post hoc* Wilcoxon's tests for **Interval** found that *Success* was lower for 500ms intervals than all others (all $z \geq 2.70$, $p \leq 0.002$) and was lower for 700ms than for 900ms ($z = 2.70$, $p = 0.03$). No other comparisons were significant.

Gesture times

Interval affects *Time-Match*, as longer movements will need more time to be matched. All timing data were normalised to an interval of 500ms to account for this. Mean normalised *Time-Match* was 2204ms (sd 1548ms); see Figure 7. Times were not normally-distributed so were transformed using the Aligned-Rank Transform [50] prior to analysis. The results from a repeated-measures ANOVA are shown in Table 2.

Post hoc t-tests for **Gesture** found the following significant differences: both C and AC took longer to match than all others (all $t \geq 5.6$, $p < 0.001$); SS took less time than FB

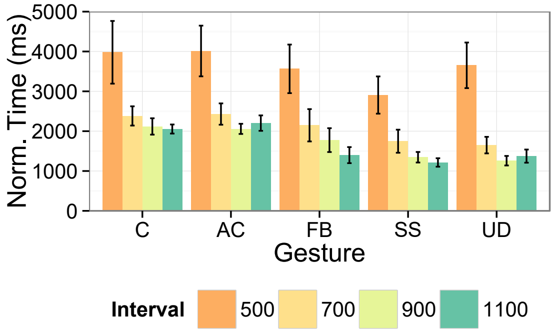

Figure 7. Mean *Time-Match* for each gesture and interval, normalised to an interval of 500ms to account for effects of interval on timing.

Effect	df	F-value	p-value
Feedback	3, 1099	0.98	0.40
Gesture	4, 1099	59.57	< 0.001
Interval	3, 1099	153.78	< 0.001
Feedback x Gesture	12, 1099	2.40	0.004
Feedback x Interval	9, 1099	2.07	0.03
Gesture x Interval	12, 1099	3.98	0.006
Feedback x Gesture x Interval	36, 1099	1.67	0.008

Table 2. Repeated-measures ANOVA results for *Time-Match*. Rows with significant effects (p < 0.05) have been highlighted .

(t = 5.6, p < 0.001) and UD (t = 2.9, p = 0.03). *Post hoc* t-tests for **Interval** found: 500ms took longer than all others (all t ≥ 13.1, p < 0.001); and 700ms took longer than 900ms and 1100ms (both t ≥ 5, p < 0.001). *Post hoc* t-tests for **Feedback x Gesture** found no differences within-Feedback or within-Gesture. *Post hoc* t-tests for **Feedback x Interval** found *Time-Match* was higher without feedback at 500ms than at 700ms. *Post hoc* t-tests for **Gesture x Interval** found that C took significantly longer to match with a 500ms interval than with a 1100ms interval (t = 4.42, p = 0.002), and SS took longer with a 500ms than with 900ms and 1100ms intervals (both t ≥ 3.9, p ≤ 0.01). Finally, *post hoc* t-tests across all three factors found no significant differences.

Difficulty ratings

Mean *Difficulty-Match* was 3.41 (sd 2.07); see Figure 8. Difficulty ratings were transformed using the Aligned-Rank Transform, meaning parametric tests could be used to analyse the non-parametric data [50]. Results from a repeated-measures ANOVA are shown in Table 3.

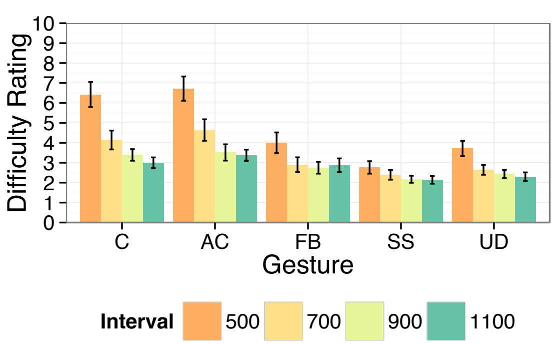

Figure 8. Mean *Difficulty-Match* for each gesture and interval.

Effect	df	F-value	p-value
Feedback	3, 1185	10.40	< 0.001
Gesture	4, 1185	146.15	< 0.001
Interval	3, 1185	138.02	< 0.001
Feedback x Gesture	12, 1185	1.73	0.06
Feedback x Interval	9, 1185	1.94	0.04
Gesture x Interval	12, 1185	14.75	< 0.001
Feedback x Gesture x Interval	36, 1185	1.02	0.42

Table 3. Repeated-measures ANOVA results for *Difficulty-Match*. Rows with significant effects (p < 0.05) have been highlighted . Reported *df*s differ from Table 2 as times were not recorded for incomplete trials.

Post hoc t-tests for **Feedback** found that: difficulty ratings were lower for Audio and Both than for None (both t ≥ 3.88, p ≤ 0.004); and ratings for Audio were lower than Tactile (t = 3.33, p = 0.005). *Post hoc* comparisons for **Gesture** were all significant (all t ≥ 5.18, p < 0.001), except between C and AC and between FB and UD (both t ≤ 2.57, p ≥ 0.08). Ratings were lower for SS than all others, while FB and UD were rated easier than C and AC. *Post hoc* t-tests for **Interval** found that all differences were significant (all t ≥ 4.55, p < 0.001), except between 900ms and 1100ms (t = 1.72, p = 0.31). Difficulty ratings were higher for the lower intervals. *Post hoc* t-tests for **Feedback x Interval** found no sig. differences.

Post hoc t-tests for **Gesture x Interval** found many significant differences; only those relevant to the research aims will be discussed here. There were no significant differences within-gestures between 700ms and 900ms, between 700ms and 1100ms, or between 900ms and 1100ms (all p > 0.05). Only SS and AC had significant differences between 500ms and 700ms (both t ≥ 3.62, p ≤ 0.04).

Discussion

Five rhythmic gesture movements were investigated in this experiment. Of these, circular ones were found to be the most difficult to use. Participants took longer to match them and gave higher difficulty ratings for them than for all other movements. Circular gestures required more complex hand movements, whereas the others only required movement in one direction. Despite the increased difficulty, users still performed them well, especially at slower intervals (≥ 900ms).

SS was the best performed gesture and also had the lowest difficulty ratings. During the experiment, users were observed to use smaller hand movements than they did for UD and FB; these smaller movements may contribute to their better performance, as users could exert more control over their gestures. UD also performed well in this experiment, suggesting that gesture systems should prioritise the use of SS and UD.

Gestures generally increased in difficulty as the interval decreased, shown through performance times and difficulty ratings. However, most of the significant differences found were between larger increases (500ms–900ms, for example) than for stepwise increases (700ms–900ms, for example). In many cases, there were no differences at all between 700ms, 900ms and 1100ms. These findings suggest that interval can be used effectively as a design parameter for rhythmic gestures, although very short intervals (e.g. 500ms) should be avoided.

When gesture-sensing systems coordinate their choice of rhythmic gestures (so that each has a unique gesture), they may benefit from using fewer types of gesture movement and more intervals, rather than selecting from more difficult movements (the circular ones, for example).

Feedback about hand movements had no effect on performance, but did make it easier to gesture (shown by lower difficulty ratings). Although participants benefitted from all types of feedback, tactile cues were less effective than audio ones. Similar findings were observed in Experiment 1 for the "there" interaction. Audio feedback may have been easier to perceive in both experiments, making it more useful during interaction. However, participants still found tactile feedback helpful and its use may be preferred in situations where audio feedback would be obtrusive or inappropriate.

EXPERIMENT 3: "DO THAT, THERE"

Our final experiment investigated the usability of "do that, there", a novel interaction technique which could be used for addressing gesture-sensing systems. This experiment evaluated the combined use of the interaction techniques, using the feedback designs described before. This experiment used the dial prototype from Experiment 2, meaning it also investigated the effectiveness of the "there" interaction with larger hand movements from across the room, for the first time.

Users were asked to "do that, there", completing tasks which combined aspects of the previous experiments. They had to locate where to gesture ("there") while matching a rhythmic gesture in that place ("do that"). Target points were positioned at one of the eight corners of a 300x150x50mm volume, centred around the hand position at the beginning of each task. Placing targets at the corners of this volume meant participants had to move their hands to find "there", while still performing rhythmic gestures within comfortable reach. The z-axis of this volume (50mm) was limited as participants would be seated during the experiment and larger variation may have required reaching too far forward or leaning backwards to find targets. Participants were seated 2.5m from the Kinect, as though controlling a device from across the room.

Once participants matched a gesture rhythm, they were to continue performing it for as long as possible. This was to investigate how feedback affected gesture performance once a gesture had been matched. Users had 12 seconds to match a gesture, as in Experiment 2. Once a gesture had been matched, users were to continue for a maximum of 10 seconds. Two gestures (Side-to-Side and Up-and-Down) were used, with a 700ms interval. These combinations performed well before. We used two gestures to see if gesture affected input; we did not use more gestures or intervals to keep the design simple. There were five conditions, one for each type of feedback described above: All, All-Short, Split, Split-Short and Split-Swap. Participants completed a block of sixteen tasks per condition, eight using each gesture. Condition order was counter balanced with a Latin square.

We measured gesture completion (*Success*) as before. We also measured the time taken to find where to gesture (*Time-Locate*) and time to match a gesture (*Time-Match*). *Time-Locate* started

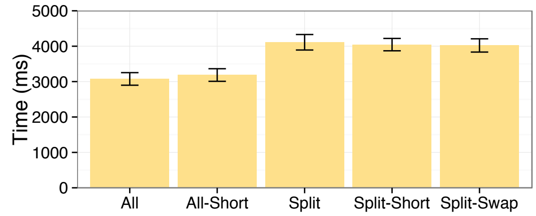

Figure 9. Mean *Time-Locate* for each feedback type.

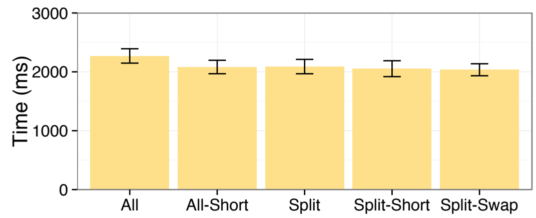

Figure 10. Mean *Time-Match* for each feedback type.

when a task began and ended once users completed three gesture movements; this was when the -Short and -Swap feedback stopped giving feedback about hand position. This indicates how long users spent finding where to gesture before trying to match the gesture movement. *Time-Match* differed from Experiment 2; it started when *Time-Locate* ended and stopped when the gesture rhythm was matched.

For each task, we also measured distance between target points and the mean hand position during a gesture, starting when first matched (*Distance*). This was calculated as in Experiment 1, using information provided by the Kinect sensor. Finally, we asked participants to rate the difficulty of finding where to gesture (*Difficulty-Locate*) and of matching a gesture (*Difficulty-Match*), using a ten-point scale as before. Twenty people took part in this study (five female) and were paid £6. Their mean age was 26.7 years (sd 3.6 years).

Results

Participants successfully matched 1598 of 1600 gestures (99.88%); both failed trials were the Side-to-Side gesture.

Mean *Time-Locate* was 3686ms (sd 1577ms), which includes at least three gesture movements (2100ms; three 700ms intervals). Figure 9 shows mean times for each **Feedback**. Times were not normally-distributed so were transformed using the Aligned-Rank Transform [50]. A repeated-measures ANOVA found that **Feedback** had a significant effect on time: $F(4, 171) = 40.69$, $p < 0.001$. **Gesture** did not have a significant effect ($F(1, 171) = 0.87$, $p = 0.35$), nor did **Feedback x Gesture** ($F(4, 171) = 0.22$, $p = 0.93$). *Post hoc* t-tests for **Feedback** found significantly lower times for All and All-Short than the other designs: all $t \geq 7.38$, $p < 0.001$.

Mean *Time-Match* was 2106ms (sd 984ms), see Figure 10. A repeated-measures ANOVA found that **Feedback** had a significant effect on *Time-Match*: $F(4, 171) = 2.86$, $p = 0.02$. **Gesture** did not ($F(1, 171) = 0.03$, $p = 0.86$), nor did the interaction ($F(4, 171) = 1.17$, $p = 0.33$). *Post hoc* t-tests for **Feedback** found one significant difference: *Time-Match* was higher for All than for Split-Swap ($t = 2.98$, $p = 0.03$).

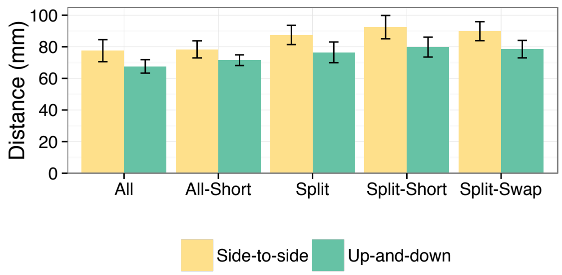

Figure 11. Mean *Distance* for each gesture and feedback.

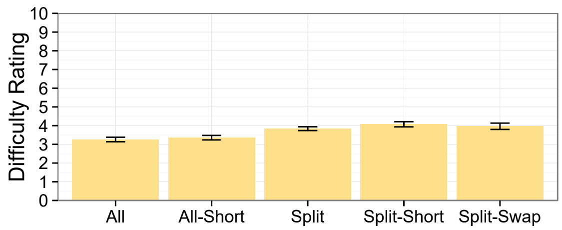

Figure 12. Mean *Difficulty-Locate* for each feedback type.

Accuracy

Mean *Distance* was 80mm (sd 34mm); see Figure 11. A repeated-measures ANOVA found that **Feedback** had a significant effect on *Distance*: $F(4, 171) = 9.17$, $p < 0.001$. **Gesture** also had a significant effect: $F(1, 171) = 30.32$, $p < 0.001$. The interaction between **Feedback** and **Gesture** did not have a significant effect: $F(4, 171) = 0.41$, $p = 0.80$.

Post hoc t-tests for **Feedback** found that participants were more accurate with All than with all three split designs (all $t \geq 3.93$, $p \leq 0.001$); they were also more accurate with All-Short than Split-Short and Split-Swap (both $t \geq 3.15$, $p \leq 0.02$). No other differences were significant. *Post hoc* t-tests for **Gesture** found that *Distance* was lower for Up-and-Down than for Side-to-Side: $t = 5.51$, $p < 0.001$.

Difficulty ratings

Mean *Difficulty-Locate* was 3.70 (sd 0.95); see Figure 12. All difficulty rating data were transformed using the Aligned-Rank Transform, as in Experiment 2. A repeated-measures ANOVA found that **Feedback** had a significant effect on *Difficulty-Locate*: $F(4, 171) = 34.85$, $p < 0.001$. **Gesture** had no significant effect: $F(1, 171) < 0.001$, $p = 0.98$. The interaction between these factors was not significant either: $F(4, 171) = 0.47$, $p = 0.76$. *Post hoc* t-tests for **Feedback** found that All and All-Short had significantly lower difficulty ratings than all other types of feedback: all $t \geq 5.99$, $p < 0.001$. No other differences were significant.

Mean *Difficulty-Match* was 3.29 (sd 0.78); see Figure 13. A repeated-measures ANOVA on the transformed difficulty ratings found that **Feedback** had a significant effect on ratings: $F(4, 171) = 56.56$, $p < 0.001$. **Gesture** had no significant effect: $F(1, 171) = 1.10$, $p = 0.30$. The interaction between these factors was not significant: $F(4, 171) = 0.09$, $p = 0.98$. *Post hoc* t-tests for **Feedback** found that difficulty ratings were lower for All and All-Short than for all other types of feedback (all $t \geq 3.91$, $p < 0.001$) and ratings were lower for Split-Swap

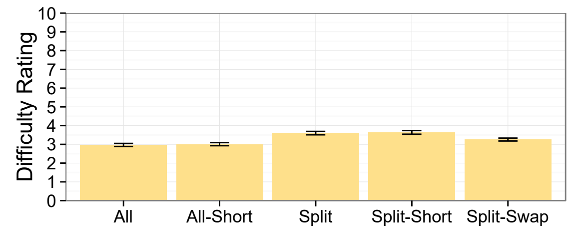

Figure 13. Mean *Difficulty-Match* for each feedback type.

than for Split and Split-Short (both $t \geq 5.95$, $p < 0.001$). No other comparisons were significant.

Comparison with Experiments 1 & 2

In Experiment 1, users searched for target points using precise finger movements, guided by multimodal feedback. Here, locating target points was only part of the task; users also had to perform a rhythmic gesture. We therefore expected users to spend less time finding "there", as that was not the focus of the experiment task. Time spent finding where to gesture was compared between experiments, using the Mann-Whitney U test. Time spent finding where to gesture was significantly lower in Experiment 3: $Z = 18.59$, $p < 0.001$ (mean time of 3686ms vs 7174ms). *Distance* was also compared using the Mann-Whitney U test. Users were significantly more accurate in Experiment 1: $Z = 9.5$, $p < 0.001$ (mean *Distance* of 51.4mm vs 80mm). However, this is unsurprising as smaller finger movements allowed greater precision than hand movements while also performing gestures.

We could not compare time taken to match rhythmic gestures (*Time-Match*) between Experiment 2 and 3 due to differences in how these were calculated (users may have begun gesturing while still searching for the target area in Experiment 3).

Discussion

Five feedback designs were compared in this experiment. Of these, All and All-Short performed the best: users gestured closer to "there", spent less time finding where to gesture, and gave lower difficulty ratings for finding where to gesture. These designs may have performed best because the visual feedback for "there" had an immediately noticeable effect on the rhythmic gesture animations. Users could see quickly if they were gesturing in a good location, or not. Experiment 1 found similar: interactive light feedback helped users initially position themselves. Unlike Experiment 1, however, there was no trade-off between speed and accuracy: All and All-Short led to greater accuracy.

Users did not locate target points as well as they did in Experiment 1; however, they also had to perform gestures and were making hand movements rather than smaller finger movements close to a device. Considering these factors, performance was very good in this experiment. Users also took less than six seconds to address the system (*Time-Locate* + *Time-Match*). These findings also show that our technique for helping users find where to gesture is effective for gesture systems where users are interacting close by (10cm over a device) and from across the room (2.5m from a device).

Feedback design had an effect on how long users spent finding where to gesture, but did not have much of an impact on rhythmic gesture performance. Difficulty ratings for matching gestures show that users gave lower ratings for All, All-Short and Split-Swap than they did for Split and Split-Short. This suggests that it was easier to match rhythmic gestures when given audio and tactile feedback about gesture movements. These findings support results from Experiment 2, where feedback did not impact performance but did lead to lower difficulty ratings. Based on this finding, we recommend always giving audio and tactile feedback about rhythmic gesture movements, even if these modalities are also used for other information (where to gesture, for example).

This experiment also suggests that feedback telling users where to gesture does not significantly affect rhythmic gesture performance. We had expected it to be distracting, although this does not appear to be the case; there were no differences in performance or difficulty ratings between All and All-Short, and between Split and Split-Short. A comparison of Experiments 2 and 3 shows increased difficulty ratings, which appears to contradict this finding. However, this increase may have been because participants also had to focus on *where* they were performing gestures. Despite this, the mean difficulty rating of 3.3 out of 10 suggests that participants did not find the interaction difficult. Based on these findings, we suggest that feedback guiding movements could be given continually while users address a gesture system, without significant detriment to usability. Users may still find this feedback helpful after they have addressed the interface, as it could help them continue gesturing in a good position.

OVERALL DISCUSSION

Experiments 1 and 3 found that users could find a good position for gesturing, using the feedback given by the gesture systems. These experiments compared two types of in-air gesturing—precise finger movements, close to a device, and less precise hand movements while gesturing further away from a device—finding the feedback effective for both. Our work in this paper focused on depth-sensing systems, which could use feedback to guide users to a good part of the field-of-view. An interesting area for future work would be to investigate how this, or similar, techniques could be used by other types of gesture sensor, like magnetic sensors [19].

Experiments 2 and 3 investigated rhythmic gestures, an interaction technique which reveals gestures to users ("do that") and allows them to direct their input when addressing gesture systems. Our findings show that users could perform rhythmic gestures successfully, even when locating where to gesture at the same time. Our research focused on the usability of this technique and provides an initial investigation of the rhythmic gesture design space. More work is needed to understand how accurately users can match a gesture rhythm, as this could inform the selection of gesture intervals.

Future work could also investigate the use of rhythmic gestures for continuous input. Experiment 3 found that users could maintain the rhythmic gestures for over nine seconds (with an upper limit of ten seconds) and difficulty ratings from Experiments 2 and 3 suggest this interaction was not too difficult. Other research [14, 30] has found rhythmic input to be an effective means of control in other input modalities and this may also be true for in-air gestures.

Design Recommendations

Use interactive light to show users where to gesture

In Experiment 3, users spent less time finding where to gesture and gestured in a better location when interactive light was used to show them where to gesture. Difficulty ratings further support the use of interactive light feedback helping users find a good position to gesture in.

Give feedback about rhythmic gestures from the start

Experiment 3 found that users rated interaction easier when given feedback about their movements from the moment they started addressing the gesture system, even if they were still finding where to gesture. Based on this finding, we recommend that audio and tactile feedback about gesture movements is always given from the start of an interaction.

Use Side-to-Side and Up-and-Down when possible

In Experiment 2, these movements were the easiest rhythmic gestures to use and also had the best performance. They were also used in Experiment 3, where they achieved high levels of performance and good difficulty ratings. These movements should be used ahead of the others investigated, if possible.

Use at least 700ms intervals, 900ms for circular gestures

Rhythmic gestures with 500ms intervals were more difficult to perform than those with slower intervals. Based on our findings, we recommend using an interval of at least 700ms as these gestures were performed well. Circular gestures were generally more difficult than the others and our results suggest they would be more usable with a minimum interval of 900ms.

CONCLUSIONS

Users must be able to address in-air gesture systems before they control them, a problem which requires finding where to gesture and how to direct input towards the intended system. Research has investigated these problems individually but existing solutions have limitations which would make them impractical in real use. We investigated a novel interaction technique, "do that, there", which uses multimodal feedback to help users find where to gesture ("there") and to show them how to direct their input ("do that"). We evaluated the effectiveness and usability of our technique in three experiments, using two small prototypes with limited visual displays.

We found that users were able to accurately (51mm–80mm) and quickly (3.7s) find "there" and achieved good performance doing "that" (93.2%–99.9%), even when combined. Our results suggest that "do that, there" could be successfully used for addressing gesture systems, as it guides movement well and users can accurately specify their intended system. Our multimodal output was also successful and we gave design recommendations for presenting feedback effectively.

ACKNOWLEDGEMENTS

This work was part-funded by Nokia Technologies, Finland. We thank the reviewers for their constructive feedback.

REFERENCES

1. Leap Motion Controller. https://www.leapmotion.com/
2. Precision Microdrives. http://www.precisionmicrodrives.com/
3. Adafruit. Adafruit NeoPixel LEDs. http://adafruit.com/products/1506
4. Alan Armstrong and Johann Issartel. 2014. Sensorimotor synchronization with audio-visual stimuli: limited multisensory integration. *Experimental brain research* 232, 11 (2014), 3453–63. DOI: http://dx.doi.org/10.1007/s00221-014-4031-9
5. Md Tanvir Islam Aumi, Sidhant Gupta, Mayank Goel, Eric Larson, and Shwetak Patel. 2013. DopLink: Using the Doppler Effect for Multi-Device Interaction. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '13*. ACM Press, 583–586. DOI: http://dx.doi.org/10.1145/2493432.2493515
6. Victoria Bellotti, Maribeth Back, W. Keith Edwards, Rebecca E. Grinter, Austin Henderson, and Cristina Lopes. 2002. Making Sense of Sensing Systems: Five Questions for Designers and Researchers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '02*. ACM Press, 415–422. DOI:http://dx.doi.org/10.1145/503376.503450
7. Peter Bennett, Kirsten Cater, and Mike Fraser. 2015. Resonant Bits: Harmonic Interaction with Virtual Pendulums. In *Proceedings of the 9th International Conference on Tangible, Embedded, and Embodied Interaction - TEI '15*. ACM Press, 49–52. DOI: http://dx.doi.org/10.1145/2677199.2680569
8. Matthias Budde, Matthias Berning, Christopher Baumgärtner, Florian Kinn, Timo Kopf, Sven Ochs, Frederik Reiche, Till Riedel, and Michael Beigl. 2013. Point & Control - Interaction in Smart Environments: You Only Click Twice. In *Adjunct Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing - UbiComp '13 Adjunct*. ACM Press, 303–306. DOI:http://dx.doi.org/10.1145/2494091.2494184
9. Alex Butler, Shahram Izadi, and Steve Hodges. 2008. SideSight: Multi-âĂIJtouchâĂİ Interaction Around Small Devices. In *Proceedings of the 21st Symposium on User Interface Software and Technology - UIST '08*. ACM Press, 201–204. DOI: http://dx.doi.org/10.1145/1449715.1449746
10. Thomas Carter, Sue Ann Seah, Benjamin Long, Bruce Drinkwater, and Sriram Subramanian. 2013. UltraHaptics: Multi-Point Mid-Air Haptic Feedback for Touch Surfaces. In *Proceedings of the 26th Symposium on User Interface Software and Technology - UIST '13*. ACM Press, 505–514. DOI: http://dx.doi.org/10.1145/2501988.2502018
11. Emiko Charbonneau, Charles E. Hughes, and Joseph J. Laviola Jr. 2010. Vibraudio Pose: An Investigation of Non-Visual Feedback Roles for Body Controlled Video Games. In *Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games - Sandbox 2010*. ACM Press, 79–84. DOI:http://dx.doi.org/10.1145/1836135.1836147
12. William Delamare, Celine Coutrix, and Laurence Nigay. 2012. Pointing in the Physical World for Light Source Selection. In *Proceedings of the Designing Interactive Lighting Workshop*. http://hal.archives-ouvertes.fr/hal-00757683/
13. William Delamare, Céline Coutrix, and Laurence Nigay. 2013. Mobile Pointing Task in the Physical World: Balancing Focus and Performance while Disambiguating. In *Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices & Services - Mobile HCI '13*. ACM Press, 89–98. DOI: http://dx.doi.org/10.1145/2493190.2493232
14. 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*. ACM Press, 289–298. DOI: http://dx.doi.org/10.1145/1518701.1518748
15. Euan Freeman, Stephen Brewster, and Vuokko Lantz. 2014. Tactile Feedback for Above-Device Gesture Interfaces: Adding Touch to Touchless Interactions. In *Proceedings of the 16th International Conference on Multimodal Interaction - ICMI '14*. ACM Press, 419–426. DOI:http://dx.doi.org/10.1145/2663204.2663280
16. Euan Freeman, Stephen Brewster, and Vuokko Lantz. 2015. Interactive Light Feedback: Illuminating Above-Device Gesture Interfaces. In *Proceedings of INTERACT '15 in LNCS 2929*. 478–481. DOI: http://dx.doi.org/10.1007/978-3-319-22723-8_42
17. William W. Gaver, Jacob Beaver, and Steve Benford. 2003. Ambiguity as a Resource for Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '03*. ACM Press, 233–240. DOI:http://dx.doi.org/10.1145/642651.642653
18. Ivan Golod, Felix Heidrich, Christian Möllering, and Martina Ziefle. 2013. Design Principles of Hand Gesture Interfaces for Microinteractions. In *Proceedings of the 6th International Conference on Designing Pleasurable Products and Interfaces - DPPI '13*. ACM Press, 11–20. DOI:http://dx.doi.org/10.1145/2513506.2513508
19. Chris Harrison and Scott E. Hudson. 2009. Abracadabra: Wireless, High-Precision, and Unpowered Finger Input for Very Small Mobile Devices. In *Proceedings of the 22nd Symposium on User Interface Software and Technology - UIST '09*. ACM Press, 121–124. DOI: http://dx.doi.org/10.1145/1622176.1622199
20. Khalad Hasan, David Ahlström, and Pourang Irani. 2013. AD-Binning: Leveraging Around Device Space for Storing, Browsing and Retrieving Mobile Device Content. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*. ACM Press, 899–908. DOI: http://dx.doi.org/10.1145/2466110.2466115

21. Khalad Hasan, David Ahlström, and Pourang Irani. 2015. SAMMI: A Spatially-Aware Multi-Mobile Interface for Analytic Map Navigation Tasks. *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '15* (2015), 36–45. DOI: http://dx.doi.org/10.1145/2785830.2785850

22. Michael J. Hove, Merle T. Fairhurst, Sonja A. Kotz, and Peter E. Keller. 2013. Synchronizing with auditory and visual rhythms: an fMRI assessment of modality differences and modality appropriateness. *NeuroImage* 67 (2013), 313–321. DOI: http://dx.doi.org/10.1016/j.neuroimage.2012.11.032

23. Michael J. Hove, Michael J. Spivey, and Carol L. Krumhansl. 2010. Compatibility of Motion Facilitates Visuomotor Synchronization. *Journal of experimental psychology. Human perception and performance* 36, 6 (2010), 1525–1534. DOI: http://dx.doi.org/10.1037/a0019059

24. Takayuki Iwamoto, Mari Tatezono, and Hiroyuki Shinoda. 2008. Non-contact method for producing tactile sensation using airborne ultrasound. In *Proceedings of EuroHaptics 2008*. Springer, 504–513. http://www.springerlink.com/index/X41J595757401387.pdf

25. Brett Jones, Rajinder Sodhi, David Forysth, Brian Bailey, and Giuliano Maciocci. 2012. Around Device Interaction for Multiscale Navigation. In *Proceedings of the 14th International Conference on Human Computer Interaction with Mobile Devices and Services - MobileHCI '12*. ACM Press, 83–92. http://dl.acm.org/citation.cfm?id=2371574.2371589

26. Rick Kjeldsen and Jacob Hartman. 2001. Design Issues for Vision-Based Computer Interaction Systems. In *Proceedings of the Workshop on Perceptive User Interfaces - PUI '01*. ACM Press, 1–8. DOI: http://dx.doi.org/10.1145/971478.971511

27. Sven Kratz and Michael Rohs. 2009. HoverFlow: Expanding the Design Space of Around-Device Interaction. In *Proceedings of the 11th International Conference on Human Computer Interaction with Mobile Devices and Services - MobileHCI '09*. ACM Press, Article 4. http://dl.acm.org/citation.cfm?id=1613864

28. David Ledo, Saul Greenberg, Nicolai Marquardt, and Sebastian Boring. 2015. Proxemic-Aware Controls: Designing Remote Controls for Ubiquitous Computing Ecologies. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '15*. ACM Press, 187–198. DOI: http://dx.doi.org/10.1145/2785830.2785871

29. Paul P. Maglio, Teenie Matlock, Christopher S. Campbell, Shumin Zhai, and Barton A. Smith. 2000. Gaze and Speech in Attentive User Interfaces. In *Proceedings of the International Conference on Multimodal Interaction in LNCS 1948 - ICMI 2000*. Springer, 1–7. DOI: http://dx.doi.org/10.1007/3-540-40063-X

30. Sylvain Malacria, Eric Lecolinet, and Yves Guiard. 2010. Clutch-Free Panning and Integrated Pan-Zoom Control on Touch-Sensitive Surfaces: The CycloStar Approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '10*. ACM Press, 2615–2624. DOI: http://dx.doi.org/10.1145/1753326.1753724

31. Nicolai Marquardt, Till Ballendat, Sebastian Boring, Saul Greenberg, and Ken Hinckley. 2012. Gradual Engagement: Facilitating Information Exchange between Digital Devices as a Function of Proximity. In *Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces - ITS '12*. ACM Press, 31–40. DOI: http://dx.doi.org/10.1145/2396636.2396642

32. Microsoft. Kinect for Xbox 360. http://www.xbox.com/en-US/xbox-360/accessories/kinect

33. Cecily Morrison, Neil Smyth, Robert Corish, Kenton O'Hara, and Abigail Sellen. 2014. Collaborating with Computer Vision Systems: An Exploration of Audio Feedback. In *Proceedings of the 2014 Conference on Designing Interactive Systems - DIS '14*. ACM Press, 229–238. DOI: http://dx.doi.org/10.1145/2598510.2598519

34. Sarah Morrison-Smith and Jaime Ruiz. 2014. Using Audio Cues to Support Motion Gesture Interaction on Mobile Devices. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*. ACM Press, 1621–1626. DOI: http://dx.doi.org/10.1145/2559206.2581236

35. Heiko Müller, Andreas Löcken, Wilko Heuten, and Susanne Boll. 2014. Sparkle: An Ambient Light Display for Dynamic Off-Screen Points of Interest. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction - NordiCHI '14*. ACM Press, 51–60. DOI: http://dx.doi.org/10.1145/2639189.2639205

36. Nest. 2012. The Nest Thermostat. (2012). http://nest.com/thermostat

37. 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 (2014), 70–77. DOI: http://dx.doi.org/10.1145/2541883.2541899

38. Qian Qin, Michael Rohs, and Sven Kratz. 2011. Dynamic Ambient Lighting for Mobile Devices. In *Adjunct Proceedings of the 24th Symposium on User Interface Software and Technology - UIST '11 Adjunct*. ACM Press, 51–52. DOI: http://dx.doi.org/10.1145/2046396.2046418

39. Bruno H. Repp. 2006. Rate Limits of Sensorimotor Synchronization. *Advances in Cognitive Psychology* 2, 2 (2006), 163–181. DOI: http://dx.doi.org/10.2478/v10053-008-0053-9

40. Dominik Schmidt, David Molyneaux, and Xiang Cao. 2012. PICOntrol: Using a Handheld Projector for Direct Control of Physical Devices through Visible Light. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology - UIST '12*. ACM Press, 379–388. DOI: http://dx.doi.org/10.1145/2380116.2380166

41. 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*. ACM Press, 3443–3452. DOI: http://dx.doi.org/10.1145/2556288.2556989

42. Jie Song, Gábor Sörös, Fabrizio Pece, Sean Ryan Fanello, Shahram Izadi, Cem Keskin, and Otmar Hilliges. 2014. In-air Gestures Around Unmodified Mobile Devices. In *Proceedings of the 27th Symposium on User Interface Software and Technology - UIST '14*. ACM Press, 319–329. DOI: http://dx.doi.org/10.1145/2642918.2647373

43. Srinath Sridhar, Anna Maria Feit, Christian Theobalt, and Antti Oulasvirta. 2015. Investigating the Dexterity of Multi-Finger Input for Mid-Air Text Entry. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '15*. ACM Press, to appear. DOI:http://dx.doi.org/10.1145/2702123.2702136

44. Breanna E. Studenka and Howard N. Zelaznik. 2011. Synchronization in repetitive smooth movement requires perceptible events. *Acta Psychologica* 136, 3 (2011), 432–441. DOI: http://dx.doi.org/10.1016/j.actpsy.2011.01.011

45. Breanna E. Studenka, Howard N. Zelaznik, and Ramesh Balasubramaniam. 2012. The distinction between tapping and circle drawing with and without tactile feedback: an examination of the sources of timing variance. *Quarterly Journal of Experimental Psychology* 65, 6 (2012), 1086–1100. DOI: http://dx.doi.org/10.1080/17470218.2011.640404

46. Colin Swindells, Kori M. Inkpen, John C. Dill, and Melanie Tory. 2002. That one there! Pointing to establish device identity. In *Proceedings of the 15th Annual Symposium on User Interface Software and Technology - UIST '02*. ACM Press, 151–160. DOI: http://dx.doi.org/10.1145/571985.572007

47. Robrecht P. R. D. van der Wel, Dagmar Sternad, and David A. Rosenbaum. 2009. Moving the Arm at Different Rates: Slow Movements are Avoided. *Journal of Motor Behavior* 26, 5 (jan 2009), 29–36. DOI: http://dx.doi.org/10.1080/00222890903267116

48. Jo Vermeulen, Kris Luyten, Karin Coninx, Nicolai Marquardt, and Jon Bird. 2015. Proxemic Flow: Dynamic Peripheral Floor Visualizations for Revealing and Mediating Large Surface Interactions. In *Proceedings of INTERACT '15 in LNCS 9299*. 264–281. DOI: http://dx.doi.org/10.1007/978-3-319-22723-8{_}22

49. 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*. ACM Press, 841–850. DOI: http://dx.doi.org/10.1145/2470654.2470774

50. Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only ANOVA Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '11*. ACM Press, 143–146. DOI: http://dx.doi.org/10.1145/1978942.1978963

51. Xing-dong Yang, Khalad Hasan, Neil Bruce, and Pourang Irani. 2013. Surround-See: Enabling Peripheral Vision on Smartphones during Active Use. In *Proceedings of the 26th Symposium on User Interface Software and Technology - UIST '13*. ACM Press, 291–300. DOI: http://dx.doi.org/10.1145/2501988.2502049