Towards the Temporally Perfect Virtual Button: Touch-Feedback Simultaneity and Perceived Quality in Mobile Touchscreen Press Interactions

TOPI KAARESOJA, Nokia Research Center
STEPHEN BREWSTER, University of Glasgow
VUOKKO LANTZ, Nokia Research Center

9

Pressing a virtual button is still the major interaction method in touchscreen mobile phones. Although phones are becoming more and more powerful, operating system software is getting more and more complex, causing latency in interaction. We were interested in gaining insight into touch-feedback simultaneity and the effects of latency on the perceived quality of touchscreen buttons. In an experiment, we varied the latency between touch and feedback between 0 and 300 ms for tactile, audio, and visual feedback modalities. We modelled the proportion of simultaneity perception as a function of latency for each modality condition. We used a Gaussian model fitted with the maximum likelihood estimation method to the observations. These models showed that the point of subjective simultaneity (PSS) was 5ms for tactile, 19ms for audio, and 32ms for visual feedback. Our study included the scoring of perceived quality for all of the different latency conditions. The perceived quality dropped significantly between latency conditions 70 and 100 ms when the feedback modality was tactile or audio, and between 100 and 150 ms when the feedback modality was visual. When the latency was 300ms for all feedback modalities, the quality of the buttons was rated significantly lower than in all of the other latency conditions, suggesting that a long latency between a touch on the screen and feedback is problematic for users. Together with PSS and these quality ratings, a 75% threshold was established to define a guideline for the recommended latency range between touch and feedback. Our guideline suggests that tactile feedback latency should be between 5 and 50 ms, audio feedback latency between 20 and 70 ms, and visual feedback latency between 30 and 85 ms. Using these values will ensure that users will perceive the feedback as simultaneous with the finger's touch. These values also ensure that the users do not perceive reduced quality. These results will guide engineers and designers of touchscreen interactions by showing the trade-offs between latency and user preference and the effects that their choices might have on the quality of the interactions and feedback they design.

Categories and Subject Descriptors: C.3 [**Computer Systems Organization**]: Special-Purpose and Application-Based Systems—*Real-time and embedded systems*; H.1.2 [**Models and Principles**]: User/Machine Systems—*Human factors*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*Auditory (nonspeech) feedback, benchmarking, evaluation/methodology, haptic I/O, input devices and strategies (e.g., mouse, touchscreen), prototyping, theory and methods*; J.7 [**Computer Applications**]: Computers in Other Systems—*Consumer products, real time*

General Terms: Human Factors

Additional Key Words and Phrases: Temporal perception, simultaneity, touch, feedback, mobile device, touchscreen, tactile, audio

Authors' addresses: T. Kaaresoja, Nokia Research Center, Otaniementie 19, 02150 Espoo, Finland; email: topi.kaaresoja@nokia.com; S. Brewster, 17 Lilybank Gardens, University of Glasgow, Glasgow, G12 8RZ, UK; email: stephen@dcs.gla.ac.uk; V. Lantz, Nokia Research Center, Otaniementie 19, 02150 Espoo, Finland; email: vuokko.lantz@nokia.com.

DOI: http://dx.doi.org/10.1145/2611387

ACM Reference Format:
T. Kaaresoja, S. Brewster, and V. Lantz. 2014. Towards the temporally perfect virtual button: Touch-feedback simultaneity and perceived quality in mobile touchscreen press interactions. ACM Trans. Appl. Percept. 11, 2, Article 9 (May 2014), 25 pages.
DOI: http://dx.doi.org/10.1145/2611387

1. INTRODUCTION

Touchscreens are becoming more and more popular in consumer products and particularly in mobile phones. A touchscreen phone is most commonly used with a finger, multiple fingers, or, in some cases, a stylus. There are many ways to interact with a touchscreen: sliding a virtual slider or flicking or panning the screen content, for example. Despite these other interaction techniques, pressing a virtual button is still the major interaction method, such as in the following everyday tasks: entering a phone number to call, entering text for a message, email or status updates in social media, entering contact information in a contact list, and entering keywords to search a topic on the Internet.

In addition to the visual feedback given for touchscreen button presses, virtual buttons can provide audio and tactile feedback to the user, to mimic physical buttons. Audio feedback has been found to improve performance, reduce errors, and make the workload lower in touchscreen button interaction [Brewster 2002]. The same effects have been found when applying tactile feedback for touchscreen virtual buttons used with a stylus [Brewster et al. 2007] and a finger [Hoggan et al. 2008]. Visual feedback may take the form of colour or shadow change of a button when pressed and when released. Audio feedback can be beeps, clicks, or other sounds from a loudspeaker. Tactile feedback often follows the characteristics of audio feedback but is provided by a rotational, linear, or piezoelectric actuator.

Although phones are becoming faster, operating systems and applications are becoming more complex. There is always latency between a finger touch on a touchscreen and the feedback given, and the amount of latency may be different for the visual, audio, and tactile modalities. In addition to software latencies from the operating systems and applications, a capacitive touch sensor causes latency to the interaction because of its function. The location of a finger is scanned through the sensor with a certain sampling rate that takes time. The feedback production also takes time in visual displays, tactile actuators, and audio buffers, for example. Ng et al. [2012] give a detailed introduction to the technical issues of touchscreen latency.

Latency can be harmful in interaction. It has been stated that latency is one of the most important problems limiting the quality, interactivity, and effectiveness of virtual and augmented reality [Miller and Bishop 2002], as well as head-mounted display systems [He et al. 2000]. It has also been shown that cursor movement latency slows down interaction performance and increases the error rate in a targeting task with a mouse [MacKenzie and Ware 1993] and joystick [Miall and Jackson 2006]. Latency in different modalities has different performance consequences: visual latency degraded the performance more than haptic latency in a reciprocal tapping task [Jay and Hubbold 2005]. Latency has also been shown to degrade subjective satisfaction in touchscreen interaction [Kaaresoja et al. 2011a; Kaaresoja et al. 2011b]. On the other hand, latency may have some benefits if used in a controlled way; latency can be used as one interaction design parameter. It has been shown that virtual buttons can be made to feel heavier when tactile feedback latency is increased [Kaaresoja et al. 2011b]. From all of this prior research, we can conclude that we need to explore latency more to fully understand its consequences on perception and interaction.

It is natural to conclude that because latency causes problems in interaction, perceived simultaneity does the opposite, enabling a natural user experience. Despite earlier research, none has systematically investigated simultaneity perception of finger touch and tactile, audio, or visual feedback to understand the effects of latency on a capacitive touchscreen virtual button interaction. Thus, our motivation was to find the simultaneity perception thresholds of touch and feedback. From these, we would then know how the different feedback modalities need to be optimised to create effective and high-quality interactions. As simultaneity perception has been widely studied in psychophysics, we took an applied psychophysical approach to the simultaneity perception of touch and feedback.

In addition, to further understand how user experience changes as a function of latency, we examined one qualitative dimension of virtual button latency: perceived quality. We hypothesized that the users might notice the degradation in quality before they perceive the nonsimultaneity of touch and feedback as the latency between them increases. No research has been carried out to investigate the effects of latency on the perceived quality of capacitive touchscreen button interactions. It is not known if the simultaneity perception threshold and the perceived quality degradation threshold are different or which one is lower. The ultimate aim was to establish latency guidelines for interaction designers, user experience experts, and hardware and software engineers. The safest choice for the longest delay recommendation would be the simultaneity perception threshold or the moment when the perceived quality starts to degrade significantly, depending on which is shorter.

In this article, we introduce a study designed to achieve the preceding goals. In this study, participants pressed simulated virtual touchscreen buttons and received feedback in a single modality at a time (visual, audio, or tactile). The length of the feedback delay was varied, and the participants' task was to judge if the feedback was simultaneous with the touch or not and to score the quality of the keys they pressed.

2. RELATED WORK

In this section, we give an overview of the key previous work in the area of latency detection and interaction.

2.1 Intramodal Asynchrony Detection

Human temporal perception has been studied for more than a century in psychology. As early as in 1875, Exner [1875] found the thresholds for simultaneity perception of two intramodal (same modality) stimuli to be 2ms for two auditory clicks and 44ms for two brief flashes of light. Wundt found very similar figures: 2ms for audio, 27ms for tactile, and 43ms for visual [Boring 1923; Levitin et al. 1999]. These values have set the baseline for human temporal perception.

2.2 Perceived Simultaneity

The perceived simultaneity of two different stimuli has been studied a great deal in psychophysics. It is usually assessed with two methods: simultaneity judgments (SJs) and temporal order judgments (TOJs). Both methods estimate a point of subjective simultaneity (PSS) and just noticeable difference (JND), but the results and the interpretation of them are usually different with the same stimulus pair. This is because SJs provide a detection threshold and TOJs provide a differentiation threshold [Harris et al. 2010; Vogels 2004]. In an SJ experiment, participants are asked to make a forced-choice decision of whether two stimuli are "simultaneous" or "not simultaneous." Generally, their decisions are reported as a frequency distribution of the simultaneous responses. This distribution tends to be Gaussian when plotted as a function of the time between two stimuli (Figure 1). A Gaussian function is usually fitted to the frequency distribution of simultaneous responses, and the peak of this fitted function indicates the time between the stimuli at which participants are most likely to respond "simultaneous." It would have been inappropriate to ask participants to judge the temporal order of touch and feedback, because in our experiment, as in real-life virtual button application, touch always came first. That is why we used the SJ method in this experiment.

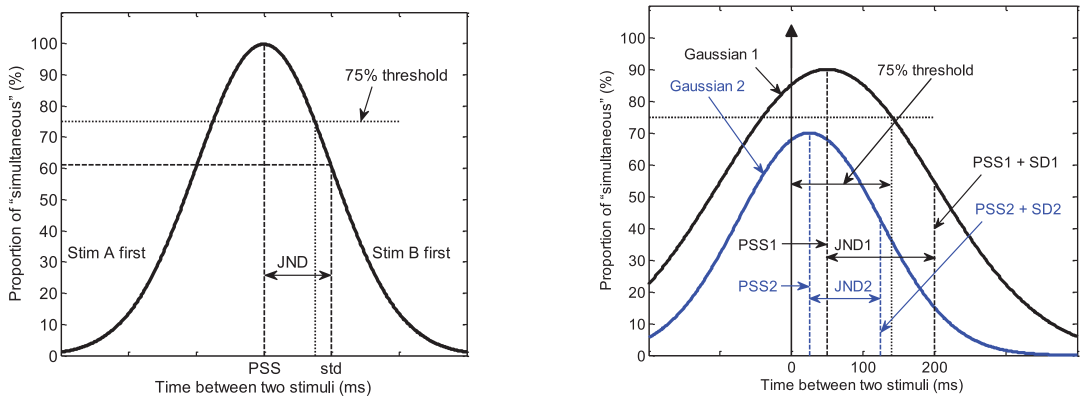

Fig. 1. Left: A Gaussian curve fitted to SJ data as a function of time between two stimuli. The point of subjective simultaneity (PSS) is the maximum of the fitted Gaussian function and states the time between two stimuli at which the participants most probably judged the two stimuli as simultaneous. The just noticeable difference (JND) is often defined to be one standard deviation (SD) of the fitted Gaussian model (61% of the maximum of the Gaussian curve), meaning the minimum time from the PSS that is needed for participants to reliably judge two stimuli as being no longer simultaneous. However, in practical applications, the 75% threshold is more useful. For clarity, the height of the Gaussian function is drawn to be in 100% in this figure. Right: The illustration of two different Gaussian curves showing the importance of the 75% threshold versus the traditional JND.

The JND is often estimated by the standard deviation (SD) of the Gaussian model in psychophysics, and the JND defined this way describes the simultaneity detection sensitivity—that is, the temporal window of simultaneity [Harris et al. 2010]. This is a convenient convention when JNDs are obtained from different conditions in a psychophysical experiment and compared with each other. However, the JND defined this way is bound to the height of the Gaussian function, but not to the actual proportion of simultaneous responses, which is the focus in practical applications. Figure 1 (right) illustrates this with two hypothetical frequency distributions of simultaneity perception modelled by Gaussian functions. It can be seen that $JND1 > JND2$, which means that the simultaneity perception threshold is smaller in the phenomenon that is modelled by the Gaussian 2 curve. However, the maximum proportion of simultaneous responses modelled by Gaussian 2 is less than Gaussian 1 and does not even touch the 75% proportion of simultaneous responses unlike Gaussian 1. That is why in practical approaches a 75% threshold is more sensible and we chose to use it (it also is used in Levitin et al. [1999] and Jota et al. [2013]). In addition, the 75% threshold is always more conservative than JND based on SD σ ($\leq 0.759 \times \sigma$, if $PSS \geq 0$ms and height of the Gaussian $\leq 100\%$), making it a stricter rule for the design guidelines (see Figure 1).

2.2.1 *Audio-Haptic Simultaneity.* In an experiment by Levitin et al. [1999], participants judged simultaneity of a mallet hit and a percussive sound. One participant hit the mallet and felt the hit haptically, whereas another visually observed the mallet being hit but did not feel it. Both of them heard an associated percussive sound from headphones. The time between the mallet hit and the sound was varied from −250ms (sound first) to 250ms (light/hit first). They found that the audio-haptic PSS was 0ms and the 75% threshold was −25ms (sound first) and 42ms (hit first) on average.

Adelstein et al. [2003] investigated the perceived asynchrony of a hammer tap and a related percussive sound. They did a comparative study where participants hit a tile with a hammer and were given a delayed sound over headphones. They had to judge which of the two hit-sound pairs had less delay. They found that the average PSS was not significantly different from zero and the average 75% threshold was 24ms, ranging from 5 to 70 ms within participants.

A hit with a mallet or hammer with an associated but delayed sound strongly relates to our practical approach to the simultaneity perception of a touch and audio feedback. These simultaneity perception threshold figures set a baseline for our hypotheses. However, in both studies discussed earlier, the hit was done with a tool in hand and the sound was provided to the headphones. We believe that it is important to investigate the simultaneity when the hit is done with a bare finger and the audio feedback is given from the same location of the hit.

2.2.2 *Audio-Visual Simultaneity.* Levitin et al. [1999] found that also the audio-visual PSS was 0ms and the 75% threshold was approximately 43ms on average and symmetrical.

Stone et al. [2001] varied the time between audio and visual stimuli from −250ms (sound first) to 250ms (light first). Their results showed that the PSS varied among the participants from −21ms (sound first) to +150ms, being 51ms on average. The average JND was 51ms. Later, Zampini et al. [2005] explored the effect of audio and visual stimuli location on perceived simultaneity. Their results suggested that the participants were more likely to report simultaneity if the stimuli came from the same spatial location. The average PSS was 19ms and the average JND was 114ms when the stimuli came from the same location. The PSS was 32ms and the JND 91ms on average when the stimuli came from different locations. In Stone's work, the light was presented in front of the participants and the sound over headphones, meaning that the stimuli came effectively from different locations. Thus, the positive thresholds (PSS + JND) found by Stone and Zampini were of the same magnitude, being 102ms and 123ms. Results of Levitin [1999] in turn showed smaller figures, because the test setup enabled participants to anticipate the event, thus making the judgment easier. In these studies, no touch or any interaction was required from the participant, but the stimuli were exogenously applied. However, an important finding of Stone and Zampini was that the proportion of simultaneity perception followed a Gaussian distribution when plotted as a function of time between the stimuli.

2.2.3 *Visual-Haptic Simultaneity.* To our knowledge, no research exists on simultaneity perception of a tactile hit and visual feedback. The nearest attempt to tackle the question of simultaneity perception between haptic and visual stimuli was by Vogels [2004]. In her experiment, participants moved a cursor on a computer screen with a force-feedback joystick and hit a horizontal line on the screen where they experienced a force representing a virtual wall. The cursor movement and the moment of the wall creation force were exposed to variable delays. The participants were asked to judge if the collision of the cursor and the line was simultaneous with the force. The results showed that the threshold for simultaneity perception was 59ms when force came first and 44ms when the cursor hit the horizontal line first. The PSS was nearly 0ms. Although the test setup and application were different from ours, we will take the findings as a reference for our study.

2.2.4 *Press-Haptic Simultaneity.* In addition to mallet or hammer interaction, perceived simultaneity has also been investigated in a physical button press setup with haptic feedback. Winter et al. [2008] varied the delay between a key press and tactile feedback. Tactile feedback could also precede the press. Participants pressed a Morse key with their index finger, and a tactile stimulus with a delay different for every key press was presented to the index finger of the opposite hand. The participants judged the simultaneity of the key press and the tactile stimulus. Like visual-audio simultaneity perception, here the results showed that the simultaneity perception followed a Gaussian function. They also showed that the average PSS was −29ms (tactile feedback first), although it was not significantly different from 0ms. This means that the point of perceived simultaneity could have been equal to physical simultaneity, which would be natural when interacting with a physical button in the real world. To be precise, a Morse key needs some time to go down and switch on after the finger has first touched the key head. In addition, the fingertip that presses the key needs some time to compress before the key goes down. This might explain the negative bias in the PSS. This consideration, and the fact that the PSS was not significantly different from zero, encouraged us to assume that the perception of simultaneity might happen when the feedback comes either at the same time or after a finger touch on a touchscreen. That is why we did not investigate the case of feedback coming before the actual key press in the experiment reported in this article. Although the Morse key is different from a touchscreen virtual button, this research motivated us to apply a psychophysical approach to understand the simultaneity perception of a button press and its associated feedback. The JND was defined to be one SD of the Gaussian function and was found to be 105ms on average in Winter et al.'s research, yielding the estimated threshold 76ms (PSS + JND). This also gave us a reference for simultaneity perception between a finger touch and tactile feedback in touchscreen virtual button interaction.

The preceding simultaneity research has concentrated on only one stimulus pair at the time, and the experimental setups were constructed to understand human perception. We constructed a setup more focused on our application domain, based around a mobile phone prototype.

2.3 Latency in Interaction

It has been shown that cursor movement latency slows down interaction performance and increases the error rate in a targeting task. MacKenzie and Ware [1993] investigated the effect of cursor movement latency on a visual targeting task with a mouse. They found that with latency of 225ms, the movement time increased 64% and error rates increased 214% compared to the minimum latency of 8.3ms. Based on their findings, they created a mathematical model between the latency and the task completion time based on Fitt's law. Miall and Jackson [2006] let participants track unpredictable targets with a handheld joystick. They found that visual feedback delay significantly reduced the performance and increased error rate.

Latency in different modalities has different performance consequences: Jay and Hubbold [2005] experimented with visual and haptic latency with a force feedback device in a reciprocal tapping task. They found that latency in visual feedback seriously degraded the performance, but haptic feedback latency had much less effect. Movement time went up significantly with visual and visual-haptic delays after 69ms, whereas with haptic feedback delay, this occurred after only 187ms. There were no more errors with the haptic feedback delay, nor did the users rate the use more difficult with haptic feedback delay. In contrast, both of these were significantly affected with visual feedback delays.

Because it seems evident that latency between a manual interaction and its feedback affects usability, it might also suggest how latency affects the overall user experience (e.g., perceived quality) in a manual interaction task. However, the participants were interacting with a device rather than a bare finger in the preceding research.

2.4 Touchscreen Feedback

There have been numerous attempts to add tactile and audio feedback to touchscreen virtual buttons to augment the visual feedback that is a standard part of the graphical design, starting from a simple click on a resistive touchscreen by Fukumoto and Sugimura [2001]. They found that tactile feedback improved the performance in a simple calculation task compared to audio feedback, especially in a noisy environment. Poupyrev et al. introduced tactile feedback for touchscreen virtual buttons using piezo technology and also expanded the tactile feedback design space from virtual buttons to more dynamic interactions. Poupyrev and Maruyama [2003] introduced a state diagram to model touchscreen interaction. They broke the interaction down in five different states where tactile feedback could be given: (1) touch-down, (2) drag, (3) hold, (4) lift-off inside a button (or other touchable item on touchscreen), and (5) lift-off outside a button (or other item). In our research, we focused on the touch-down phase as a first step. Poupyrev et al. [2004] also explored different touchscreen graphical user interface elements that could benefit from tactile feedback, in addition to buttons, such as sliders and text selection. They conducted informal evaluations on these concepts and received positive feedback. However, no controlled and detailed study was conducted, whereas our study focuses on the details of the touchscreen button feedback. Kaaresoja et al. [2006] also introduced and demonstrated touchscreen virtual buttons, text selection, as well as scrolling and drag & drop enhanced with tactile feedback implemented with piezo technology. It also has been shown that audio and tactile feedback significantly increased performance and reduced errors in virtual button interaction. Brewster [2002] found that adding sounds to touchscreen virtual buttons increased performance and reduced workload when used with a stylus. Tactile feedback added to touchscreen buttons also increased the performance and reduced error rate when used with a stylus [Brewster et al. 2007] as well as with a finger [Hoggan et al. 2008]. None of these studies, however, considered latency. They did not measure the latency between the finger or stylus touch and the associated feedback, report the latency of the feedback, or assess the effect of the latency on their results.

2.5 The Structure of Touchscreen Button Presses

A touchscreen button press involves a complex sequence of actions. Kaaresoja and Brewster [2010] presented a model to help understand the steps involved. The touchscreen display is touched with a finger or a stylus and feedback is given for this touch after some time has passed. This time is the latency between touch and feedback and it is distinguished from the latency between release and feedback. As a first step and for the sake of simplicity, we focus only on the latency between touch and feedback in this paper and leave the investigation of latency between release and feedback for future study. We therefore use the term "feedback" to refer to feedback associated with touch, "feedback latency" to latency between touch and feedback. Finally we define "touch-feedback simultaneity" to mean the simultaneity of touch and its associated feedback.

The feedback can be separated into the different modalities. After a finger or stylus has touched the screen, the different feedback elements (visual, audio, tactile and action feedback) are initiated after their individual latency periods. Visual feedback may be a colour change of the button pressed or a popup to help the user see what was actually pressed. Audio feedback can be an audible click and tactile feedback a short vibration, both confirming that a button was successfully pressed.

2.6 Feedback Latency in Touchscreen Interaction

Researchers have begun to investigate the effects of latency in touchscreen virtual button interaction. Kaaresoja and Brewster [2010] built a multimodal latency measurement tool and measured the tactile, audio, and visual latencies in various mobile phones. The tool consisted of an accelerometer, a microphone, and a high-speed camera. The tactile and audio feedback latency was assessed by measuring the time between the touch and feedback events in a sound editor. The visual feedback latency was determined by calculating the number of frames with a special high-speed video editor and multiplying it with the duration of one frame (3.33ms). They did not perform any user studies, so we do not know the effects of latency on the interaction and its consequences to the user.

Kaaresoja et al. [2011a, 2011b] studied the effects of differing tactile latencies on performance, error rate, and user preference in text entry with touchscreen virtual buttons. They found that the text entry and error rates were not affected when the latency between finger touch and tactile feedback was constant and in the range between 18 and 118 ms. However, there was a trend that the higher latencies were subjectively rated lowest. The subjective satisfaction dropped most when a virtual QWERTY keyboard was used where the latency was different on every key press. This study was the first attempt to understand the effect of latency on the touchscreen virtual button interaction, but the latency range used was too narrow to cause performance degradation. In addition, their device featured a resistive touchscreen, which is not the technology utilised in most contemporary mobile phones. Capacitive touchscreens are different from resistive ones, as the user only needs to touch lightly, without the larger force required by resistive panels, potentially causing a different level of latency. In this article, we use a capacitive device to give data useful for current mobile phone designs. Previous research only investigated tactile feedback latency and ignored the audio and visual components, which we focus on in this article in addition to tactile feedback, as they are common forms of feedback in mobile devices.

Latency may have some benefits if used in a controlled way, as it can be used as one interaction design parameter. Kaaresoja et al. [2011b] showed that virtual buttons could be made to feel heavier when tactile feedback latency was increased. Participants were asked to estimate the weight of a button in relation to a reference button featuring the minimum latency of the system. A positive significant correlation was found between latency and perceived weight: 78ms tactile feedback latency was rated significantly heavier than the reference, and 118ms latency was rated significantly heavier than 78ms. A resistive touchscreen was again used, and visual feedback latency was not controlled nor reported.

Ng et al. [2012] investigated latency perception in a dragging task on a touchscreen. They constructed a proprietary system capable of producing very low latency visual response for gestures on a touchscreen. They let participants to drag their finger on a touchscreen display, and a small square following their finger was presented as visual feedback. The participants judged which of the two conditions, the reference (1ms latency) or the probe (1 to 65 ms latency), was faster. They found that the participants were able to perceive latencies far below what the current commercial touchscreen devices offer. They found that the 75% threshold for latency perception in dragging task varied from 2.4 to 11.4 ms, being 6.0ms on average. The perception threshold was gained by a comparison method commonly used in a laboratory, not ecologically valid, psychophysics study. Users mainly use one touchscreen device at a time and may adapt to the latencies on that particular device. Comparison might happen when purchasing a touchscreen device, however. Ng et al.'s paper focused on the technical details of touchscreen latencies and solutions to overcome the challenges of reducing touch-to-display latency; in addition to dragging, no other interaction techniques were tested.

Jota et al. [2013] continued to investigate latency in direct-touch input on a touchscreen. They used a similar hardware setup to Ng et al. and found that performance in a visual targeting task degraded as latency increased. The results showed that there was no significant difference in performance between touch and feedback latencies, 1ms and 10ms, although further analysis showed that there might not be any floor effect of latency on performance. This would mean that the performance would always be better as latency goes towards zero. They also experimented with latency between finger touch and visual on-screen feedback, studying feedback latency detection with comparison (a probe against a reference). Their results showed that the 75% latency detection threshold varied from 20 to 100 ms depending on the participant, with the average being 64ms. They concluded that although the users could detect the latencies below 10ms, optimizing latency below 25ms gives little advantage in a pointing task. This value is even higher—40ms—for a tapping task. This gives an important baseline for current research, although it included only visual feedback for a touch input. So the perception thresholds for touch and audio or tactile feedback remain unknown.

Our article presents research that fills gaps in the literature regarding touchscreen feedback and latency. The device used in our experiment was designed to have a similar form factor and size as a typical mobile phone and featured capacitive switches and means to provide tactile, audio, and visual feedback. Tactile, audio, and visual feedback modalities were included in the same experimental session, albeit not provided together, to get full insight into the effects of latency on the modalities most commonly used for feedback in mobile phones. Based on the simultaneity research mentioned

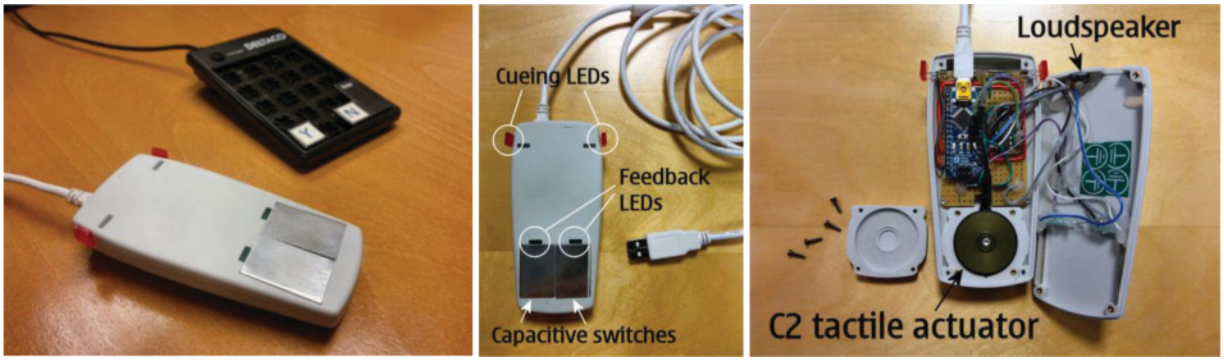

Fig. 2. Left: The Virtual Button Simulator (white) with the response pad for the experiment (black). Middle: The Virtual Button Simulator and the USB cable used for connecting to the laptop. Two capacitive switches were located at the bottom of the device. Above the switches were two green LEDs for visual feedback. At the top of the device were two red LEDs for the cueing purposes. Right: The opened enclosure of the Virtual Button Simulator. The USB cable was connected to Arduino Nano, and the tactile and audio driver was located next to Arduino. The C2 was located in its own enclosed cavity on the bottom of the device (cover open). The loudspeaker was attached inside the cover on the top of the device.

earlier, the latency ranged from 0 to 300 ms. As stated previously, our ultimate aim was to find practical guidelines for designers. We also tested five contemporary touchscreen mobile phones and measured the range of their latencies with respect to our guidelines.

3. EXPERIMENT

3.1 Experiment Methodology

A within-subjects design with the method of constant stimuli [Coren et al. 2003] was chosen with a forced-choice SJ task for all three different feedback modalities and nine latency conditions. Each participant went through all the feedback latency conditions and were instructed to respond either "yes" ("simultaneous") or "no" ("not simultaneous") for each.

3.2 Participants

Twenty four (12 female) volunteer participants aged 26 to 50 years (mean 36.4, SD 6.3) took part in the experiment. Three were left-handed. All filled in a consent form at the start of the experiment and were given a movie ticket and a chocolate bar as a reward for their participation.

3.3 Equipment

Current commercial mobile phones cannot provide feedback latencies near zero with low variance. Therefore, we built a proprietary research device resembling a mobile phone as much as possible. We called the research device the *Virtual Button Simulator*. The size and weight of the Virtual Button Simulator were similar to a small mobile phone: $54 \times 112 \times 21$ mm (max width $\times$ height $\times$ thickness) and 83g (Figure 2). In order to feature capacitive sensing, but to keep the sensing latency as low as possible, we used two metallic capacitive buttons at bottom on the front of the device instead of using a full touch sensor, which would have caused extra latency (see Figure 2, left). One button would have caused still less latency, but it would have been difficult to set up a reasonable task for the participants. Visual feedback was provided by two green LEDs (HLMP-0504, 565nm, 2.5×7.6 mm) placed just above the key area for giving visual feedback to imitate a key popup (see Figure 4). Audio feedback was played through a miniature loudspeaker ($9 \times 9 \times 3$ mm) located inside the cover on top of the device like in a real mobile phone. Tactile feedback was provided by a C2 Tactor by Engineering

Acoustics (www.eaiinfo.com), which has been used in several mobile experiments in the past (e.g., Brewster et al. [2007]; Hoggan et al. [2008]) and was located inside the device in its own covered cavity. Two red LEDs (HLMP-0301, 635nm, 2.5 × 7.6 mm) were located on top of the device to give cueing information.

To minimize latencies, all processing of button presses and feedback generation happened in an Arduino Nano (http://arduino.cc) microcontroller inside the Virtual Button Simulator instead of the controlling PC. The metallic capacitive buttons were connected directly to the Arduino Nano input pins, and the capacitive sensing was implemented with the help of a piece of open source software (http://playground.arduino.cc/Code/CapacitiveSensor). Since the Arduino was not capable of driving strong enough signals to the loudspeaker or the tactile actuator C2, a Texas Instruments L293DN digital switch was used as a driver between the Arduino and the loudspeaker and the C2. According to the specifications, the L293DN added less than 1ms latency to the circuit. The LEDs were connected directly to the Arduino's output pins. The Virtual Button Simulator was connected to a laptop PC via USB, which powered the Arduino and enabled communication between the Arduino and the PC. With the LEDs, loudspeaker, and C2 tactile actuator, the Virtual Button Simulator was able to provide visual, audio, and tactile feedback with less than 4ms baseline latency between finger touch and feedback. Above the baseline, the latency was fully controllable in millisecond resolution. The system baseline latency of the Virtual Button Simulator was measured with the latency measurement tool [Kaaresoja and Brewster 2010]. Each feedback modality and latency condition was measured seven times. The average baseline latency was 2.81ms for tactile, 0.65ms for audio, and 3.92ms for visual feedback, and the mean SD was 0.41, 0.46, and 1.6 ms, respectively. The audio and tactile latency were the time between the first moment of the finger touch and the first local intensity maximum of the feedback. The visual feedback latency was the time between the first moment of the touch and the moment when the green LED was fully switched on. The measurements proved us that the performance of Virtual Button Simulator allowed us to control latencies across the modalities at levels below human perception.

3.4 Experiment Software

The experiment software ran on a laptop PC and was programmed with Presentation® (www.neurobs.com), a software package designed specifically for programming and running experiments. A Presentation application was programmed to randomize the stimuli, ask the task-related questions, and receive the participants' response. The Virtual Button Simulator and the Presentation application communicated via a serial communication protocol through USB.

3.5 Stimuli

There were two independent variables in the experiment: feedback modality and feedback latency. Feedback modality had three types: tactile, audio, and visual. There were nine latency levels: 0, 10, 20, 30, 50, 70, 100, 150, and 300 ms. This led to 27 different conditions, and every condition was repeated four times in addition to 36 training stimuli, giving a total of 144 individual stimuli for each participant in the simultaneity perception part. The perceived quality part consisted of one repetition of each latency and feedback modality condition without training leading to 27 additional stimuli.

3.5.1 *Tactile Feedback.* The tactile feedback was designed to be a short tactile click (Figure 3, left) mimicking a tactile feedback of a physical button. It was produced by sending a 1ms pulse of 5V to the C2, resulting in a click with 1.5ms rise time and 13ms fall time (50%) (see Figure 3). The acceleration level of the tactile click was 2.2g peak to peak. The sound level of the tactile feedback was 40dB (A) measured at a 30cm distance from the Virtual Button Simulator.

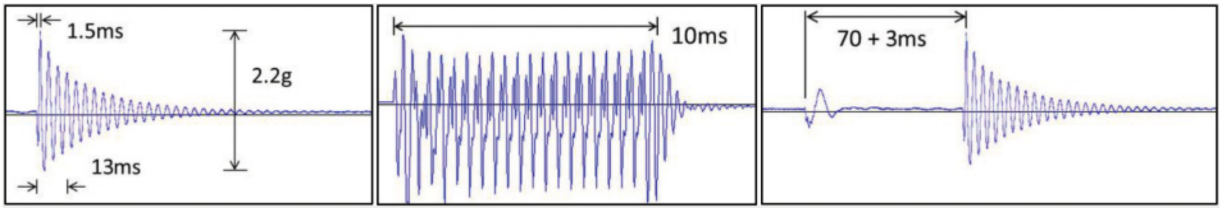

Fig. 3. Left: The acceleration and timing of the tactile click used as the tactile feedback in the experiment. The time between the start of the feedback and the peak was 1.5ms, and the fall time to the 50% level was 13ms. The acceleration level was 2.2g. Middle: The recorded waveform and the timing of the audio click used as the audio feedback. Right: The 70ms latency for tactile feedback. The 70ms latency is added to the 3ms system baseline (measured 2.81ms on average for the tactile feedback).

Fig. 4. A text entry popup in the Nokia Lumia and the Apple iPhone, and the simulated one in the Virtual Button Simulator.

3.5.2 *Audio Feedback.* The short audible click used in Apple iPhone virtual buttons was used as the basis for the audio feedback design. Figure 3 (middle) shows the recorded waveform from the Virtual Button Simulator. It was an audible click with a duration of 10ms and a frequency of 2,033Hz. The sound level of the audio feedback was 60dB (A) measured at a 30cm distance from the Virtual Button Simulator.

3.5.3 *Visual Feedback.* The visual feedback was designed to mimic a text entry popup that occurs when a key is pressed on a phone keyboard (Figure 4). The metallic buttons used in the Virtual Button Simulator could not change colour or shape; they were primarily designed to be as low latency as possible. Therefore, we used green LEDs that highlighted just above the finger position (like the key popups shown in Figure 4). We could not use a proper LCD display, as it would not have had a low enough latency for our study design. The green feedback LED glowed as long as the button was pressed. However, to tackle bouncing effects, an 8ms dead period was added after the release, which meant that the LED actually glowed 8ms after the key was released. This did not cause any problems, because 8ms is a short time compared to the time that the user presses the key and the LED is on. Based on earlier research on tap and audio feedback [Adelstein et al. 2003], we also believe that the duration of the stimulus does not affect the touch-feedback simultaneity perception. We did not attempt to equalize the intensity of the different feedback stimuli. However, they were all clearly over the perception thresholds.

3.5.4 *Latency Conditions.* We varied the latency between the first moment of finger touch and the feedback from 0 to 300 ms in all modality conditions in addition to the system baseline latency. Nine

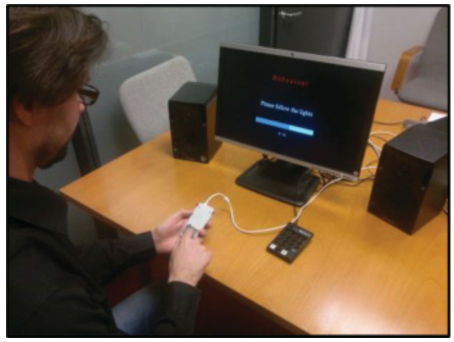

Fig. 5. Experiment setup. Participants held the Virtual Button Simulator in their nondominant hand and pressed the keys with their dominant hand. They responded with a modified keypad connected to a PC.

different latency conditions were selected for the method of constant stimuli: 0, 10, 20, 30, 50, 70, 100, 150, and 300 ms. These were added to the Virtual Button Simulator's measured baseline for each of the modalities (an example is shown in Figure 3, right). The selection of the latency values was based on earlier work introduced in Sections 2.2.4. and 2.6. The baseline latency is usually added to the latency conditions (e.g., Adelstein et al. [2003]), since it makes the mathematical analysis simpler and low-latency conditions can be selected evenly.

3.6 Hypotheses

The experiment hypotheses for each modality were mainly based on earlier work as follows.

3.6.1 *Perceived Simultaneity*

(H1) The distribution of simultaneous responses will follow a Gaussian distribution (e.g., Stone et al. [2001]);

(H2) The PSS will not be significantly different from 0ms (e.g., Levitin et al. [1999]; Winter et al. [2008]);

(H3) The 75% simultaneity perception threshold of touch and tactile feedback will be near 60ms ($(PSS + JND) \times 0.758 = 58$ms [Winter et al. 2008]), audio feedback 42ms [Levitin et al. 1999], and visual feedback 45ms [Jota et al. 2013].

3.6.2 *Perceived Quality*

(H4) The perceived quality score for the buttons will drop when latency is higher than 70ms ([Kaaresoja et al. 2011a]);

(H5) The participants would perceive a drop in quality earlier than the simultaneity perception threshold (based on pilot studies).

3.7 Procedure

Participants sat at a desk in a quiet office room, read the experiment instructions, and filled in a background questionnaire and consent form. They were instructed to hold the Virtual Button Simulator in their nondominant hand and asked to press the capacitive keys with the index finger of their dominant hand (Figure 5).

We designed the task to be simple, realistic, and feasible to give meaningful results. The goal was to get participants to press the two buttons several times but not to spend too much time on one press;

otherwise, we could not control the length of the experiment session. We could not ask participants to write text with just two buttons. However, we wanted the task to contain several button presses to mimic text entry without a need to remember arbitrary sequences composed of two letters, numbers, or symbols mapped to the buttons, for example. Since short-term memory can only contain limited number of items, the participants might not be able to remember the sequences properly [Miller 1956]. That could have slowed down the task, affecting the simultaneity or the perceived quality judgment and reliability of results. One choice would have been to let participants just press the buttons at their own pace. It turned out in the pilot studies that a participant started to explore button presses very slowly and carefully, which both took time and was unnatural. To overcome these challenges, we ended up having two cueing LEDs at the top of the device, one at each side as described in Section 3.3. These LEDs caused visual and cognitive load on the participant during the button presses, but that was an ecologically valid solution, since they simulated the visual load caused by looking at text and icons at the top of the screen on a mobile phone.

The participants' task was to follow the flashing red cueing LEDs by pressing the keys according to the side of the flash: if the right red LED flashed, participants were to press the right capacitive key and vice versa. If they made a mistake, they were instructed to continue the task without interruption. The cueing flash was designed to be as short as possible but still clearly perceivable. The interval between the flashes needed to be as short as possible to keep the task realistic, not to make the experiment unnecessarily long, but long enough so that the participants had time to react to the cue, press the button, and wait for the maximum feedback latency before the next cue. After a little iteration, we chose the length of the cueing flash as 50ms and a flash interval of 1s. Cueing like this ensured the control over the length of the experiment session and the time spent on one stimulus set while giving each participant good exposure to the latency stimuli.

Feedback was given depending on the modality and latency condition for each button press. One stimulus set consisted of seven cueing flash and key press pairs, within which the modality of feedback and the latency of the feedback were kept constant. After these seven flash-press pairs, the participant was asked a question: "Was the feedback simultaneous with your touch?" The participant responded "Y" or "N" on the response pad according to her or his perception. The response pad was a modified number keypad connected to the experiment PC containing only two keys, one for "no" and one for "yes" responses (see Figures 2 and 5). After the response, another stimulus set was presented to the participant. Background noise was played from two external active loudspeakers (Genelec 2029AL Digital) during flashes and presses to prevent the possible sound from the tactile actuator being audible to the participants. To equalize the conditions, the noise was also played in the audio and visual feedback conditions. Brown noise was chosen for the background, since it successfully masked the tactile feedback frequency, but not the audio feedback from the experiment. The noise level was 64dB (A), measured 60cm from the midpoint of the loudspeakers. The room background noise level was 39dB (A).

Before the actual experiment, the participant went through a training period of 12 flash-press stimulus sets for each modality using the latency conditions 0, 150, and 300 ms. These conditions were selected for the training period to ensure that the participant understood the tasks properly. All nine latency conditions were repeated four times in one feedback modality condition, meaning that there were 36 flash-press-response sequences in the real experiment for each of the three modalities. There were $3 \times (12 + 36) = 144$ flash-press-response sequences for SJ altogether for one participant.

After the simultaneity perception phase was completed, a perceived quality questionnaire was administered for each stimulus. The participants experienced the nine latency conditions again without training or repetition in a randomized order for each modality. The task was exactly the same as in the previous part of the experiment: to follow the flashing red cueing LEDs by pressing the keys according to the side of the flash. After the seven flash-press pairs, the following question was presented to the participants: “How would you rate the quality of the keys?” They responded on 1-to-7 scale on the perceived quality questionnaire with a pen, “1” meaning low quality and “7” high quality. There were $3 \times 9 = 27$ flash-press-response sequences for quality scoring altogether for one participant.

The feedback latency conditions were randomized, and the feedback modality conditions were counterbalanced during both parts of the experiment. The experiment took approximately 1 hour.

3.8 Analysis Methods

There were $n = 9 \times 4 \times 24 = 864$ binary responses altogether for each modality condition. Earlier work shows that the probability of simultaneity perception can be modelled with a Gaussian function [Stone et al. 2001; Zampini et al. 2005]. Thus, according to Stone et al., the probability p_1 of observing a “simultaneous” response $r_i = 1 (i = [1, n])$ at feedback latency equal to LAG_i ms is

$$p_1 (r_i = 1 | LAG_i, \mu, \sigma, a) = ae^{-\frac{1}{2}\left(\frac{LAG_i - \mu}{\sigma}\right)^2}, \tag{1}$$

where μ is the feedback latency at which the “simultaneous” answer is most likely to happen, a is the maximum probability of a simultaneous answer at the feedback latency $LAG = \mu$, and σ is the SD associated with responses determining the width of the Gaussian function. Probability p_0 of a “not simultaneous” response $r_i = 0$ at a latency equal to LAG_i ms is $(1 - p_1)$

$$p_0(r_i = 0 | LAG_i, \mu, \sigma, a) = 1 - ae^{-\frac{1}{2}\left(\frac{LAG_i - \mu}{\sigma}\right)^2}. \tag{2}$$

We fitted the probabilities p_1 and p_0 defined previously jointly to all the observed responses—that is, to all “simultaneous” and “not simultaneous” responses by all participants in each and every latency condition. The fitting was implemented separately for each feedback modality using the maximum likelihood estimation (MLE) method. The MLE method estimates the model parameters so that the probability of the observed data is maximized [Millar 2011]. We assume that the responses were made independently from each other, thus the likelihood function $L(\mu, \sigma, a)$ is of a product form

$$\begin{aligned} L(\mu, \sigma, a) &= \prod_{i=1}^{n_1} ae^{-\frac{1}{2}\left(\frac{LAG_i - \mu}{\sigma}\right)^2} \times \prod_{i=1}^{n_0} \left(1 - ae^{-\frac{1}{2}\left(\frac{LAG_i - \mu}{\sigma}\right)^2}\right) \\ &= \prod_{i=1}^{n} \left(ae^{-\frac{1}{2}\left(\frac{LAG_i - \mu}{\sigma}\right)^2}\right)^{r_i} \times \left(1 - ae^{-\frac{1}{2}\left(\frac{LAG_i - \mu}{\sigma}\right)^2}\right)^{(1 - r_i)}, \end{aligned} \tag{3}$$

where $n = (n_1 + n_0)$ (n_1 “simultaneous” and n_0 “not simultaneous” responses). This likelihood function was exactly the same as introduced by Stone et al. [2001]. However, in this experiment, we observed only positive feedback latencies; in other words, the feedback always came after the touch. For a realistic key press task, it would be unnatural and thus irrelevant to observe the negative touch feedback latencies.

The MLE estimates $\hat{\mu}$, $\hat{\sigma}$, and $\hat{a}$ of the parameters μ, σ, and a were obtained for each modality condition by minimizing the negative log-likelihood function. This minimization was done with Matlab function *fminsearch*, which is based on Nelder-Mead simplex algorithm (www.mathworks.se). Function *fminsearch* needs an initial starting point set for the parameter optimization, and it was obtained by fitting curves with the Matlab Curve Fitting Tool *cftool*, which is based on least square estimation. This initial estimate for the parameter values (μ, σ, a) was (50, 50, 0.7) for all modality conditions, and there were no constraints involved in the minimization procedure.

Table I. The Gaussian Curve Fitting Results for the Probability p_1

Feedback Modality	$\hat{\mu}$	95%CI$_{\hat{\mu}}$	$\hat{\sigma}$	95%CI$_{\hat{\sigma}}$	$\hat{a}$	95%CI$_{\hat{a}}$
Tactile	2.5	−5.9–11	78	70–87	0.90	0.85–0.93
Audio	18	7.5–29	94	84–106	0.92	0.89–0.95
Visual	28	16–39	97	85–110	0.88	0.84–0.91

Note: $\hat{\mu}$ is the MLE estimate for μ, $\hat{\sigma}$ is the MLE estimate for σ, and $\hat{a}$ is the MLE estimate for a. All the times are in milliseconds (ms), and all the quantities are MLE estimates and their 95% confidence intervals. Note that the 95% confidence intervals are asymmetric around MLE estimates due to nonnormal distribution of the parameters.

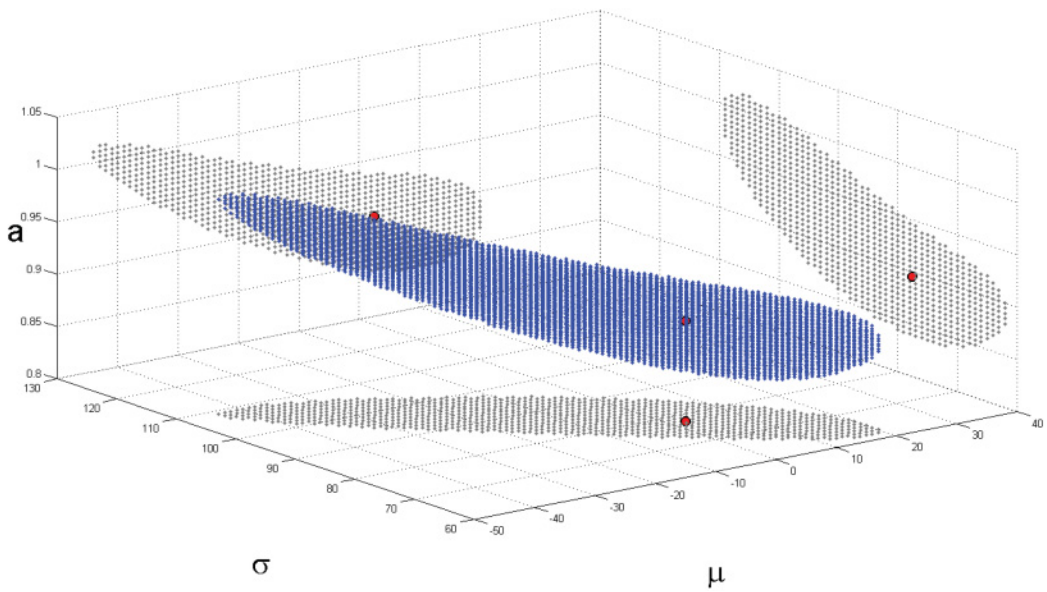

Fig. 6. The 3D confidence body, and its 2D projections, of the MLE of Gaussian function parameter estimates $\hat{\mu}$, $\hat{\sigma}$, and $\hat{a}$ for the simultaneity perception in touch and tactile feedback condition. The MLE points are marked as red dots. The confidence bodies for audio and visual feedback conditions were similar in shape (i.e., not ellipsoids), and the violation of the normality of the parameter estimate distributions was similarly evident. That is why the LRT for the uncertainty analysis of the individual parameters was used instead of Wald's test (see text). This confidence body and the corresponding ones for audio and visual feedback conditions were also used to calculate the 95% confidence intervals for the Gaussian model values.

4. RESULTS

4.1 Simultaneity Perception

The results of the Gaussian model fitting for the probability p_1 including the model parameter MLE estimates and their joint likelihood ratio tests (LRTs) 95% confidence intervals of the parameters are summarized in Table I. The LRTs of all three parameters of all feedback-specific Gaussian models were implemented against $\chi^2_3(0.95)$. Figure 6 shows the three-dimensional confidence body with its two-dimensional projections of the MLE of the Gaussian model parameters for the tactile feedback modality. It can be seen that the projections are not ellipsoids and that the MLE is in the middle of them. This indicates that the distribution of the parameter estimates was not normal. This was also the case when considering the Gaussian models for audio and visual modality feedback conditions and their confidence bodies. Stone et al. used Wald's test to determine the uncertainty of the MLE parameters as 95% confidence intervals. This method assumes a normal distribution of the estimated parameters. However, it is advisable to use LRT statistics instead for finding the confidence intervals if the assumption is not valid or is inaccurate [Millar 2011]. Thus, we implemented the restricted LRT against $\chi^2_2(0.95)$ statistics for each parameter estimate for each modality condition. The 95% confidence intervals for the probability p_1 for all feedback modality conditions were calculated by going through the parameter triplets within the whole three-dimensional confidence body and finding the minimum and the maximum values of the probability p_1 at each *LAG* running from 0 to 300 ms (1ms resolution).

The goodness of a Gaussian fit was tested with Chi-square and Kolmogorov-Smirnov goodness-of-fit tests. The proportion of simultaneous responses was compared with the modelled proportions at the latency conditions. All the fits passed these two tests. This proves that the experimental data support (H1)—the distribution of “simultaneous” responses will follow a Gaussian distribution.

The PSS was calculated as $\hat{\mu}$ + system baseline latency for each modality. For simultaneity perception of touch and tactile feedback, the PSS was 5ms with the 95% confidence interval being −3.1 to 14 ms, touch and audio feedback 19ms with a 95% confidence interval of 8.2 to 30 ms, and touch and visual feedback 32ms with a 95% confidence interval of 20 to 43 ms. The PSS of touch and tactile feedback did not differ statistically significantly from physical simultaneity, as 0ms was within the 95% confidence interval. However, the PSS of touch and audio, as well as touch and visual feedback, were significantly different from physical simultaneity, because 0ms was not within the 95% confidence intervals. Thus, (H2)—the PSS will not be significantly different from 0ms—was partially supported.

A pairwise Chi-square test of proportion was conducted between the observations to see when the proportion of simultaneity perception drops significantly. A Bonferroni correction was applied, resulting in a significance level set at $p < 0.0056$. The test showed that the proportion of simultaneity perception of touch and tactile feedback was not significantly different when the latency condition was 0, 10, 20, or 30 ms, but was significantly higher at the latency condition 20ms than at 50ms ($\chi^2_1 = 10.074, p < 0.0015$), meaning a significant drop between 20 and 50 ms. The proportion of the simultaneity perception of touch and audio feedback was not significantly different when the latency condition was 0, 10, 20, 30, 50, or 70 ms, but it dropped significantly between 50 and 100 ms ($\chi^2_1 = 9.8091, p < 0.0017$). The proportion of the simultaneity perception of touch and visual feedback was not significantly different when the latency condition was 0, 10, 20, 30, 50, or 70 ms, but it dropped significantly between 70 and 100 ms ($\chi^2_1 = 9.9187, p < 0.0016$).

The proportions of “simultaneous” responses and the MLE probability p_1 models with 95% confidence intervals are plotted in Figure 7. The figure also shows also the uncertainty (95% confidence intervals) of the values of the Gaussian models. This plot can be used to find the practical 75% simultaneity perception thresholds, which can be used as guidelines.

It can be seen that the 75% simultaneity perception threshold for touch and tactile feedback is 52ms with the 95% confidence interval being 40 to 62 ms. For touch and audio feedback, the threshold is 80ms with a 95% confidence interval of 65 to 90 ms. For touch and visual feedback, the threshold is 85ms with a 95% confidence interval of 70 to 100 ms.

Thus, our hypothesis about the 75% threshold (H3)—tactile 60ms, audio 42ms, and visual 45ms—was partially supported: the hypothesized 75% threshold for tactile feedback was within the confidence interval, but was higher for audio and visual feedback. These values fell within the time windows found in the statistical inference of the preceding observations.

4.2 Perceived Quality

A boxplot with the medians and means with trendlines of the scores from the perceived quality questionnaire are shown in Figure 8. A Friedman test showed significant differences in perceived quality depending on latency and feedback modality ($\chi^2 = 223.24, p < 0.001, df = 26$). Post hoc analysis with

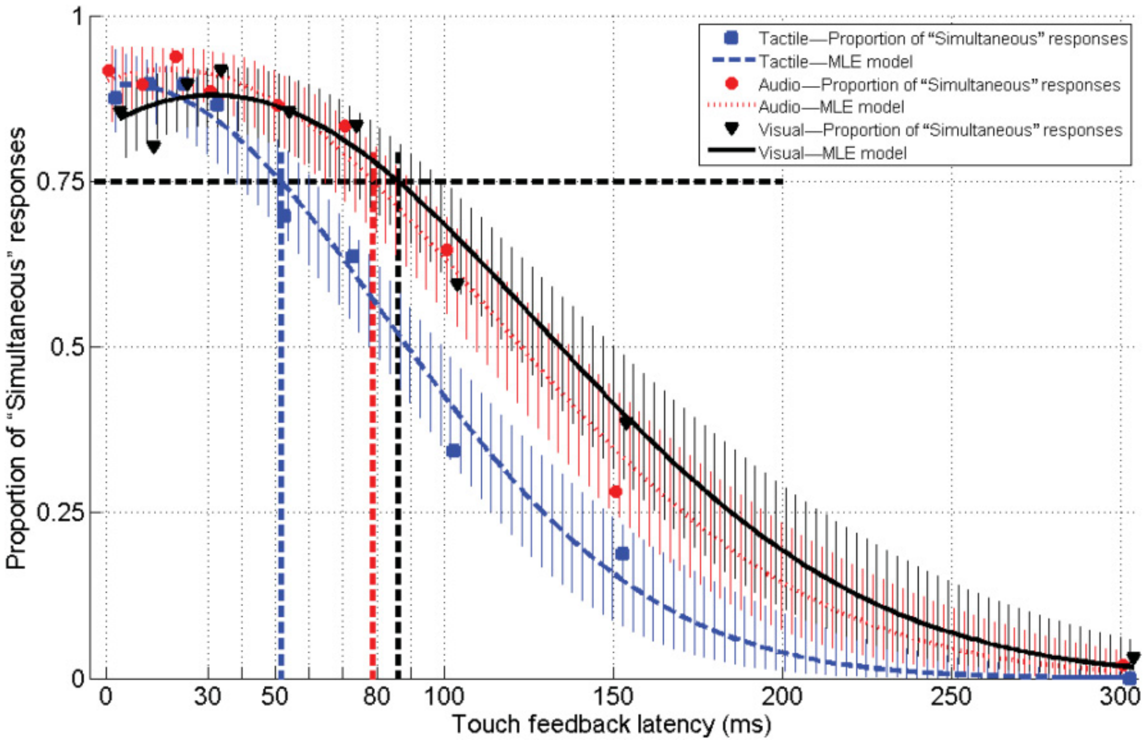

Fig. 7. Proportion of "simultaneous" responses and the corresponding MLE Gaussian functions with the 95% confidence intervals (the line clouds around the Gaussian functions). Vertical dashed lines show the 75% simultaneity perception thresholds. The system baseline latencies have been added to all latency values.

Wilcoxon Signed-Rank tests was conducted with a Bonferroni correction applied, resulting in significance levels set at $p < 0.0019$ and $p < 3.7 \times 10^{-5}$ (corresponding significance levels 5% and 0.1%). The post hoc analysis results are introduced in the significance maps shown later in Figure 10. Significance maps are our way to visualize a relatively complex set of condition comparisons. An example of a significance map is illustrated in Figure 9. The black square means the current feedback condition (modality and latency)—the condition under comparison with the other conditions. If the average quality score of the current combination is statistically significantly higher on a level 5% than of another condition, the other condition is marked green and with a "+." Significance level 0.1% is marked with dark green and an "X." If the average quality score of the current combination is statistically significantly lower on a level 5% than of another condition, the other condition is marked red and marked with an "o." Significance level 0.1% is marked with dark red and an "O." The difference with no statistical significance is coloured gradients either between yellow and green or yellow or red depending on whether the average quality score of the current combination is higher or lower than of another condition. This colouring highlights the relative quality of the current condition.

From the maps, it is easy to see that there was a significant drop in perceived quality between 70 and 100 ms in tactile and audio feedback conditions. The visual modality condition differed from the tactile and audio conditions; the perceived quality dropped significantly only between 100 and 150 ms. The buttons with any feedback with a latency of 300ms were rated significantly lower than the buttons with any feedback with latency from 0 to 150 ms. It also can be seen that the modality conditions did not

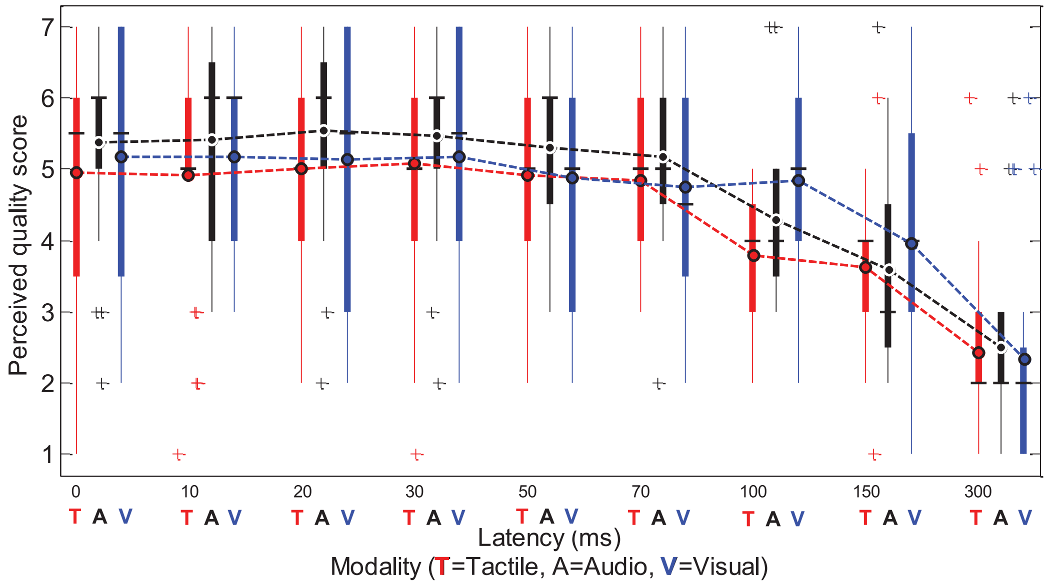

Fig. 8. A boxplot showing medians and the distribution of the scores from the perceived quality questionnaire. The horizontal black lines inside or on the edge of the boxes show medians for each latency and modality condition. The edges of the boxes show the 25th and 75th percentiles of the data, and the whiskers show the most extreme data points not considered outliers [Tukey 1977]. Outliers are presented as "+" marks and are considered only in this visualization. The "o" markers show the means of the data for each latency and modality condition, and the dashed lines show the trendlines.

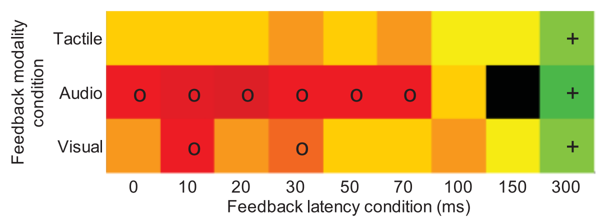

Fig. 9. An example of a significance map used to illustrate the statistical significance differences in perceived quality scores. This figure shows the audio feedback modality and 150ms feedback latency conditions visualizing quality in relation to the other condition combinations. The black square marks the current condition combination (audio, 150ms). A red square with an "o" means that the current condition is statistically significantly lower than the condition marked with red. A green square with a "+" means that the current condition is statistically significantly higher than the condition marked with green. The squares without any mark mean that there is no significant difference. See the text for a more detailed description.

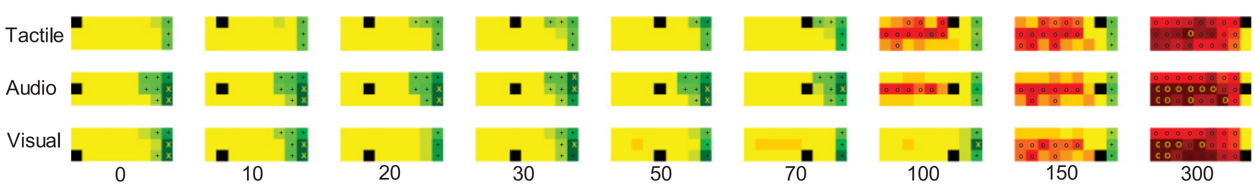

Fig. 10. Significance maps of all feedback modality and latency conditions. The black square means the current condition combination labelled on the horizontal and vertical axes. Each map follows the scheme introduced in Figure 9. See the text for a more detailed description.

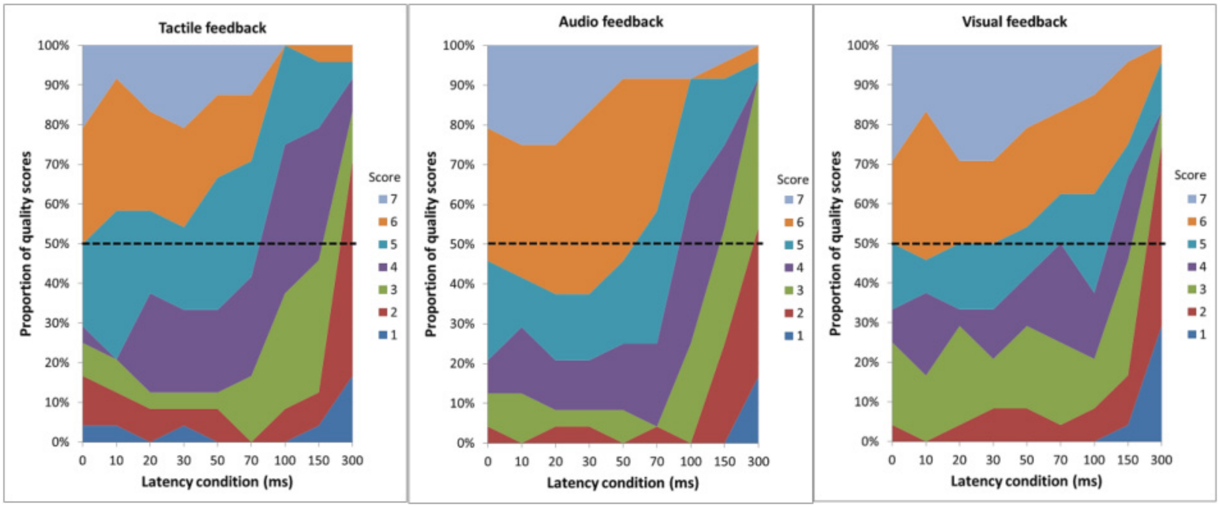

Fig. 11. Proportion of quality scores as a function of latency conditions for each feedback modality condition. The dashed black line shows a 50% threshold. It can be seen that the proportion of favourable ratings (scores from 5 to 7) is more than 50% until the perceived quality is degraded (tactile and audio 100ms and visual 150ms).

differ significantly from each other in any latency condition, even though the mean trendline of audio feedback condition seems to go higher than the tactile or visual feedback (see Figure 8). Figure 11 shows the proportion of each score level as a function of feedback modality and latency conditions. It can be seen that the proportion of favourable ratings (scores from 5 to 7) is more than 50% until the perceived quality is degraded (tactile and audio 100ms and visual 150ms).

These results support (H4)—the perceived quality score for the buttons will drop when latency is higher than 70ms—although the quality dropped even later than hypothesized with visual feedback modality.

5. DISCUSSION

We hypothesized that the distribution of "simultaneous" responses would follow a Gaussian function. We wanted to achieve a general model of touch-feedback simultaneity perception to derive practical design guidelines. Our experimental data and statistical analysis show that the hypothesised Gaussian model was a feasible choice for that purpose. Our results confirm that touch-feedback simultaneity perception behaves in similar manner to the simultaneity perception of exogenously applied stimuli in earlier work (e.g., Stone et al. [2001] and Winter et al. [2008]). In these earlier studies, the model fitting was implemented for individual participants' data. In the current study, we made a practical choice to keep the duration of the test reasonable, because we wanted to inspect the touch-feedback simultaneity, in addition to the perceived quality assessment, with all feedback modalities in the same experiment. More importantly, our objective was to define general design guidelines for the feedback latencies. Thus, we were interested in the general model of touch-feedback simultaneity instead of accurately modelling simultaneity perceptions of individual participants and understanding the differences between them.

We also hypothesized that the PSS would not differ significantly from the actual physical simultaneity (i.e., when feedback comes exactly at the same time as the touch). The results partially supported this. The PSS of touch and tactile feedback was 5ms and did not differ significantly from 0ms. However, the PSS of touch and audio feedback was 19ms, and physical simultaneity was not within the

95% confidence interval, meaning that the PSS was significantly different from 0ms. The PSS of touch and visual feedback latency was 32ms and significantly different from 0ms as well. Since in the case of touch and audio or visual feedback the simultaneity perception happens most likely when there is some latency between the touch and the feedback, it is not necessary to reach zero latency. This is good news for the hardware and software engineers aiming to minimize the touchscreen device latencies; 19ms is enough for touch and audio feedback, and 32ms is enough for touch and visual feedback.

Although we proved that the fit of the Gaussian model was successful, the statistical analysis of the observations did not show any significant peak in proportions of simultaneity perception. Instead, at 0, 10, 20, and 30 ms, the proportion of simultaneity perception of touch and tactile feedback was not significantly different. Similarly, at 0, 10, 20, 30, 50, and 70 ms, the proportion of simultaneity perception of touch and audio or visual feedback was not significantly different. However, it is assumed that the Gaussian function models the simultaneity perception and that the observations would converge to the model if the sample size were large enough. The significant PSS shift from 0ms shown by the Gaussian functions is supported by an additional finding; participants verbally reported in 26% (19/72) of all modality conditions that in some latency conditions, it felt like the feedback was coming *before* the touch. These comments were spontaneous, so the number of this kind of perception could have been higher if we had explicitly asked about it. There might be multiple reasons for this PSS shift. One might be that the participants had certain expectations of the characteristics of a button based on their previous experiences with real buttons. The feedback of a physical button always comes later than the first touch of the finger: the finger compresses before the button goes down and triggers the mechanical feedback. In this experiment, a very slight touch on the button was sufficient to trigger the feedback. The participants might not register the actual press until the finger has compressed and the receptors have been activated at the fingertip. When a feedback is presented to the participant exactly at the same moment that the finger first touches the touchscreen, the expectations are not met and the participant perceives the feedback before registering the actual press. This causes an unnatural button press experience.

Related to the expectations, still another reason might be an adaptation issue. It has been proved that adaptation to certain latencies causes a shift in the PSS [Fujisaki et al. 2004; Harrar and Harris 2005]. The participants have been exposed to the latencies of their own mobile devices. If not too long, they have accustomed to virtual buttons with certain latency, and that is why buttons with shorter latencies, especially 0ms, feel unnatural and can even cause the feeling that the feedback comes earlier than the touch. There might be several reasons why the PSS of touch and tactile feedback is *not* significantly different from 0ms. One explanation might be that tactile feedback is special: it comes to the same finger and receptor cells that feel the touch event and the compression. When the latency is 0ms, tactile feedback most probably goes unnoticed because the compression sensation masks the tactile feedback. When the latency increases, the tactile feedback is still felt in the same finger, but at some point, when the finger is released and is not touching the surface anymore, the tactile feedback is felt only in the hand that holds the device. So, the judging the simultaneity can also be based on these differences rather than the true SJ, as it would be the case when the tactile feedback came in the other hand only, like in the research of Winter et al. [2008].

The practical simultaneity perception thresholds were obtained both by examining the 75% level in the Gaussian models and also by conducting statistical significance analysis of the observations (see Section 4). These results are collected in Table II. We hypothesized (H3) that the touch feedback simultaneity perception 75% threshold will be near 60ms for tactile, 42ms for audio, and 45ms for visual feedback. The derived threshold did not differ significantly from the hypothesized one only when the feedback was tactile (52ms, with a 95% confidence interval of 40 to 62 ms), thus supporting the hypothesis (H3) only partially. The threshold was higher when the feedback was audio (80ms, with

Table II. Summary of the Simultaneity Perception Thresholds and Drops in the Perceived Quality Scores

	Significant Drop in the Proportion of "Simultaneous" Responses	75% Threshold of the Model	Significant Drop in the Perceived Quality Scores
Tactile	20–50ms	52ms	70–100ms
Audio	50–100ms	80ms	70–100ms
Visual	70–100ms	85ms	100–150ms

a 95% confidence interval of 65 to 90 ms) or visual (85ms, with a 95% confidence interval of 70 to 100 ms). These thresholds will be used for deriving guidelines.

There were no significant peaks in the perceived quality scores at latency conditions between 0 and 70 ms when the feedback was tactile or audio, and between 0 and 100 ms when the feedback was visual. The perceived quality score dropped significantly for tactile and audio feedback latencies between 70 and 100 ms and for visual feedback latencies between 100 and 150 ms. This result partially supported the hypothesis (H4)—the perceived quality score for the buttons would drop when latency is larger than 70ms; the quality score dropped only after 100ms when the feedback modality was visual.

From the results, we can conclude that our last hypothesis (H5)—the participants would perceive a drop in quality earlier than the simultaneity perception threshold—was not supported for tactile or visual feedback conditions. The significant drop in the proportion of "simultaneous" responses was before the significant drop in the perceived quality scores in those feedback modalities. It seems that the audio feedback condition was different; the time window where the proportion of the simultaneity perception of touch and audio feedback dropped significantly overlapped with the time window where the perceived quality dropped significantly. In addition, the 75% threshold obtained from the model was indeed inside the time window where the perceived quality dropped significantly. The reason for the difference between audio and the other modalities remains unclear and needs further investigation.

In addition to the thresholds and recommendations, the results can be used to assess the possible simultaneity and quality perception of a virtual buttons in a mobile phone product. The latencies can be measured with a similar tool to that in Kaaresoja and Brewster [2010], and the simultaneity perception models and the perceived quality scores can be used to investigate the possible perceptual consequences of those measured latencies. Our results might also be applied to any programmable buttons that can provide tactile, audio, or visual feedback or to other touchscreen devices such as tablets or tabletop computers.

5.1 Latency Guideline

We have investigated the temporal aspects of touch and feedback from two different angles: simultaneity perception and perceived quality. To summarize the results as a guideline, the recommended minimum latency was selected to be the PSS of the touch and feedback as explained earlier. Since the models were proved to be reliable, the maximum recommended latency was selected both from the models and the significant drop in the perceived quality score: the smaller of either the 75% simultaneity perception threshold or the latency when the perceived quality started to drop. For tactile and visual feedback, the 75% threshold was smaller; for audio feedback, the latency when the perceived quality started to drop was smaller. As the guideline (results rounded to the nearest 5ms), tactile feedback latency should be between 5 and 50 ms, audio feedback latency between 20 and 70 ms, and visual feedback latency between 30 and 85 ms. It must be noted that because these guidelines are based on user preferences, they may change when the technology develops towards virtual buttons with less latency in the future.

Table III. Feedback Latencies for Virtual Buttons in the Default Messaging Application in Five Touchscreen Mobile Phones

Mobile Phone (Operating System)	Tactile Feedback Latency (guideline: 5–50ms)	Audio Feedback Latency (guideline: 20–70ms)	Visual Feedback Latency (guideline 30–85ms)
Nokia Lumia 800 (Windows Phone)	Not supported	37ms (sd 2.2)	53ms (sd 7.1)
Nokia N9 (MeeGo)	35ms (sd 1.7)	38ms (sd 4.7)	110ms (sd 14.5)
Apple iPhone 4S (iOS)	Not supported	102ms (sd 9.1)	83ms (sd 7.4)
HTC Wildfire S (Android)	74ms (sd 4.5)	149ms (sd 10.8)	140ms (sd 9.4)
Samsung Galaxy Note (Android)	123ms (sd 6.0)	172ms (sd 11.9)	197ms (sd 23.5)

Note: The table is sorted according to the average latency of all the feedback. The green highlight shows that the latency was within the guideline set in this study.

5.2 Evaluation of Mobile Devices Latencies

To show how our latency guideline can be put in practice, the latencies of five contemporary mobile phones were measured with the tool introduced by Kaaresoja and Brewster [2010]: HTC Wildfire S running Android, Apple iPhone 4S running iOS, Nokia Lumia 800 running Windows Phone, Nokia N9 running MeeGo, and Samsung Galaxy Note running Android. All wireless functions were switched off in the phones during the measurement to avoid extra variance in latencies. The default text message application was opened, and for the measurement, the "g" key was pressed 20 times. The audio and tactile latencies were measured as the time between the first moment of the finger touch and the first local intensity maximum of the feedback. The visual feedback latency was the time between the first moment of the finger touch and the moment when the visual popup of the key was fully drawn on the screen. The measurement results were reflected against the guideline just introduced. The results can be seen in Table III. Some of the phones perform very well according to our guidelines. Some phones have latencies higher than the guidelines, meaning that many users would perceive the latency between the touch and feedback or rate the quality of the buttons interaction as lower, both of which are undesirable when producing a high-quality product. The results show that the Nokia Lumia 800 had audio and visual feedback latencies within our guideline. The Nokia N9 had tactile and audio feedback latencies within the guideline. The visual feedback latency in the Apple iPhone 4S was also within the guideline. These results are shown in Table III in green. The rest of the feedback had longer latencies than recommended in the guidelines. None of the phones that provided all three forms of feedback did so within our latency guidelines for each modality.

6. CONCLUSIONS AND FUTURE WORK

Our research shows for the first time that the perception of simultaneity of touch and tactile, touch and audio, and touch and visual feedback in a realistic setup can all be modelled with a Gaussian function. This confirms the results of Winter et al. [2008] and suggests that the simultaneity perception of an action and passive event follows a Gaussian function just like the simultaneity perception of two passively received events, as is usually investigated in simultaneity perception research. In this work, we wanted to understand simultaneity perception in a particular context and task with practical interactions; the research device and task were designed to be as mobile-phone-like as possible to ensure that the results would be usable for touchscreen mobile device designers. Our approach was to ensure perceived simultaneity of touch and feedback to make the users' experience as natural as possible, mimicking the physical buttons users are accustomed to. The participants pressed capacitive buttons, and the associated feedback was provided from the same device as in a real mobile phone. Next, we asked participants to judge if the feedback was simultaneous with the touch. The Gaussian models were convenient tools for finding parameters for applicable guidelines. It was found that the PSS according to the Gaussian models were not the same as physical simultaneity; the PSS of touch and tactile feedback was 5ms, touch and audio feedback 19ms, and touch and visual feedback 32ms. To establish practical guidelines, the 75% thresholds were obtained from the Gaussian models: 52ms for tactile feedback, 80ms for audio feedback, and 85ms for visual feedback.

To further understand the effect of latency to the user experience, we asked the participants to score the perceived quality of the buttons. We found that the scores dropped between latency conditions 70 and 100 ms when the feedback modality was tactile and audio, and between 100 and 150 ms when feedback was visual. Although we did not perform any correlation statistics, these results suggest that simultaneity perception reflects perceived quality: on average, when the participants perceived touch and feedback as simultaneous, they also scored the quality higher than when they perceived the touch and feedback as nonsimultaneous. Thus, the initial quality perception assessment reinforced the simultaneity perception findings in this study.

Practical guidelines for interaction designers were established for the first time. The guidelines recommend that (rounded to the nearest 5ms) tactile feedback latency should be between 5 and 50 ms, audio feedback latency between 20 and 70 ms, and visual feedback latency between 30 and 85 ms in capacitive touchscreen virtual button interaction. These guidelines have a two-fold importance to the field. First, hardware and software engineers do not need to optimize the latency between touch and feedback towards 0ms. Second, these numbers ensure that the majority of users will either feel the feedback as simultaneous with their touch or feel no degradation in quality of the buttons, ensuring a good user experience.

The natural continuation of this work is to provide feedback consisting of two or three modalities to further specify the latency guidelines by finding out the thresholds for the simultaneity perception and perceived quality. Testing more modality combinations for feedback is valuable because virtual buttons in mobile phones often include two or even three modalities. Using the Virtual Button Simulator would be still necessary, as the latencies are usually long and variable in real touchscreen phones. However, conducting these experiments with real mobile phones would further validate the results achieved in this work when taking their limitations into account. As stated earlier, in our study, we did not model the simultaneity perception for individual participants as done usually in pure psychophysical experiments (as we had a more practical application for our work). Future work in psychophysics should include experiments collecting more data per feedback modality so that the simultaneity perception of each participant can be modeled, PSS and JND derived, and statistics done. It would be interesting to see the differences between different modalities and the distribution of PSS and JNDs in this kind of ecologically valid but unexplored context.

In conclusion, our results provide valuable guidance for touchscreen interaction design and enable the creation of better user interfaces for this rapidly growing area of human-computer interaction.

ACKNOWLEDGMENTS

We thank Craig Stewart, Antti Rönkkö, and Tom Ahola for helping to build the Virtual Button Simulator, and Teemu Ahmaniemi, Johan Kildal, and Hong Z. Tan for valuable feedback, support, and comments on the manuscript.

REFERENCES

B. D. Adelstein, D. R. Begault, M. R. Anderson, and E. M. Wenzel. 2003. Sensitivity to haptic-audio asynchrony. In *Proceedings of the 5th International Conference on Multimodal Interfaces*. ACM Press, New York, NY, 73–76. DOI:http://dx.doi.org/10.1145/958432.958448

E. Boring. 1923. *A History of Experimental Psychology*. Pendragon, New York, NY.

S. Brewster. 2002. Overcoming the lack of screen space on mobile computers. *Personal and Ubiquitous Computing* 6, 188–205. DOI:http://dx.doi.org/10.1007/s007790200019

S. Brewster, F. Chohan, and L. Brown. 2007. Tactile feedback for mobile interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'07)*. ACM Press. DOI:http://dx.doi.org/10.1145/1240624.1240649

S. Coren, L. M. Ward, and J. T. Enns. 2003. *Sensation and Perception*. Wiley & Sons.

S. Exner. 1875. Experimentelle Untersuchung der einfachsten psychischen Processe. *Archiv für die gesamte Physiologie des Menschen und der Tiere* 11, 403–432.

W. Fujisaki, S. Shimojo, M. Kashino, and S. Y. Nishida. 2004. Recalibration of audiovisual simultaneity. *Nature Neuroscience* 7, 773–778.

M. Fukumoto and T. Sugimura. 2001. Active click: Tactile feedback for touch panels. In *Proceedings of Extended Abstracts on Human Factors in Computing (CHIEA'01)*. ACM Press, New York, NY, 121–122. DOI:http://dx.doi.org/10.1145/634067.634141

V. Harrar and L. R. Harris. 2005. Simultaneity constancy: Detecting events with touch and vision. *Experimental Brain Research* 166, 465–473. DOI:http://dx.doi.org/10.1007/s00221-005-2386-7

L. R. Harris, V. Harrar, P. Jaekl, and A. Kopinska. 2010. Mechanisms of simultaneity constancy. In *Space and Time in Perception and Action*, R. Nijhawan (Ed.). Cambridge University Press, Cambridge, UK, 232–253. DOI:http://dx.doi.org/10.1017/CBO9780511750540.015

D. He, F. Liu, D. Pape, G. Dawe, and D. Sandin. 2000. Video-based measurement of system latency. In *Proceedings of the IPT2000 International Immersive Projection Technology Workshop*.

E. Hoggan, S. A. Brewster, and J. Johnston. 2008. Investigating the effectiveness of tactile feedback for mobile touchscreens. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'08)*. 1573–1582. DOI:http://dx.doi.org/10.1145/1357054.1357300

C. Jay and R. Hubbold. 2005. Delayed visual and haptic feedback in a reciprocal tapping task. In *Proceedings of the 1st Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*. IEEE Computer Society, Washington, DC, 655–656.

R. Jota, A. Ng, P. Dietz, and D. Widgor. 2013. How fast is fast enough?: A study of the effects of latency in direct-touch pointing tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM Press, New York, NY, 2291–2300. DOI:http://dx.doi.org/10.1145/2470654.2481317

T. Kaaresoja, E. Anttila, and E. Hoggan. 2011a. The effect of tactile feedback latency in touchscreen interaction. In *Proceedings of the World Haptics Conference (WHC)*. IEEE Computer Society, Washington, DC, 65–70.

T. Kaaresoja and S. Brewster. 2010. Feedback is... late: Measuring multimodal delays in mobile device touchscreen interaction. In *Proceedings of the International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction (ICMI-MLMI'10)*. ACM Press, New York, NY, Article 2. DOI:http://dx.doi.org/10.1145/1891903.1891907

T. Kaaresoja, L. M. Brown, and J. Linjama. 2006. Snap-crackle-pop: Tactile feedback for mobile touch screens. In *Proceedings of Eurohaptics 2006*. 565–566.

T. Kaaresoja, E. Hoggan, and E. Anttila. 2011b. Playing with tactile feedback latency in touchscreen interaction: Two approaches. In *Proceedings of the 13th IFIP TC 13th International Conference on Human-Computer Interaction—Volume Part II (INTERACT'11)*. Springer-Verlag, Berlin, Heidelberg, 554–571.

D. J. Levitin, K. Maclean, M. Mathews, and L. Chu. 1999. The perception of cross-modal simultaneity. In *Proceedings of the 3rd International Conference on Computing Anticipatory Systems*. 1999.

S. MacKenzie and C. Ware. 1993. Lag as a determinant of human performance in interactive systems. In *Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems (CHI'93)*. ACM Press, New York, NY, 488–493.

R. C. Miall and J. K. Jackson. 2006. Adaptation to visual feedback delays in manual tracking: Evidence against the Smith Predictor model of human visually guided action. *Experimental Brain Research* 172, 1, 77–84.

R. B. Millar. 2011. *Maximum Likelihood Estimation and Inference: With Examples in R, SAS and ADMB.* John Wiley & Sons, West Sussex, United Kingdom.

D. Miller and G. Bishop. 2002. Latency meter: A device for easily monitoring VE delay. In *Proceedings of SPIE Vol. #4660 Stereoscopic Displays and Virtual Reality Systems IX*, San Jose, CA.

G. A. Miller. 1956. The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63, 81–97.

A. Ng, J. Lepinski, D. Widgor, S. Sanders, and P. Dietz. 2012. Designing for low-latency direct-touch input. In *Proceedings of the UIST'12*, St Andrews, UK, 2012, ACM, 453–464. DOI: http://dx.doi.org/10.1145/2380116.2380174

I. Poupyrev and S. Maruyama. 2003. Tactile interfaces for small touch screens. In *Proceedings of the 16th Annual Symposium on User Interface Software and Technology (UIST'03)*. ACM Press, New York, NY, 217–220. DOI:http://dx.doi.org/10.1145/964696.964721

I. Poupyrev, M. Okabe, and S. Maruyama. 2004. Haptic feedback for pen computing: Directions and strategies. In *Proceedings of Extended Abstracts on Human Factors in Computing Systems (CHIEA'04)*. ACM Press, New York, NY, 1309–1312. DOI:http://dx.doi.org/10.1145/985921.986051

J. V. Stone, N. M. Hunkin, J. Porrill, R. Wood, V. Keeler, M. Beanland, M. Port, and N. R. Porter. 2001. When is now? Perception of Simultaneity. *Proceedings of the Royal Society of London: Series B*, 31–38.

J. W. Tukey. 1977. *Exploratory Data Analysis*. Addison-Wesley.

I. M. L. C. Vogels. 2004. Detection of temporal delays in visual-haptic interfaces. *Human Factors* 46, 118–134.

R. Winter, V. Harrar, M. Gozdzik, and L. R. Harris. 2008. The relative timing of active and passive touch. *Brain Research* 1242, 54–58.

M. Zampini, S. Guest, D. I. Shore, and C. Spence. 2005. Audio–visual simultaneity judgments. *Perception and Psychophysics* 67, 531–544.

Received June 2013; revised April 2014; accepted April 2014