BEHAVIOUR & INFORMATION TECHNOLOGY, 2016
VOL. 35, NO. 12, 1044–1062
http://dx.doi.org/10.1080/0144929X.2016.1194477

Taylor & Francis
Taylor & Francis Group

Effects of auditory, haptic and visual feedback on performing gestures by gaze or by hand

Anne Köpsel[a], Päivi Majaranta[b], Poika Isokoski[b] and Anke Huckauf[a]

[a]General Psychology, Ulm University, Ulm, Germany; [b]School of Information Sciences, University of Tampere, Tampere, Finland

ABSTRACT

Modern interaction techniques like non-intrusive gestures provide means for interacting with distant displays and smart objects without touching them. We were interested in the effects of feedback modality (auditory, haptic or visual) and its combined effect with input modality on user performance and experience in such interactions. Therefore, we conducted two exploratory experiments where numbers were entered, either by gaze or hand, using gestures composed of four stroke elements (up, down, left and right). In Experiment 1, a simple feedback was given on each stroke during the motor action of gesturing: an audible click, a haptic tap or a visual flash. In Experiment 2, a semantic feedback was given on the final gesture: the executed number was spoken, coded by haptic taps or shown as text. With simultaneous simple feedback in Experiment 1, performance with hand input was slower but more accurate than with gaze input. With semantic feedback in Experiment 2, however, hand input was only slower. Effects of feedback modality were of minor importance; nevertheless, semantic haptic feedback in Experiment 2 showed to be useless at least without extensive training. Error patterns differed between both input modes, but again not dependent on feedback modality. Taken together, the results show that in designing gestural systems, choosing a feedback modality can be given a low priority; it can be chosen according to the task, context and user preferences.

ARTICLE HISTORY
Received 2 January 2015
Accepted 20 May 2016

KEYWORDS
Gaze input; hand gestures; feedback

Introduction

Mobile and pervasive computing environments are moving from laboratories to daily use. The scenario known as the Internet of Things where most tools and devices include computers that are all interconnected through the Internet is slowly becoming reality. In environments where many objects are computers, the need to interact with them often arises. Unlike physical switches and controllers, computerised devices with the same function can be also controlled remotely. Such situations require input methods that allow easy interaction with distant objects and displays without touching. Two solutions for remote control are already in widespread use. Depth cameras with models for tracking body movements such as the Microsoft's Kinect and Leap Motion input devices are widely deployed. Similarly, handheld devices such as mobile phones and tablet computers can be used as universal remote controllers. They are almost exclusively built with touch displays, which allow touch gestures to be used as an input method.

Other possibilities for remote control include gaze gestures (Dybdal, Agustin, and Hansen 2012), which are eye movement sequences that an eye-tracking device can recognise. Since we usually look at the things we operate, gaze provides a natural alternative for interaction. Since the price for the hardware dropped significantly recently (e.g. EyeTribe[1] and Tobii EyeX[2]) and remote calibration-free gaze control of smart screens is possible (Vidal et al. 2013; Zhang, Bulling, and Gellersen 2013), gaze input should be taken into consideration as a possible input modality.

Conventional use cases for gaze gestures include alternative control methods for people with extreme motor impairments such as the locked-in syndrome (e.g. in progressed amyotrophic lateral sclerosis or paraplegia). Also avoiding manual touch to prevent influenza and other infections is sometimes cited as a motivation for using gaze and other non-contact input techniques. Frequently used touch displays or keyboards can spread microbes from one user to the other. The gaze orientation is more difficult to observe than hand gestures and, therefore, useful for handling private information in public settings. In addition, widespread use of gaze input in smart glasses is a plausible future scenario.

Regardless of the input method, the feedback given to an executed gesture plays a crucial role. In touch interaction, such as using a keyboard, the user gets multimodal feedback on both the action and its (semantic) result:

CONTACT Anne Köpsel anne.koepsel@uni-ulm.de

tactile feeling of pressing the key and auditory click on the key activation, as well as the visual result shown on the screen. In remote operation of distant screens, all tactile feedback except for proprioception (i.e. the sense of the position of one's body) is missing. In accordance with good usability engineering practice, 'the system should continuously inform the user about what it is doing and how it is interpreting the user's input' (Nielsen 1993, 134). Hence, if a gesture is compounded of several related actions, it may be useful to inform the user about what is happening already during the gesturing process (Cassidy, Hook, and Baliga 2002; Kangas, Akkil, et al. 2014). However, sometimes observing the feedback may actually slow down interaction (Clawson et al. 2005). Besides, it may be too late to make any corrections if the feedback is given after a false input had already been made.

Pervasive and mobile environments introduce extra challenges for interaction. For example, visual feedback given on a specific spot on a large display may go unnoticed. Auditory feedback might support or even replace visual feedback (Zhao et al. 2007). If the feedback should be given privately, without others seeing or hearing it, haptic feedback could be used. That, of course, has the disadvantage of requiring a physical contact with the user and the haptic device (unless ultrasound or air pressure-based haptics are used; Iwamoto, Tatezono, and Shinoda 2008). Therefore, the question at issue is whether and to what extent the feedback modality affects performance and how this relates to the input modality.

Current research reports heterogeneous findings with respect to feedback given for gesture control. For example, visual feedback on the sequence of a composing process can aid users in performing gaze gestures (Porta and Turina 2008). On the other hand, continuous visual feedback on hand gestures may steal attention from the primary task (Witt, Lawo, and Drugge 2008). Whereas haptic feedback may interfere with other senses with no benefit of using it with hand gestures (Foehrenbach et al. 2009), some other studies (Kangas, Akkil, et al. 2014) report significant benefits for using haptic feedback with gaze gestures.

Regarding the potential effects of feedback modality on performance, one might consider psychological assumptions concerning effects of perception on action: respective approaches referred to as common coding (Prinz 1997; Prinz and Hommel 2002) maintain that action and perception are closely linked. For example, it is easier to point rightwards when sensory information also refers to the right (Prinz and Hommel 2002). In addition, sharing modalities between perception and action might be of special relevance. That is, concerning effects of feedback modality, one might suspect that the quality of a certain feedback modality depends on the currently used action mode (gaze or hand). However, whether these theoretically derived claims hold true for relatively complex realistic tasks is still unclear.

Our contribution

Previous research has mainly focused on a specific input method and a set of output modalities, but has neglected potential interactions between input and output modalities. To our knowledge, our research is the first that compares auditory, haptic and visual feedback modalities for both gaze gestures and freehand gestures. Furthermore, previous research reports heterogeneous findings on the – supporting or interfering – effects of feedback during gesturing. We explored both the effect of (semantic) feedback on the *final* gesture and the effect of (simple) feedback given *during* gesturing. The semantic feedback is representative of an action triggered by an executed gesture. We wanted to emulate the situation in which a successful gesture results in a change in the application state (e.g. initiates a command to stop/start an action). Thus, this could also be seen as no specific feedback, but executing the command the system has recognised. Experiment 2 let us study if the end result is enough as 'feedback' or if the execution in detail should be interlaced with information about what the system recognised (i.e. the simple feedback).

On the other hand, even if the feedback is given in the very end and will hence not interfere or support the action itself, perhaps using the same modality for both input and output still has an effect of some kind. As justified above, we assume that users prefer a different modality for feedback as the input mode: when using gaze as input, auditory or haptic feedback might be preferred; and for hand input, visual and auditory feedback might be preferred. On the contrary, using the same modality for input and feedback might lead to interferences impairing performance as well as satisfaction. For example, since visual feedback requires explicit visual attention and might easily be missed due to a blink or a saccade (sudden eye movement), perhaps other modalities are preferred for gaze input.

This interference between input and feedback modality should be more pronounced when feedback is used directly for motor commands in a certain modality, and less pronounced when it is used for higher level abstract processing. Hence, we expect larger effects of feedback modality with simple feedback relative to semantic feedback (apart from the time required for decoding the semantic feedback).

Since straight strokes can be performed quite fast (especially with gaze; Heikkilä and Räihä 2009), only simple, brief feedback can be given. After the whole gesture is finished, the effect of the action can be communicated in a more abstract, semantic form. Thus we will answer the question: What are the differences in effects of feedback on user performance and preference between stroke-by-stroke feedback concerning single motor commands and semantic feedback concerning the meaning of a gesture provided at the end of the gesture?

Our approach

To study these questions, we compared auditory, haptic and visual feedback for both hand and gaze gestures in two experiments. In Experiment 1, stroke-based simple feedback was given during the execution of the gesture, on each of the four strokes that a gesture consisted of. The feedback was an auditory 'click', a haptic 'tap' or a visual 'flash'. In Experiment 2, the feedback was provided after the execution of the full gesture, either by hand or by gaze. The result of the action was either spoken, coded with haptic taps or shown as written text. The task in both experiments was the same: to enter gestures that are formed of four straight horizontal or vertical strokes. To generalise the effects of feedback modalities for various input modes, we included gaze and hand gestures in each experiment.

Below, we first review related research on gesture made by hand or gaze, as well as studies on the effects of feedback. We then introduce the method that was used in both experiments. We first report the results of Experiment 1 on the effects of simple feedback given during gesturing, followed by the results of Experiment 2 on the effects of semantic feedback given at the end of the gesture. We then discuss the implications of both experiments and finish with conclusions and future work.

Previous work

Research on hand gestures in human–computer interaction is comprehensive and spans over 30 years. Already in 1980 Bolt (1980) developed a simple static hand gesture system to move ships on a screen. Nowadays, hand gestures are usually dynamic, which means they do not consist of static pointing only (Köpsel and Huckauf 2013; Villamor, Willis, and Wroblewski 2010; Wobbrock, Morris, and Wilson 2009). Gestures have been used for controlling smartphones, home electronics (Lenman, Bretzner, and Thuresson 2002; Shan 2010), factory automation (Heimonen et al. 2013), human–robot interaction (Alvarez-Santos et al. 2014), multi-touch surfaces (Wobbrock, Morris, and Wilson 2009) and many other electronic devices (Baudel and Beaudouin-Lafon 1993; Bhuiyan and Picking 2009; Garzotto and Valoriani 2012), even head-up displays (Saxen et al. 2012). Whereas earlier systems needed utilities such as gloves (Baudel and Beaudouin-Lafon 1993) or other tracking targets (Tsukadaa and Yasumura 2002; Zimmerman et al. 1987), these days no equipment attached to the body is necessary (Shan 2010).

Systems with gaze gestures are not yet widely used. However, there is a lot of research on gaze gestures demonstrating its potential for various scenarios and tasks. For example, gaze gestures can be used for text entry (Wobbrock et al. 2008), computer control (Porta and Turina 2008), gaming (Vickers, Istance, and Hyrskykari 2013) or drawing (Heikkilä 2013). People can also browse objects on a remote smart screen by gazing sideways (Zhang, Bulling, and Gellersen 2013), or select moving objects by matching their gaze movement with the movement of the object (Vidal et al. 2013). If only relative changes in the direction are analysed, gestures are independent of location (Drewes and Schmidt 2007) and can thus be used in mobile situations, for controlling a mobile phone or music player, or to play games while on the move (Bulling and Gellersen 2010; Bulling, Roggen, and Tröster 2008; Dybdal, Agustin, and Hansen 2012). Hence, there is a wide scope of potential applications that could benefit from hand or gaze gestures.

Gestures

We are interested in stroke gestures (Karam and Schraefel 2005) where a sequence of individual strokes is interpreted as a command. As defined by Istance et al. (2010), such gaze (or hand) gesture is

> a definable pattern of eye movements performed within a limited time period, which may or may not be constrained to a particular range or area, which can be identified in real time, and used to signify a particular command or intent. (Istance et al. 2010, 323)

Gestures can be absolute, for example, linked to a screen, where the gesture is performed by executing a sequence of strokes on absolute positions on screen, or relative, so that they can be performed anywhere in space. Gestures can also be classified into single stroke gestures or more complex gestures (Møllenbach, Hansen, and Lillholm 2013). Single stroke gestures are fast, but can be confused with normal gaze behaviour (e.g. quickly glancing the road from left to right); thus, often gestures consisting of several strokes are preferred. In addition to discrete, straight strokes, also continuous, curved movements can be used. In gaze interaction, stroke-based gestures seem to be more efficient than continuous 'gliding' based on smooth pursuit (Rozado et al. 2012).

Gestures can also provide a device independent input method of interaction (Isokoski and Raisamo 2000). To make it an effective method, the ease and intuitiveness of the gesture should be considered (Atia et al. 2009; Köpsel and Huckauf 2013; Wu and Wang 2013). Proper feedback can assist both in learning to perform gestures and in interpreting their effect (Zhai et al. 2012).

Feedback for different input/interaction modalities

Feedback plays an important role in informing the user how the system is interpreting the user's gaze (Majaranta et al. 2006), especially since the eyes are primarily a perceptual organ: if unintentional gazes are falsely interpreted as commands, the interaction may be impaired by the so-called Midas touch problem (Jacob 1991). Gaze gestures can be combined by a sequence of movements that have little change in appearing in nature. Thus, gaze gestures are not so easily impeded by the Midas touch problem (Drewes and Schmidt 2007). Kangas, Akkil, et al., 2014 found that intermediate haptic feedback improved the performance for gaze gestures. Here, the task was to browse a list and make selections on a (handheld) mobile phone by quick gaze gestures outside of the screen. Obviously, visual feedback was not given outside of the screen, so haptic feedback was given to confirm the off-screen stroke. Haptic feedback significantly improved performance and user experience. Since eye movements are very fast (Ware and Mikaelian 1987), one should pay attention to the timing of the feedback; if the feedback on the action is delayed, users may not be able to associate the feedback with the action (Kangas, Rantala, et al. 2014).

When freehand gestures are used for interaction, there is little natural feedback. Only visual perception of the body parts that happen to be within the present field of view and the proprioceptive feedback through muscles and joints are available. In general, gestures lack critical cues and requisite feedback (Norman 2010). There should be some notification of an immediate system response (Shneiderman 1998) as well as the success of the action (Wu and Wang 2013). Without feedback the user will not know if there is a system error (e.g. the tracking is not receiving data), or if the user's gesture was not understood, or if the user made a wrong gesture. Furthermore, if the system provides feedback on its tracking status and inaccuracies in detecting the gesture, users can use this information in improving their input. This, in return, can improve gesture recognition rates (Kratz and Ballagas 2009).

Vogel and Balakrishnan (2005) found that subtle auditory and visual feedback supported freehand interaction with a large, distant screen. Similar results were found by Chen et al. (2005) on their experiments on interaction with large screens via multimodal input by gestures and speech. They found that both audio and visual cues enhanced gesture input. Audio feedback was especially useful. Participants found visual feedback somewhat distracting, but had no issues with the audio feedback.

Kajastila and Lokki (2013) found that a circular auditory menu reproduced with headphones can be as good as or even slightly better than a visual menu in a task where items are selected from a virtual menu surrounding the user's head. The circular menu was controlled by smooth round gestures around a centre point, or by using a direct stroke from the centre towards the menu item (a sector in the circular menu). As the user browsed the menu, the focused item was spoken. The spatial sound coming from the corresponding direction helped the user to associate the sound with the specific item location in the menu.

Feedback is also useful in providing a reference point when gestures are performed into free air without specific visual targets to point at (Hinckley, Pausch, and Proffitt 1997). If two hands are used, one hand can be used to make the gesture and the other hand can provide the reference point (Gustafson 2012).

Lee, Li, and Starner (2011) found that tactile feedback given on the users' wrist not only improved performance in a condition where visual feedback was available, but the tactile feedback could also compensate for the lack of visual feedback (e.g. while walking). Users were able to mentally synchronise the freehand or mid-air gesture with on-body tactile feedback.

Even though proper feedback improves performance and is often crucial for successful interaction, it may also hinder interaction. For example, if the feedback steals attention from the task, it may impair performance. Witt, Lawo, and Drugge (2008) found that even though the users did not need the feedback for successful gesture interaction, they were not able to ignore the continuous feedback that was shown in the result; this prevented them from fully focusing on the primary task. Reducing visual feedback may speed up also manual gestures: 'with less on-going feedback to monitor, there is a tendency for the user to enter gestures more quickly' (MacKenzie and Castellucci 2012, 2585). However, even if extra feedback slows down performance, it may increase accuracy as users pay more attention to the careful execution of the task (Lécuyer et al. 2002). In addition, sometimes there are no statistically significant differences on performance for feedback, but the users' subjective experience may be significantly different (MacGregor and Thomas 2001). Finally, there may also be interference between different modalities. For example, Foehrenbach et al. (2009) found tactile feedback to increase error rates, implicating that 'tactile feedback can interfere with other senses' (341). Other factors that affect the usefulness of feedback include workload and task (Burke et al. 2006).

In summary, the previous work was fragmented. It was difficult to draw conclusions regarding the relative efficiency of different feedback modalities on gaze and hand gestures.

Method

The same experimental method was used in both experiments. Hence, we first introduce the method before going into the details of each experiment.

Gestures

Since we wanted to study feedback on both eye gestures and hand gestures, we decided to use dynamic gestures which consisted of single strokes (Karam and Schraefel 2005). Single straight strokes are easy to perform with both input methods.

The stroke gestures we used represented numbers from 1 to 12, following a paradigm based on the clock face and a pie menu (see Figure 1(a)). A number can be selected by making a stroke from the centre of the pie to the sector where the number is located. Similarly to the 'crossing borders' technique (Urbina and Huckauf 2010), we registered a stroke as soon as it crossed an invisible border surrounding the centre.

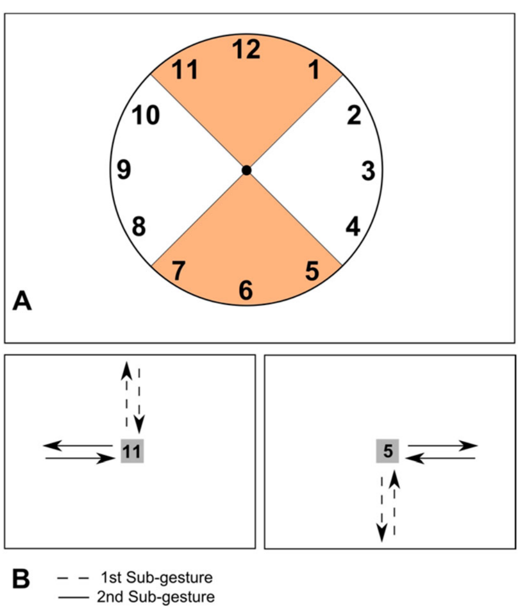

Figure 1. (a) Clock, used for the 12 gestures. The colour marking the four quarters used for the first sub-gesture. The clock was not presented during the experiment and was just used before to practise the gestures. (b) Two examples for a complete gesture, consisting of two sub-gestures, or more detailed four strokes.

Pie menus have been found to be quick and efficient in comparison with linear menus (Callahan et al. 1988). However, as the number of sectors increases, the accuracy decreases. Hence, we decided to adopt the 'hybrid' clock face paradigm (Isokoski and Käki 2002) where only horizontal and vertical stokes are used. Numbers located in the diagonal sectors are selected by a combination of two sub-gestures.

In our study, each gesture consisted of two sub-gestures. Each sub-gesture consisted of two strokes. Thus, there were always four strokes in each top-level gesture. With the first sub-gesture, one had to choose one quadrant (up, down, left or right); with the second sub-gesture, the number inside that quadrant was defined, also by its spatial position (see Figure 1(a)). Each gesture as well as each sub-gesture started from the centre. Therefore, each sub-gesture consisted of a movement in the intended direction and back to the centre.

For example, in order to enter number '5' (see Figure 1 (b)), the participant had to first move, either by gaze or hand, from the centre downwards and back to the centre to choose the correct quadrant consisting of numbers 5, 6 and 7. Thus, a sub-gesture ended and the next sub-gesture will be recognised as one to choose one digit in the specific quadrant, which is again spatially coded. Hence, to finally select '5', located on the right side of the lower quadrant, the participant had to move from the centre to the right, and back. Similarly, to select '12', for example, a sub-gesture of 'move up and back' had to be performed twice.

The clock was shown to the participants only during instruction. That is, the clock visualisation given in Figure 1(a) was used to explain the gestures to the participants and to facilitate their recall. It was not presented during the experiment.

Apparatus

The stimuli and the visual feedback were presented on a 27″ (68.6 cm) Acer T272HLbmidz screen (with 1680 × 1050 pixel resolution). Programs for running the experiment were written in Visual Studio 2012 using C#.

SMI RED eye tracker (120 Hz) with iViewX and the COGAIN Eye-Tracking Universal Driver was used for tracking the gaze binocularly. For hand gestures, a Microsoft Kinect (first generation) with 'Kinect for Windows SDK Version 1.8' and the Developer Toolkit was used.

Procedure

Participants were informed about the purpose of the experiment and filled in an informed consent form.

After instruction, the participants were informed about the experimental task as explained above using an illustration similar to Figure 1(a), then they practised with the Kinect and the eye tracker using a screen where the quadrants' colours changed (from black to light blue) when they moved over it either by hand or by gaze.

Each participant completed six experimental blocks: 3 feedback types (auditory, haptic and visual) × 2 types of input (gaze versus hand gestures). The block order was counterbalanced over participants. Within each block, all 12 numbers were presented 4 times permutated in 4 closed sets, resulting in 48 trials. Numbers were presented so that the same number would not appear multiple times in a row. All 288 trials for the whole experiment lasted about 60 minutes per participant.

Each trial started with a presentation of the number in the centre of the screen. The participants' task was to perform the gesture corresponding to the number. As soon as the gesture started, the number on the screen disappeared. After finishing the first sub-gesture, the participants had 2 seconds to finish the gesture. When having completed the trial, the next target was presented.

After 12 trials, an orange circle was displayed, encouraging the participant to take a short break. To continue with the next set of 12 trials, the participants had to dwell on the circle by hand or gaze for 3 seconds.

While testing with freehand gestures, a hand-shaped cursor was shown in order to give feedback about the relative position of the hand. This was necessary since the Kinect produces some spatial inaccuracy which an informed user can easily compensate.

We decided not to display a cursor in the gaze condition because a visible cursor can disturb gaze interaction (Jacob 1991). In addition, the targets were presented at the centre so that there was no need for a cursor, because the next gaze gesture naturally started from the correct position in the centre. (With hand-pointing, the participant could look at the stimuli while the hand could point at something else.) Moreover, also to facilitate easy return to the centre of the screen, a grey square was also shown in the centre of the display – exactly on the same place where the targets were shown. This central square appeared as soon as the participant started the first stroke by crossing one of the invisible borders. After the gesture was finished, there was a small pause (1 second) and then the next stimuli replaced the central square. This central grey square was shown during the execution of the gesture in all conditions.

When testing with the eye tracker, the participants sat in front of the screen at a viewing distance of about 60 cm. The SMI eye tracker was mounted below the screen. At the beginning of each block consisting of gaze gestures, the eye tracker was calibrated using a 9-point automatic calibration followed by a validation.

For freehand input, participants stood at 1.5 m distance in front of the screen. The Kinect was mounted below the screen (see Figure 2). For Kinect, we verified at the beginning of each block that the system was correctly following the participants' hand movement.

The aim was to investigate the effects of feedback modalities in common conditions of various input modes. Therefore, the experiment was performed using both hand and gaze gestures. For both input modes, we used typical conditions (e.g. standing position with hand input versus sitting position with gaze input). Hence, one has to carefully bear in mind that absolute performance with hand and gaze gestures can be compared only if these different set-ups are real alternatives in a planned design. In most cases, this will not be the case because in different physical set-ups the results may be different regarding the hand–gaze comparison. We believe that within each input modality, the comparison of the feedback modalities is more generalisable.

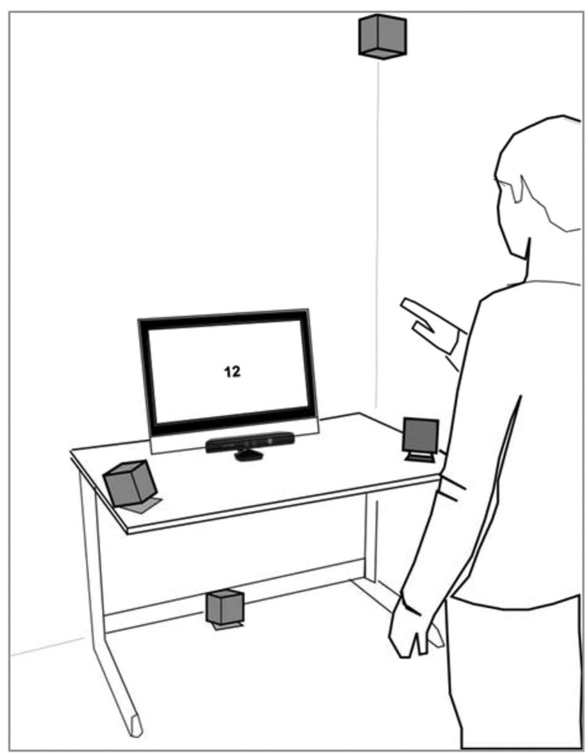

Figure 2. Setup for hand gesture input. The participant is standing 1.5 m in front of the desk while executing the gestures. Auditory feedback was given via four speakers (up, down, left and right). When giving haptic feedback, a small table with a pillow was placed left of the participant and at a height to easily place the hand on it while staying. Gestures were tracked using a Microsoft Kinect for Windows. For gaze input, the participant sat in front of the screen and the Kinect was replaced by the eye tracker.

After having performed the gesture task, the participants answered a questionnaire about their experience. For the different input modalities, they were asked how helpful they found each of the different feedback modalities to be. Also, they were asked to evaluate the feedback according to reliability, comprehensiveness, preference and dominance. In pilot tests, we tried to adjust the parameters of each feedback mode (such as volume level, amplitude of the haptic tap and tone of the colour) so that these different types of feedback modalities would be experienced as equal as possible. Since this is subjective, we wanted to ask the participants to rate if they felt any of the feedback modes was more intensely experienced than the others and thus 'dominant'.

Stimuli

Numbers 1–12 were used as targets. The targets were always presented as a digit(s) in the centre of the screen (Microsoft Sans Serif, Bold, 16pt).

Feedback

We conducted two experiments with different kinds of feedback. In Experiment 1, the feedback signalled the direction of each stroke; that is, it was given during gesturing. This is what we refer to as 'simple' feedback. In Experiment 2, the feedback informed about the recognised number and, therefore, was given after the gesture was completed. This we call the 'semantic' feedback. In Experiment 2, we wanted to emulate the situation in which a successful gesture results in a change in the application state (e.g. initiates a command to stop/start an action). For simplicity, we chose the semantic form of number as it also gives feedback of the success of the command (i.e. if the gesture was correctly recognised).

Experiment 1: Simple feedback

Participants

Twelve participants took part in Experiment 1 in exchange of course credit. All participants were students or employees of the Ulm University, Germany. Seven of them were female. Ages ranged between 21 and 44 years, with an average of 29 years. All of them were familiar with multi-touch gestures on handheld surfaces such as smartphones. Two of them had a technical background, whereas the rest were psychology students. One of the participants had experience with the Kinect; all of them had experience as participants in eye tracker experiments, but not in using gaze gestures.

Feedback design

Based on previous research (Kangas, Akkil, et al. 2014), we gave feedback after each stroke. That is, the feedback specified which of the possible strokes (up, down, left or right) – or a stroke back to the original centre – was recognised. Since each gesture consisted of a combination of four strokes (see Figure 1(b)), the user received four feedback signals for one complete gesture. Every time a stroke was made, we gave feedback to the stroke, using one of the three modalities described below.

Of course, the design of the feedback features is crucial to all results. Therefore, we extensively piloted respective stimuli and carefully checked for various potential side effects. In general, we used the weakest signal as baseline and tried to make stimuli in the other modalities as comparable in terms of recognition frequency as possible.

The weakest signal in the study was haptic vibrations. By pilot testing with three experts and student raters who all worked in our laboratory, we arrived at the feedback settings described below.

Auditory feedback

During the auditory block, the feedback was realised with four desk speakers (see Figure 2). Their spatial positions represented the stroke directions. That is, one was placed to the left of the participant and one to the right. One was situated below the table and one was mounted on the wall and placed above the participant's head. The sound was a standard wave file producing a 55 ms 'tap' sound. The loudness and duration of the auditory feedback were set so that the auditory feedback would be matching the subjective intensity of the other feedback modalities.

Haptic feedback

The haptic feedback was given via four mini-speakers (Tectonic HIHX14C02-8 compact audio exciters, with linear movement powered with a Gigaport amplifier to produce a strong enough effect), one for each direction. The mini-speaker acted as a haptic 'actuator' that produced a short haptic 'tap' when a standard wave file was played through it. The effect (heard or felt) is different depending on the sound file. We used a single sine wave (90 Hz, duration 55 ms), which was felt as a single 'tap' (instead of rapid vibration often used in cell phones). All four speakers could be controlled separately. It was possible to also execute them all at the same time, but during pilot tests we found that it is hard to perceive simultaneous taps (e.g. exactly how many actuators were used at a time). There is some latency for the rise time of the haptic actuator. However, with other sources for

Figure 3. (a) Participant in front of the screen with the eye tracker below and the hand on the pillow for the haptic feedback. (b) Pillow with four speakers providing haptic feedback on the participant's hand. (c) Screen with the four possible positions to present the visual feedback depending on the stroke made by the participant.

latencies such as eye tracking sampling rate, our total system delay remained below 150 ms, which is similar to Kangas, Akkil, et al., 2014.

The participants had to place their left hand on a pillow where the speakers were attached (see Figure 3(a), the pillow is shown detailed in Figure 3(b)). The pillow was self-made of foam plastic, covered with a piece of cotton fabric. The pillow prevented any vibrating sounds from the speakers. The relative position and angle of the speakers were tested during pilot tests, to make sure people would understand feedback resembling up, down, left and right. The pillow was placed on a tall chair so that the left hand could comfortably rest there. For each recognised gesture part (up, down, left and right), feedback was given on the corresponding speaker, as well as on the stroke returning back to the middle. So for each number from 1 to 12, feedback was given four times. Before the experiment started, we assured that each participant could discriminate the taps produced by the haptic actuator (the mini-speaker).

Visual feedback

Visual feedback was displayed on the screen by four small squares (30×30 pixels, approximately 10×10 mm) left, right, above and below the centre, 175 pixels (approximately 6 cm) away from the centre (see Figure 3 (c)). The square corresponding to the movement direction was displayed as soon as a stroke was recognised. The squares coloured light grey (#D3D3D3) were presented for 100 ms. A longer duration – relative to haptic and audio feedback – had to be used because visual stimuli of 55 ms were easily missed during a movement. Subjective experience on the strength and length also supported longer visual stimuli, in comparison to auditory and haptic stimuli.

Results

We measured the number of correct gestures as well as their durations. The gestures that were not executed correctly were classified into four categories:

1. Any unintended gesture was counted as a *wrong* gesture (e.g. a gesture for '5' instead of '4').
2. A *not-recognised gesture* was a gesture composed of two correct sub-gestures but in not existing combination (e.g. up-down + down-up).
3. A *timeout* was a gesture that lasted longer than the two-second time limit after the first sub-gesture.
4. *Off-crossings* represent spatial inaccuracies: each sub-gesture requires two strokes, one going out (left, right, up or down) and another coming back to the centre. In order to detect cases when a movement forth does not return back to the centre but goes directly towards the next goal, the screen was virtually divided into a 3×3 matrix, and crossing a border other than the one towards the centre was counted as an off-crossing.

Correct gestures as well as their durations were analysed using 3 (feedback: auditory, haptic, visual) $\times$ 2 (input: gaze, hand) repeated-measures analyses of variance. When necessary, violations of the sphericity assumption were corrected for with the Greenhouse–Geisser ε (Winer 1971). The data for correctly executed gestures were arcsine-transformed to make the distributions approximately normal.

Arcsine-transformed error rates were analysed using a $4 \times 3 \times 2$ (feedback: auditory, haptic, visual) $\times$ 4 (error: wrong gesture, timeout, off-crossing, not-recognised) $\times$ 2 (input: gaze, hand) repeated-measures analysis of variance (ANOVA). The correctly executed gestures were not considered in the error type analysis, but they were counted in the calculation of proportions to be sure the weighting of all errors reflects the ratio to all trials. The reader should keep in mind that, of course, the comparison between gaze and hand input is scarcely possible because of various differences between them. Our focus was on the effects of feedback modes.

Accuracy and task time

Overall, 78.21% (SE = 0.7%) gestures were executed correctly. Hand gestures (85.03%; SE = 0.86%) were

Table 1. Accuracy in percentage of correct responses and task time in milliseconds separately for gaze and hand input, presented separately for feedback modalities.

	Gaze			Hand		
	Auditory	Haptic	Visual	Auditory	Haptic	Visual
Accuracy in %	73.09	65.63	75.52	86.63	86.74	81.71
(SE)	(1.85)	(1.98)	(1.79)	(1.42)	(1.42)	(1.62)
Task time in ms	2812	2793	3253	4498	4691	4556
(SE)	(296)	(54)	(166)	(42)	(38)	(39)

more often correct than gaze gestures (71.41%; SE = 1.09%; $F[1,11] = 13.113$; $\eta_p^2 = .54$; $p < .01$).

In general, all feedback types produced comparable performance (auditory feedback: $M = 79.86\%$, SE = 1.18%; haptic feedback: $M = 76.15\%$, SE = 1.26%; visual feedback: $M = 78.61\%$, SE = 1.21%; $p < .56$).

The interaction between feedback modality and input modality was marginally significant ($F[2,22] = 3.376$; $\eta_p^2 = .24$; $p < .054$). Table 1 suggests that this marginal effect was produced by the haptic feedback for gaze input, where the haptic feedback for gaze led to a lower percentage of correct gestures without a similar dip in correct gestures for hand gesturing.

The percentage of correct gestures executed separately for each number from 1 to 12 can be seen in Figure 4. As can be seen, this detailed analysis indicates that the larger portion of errors with gaze input occurs mainly for diagonally placed numbers. Common to all these numbers is that they have to be gestured via two different movement directions.

It should be noted that the task time calculation starts from the moment when the stimuli are shown and ends when the last stroke returns to the centre. Thus, the task time includes the time required to perceive the stimulus and the time for planning the gesture. As expected, gaze was (2961 ms; SE = 118 ms) significantly faster than hand input ($M = 4582$ ms; SE = 22 ms; $F[1,11] = 105.20$; $\eta_p^2 = .91$; $p < .001$). As in correct responses, other effects, such as feedback modality, and the interaction between feedback modality and input (see Table 1), were not significant (all p-values > .05).

Results for the different error types

Of all errors, independent of the input, 6.7% (SE = 0.9%) could not be classified (not-recognised); 19.4% (SE = 3.4%) errors were made by the participants (wrong gesture); 34.39% (SE = 4.8%) of these erroneous gestures were timeouts, and 38.61% (SE = 5.8%) of all errors were consecutive off-crossings. The differences between the four error types were statistically significant ($F[3,33] = 13.487$; $\eta_p^2 = .55$; $p < .001$). Also, the amount of errors made for the different input modalities differs significantly ($F[1,11] = 18.081$; $\eta_p^2 = .62$; $p < .01$), as already shown in correct gestures.

For gaze input, errors mainly consisted of consecutive off-crossings, whereas for hand input, most of the errors were timeouts. The interaction effect for input modality and error type was significant ($F[3,33] = 50.958$; $\eta_p^2 = .82$; $p < .001$). The pairwise comparisons for differences between error types can be seen in Figure 5 (*post hoc* one-way ANOVA; pairwise comparison adjustment: Bonferroni). There were significant differences for each input modality and for each error type ($p < .05$ for not-recognised, $p < .001$ for all other cases). No other statistically significant effects were found (all p-values > .05).

User experience

After each block, the participants answered a questionnaire. They had to decide how helpful (1 = not at all helpful, 7 = very helpful) they found the presented feedback in the preceding block to be. The results, separated for

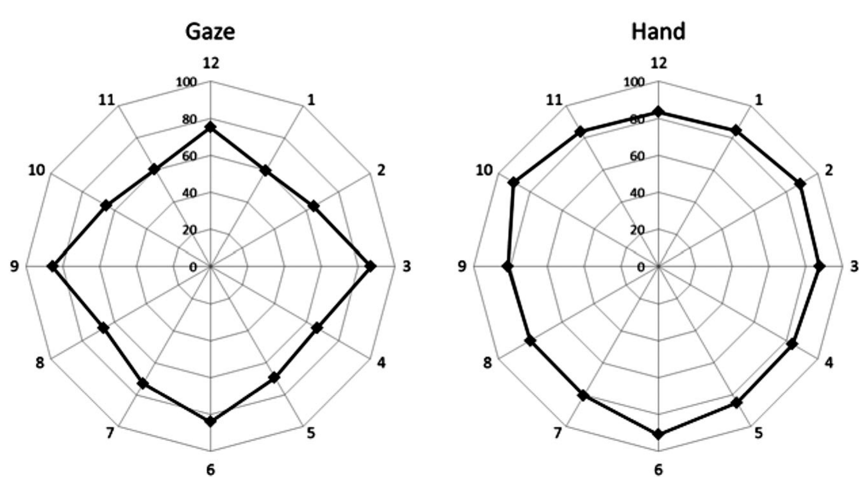

Figure 4. Correctly executed gestures for gaze (left) and hand (right) input presented for each number from 1 to 12 for the simple feedback.

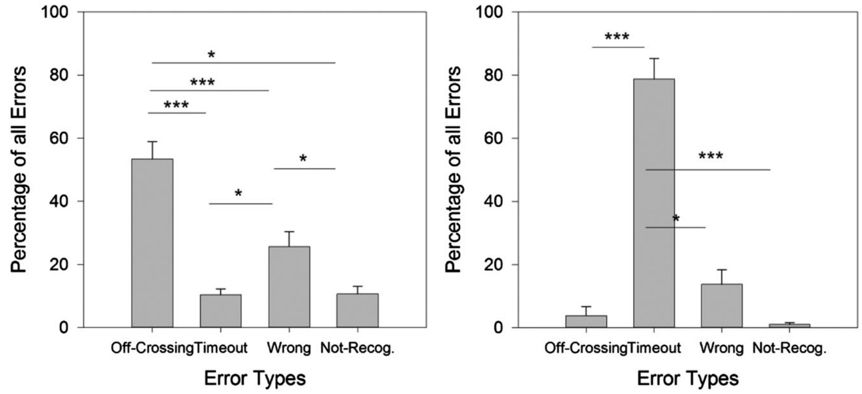

Figure 5. Error types in gaze (left) and hand (right) control. Error bars correspond to ±1 error of the mean. The stars mark the significant pairwise comparisons. The number of stars corresponds to the degree of significance (* < .05; ** < .01; *** < .001). Multiple comparisons adjustment: Bonferroni.

gaze and hand input, can be seen in Figure 6. Feedback modality did not affect the ratings, although visual feedback was rated somewhat less helpful than other modalities with both input methods.

After finishing the experiment, the participants were asked to rank the different feedback modalities regarding the following dimensions: the participant's personal preference, how dominant the different feedback modalities felt, how easy the feedback was to comprehend and, finally, how reliable the feedback appeared.

These results are presented in Figure 7. In general, the participants' opinions varied a lot. Auditory and haptic feedback were preferred somewhat more than visual feedback, but there was no winner in reliability and comprehensibility. There were no significant differences between the helpfulness of the different feedback modalities ($p > .05$). However, the difference between the two input modalities in the participants' subjective evaluation for helpfulness was significant ($F[1, 10] = 6.106$; $\eta_p^2 = .38$; $p < .05$). The interaction between feedback and input was not significant ($p > .05$).

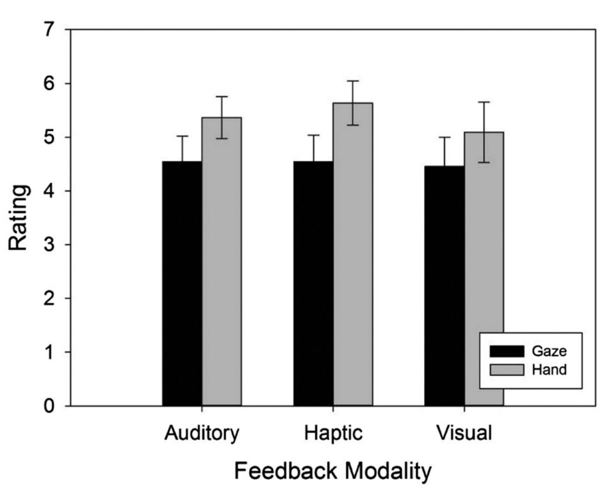

Figure 6. Users rating of the helpfulness of a certain feedback, separately for gaze and hand input. Error bars correspond to ±1 error of the mean.

Experiment 2: Semantic feedback

In Experiment 2, we wanted to emulate the situation in which a successful gesture results in a change in the application state (e.g. initiates a command to stop/start an action). For simplicity, we chose to use numbers as the format of feedback. The numbers carried natural semantics in the task and they also gave feedback of the success of the command (i.e. if the gesture was correctly recognised).

Participants

Twelve participants took part in Experiment 2 in exchange of course credits. All participants were students of the Ulm University, Germany. Seven of them were female. Age ranged between 19 and 34 years, with an

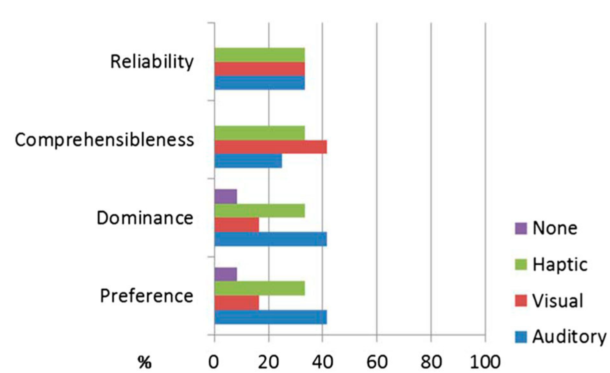

Figure 7. Preference, dominance, comprehensibleness and reliability for the simple feedback. 'None' was chosen by the participants, if they could not decide for one of the modalities.

average of 23 years. All of them were familiar with multi-touch gestures on handheld surfaces such as smartphones. All of them had experience as participants in eye tracker experiments, but not in gaze gestures.

Feedback design

In Experiment 2, feedback was given in a semantic format after the full gesture had been made; that is, each feedback informed about the recognised number.

Auditory feedback

In the auditory block, the feedback was given via one speaker placed directly in front of the participant. The feedback consisted of the spoken number word (e.g. 'drei'; German for 'three').

Haptic feedback

Since there is no common way to present the haptic feedback the same way as the visual or auditory feedback, the haptic feedback was more unfamiliar than the visual or the auditory. A fairly easy way to encode numeric feedback in haptics would be to provide a series of short pulses according to the number, but that would take a considerable amount of time and may still not be easy to perceive (Pakkanen et al. 2010). Given that we had four haptic actuators, a number of shorter encoding alternatives exist. We considered various options (such as location-based tactile taps and combined tap series) for semantic haptic feedback, but did not find any alternative that would be comparable with the clarity and speed of spoken or written numbers. Some options would also require long learning (such as Morse code or Braille-like coding). In order to minimise learning with the participants, we chose to encode the number according to the gestures. Thus, the feedback consisted of the representation of the gesture as two 55 ms pulses, with a 250 ms delay between them (see Figure 8(a)). The first pulse represented the first direction of the gesture and the second represented the second direction.

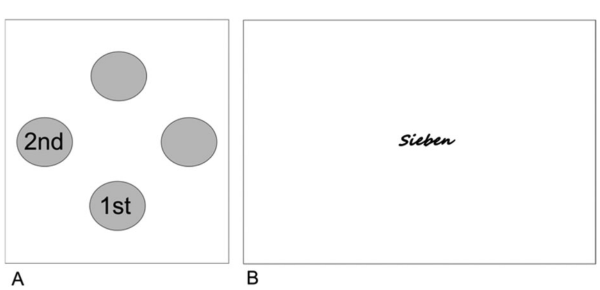

Figure 8. (a) The sub-gestures made for number '7' are given as feedback in the haptic modality. The graphics here represent the firing order of the actuators. (b) Screen with the semantic feedback, given after executing the relevant gesture correctly.

Visual feedback

For this feedback, the number was written on the centre of the screen in Word format, using cursive font style (Segoe Script, Bold, Italic, 21pt [visually about the same size as the stimuli], 1000 ms). We used visually different styles for representing the visual feedback on the number in order to clearly separate it from the stimuli. In addition, we tried to make sure none of the feedback modalities were more dominant than the other (both haptic and spoken auditory feedback take some time to produce; cursive writing of the number provides a better match than the digit format). An example of the semantic feedback can be seen in Figure 8(b).

Results

The analysis of the results was done the same way as described above for Experiment 1.

Accuracy and task time

Overall, 80.96% (SE = 0.67%) of the gestures were executed correctly, independent of feedback or input modality. For gaze, 80.99% (SE = 0.95%) of the gestures were executed correctly; for hand, 80.94% (SE = 0.95) of the gestures were executed correctly. This difference was not statistically significant ($p > .05$).

Also, the interaction between feedback modality and input modality was not statistically significant ($p < .15$). The results for the different feedback modalities for gaze and hand can be seen in Table 2. The percentage of correct gestures executed for each number from 1 to 12 can be seen in Figure 9.

Similar to the Experiment 1 the task time calculation starts from the appearance of the stimuli and ends when the last (fourth) stroke returns to the centre. Only the gestures executed correctly were included in the task time analysis: again, the difference between gaze and hand input is statistically significant (input: $F[1,11] =$

Table 2. Accuracy in percentage of correctly executed gestures and task time in milliseconds separately for gaze and hand input, presented separately for the different feedback modalities.

	Gaze			Hand		
	Auditory	Haptic	Visual	Auditory	Haptic	Visual
Accuracy in %	81.25	78.02	83.62	84.20	74.61	84.00
(SE)	(1.15)	(1.22)	(1.09)	(1.52)	(1.82)	(1.53)
Task time in ms	2959	3118	2790	4901	5633	4575
(SE)	(51)	(59)	(122)	(84)	(306)	(39)

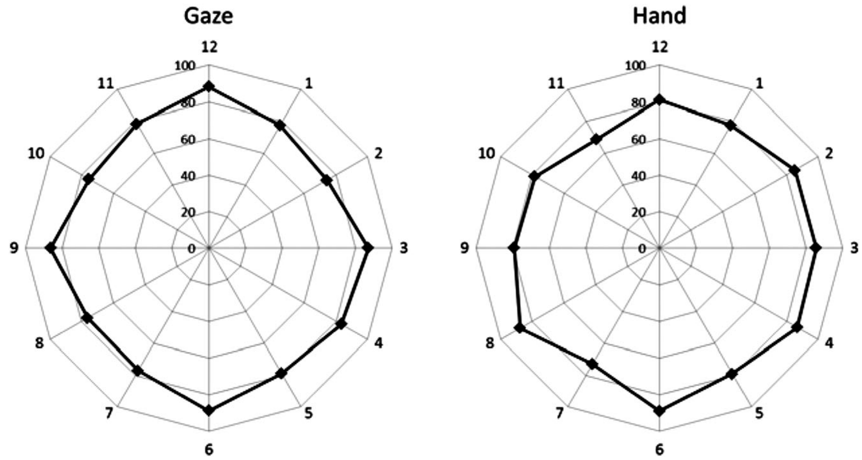

Figure 9. Correctly executed gestures for gaze (left) and hand (right) input presented for each number from 1 to 12 for the semantic feedback.

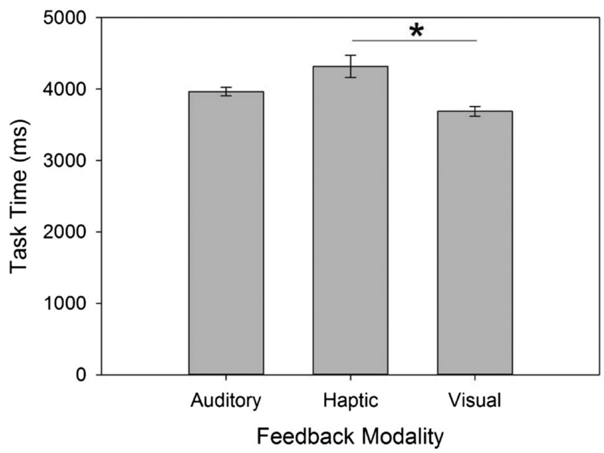

Figure 10. Mean time needed to execute a correct gesture for the three feedback types. The number of stars corresponds to the degree of significance (* < .05; ** < .01; *** < .001). Error bars correspond to ±1 error of the mean.

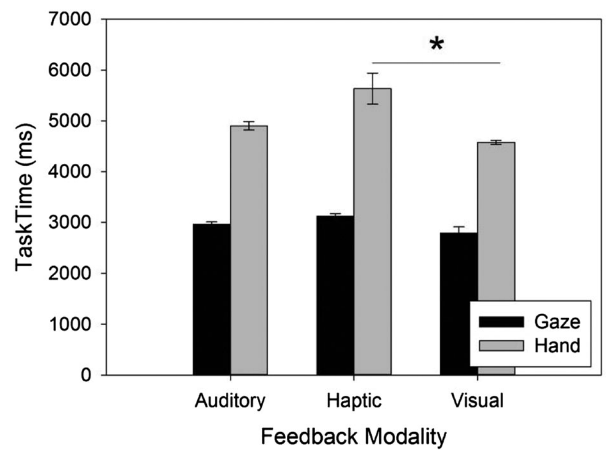

Figure 11. Mean time to execute a correct gesture for each feedback type separately for the two input types. The number of stars corresponds to the degree of significance (* < .05; ** < .01; *** < .001). Error bars correspond to ±1 error of the mean.

69.12; $\eta_p^2 = .86$; $p < .001$). For the feedback modalities, the main effect was also significant with $F(2,22) = 5.00$; $\eta_p^2 = .31$; $p < .05$. Pairwise comparisons showed a significant difference for haptic and visual feedback ($p < .05$) (Figure 10). The interaction between feedback modality and input modality was marginally significant ($F[1.336,14.697] = 3.89$; $\eta_p^2 = .26$; $p < .06$). For hand input here, there is a difference for the haptic and visual feedback ($p < .05$) (Figure 11).

Results for the different error types

Error patterns replicated the findings of Experiment 1: of all wrong gestures, independent of the input, 10.24% (SE = 1.37%) could not be classified (not-recognised). In all, 25.91% (SE = 3.27%) of the erroneous gestures were due to wrong gestures; 29.50% (SE = 3.41%) of these erroneous gestures were timeouts and 33.76% (SE = 2.36%) of all errors were consecutive off-crossings. There were statistically significant differences between error types ($F[3,33] = 9.138$; $\eta_p^2 = .45$; $p < .001$).

For gaze input, mainly consecutive off-crossings were made, whereas most of the errors made with hand input were timeouts (see Figure 12). The interaction effect for error type and input modality was also statistically significant with $F(3,33) = 40.052$; $\eta_p^2 = .76$; $p < .001$). The statistically significant differences (*post hoc* one-way ANOVA; pairwise comparison adjustment: Bonferroni) are shown in Figure 12. Between input modalities, the differences in off-crossings and timeout were statistically significant ($p < .001$). No other effects were statistically significant (all p-values > .05).

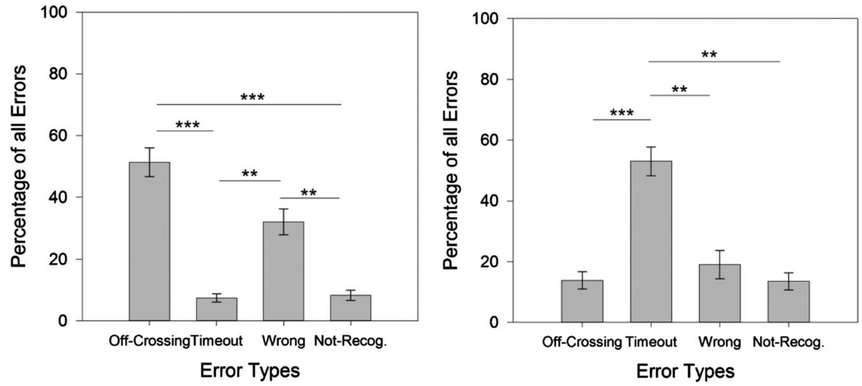

Figure 12. The different errors made for gaze (left) and hand (right) separately. Error bars correspond to ±1 error of the mean. The stars mark the significant pairwise comparisons. The number of stars corresponds to the degree of significance. Multiple comparisons adjustment: Bonferroni.

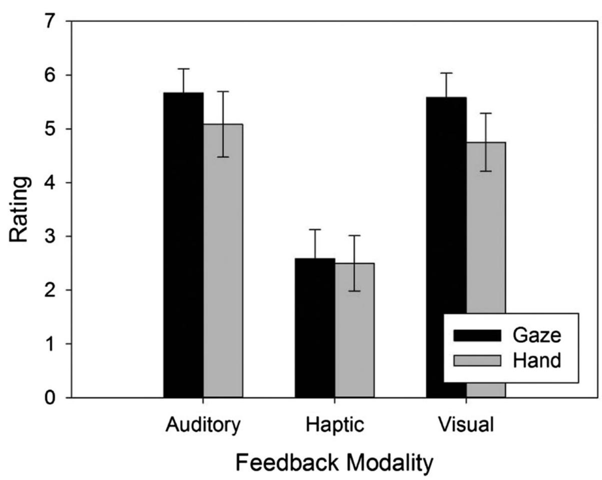

Figure 13. Users rating about the helpfulness of a certain feedback, separately for gaze and hand input. Error bars correspond to ±1 error of the mean.

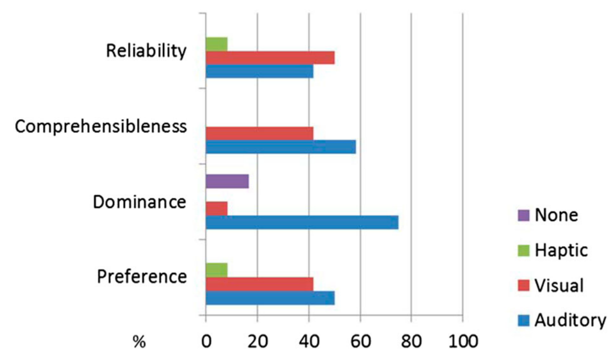

Figure 14. Preference, dominance, comprehensibleness and reliability for the semantic feedback. 'None' was chosen by the participants, if they could not decide for one of the modalities.

User experience

The participants answered how helpful they had found the presented feedback after each block to be. The results, separated for gaze and hand input, can be seen in Figure 13. After finishing the experiment, the participants also evaluated the different feedback modalities regarding preference, dominance, comprehensibleness and reliability. These results are presented in Figure 14. Different from Experiment 1, there was a clear effect of feedback modality in that auditory and visual feedback were rated to be more helpful, more comprehensible and more reliable, and were also preferred more by the participants than the haptic feedback. Auditory feedback that was given as a speech was found to be more dominant than the other modalities. This time the difference between the input modalities was not significant anymore ($p > .10$), but the differences for the feedback were ($F[2,22] = 17.484$; $\eta_p^2 = .61$; $p < .001$). The difference was for auditory and visual feedback ($p < .01$) and auditory and haptic feedback ($p < .01$). The interaction was as well not significant ($p > .48$).

General discussion

The kinds of errors varied systematically for the two input types in both experiments. The errors made by gaze were mostly off-crossings; hand input produced more timeout errors.

The biggest difference in the participants' experience was shown for the haptic feedback. Whereas it was evaluated as dominant, comprehensible and reliable by one-third of the participants in Experiment 1, it was none of these in Experiment 2. One reason might be that the haptic semantic feedback had to be decoded by the participants, whereas the fairly simple feedback was easy to understand in Experiment 1. In Experiment 2, the visual and the auditory feedback were given in the form of a common semantic meaning (shown or spoken number), whereas the two-part haptic feedback consisted of the two separate strokes executed by the participant.

In Experiment 1, all feedback modalities provided equally simple and short directional feedback that was reasonably easy to decode. In Experiment 2, the participants were familiar with the auditory (spoken) and visual (written) feedback modalities, but the haptic feedback (coded by haptic tabs) probably required more cognitive effort to decode.

Overall, the participants produced significantly more correct gestures in Experiment 1.

In relation to the correctly executed gestures, 17% in Experiment 1 and 10% in Experiment 2 were off-crossings made with gaze input.

Effects on feedback modalities

Our focus was on the effects of feedback modalities for each input type. With regard to our assumptions derived by common coding approaches of a special connection between feedback and input modality, there was no indication for a special suitability of a certain feedback modality for a certain input mode: neither did same modalities (i.e. visual feedback for gaze gestures or haptic feedback for hand gestures) produce especially good or especially bad performance. This was the case in Experiment 2 using abstract semantic feedback, as well as in Experiment 1 using stroke-by-stroke simple feedback. Of course, we cannot exclude that facilitating and impeding effects of identical modalities in feedback and input exist and have counteracted each other in the current experimental settings. Checking this possibility has to be subject to further research. Nevertheless, based on the current observations, one might conclude that there is no need to adapt the feedback modality depending on the input modality.

In Experiment 1, participants were able to adjust the amplitude of the stroke based on the location of the feedback that confirmed the stroke had been recognised. This was observed especially with hand gestures. As soon as the participants got the feedback, they aborted the stroke heading from the centre towards one of the four target directions (left, right, up or down) and initiated the counter-stroke back to the centre. This was possible with the continuous movement of the hand. In the visual feedback condition, the participants adjusted the length of the gaze stroke based on the location of the visual feedback on the screen. Having the visual flash fairly near to the centre of the screen (see Figure 3(b)) gave a clear indication that the participant does not need to make a long stroke outside of the screen. Similar effects of the feedback on the amplitude of the gestures have been observed earlier on gestures made by hand on touch surfaces (Zhai et al. 2012): with visual feedback, the participants used smaller hand movements and shorter strokes to draw the gestures.

Auditory feedback

Overall, auditory feedback was perceived well by the participants and did not cause any major problems in either of the experiments.

Haptic feedback

Haptic feedback was found hard to decode by the participants in both experiments. In Experiment 1, some of them commented that they totally ignored the directions but only paid attention to 'a feedback', confirming that 'some' stroke had been recognised. This may partly be due to the technique we used. We chose to give the haptic feedback on the palm of the hand because the hand is highly sensitive and a handheld device can simply be picked up when desired (Rantala et al. 2011). We also foresee that in future, mobile phones and other handheld devices can provide spatial haptic cues, similar to the feedback used in this experiment (Turunen et al. 2012). However, it was not easy to separate the feedback given by actuators placed at close regions in the palm. One could argue that spatial information via haptic feedback might have been easier to interpret if the actuators were spread sparsely on a larger area on the user's body, for example, on the back through a vest (Piateski and Jones 2005), or the head area, attached to, for example, glass frames (Rantala et al. 2014).

The plane of the haptic device was horizontal, which is different from the (vertical) plane of the other feedback modalities and the stimuli. This may have added an extra challenge to the coding of the feedback.

Our findings for haptic feedback, at least in Experiment 2, are contrary to the results by Kangas, Akkil, et al., 2014. They found haptic feedback on gaze strokes to improve user performance and satisfaction. The difference may be explained by the fact that their task was simpler, including only horizontal and vertical gestures consisting of two strokes. Also the feedback was simpler, it only confirmed that a stroke was recognised, but did not include any information on the direction. Obviously, the task in our study was fairly difficult, as reflected in the relatively high error rates (>20%). On the other hand, Foehrenbach et al. (2009) found haptic feedback to interfere with other modalities in hand gesturing tasks. They discuss that the technical implementation of the tactile feedback may play an important role. In addition, the requirements of the task and workload should also be taken into account when planning the feedback, be it auditory, haptic, visual or multimodal.

Taken together with prior work, the results of Experiments 1 and 2 seem to suggest that haptic feedback is best suited for simple feedback on motor acts. We measured lower performance for haptic feedback than for visual feedback in Experiment 2, where semantic information was encoded in haptic pulses. These pulses were similar to the simple feedback given in Experiment 1. However, instead of pulses given after a sub-gesture in Experiment 1, here they were given together at the end of one complete gesture.

We acknowledge the limitations of the haptic feedback in Experiment 2. Our results are very sensitive to the particular encoding used in presenting semantic feedback through haptics. If encodings that are much simpler to understand can be constructed, different results regarding semantic feedback may be possible.

Visual feedback

Visual feedback slowed down the interaction somewhat, but the slower speed was compensated by the smaller gesture size. With hand gesturing the compensatory effect can be noticeable. However, gaze strokes are so quick that the length of the stroke does not have much practical influence on the speed of interaction – other than the consequential effect of longer strokes being more tiring to the eyes. In general, it can also be noted that fast discrete actions may suffer from continuous feedback, especially if the user is caught in a feedback loop that requires attention and thus slows down the task performance (Witt, Lawo, and Drugge 2008).

Different types of errors

Off-crossings were errors where the participants did not follow the required sub-gesture pattern of always first making a horizontal (left, right) or vertical (up, down) stroke and then returning back to the centre. These errors cannot be explained by technical inaccuracy, since hand recognition suffers from the inaccuracy as much, or even more than gaze (Khoshelham and Elberink 2012). This type of error was mainly produced by participants taking shortcuts; that is, a movement from one quarter of the clock to another (see Figure 1(a)) without returning back to the centre in between the two sub-gestures. Participants arrived thus faster with the eye on their destination instead of taking the right way through the correct sequence of strokes. This did not happen for numbers that required entering the same sub-gesture twice (i.e. 3, 6, 9 and 12), since these off-crossings seem to have happened mainly when the movement directions changed between two sub-gestures, suggesting that controlling the movement direction seems to be of major importance in gaze. Also, the amount of wrong gestures was larger for gaze input, suggesting that gaze input requires more cognitive resources, probably due to its unconscious nature.

The decreased accuracy for numbers located in the diagonal positions was also observed by Isokoski and Käki (2002) in their study where they experimented with the 'hybrid' clock face metaphor using gestures on a touchpad. The large number of off-crossings may suggest that the design by Isokoski and Käki where the stroke continued directly to the second segment without returning to the centre may be more intuitive to use. Note, however, that in touch input, Isokoski and Käki had the additional dimension of touch/no-touch detection that does not exist in free hand gesturing or in eye tracking. Thus, the same design could not be implemented in this experiment.

The complexity of the gesture in general also probably has an effect. This is in line with previous research (Istance et al. 2010), where participants did more errors when the complexity of the gesture increased by skipping one stroke in a three-stroke gesture.

For hand input the main error type was timeout. Since the task time included the time required to observe the stimuli and plan for the action, the timeout threshold was only used for the second sub-gesture. The chosen timeout (2 seconds after the first sub-gesture) was long enough for gaze, but apparently the hand might benefit from a bit longer time limit. In addition to physiological differences, the sample rate and image processing speed may also have an effect on the responsiveness of the system: if there is any delay in the movement of the visible hand cursor, it may encourage the user to slow down accordingly. Furthermore, our experiment used location-based (i.e. boundary crossing) algorithm to recognise the strokes. Other relative methods may provide more robust gesture recognition and better user satisfaction (Rozado et al. 2012).

Correct gesture distribution per number

To study the error patterns further, we analysed the gesturing accuracy (mean correct in %) for each number. Figure 4 shows that in Experiment 1 the numbers located in the horizontal or vertical positions produced more accurate results than numbers located in the diagonal positions – when gaze was used for input. Hand input produced more consistent accuracy. This effect disappears in the Experiment 2 where semantic feedback is given only for the final result.

A potential affecting factor may have been the absence of the visible cursor in the gaze condition. Perhaps with a visible cursor, the participants would have been more aware of their gaze location and movements. We did not show the cursor in the gaze condition, because it might disturb and attract attention (Jacob 1991). Incidentally, a very recent research (Graupner and Pannasch 2014) indicates that showing the gaze cursor may not be as disturbing as it has been assumed. This can be taken into account in future studies. The fact that a cursor was shown for the hand, but not for the gaze clearly may have had an effect on our results.

User experience

In the Experiment 1 participants' experience and preferences varied a lot. None of the feedback modalities was superior to the others. Combined with the fact that there were no statistically significant differences in the performance metrics either, we can conclude that the feedback modality can be chosen according to the technology available, context, user preferences, etc. However, in Experiment 2 the semantic haptic feedback was not comprehensible and reliable enough to be really useful for the participants. In our task, this did not have a significant effect on performance as the next task was not dependent on the result of the previous task and there was no penalty in wrong interpretation. Obviously, if the feedback should deliver a semantic meaning that is required for a successful execution of the task, one should make sure the feedback is comprehensible to make it really helpful for the users.

Conclusion and future work

The effects of feedback modalities on task completion times and error rates were rather small. In error rates, there were no statistically significant differences between modalities. Completion times of correct gestures were faster for gaze input than for hand gesture input.

Error rates on all numbers were about the same for hand input. Gaze input led to more errors on numbers that consisted of both horizontal and vertical gestures, especially with the simple feedback. Since there were no statistically significant differences between feedback modalities, there was also no evidence of support or interference of using the same channel (hand or eye) for both input and output. The simple feedback given during gesturing seemed to provide some benefits for hand, but disturbed gaze input. The differences in error patterns per number were not significant anymore in Experiment 2.

If we interpret the results from the user interface designer's point of view, a clear finding is that the feedback modality does not matter much. This is a good thing. It means interaction designers will be able to implement the feedback based on available technologies, to match the needs of the user and whatever is suitable for the current task and context. However, the gesture design, input modality and the feedback type all seem to have an impact on usability. Gestures consisting of segments in horizontal and vertical directions led to many off-crossing errors with gaze. The participants seemed to skip the visit to the centre between the sub-gestures. A three-segment gesture that takes this path would perhaps be a better design for gaze-based use.

Our experiment was conducted with no particular emphasis on gesturing speed. In addition to such self-paced gesturing, future work should investigate maximum performance situations. Gaze gestures are best used in a desktop PC usage situation or when wearing a personal head-mounted gaze tracker. Hand gestures have better potential in walk-up scenarios without per-participant calibration and with some distance between the sensor and the user. These differences in ideal usage contexts can be further explored to see whether the hand and gaze gesturing techniques are competitive or complementary. Whether gestures are location based (as in our experiment) or location free is an important design consideration that was also left for future work.

For the usage of different feedback types, a more appropriate realisation for the haptic feedback needs to be elaborated. In our experiment, it was hard to decode the given feedback, which had the consequence that participants did not concentrate on the direction of the feedback but simply noticed there was some feedback. Thus, instead of complex encoding, haptic feedback may better function as a simple notification. In our implementation, the haptic actuators were placed close to each other, which may also have hindered the perception of the direction.

Notes

1. http://www.theeyetribe.com
2. http://www.tobii.com

Acknowledgements

We wish to thank Jussi Rantala for building the haptic feedback device for us and other members of the HAGI project for comments on the experimental set-up.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding information

The work reported is performed within the Transregional Collaborative Research Centre SFB/TRR 62 'Companion – Technology for Cognitive Technical Systems' funded by the

German Research Foundation (DFG), Transfer project (T01). The work of Päivi Majaranta was funded by the Academy of Finland, project Haptic Gaze Interaction (#260179). The work of Poika Isokoski was funded by the Academy of Finland: Mind, Picture, Image (#266285).

References

Alvarez-Santos, V., R. Iglesias, X. M. Pardo, C. V. Regueiro, and A. Canedo-Rodriguez. 2014. "Gesture-based Interaction with Voice Feedback for a Tour-guide Robot." *Journal of Visual Communication and Image Representation* 25 (2): 499–509.

Atia, A., S. Takahashi, K. Misue, and J. Tanaka. 2009. "UbiGesture: Customizing and Profiling Hand Gestures in Ubiquitous Environment." In *Proceedings of the 13th International Conference on Human–Computer Interaction. Part II: Novel Interaction Methods and Techniques*, 141–150. Berlin: Springer-Verlag.

Baudel, T., and M. Beaudouin-Lafon. 1993. "Charade: Remote Control of Objects Using Free-hand Gestures." *Communications of the ACM* 36 (7): 28–35.

Bhuiyan, M., and R. Picking. 2009. "Gesture-controlled User Interfaces, What Have We Done and What's Next." In *5th Collaborative Research Symposium on Security, E-Learning, Internet and Networking.*

Bolt, R. A. 1980. "'Put-that-there': Voice and Gesture at the Graphics Interface." In *Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '80)*, Vol. 14, 262–270. New York, NY: ACM.

Bulling, A., and H. Gellersen. 2010. "Toward Mobile Eye-based Human–Computer Interaction." *IEEE Pervasive Computing* 9 (4): 8–12.

Bulling, A., D. Roggen, and G. Tröster. 2008. "EyeMote – Towards Context-aware Gaming Using Eye Movements Recorded from Wearable Electrooculography." In *Proceedings of the 2nd International Conference on Fun and Games*, 33–45. Berlin: Springer Berlin Heidelberg.

Burke, J. L., M. S. Prewett, A. A. Gray, L. Yang, F. R. Stilson, M. D. Coovert, L. R. Elliot, and E. Redden. 2006. "Comparing the Effects of Visual–Auditory and Visual-Tactile Feedback on User Performance: A Meta-analysis." In *Proceedings of the 8th International Conference on Multimodal Interfaces (ICMI '06)*, 108–117. New York, NY: ACM.

Callahan, J., D. Hopkins, M. Weiser, and B. Shneiderman. 1988. "An Empirical Comparison of Pie vs. Linear Menus." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 95–100. New York, NY: ACM.

Cassidy, A., D. Hook, and A. Baliga. 2002. "Hand Tracking Using Spatial Gesture Modeling and Visual Feedback for a Virtual DJ System." In *Proceedings of the 4th IEEE International Conference on Multimodal Interfaces (ICMI '02)*, 197–202. Washington, DC: IEEE Computer Society.

Chen, F., E. Choi, J. Epps, S. Lichman, N. Ruiz, Y. Shi, Chen, F., E. Choi, J. Epps, S. Lichman, N. Ruiz, Y. Shi, R. Taib, and M. Wu. 2005. "A Study of Manual Gesture-based Selection for the PEMMI Multimodal Transport Management Interface." In *Proceedings of the 7th International Conference on Multimodal Interfaces (ICMI '05)*, 274–281. New York, NY: ACM.

Clawson, J., K. Lyons, T. Starner, and E. Clarkson. 2005. "The Impacts of Limited Visual Feedback on Mobile Text Entry for the Twiddler and Mini-QWERTY Keyboards." In *Proceedings of the Ninth IEEE International Symposium on Wearable Computers (ISWC '05)*, 170–177. Washington, DC: IEEE Computer Society.

Drewes, H., and A. Schmidt. 2007. "Interacting with the Computer Using Gaze Gestures." In *Proceedings of the 11th IFIP TC 13 International Conference on Human–Computer Interaction – Volume Part II (INTERACT '07)*, 475–488. Berlin: Springer-Verlag.

Dybdal, M. L., J. S. Agustin, and J. P. Hansen. 2012. "Gaze Input for Mobile Devices by Dwell and Gestures." In *Proceedings of the Symposium on Eye Tracking Research and Applications*. New York, NY: ACM.

Foehrenbach, S., W. König, J. Gerken, and H. Reiterer. 2009. "Tactile Feedback Enhanced Hand Gesture Interaction at Large, High-resolution Displays." *Journal of Visual Languages & Computing* 20: 341–351.

Garzotto, F., and M. Valoriani. 2012. "'Don't Touch the Oven': Motion-based Touchless Interaction with Household Appliances." In *Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI '12)*, 721–724. New York, NY: ACM.

Graupner, S.-T., and S. Pannasch. 2014. "Continuous Gaze Cursor Feedback in Various Tasks: Influence on Eye Movement Behavior, Task Performance and Subjective Distraction." In *HCI International 2014 – Posters' Extended Abstracts*, 323–329. Cham, Switzerland: Springer International Publishing.

Gustafson, S. 2012. "Imaginary Interfaces: Touchscreen-like Interaction Without the Screen." In *CHI '12 Extended Abstracts on Human Factors in Computing Systems (CHI EA '12)*, 927–930. New York, NY: ACM.

Heikkilä, H. 2013. "Tools for a Gaze-controlled Drawing Application – Comparing Gaze Gestures Against Dwell Buttons." In *Proceedings of INTERACT 2013, Part II, LNCS 8118*, 187–201. Berlin: Springer.

Heikkilä, H., and K.-J. Räihä. 2009. "Speed and Accuracy of Gaze Gestures." *Journal of Eye Movement Research* 3 (2): 1–14.

Heimonen, T., J. Hakulinen, M. Turunen, J. P. Jokinen, T. Keskinen, and R. Raisamo. 2013. "Designing Gesture-based Control for Factory Automation." In *14th IFIP TC 13 International Conference on Human–Computer Interaction – INTERACT 2013*, 202–209. Berlin: Springer.

Hinckley, K., R. Pausch, and D. Proffitt. 1997. "Attention and Visual Feedback: The Bimanual Frame of Reference." In *Proceedings of the 1997 Symposium on Interactive 3D Graphics (I3D '97)*, 121–126. New York, NY: ACM.

Isokoski, P., and M. Käki. 2002. "Comparison of Two Touchpad-based Methods for Numeric Entry." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '02)*, 25–32. New York, NY: ACM.

Isokoski, P., and R. Raisamo. 2000. "Device Independent Text Input: A Rationale and an Example." In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '00)*, 76–83. New York, NY: ACM.

Istance, H., A. Hyrskykari, L. Immonen, S. Mansikkamaa, and S. Vickers. 2010. "Designing Gaze Gestures for Gaming: An

Investigation of Performance." In *Proceedings of the 2010 Symposium on Eye-Tracking Research and Applications (ETRA)*. New York, NY: ACM.

Iwamoto, T., M. Tatezono, and H. Shinoda. 2008. "Non-contact Method for Producing Tactile Sensation Using Airborne Ultrasound." In *Proceedings of EuroHaptics 2008. Haptics: Perception, Devices and Scenarios*, 504–513. Berlin: Springer.

Jacob, R. J. 1991. "The Use of Eye Movements in Human–Computer Interaction Techniques: What You Look At Is What You Get." *ACM Transactions on Information Systems* 9 (2): 152–169.

Kajastila, R., and T. Lokki. 2013. "Eyes-free Interaction with Free-hand Gestures and Auditory Menus." *International Journal of Human–Computer Studies* 71 (5): 627–640.

Kangas, J., D. Akkil, J. Rantala, P. Isokoski, P. Majaranta, and R. Raisamo. 2014a. "Gaze Gestures and Haptic Feedback in Mobile Devices." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2014)*, 435–438. New York, NY: ACM.

Kangas, J., J. Rantala, P. Majaranta, P. Isokoski, and R. Raisamo. 2014b. "Haptic Feedback to Gaze Events." In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '14)*, 11–18. New York, NY: ACM.

Karam, M., and M. C. Schraefel. 2005. *A Taxonomy of Gestures in Human Computer Interactions*. Southampton: University of Southampton, Computing Service. http://eprints.soton.ac.uk/261149/.

Khoshelham, K., and S. O. Elberink. 2012. "Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications." *Sensors* 12: 1437–1454.

Köpsel, A., and A. Huckauf. 2013. "Evaluation of Static and Dynamic Freehand Gestures in Device Control." In *Proceedings of the Tilburg Gesture Research Meeting (TiGeR 2013)*, Tilburg.

Kratz, S., and R. Ballagas. 2009. "Unravelling Seams: Improving Mobile Gesture Recognition with Visual Feedback Techniques." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*, 937–940. New York, NY: ACM.

Lécuyer, A., C. Megard, J. M. Burkhardt, T. C. Lim, P. Coiffet, and L. Graux. 2002. "The Effect of Haptic, Visual and Auditory Feedback on an Insertion Task on a 2-Screen Workbench." In *Proceedings of Immersive Projection Technology Symposium*, 12–18.

Lee, S. C., B. Li, and T. Starner. 2011. "AirTouch: Synchronizing In-air Hand Gesture and On-body Tactile Feedback to Augment Mobile Gesture Interaction." In *Proceedings of the 2011 15th Annual International Symposium on Wearable Computers (ISWC '11)*, 3–10. Washington, DC: IEEE Computer Society.

Lenman, S., L. Bretzner, and B. Thuresson. 2002. *Computer Vision Based Hand Gesture Interfaces for Human–Computer Interaction*. Schweden: Royal Institute of Technology.

MacGregor, C., and A. Thomas. 2001. "Does Multi-modal Feedback Help in Everyday Computing Tasks?" In *Proceedings of the 8th IFIP International Conference on Engineering for Human–Computer Interaction (EHCI '01)*, 251–262. Berlin: Springer-Verlag.

MacKenzie, I. S., and S. J. Castellucci. 2012. "Reducing Visual Demand for Gestural Text Input on Touchscreen Devices." In *Extended Abstracts of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*, 2585–2590. New York, NY: ACM.

Majaranta, P., I. S. MacKenzie, A. Aula, and K. J. Räihä. 2006. "Effects of Feedback and Dwell Time on Eye Typing Speed and Accuracy." *Universal Access in the Information Society* 5 (2): 199–208.

Møllenbach, E., J. P. Hansen, and M. Lillholm. 2013. "Eye Movements in Gaze Interaction." *Journal of Eye Movement Research* 6 (2): 1–15.

Nielsen, J. 1993. *Usability Engineering*. Boston, MA: Academic Press.

Norman, D. A. 2010. "The Way I See It: Natural User Interfaces Are Not Natural." *Interactions* 17 (3): 6–10.

Pakkanen, T., R. Raisamo, K. Salminen, and V. Surakka. 2010. "Haptic Numbers: Three Haptic Representation Models for Numbers on a Touch Screen Phone." In *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction (ICMI-MLMI '10)*, 4. New York, NY: ACM, Article 35.

Piateski, E., and L. Jones. 2005. "Vibrotactile Pattern Recognition on the Arm and Torso." In *Proceedings of the First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC '05)*, 90–95. Washington, DC: IEEE Computer Society.

Porta, M., and M. Turina. 2008. "Eye-S: A Full-screen Input Modality for Pure Eye-based Communication." In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications (ETRA '08)*, 27–34. New York, NY: ACM.

Prinz, W. 1997. "Perception and Action Planning." *European Journal of Cognitive Psychology* 9: 129–154.

Prinz, W., and B. Hommel. 2002. *Common Mechanisms in Perception and Action: Attention and Performance XIV*. Oxford: Oxford Press.

Rantala, J., J. Kangas, D. Akkil, P. Isokoski, and R. Raisamo. 2014. "Glasses with Haptic Feedback of Gaze Gestures." In *Proceedings of the Extended Abstracts of the 32nd Annual ACM Conference on Human Factors in Computing Systems (CHI EA '14)*, 1597–1602. New York, NY: ACM.

Rantala, J., R. Raisamo, J. Lylykangas, T. Ahmaniemi, J. Raisamo, J. Rantala, K. Mäkelä, K. Salminen, and V. Surakka. 2011. "The Role of Gesture Types and Spatial Feedback in Haptic Communication." *IEEE Transactions on Haptics* 4 (4): 295–306.

Rozado, D., J. San Agustin, F. B. Rodriguez, and B. Varona. 2012. "Gliding and Saccadic Gaze Gesture Recognition in Real Time." *ACM Transactions on Interactive Intelligent Systems* 1 (2): 1–27 (Article No. 1).

Saxen, F., O. Rashid, A. Al-Hamadi, S. Adler, A. Kernchen, and R. Mecke. 2012. "Image-based Gesture Recognition for User Interaction with Mobile Companion-based Assistance Systems." In *Proceedings of the 4th International Conference of Soft Computing and Pattern Recognition (SoCPaR)*, Brunei.

Shan, C. 2010. "Gesture Control for Consumer Electronics." In *Multimedia Interaction and Intelligent User Interfaces*, edited by L. Shao, C. Shan, J. Luo, and M. Etoh, 107–128. London: Springer.

Shneiderman, B. 1998. *Designing the User Interface*. Reading, MA: Addison-Wesley.

Tsukadaa, K., and M. Yasumura. 2002. "Ubi-Finger: Gesture Input Device for Mobile Use." In *Proceedings of the*

APCHI 2002 (5th Asia Pacific Conference on Human Computer Interaction): User Interaction Technology in the 21st Century. Beijing: Science Press.

Turunen, S., H. Vilpponen, A. Vänskä, J. Raisamo, R. Raisamo, and J. Rantala. 2012. US Patent No. 20120206371.

Urbina, M. H., and A. Huckauf. 2010. "Alternatives to Single Character Entry and Dwell Time Selection on Eye Typing." In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications (ETRA '10)*, 315–322. New York, NY: ACM.

Vickers, S., H. Istance, and A. Hyrskykari. 2013. "Performing Locomotion Tasks in Immersive Computer Games with an Adapted Eye-tracking Interface." *ACM Transactions on Accessible Computing (TACCESS)* 5 (1): 1–33 (Article No. 2).

Vidal, M., K. Pfeuffer, A. Bulling, and H. W. Gellersen. 2013. "Pursuits: Eye-based Interaction with Moving Targets." In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, 3147–3150. New York, NY: ACM.

Villamor, C., D. Willis, and L. Wroblewski. 2010. *Touch Gesture Reference Guide*. LukeW Ideation + Design. http://www.lukew.com/ff/entry.asp?1071

Vogel, D., and R. 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. New York, NY: ACM.

Ware, C., and H. H. Mikaelian. 1987. "An Evaluation of an Eye Tracker as a Device for Computer Input." In *CHI '87 Proceedings of the SIGCHI/GI Conference on Human Factors in Computing Systems and Graphics Interface*, 183–188. New York, NY: ACM.

Winer, B. 1971. *Statistical Principles in Experimental Design*. New York: McGraw-Hill.

Witt, H., M. Lawo, and M. Drugge. 2008. "Visual Feedback and Different Frames of Reference: The Impact on Gesture Interaction Techniques for Wearable Computing." In *Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileCHI '08)*, 293–300. New York, NY: ACM.

Wobbrock, J. O., M. R. Morris, and A. D. Wilson. 2009. "User-defined Gestures for Surface Computing." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*, 1083–1092. New York, NY: ACM .

Wobbrock, J. O., J. Rubinstein, M. W. Sawyer, and A. T. Duchowski. 2008. "Longitudinal Evaluation of Discrete Consecutive Gaze Gestures for Text Entry." In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications (ETRA '08)*, 11–18. New York, NY: ACM.

Wu, H., and J. Wang. 2013. "Understanding User Preferences for Freehand Gestures in the TV Viewing Environment." *International Journal on Advances in Information Sciences and Service Sciences* 5 (4): 709–717.

Zhai, S., P. O. Kristensson, C. Appert, T. H. Andersen, and X. Cao. 2012. "Foundational Issues in Touch-surface Stroke Gesture Design – An Integrative Review." *Foundations and Trends in Human–Computer Interaction* 5 (2): 97–205.

Zhang, Y., A. Bulling, and H. Gellersen. 2013. "Sideways: A Gaze Interface for Spontaneous Interaction with Situated Displays." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 851–860. New York, NY: ACM.

Zhao, S., P. Dragicevic, M. Chignell, R. Balakrishnan, and P. Baudisch. 2007. "Earpod: Eyes-free Menu Selection Using Touch Input and Reactive Audio Feedback." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*, 1395–1404. New York: ACM.

Zimmerman, T. G., J. Lanier, C. Blanchard, S. Bryson, and Y. Harvill. 1987. "A Hand Gesture Interface Device." In *ACM SIGCHI Bulletin* (In Vol. 18, No. 4), 189–192 New York, NY: ACM.