Latency Guidelines for Touchscreen Virtual Button Feedback

Topi Johannes Kaaresoja
Master of Science

Submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy

School of Computing Science
College of Science and Engineering
University of Glasgow

March 2015

Abstract

Touchscreens are very widely used, especially in mobile phones. They feature many interaction methods, pressing a virtual button being one of the most popular ones. In addition to an inherent visual feedback, virtual button can provide audio and tactile feedback. Since mobile phones are essentially computers, the processing causes latencies in interaction. However, it has not been known, if the latency is an issue in mobile touchscreen virtual button interaction, and what the latency recommendations for visual, audio and tactile feedback are.

The research in this thesis has investigated multimodal latency in mobile touchscreen virtual button interaction. For the first time, an affordable, but accurate tool was built to measure all three feedback latencies in touchscreens. For the first time, simultaneity perception of touch and feedback, as well as the effect of latency on virtual button perceived quality has been studied and thresholds found for both unimodal and bimodal feedback. The results from these studies were combined as latency guidelines for the first time. These guidelines enable interaction designers to establish requirements for mobile phone engineers to optimise the latencies on the right level.

The latency measurement tool consisted of a high-speed camera, a microphone and an accelerometer for visual, audio and tactile feedback measurements. It was built with off-the-shelf components and, in addition, it was portable. Therefore, it could be copied at low cost or moved wherever needed. The tool enables touchscreen interaction designers to validate latencies in their experiments, making their results more valuable and accurate. The tool could benefit the touchscreen phone manufacturers, since it enables engineers to validate latencies during development of mobile phones. The tool has been used in mobile phone R&D within Nokia Corporation and for validation of a research device within the University of Glasgow.

The guidelines established for unimodal feedback was as follows: visual feedback latency should be between 30 and 85 ms, audio between 20 and 70 ms and tactile between 5 and 50 ms. The guidelines were found to be different for bimodal feedback: visual feedback latency should be 95 and audio 70 ms when the feedback was visual-audio, visual 100 and tactile 55

ms when the feedback was visual-tactile and tactile 25 and audio 100 ms when the feedback was tactile-audio. These guidelines will help engineers and interaction designers to select and optimise latencies to be low enough, but not too low. Designers using these guidelines will make sure that most of the users will both perceive the feedback as simultaneous with their touch and experience high quality virtual buttons.

The results from this thesis show that latency has a remarkable effect on touchscreen virtual buttons, and it is a key part of virtual button feedback design. The novel results enable researchers, designers and engineers to master the effect of latencies in research and development. This will lead to more accurate and reliable research results and help mobile phone manufacturers make better products.

Table of Contents

List of Tables

List of Figures

Acknowledgements

First of all, I want to give my thanks to Professor Stephen Brewster for excellent supervision, encouragement, leadership and invaluable advice during all these years. Without him, I could never had finished this project. I would also like to thank my second supervisor, Professor Roderick Murray-Smith. I also want to say thanks to Adjunct Professor Vuokko Lantz in Nokia Research Center, who encouraged me to kick off the whole PhD process.

Although being a remote member of the Multimodal Interaction Group at Glasgow, I want to say thanks to all group members, in particular, Dr Graham Wilson kindly reviewing drafts of this thesis' chapters and Craig Stewart introducing me the world of Arduino.

I want to thank my dear colleagues Dr Teemu Ahmaniemi and Dr Johan Kildal in Nokia for support and encouragement that have been priceless during ups and downs. Many thanks to my colleagues Tom Ahola and Antti Rönkkö helping me and participating in building the research gear. I would also like to thank the following Research Leaders and Directors in Nokia, who enabled the funding and the freedom to conduct my research in the corporate research environment: Viljakaisa Aaltonen, Dr Jyri Huopaniemi, Dr Tapani Ryhänen, Dr Henry Tirri and Marja Salmimaa. I want also to say thanks to a countless number of colleagues with whom I have had privilege to work with in Nokia.

My special thanks go to my dear wife, Kati, for all loving, supporting and encouragement. I also would like to thank all my relatives and friends, who have given helping hand and support. Finally, I want to express the greatest gratitude go to my Highest Supervisor, God – Heavenly Father, Jesus and Holy Spirit – for Love, Caring, Help and Peace.

This research was funded by Nokia Research Center and Nokia Foundation. Final thesis printing and binding were funded by Ilves-Paino Oy, Hämeenlinna, Finland and Kiinteistömaailma | LKV Autero Oy, Helsinki, Finland.

Dedication

To Kati, my Love, wife and best friend. You are and will always be the number one to me.

To my lovely daughters Meri and Helmi. You are and will always be my little princesses.

Minä rakastan teitä!

Declaration

The material presented in this thesis is the result of my own research carried out at Nokia Research Center in Helsinki and Espoo, Finland and the Department of Computing Science at the University of Glasgow, Scotland, UK working under the supervision of Research Leader and Adjunct Professor Vuokko Lantz, Research Leader Marja Salmimaa and Research Leader Viljakaisa Aaltonen in Nokia and of Professor Stephen Brewster and Professor Roderick Murray-Smith in the University of Glasgow.

This thesis exploits only the parts of these papers that are directly attributable to the author:

The multimodal latency measurement tool in Chapter 3 has been reported in International Conference of Multimodal Interaction (ICMI) 2010 (Kaaresoja and Brewster, 2010):

Kaaresoja, T. and Brewster S. (2010). Feedback is… Late: Measuring Multimodal Delays in Mobile Device Touchscreen Interaction. *ICMI'2010, Beijing, China*, ACM.

Experiment 1 in Chapter 4 has been published in the ACM Transactions of Applied Perception (TAP) journal (Kaaresoja *et al.*, 2014):

Kaaresoja, T., Brewster, S. and Lantz V. (2014). "Towards the Temporally Perfect Virtual Button: Touch-Feedback Simultaneity and Perceived Quality in Mobile Touchscreen Press Interactions." *ACM Trans. Appl. Percept.* 11(2): 1-25.

Experiment 2 in Chapter 5 has been submitted and reviewed for publication in the ACM Transactions of Applied Perception (TAP) journal:

Kaaresoja, T., Brewster, S. and Lantz V. (2015). "Guidelines for Feedback Latency in Touchscreen Press Interactions" submitted to *ACM Trans. Appl. Percept* and reviewed.

Chapter 1 Introduction

1.1 Motivation

Touchscreens are very widely used. One can find touchscreens in wrist devices[1], mobile phones[2], tablets[3], tabletop computers[4], copying machines[5], vending machines[6], car navigation and entertaining systems[7], dental equipment[8] and even power plants (Carvalho *et al.*, 2011). However, the vast majority of touchscreen devices are consumer mobile products; 1.3 billion mobile touchscreen devices were shipped in 2012 and it is predicted that this number will double by 2016 (Wu and Yi, 2013). Most of these devices are mobile phones, making them the most familiar device in users' hands. A touchscreen phone is most commonly used with a single finger, multiple fingers or sometimes a stylus. There are various ways of interacting with a touchscreen: sliding a virtual slider or flicking or panning the screen content, for example. Despite these 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 and entering keywords to search a topic on the Internet.

In addition to the visual feedback given for touchscreen button presses, virtual buttons can also provide audio and tactile feedback to the user, to mimic physical buttons. Audio feedback has been found to improve performance, reduce errors and reduce workload in touchscreen button interaction (Brewster, 2002). The same effects have been found when applying tactile feedback used with a stylus (Brewster *et al.*, 2007) and finger (Hoggan *et al.*, 2008a). Visual feedback may take the form of colour or shadow change of a button when

---

[1] www.samsung.com/us/mobile/wearable-tech

[2] www.apple.com/iphone

[3] www.microsoft.com/Surface

[4] www.microsoft.com/en-us/pixelsense

[5] www.office.xerox.com/multifunction-printer/color-multifunction/workcentre-7800-series

[6] http://www.wordpress.tokyotimes.org/a-touchscreen-vending-machine-in-tokyo

[7] http://www.toyota.co.uk/owners-info/touch-and-go

[8] http://www.planmeca.com/

pressed and released. Audio feedback can be beeps, clicks, or other short sounds from a loudspeaker. Tactile feedback often follows the characteristics of the audio being a short click or vibration and is provided by a rotational, linear, or piezoelectric actuator.

Although devices are becoming faster, operating systems and applications are becoming more complex. There is always latency, or delay, 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. The sources of latency include the time needed to recognize the interaction technique, or touchscreen gesture, the user intended, and processing time to interpret the input and calculate the response (Anderson *et al.*, 2011). In addition to these latencies, a capacitive touch sensor causes latency because of its function. The location of a finger is scanned by the sensor with a certain sampling rate which takes time, and often several scanning cycles are needed in order to reliably recognize the finger position. The feedback production also takes time. It takes time for the visual display to change from one colour to another for visual feedback. The audio generation pipeline usually includes buffers, which again can cause latency, if the audio data used for the feedback is not stored but generated for every button press. Commonly used rotational tactile actuators can suffer from a slow startup time because of the inertia of the weight to be moved to generate mechanical movement.

Latency affects a device's responsiveness and the perceived ability of the device to react to the user's input (Anderson *et al.*, 2011; Ng *et al.*, 2014). That is why latency can be harmful in interaction. It has been stated that latency is one of the major issues limiting the quality, interactivity, and effectiveness of virtual and augmented reality (Allison *et al.*, 2001; 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; Pavlovych and Stuerzlinger, 2009) or 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). As Hinckley and Widgor (2012) state, latency can be especially harmful in direct input devices such as touchscreens used with a finger or stylus. Latency has been shown to degrade subjective satisfaction in touchscreen interaction (Kaaresoja *et al.*, 2011a; Kaaresoja *et al.*, 2011b) as well as user performance (Jota *et al.*, 2013). Latency also increases user annoyance (Anderson *et al.*, 2011). From all this prior research it can be concluded that latency needs to be explored to fully understand its consequences on perception and interaction in touchscreens.

In order to understand if latency is an issue or not and, if so, take corrective actions, it has to be possible to measure the latencies between user action and device response in the visual, audio and haptic modalities. There are several research prototypes and methods (He *et al.*, 2000; Miller and Bishop, 2002; Lehtosalo, 2009; Montag *et al.*, 2011) for latency measurements in different contexts. However, these are all unimodal, meaning that only one of the visual, audio or tactile latencies is measured at a time. Commercial products[9] exist for multimodal latency measurements in different contexts, but they are big, expensive and clumsy. Therefore, an affordable, portable, but still accurate multimodal latency measurement tool for touchscreen interaction is introduced in this thesis. The use of the tool requires no changes or modifications to the device being measured. It uses mostly off-the-shelf components and free software and is capable of measuring latencies accurately between different events in different modalities. The target devices were mobile touchscreen devices, but with minor modifications the tool could also be used in other domains.

As latency causes a system to be slower, degrading the user experience, it is natural to conclude that simultaneity, where there is no latency, would enable an improved user experience through responsiveness. 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, the motivation of the research in this thesis was to find the simultaneity perception thresholds of touch and feedback. From these, it can be derived how the different feedback modalities need to be optimized to create effective and high-quality interactions. As simultaneity perception has been widely studied in psychophysics, an applied psychophysical approach to the simultaneity perception of touch and feedback was taken. In addition, to further understand how user experience changes as a function of latency, one qualitative dimension of virtual button latency was examined: perceived quality. 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 is lower. The ultimate aim was to

[9] www.blackboxtoolkit.com, www.optofidelity.com

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.

Therefore, two experiments were designed to achieve the goals described above. In these studies, participants pressed simulated virtual touchscreen buttons with a finger and received either unimodal or bimodal feedback in Experiment 1 and Experiment 2, respectively. The amount of 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. The results were combined as guidelines.

1.2 Thesis Statement

It is not known what the recommended latency for virtual button feedback in touchscreen is. An affordable latency measurement tool was built and two extensive experiments conducted to find out simultaneity and quality perception thresholds. From the thresholds, latency guidelines were derived and commercial touchscreen products were measured with the tool and validated against the guidelines.

1.3 Research Questions

This thesis aims to answer the following research questions:

RQ1: Can an affordable and accurate touchscreen latency measurement tool be built?

RQ2: What are the touch-feedback simultaneity perception thresholds in virtual button interaction?

RQ3: How does the perceived quality of a virtual button change when latency between touch and feedback changes?

RQ 4: What are the latency guidelines recommended for visual, audio and tactile feedback for a virtual button press?

1.4 Terminology

Feedback. A response to an input (Hinckley and Wigdor, 2012). The British Standard ISO 9241-400 (ISO, 2007) definition is as follows: "*An input device shall provide effective feedback, i.e. the user is given an immediately perceptible and easily understandable indication that the device is responding to user actuation*". Touchscreen virtual button feedback can consist of visual (Sears, 1991), audio (Sears, 1991) and tactile (Fukumoto and Sugimura, 2001) events. In this thesis, the feedback is defined as the response associated with a finger touch on a surface of a touchscreen (not finger release).

Latency. The time between two physical events. Walker (1995) defines latency in computer science as "*Delay, in digital computers, between the initiation of the call for data and the start of the transfer*". MacKenzie and Ware (1993) refine this definition into the HCI field: "*The delay between input action and output response*". Hinckley and Widgor (2012) define latency broadly as an "*end-to-end measure of the time elapsed between the moment a physical action is performed by the user, versus the moment the system responds to it with feedback that the user can perceive*". Jota *et al.* (2013) define latency as "*the lag between a finger touch and the on-screen response*". These definitions are about ***physical latency*** (separate from ***perceived latency***) which can be measured by timekeeping. Based on the definitions above, latency is defined in this thesis as follows: ***Time between the first moment of touch and the first intensity maximum of the feedback.*** In that definition it must be assumed that the feedback is measurable and perceivable.

Simultaneity. ***Physical simultaneity*** occurs, when two or more events happen exactly at the same time. Jammer (2006) defines this broadly as the "*Temporal coincidence of events*". Power (2011) defines ***perceived simultaneity*** as follows: "*A and B appear to be simultaneous if they seem to happen at the same time. So, there will be an illusion of simultaneity if two perceived events seem to be happening at* *the same time but are not happening at the same time.*” Power defines the simultaneity of two events, but more than two can be perceived as simultaneous.

Touch. A light stroke, tap or push (Merriam-Webster, 2015). Touch is defined in this thesis as the first moment of the finger or stylus tap contacting the surface of a touchscreen.

Multimodal. Charwat (Charwat, 1992; Schomaker *et al.*, 1995) defines *modality* as follows: “Perception via one of the three perception-channels. You can distinguish the three modalities: visual, auditive and tactile (physiology of senses)”. Schomaker *et al.* (1995) notes that “whenever more than two of these modalities are involved, we will speak of *multimodality*. To be more precise, in some cases we will also use the term *bimodal* (or bimodality) to denote the usage of exactly two different modalities”. Vitense *et al.* (2003) use the terms *unimodal*, bimodal and *trimodal* feedback to mean feedback consisting of one, two and three modalities. In the research in this thesis, multimodal feedback consists of visual, audio and tactile modalities. It can be unimodal, bimodal or trimodal. Other modalities, such as smell and taste are beyond the scope of this thesis.

1.5 Thesis Overview

Chapter 2, *Literature Review*, reviews the literature on the latency measurement methods in various domains such as virtual environments, computer music systems and touchscreen devices. In this chapter, earlier work in simultaneity perception and effect of latency on indirect and touchscreen interaction are introduced both from psychophysics and human-computer interaction perspectives.

Chapter 3, *Multimodal Latency Measurement Tool*, introduces the design and implementation of the multimodal latency measurement tool, which was used to assess the latency between touch and visual, audio and tactile feedback.

Chapter 4, *Latency Guidelines for Unimodal Feedback*, first reports the design and implementation of Virtual Button Simulator, the research tool developed for latency research in touchscreen virtual button interaction. Then Experiment 1 is reported, which tested the perceived simultaneity and quality of touch and unimodal feedback. Based on the results, latency guidelines for unimodal feedback were established.

Chapter 5, *Latency Guidelines for Bimodal Feedback*, reports Experiment 2, which was the direct continuation to the previous Experiment. In Experiment 2 the perceived simultaneity and quality perception of touch and bimodal feedback was examined. Based on the results, latency guidelines for bimodal feedback were established.

Chapter 6, *Discussion and Conclusions*, introduces the review and summary of the thesis, including the novel contributions and the answers for the research questions. Limitations of the research are discussed as well as the possibilities for the future research, such as simultaneity and quality perception of touch and trimodal feedback. Based on the results, latency guidelines for trimodal feedback could be established.

Chapter 2 Literature Review

2.1 Introduction

The aims of the research in this thesis were to investigate and understand the measurement and perceptual consequences of multimodal latency in mobile touchscreen virtual button interaction. This chapter introduces the literature related to these topics, including latency measurement technology and methods, temporal perception research and the effects of latency on usability and user experience, especially in touchscreen interaction. The purpose of this chapter is to give an overview of the field in order to place the contributions of this thesis in context. Figure 2-1 shows the overview of the context of this research and the literature review. It also shows the focus of the research in yellow.

The review begins with an introduction to touchscreens and touchscreen interaction followed by the fundamental touchscreen virtual button feedback modalities – visual, audio and tactile – including their design and benefits, especially in mobile touchscreen interaction. The anatomy of a multimodal virtual button press is also described and explained in detail. It is important to understand the fundamental structure of a multimodal virtual button press before planning any measurements or building experiments on that. There are several phases included in the simple press of a virtual button and these events also tend to have different temporal characteristics. One of the goals of the research in this thesis is to isolate them one by one and test the effect of their latency on perceived simultaneity and quality. In addition to the perception studies, the latency measurement method will be different for the different modalities. One of the aspects of taking apart the virtual button press is to delimit this work context.

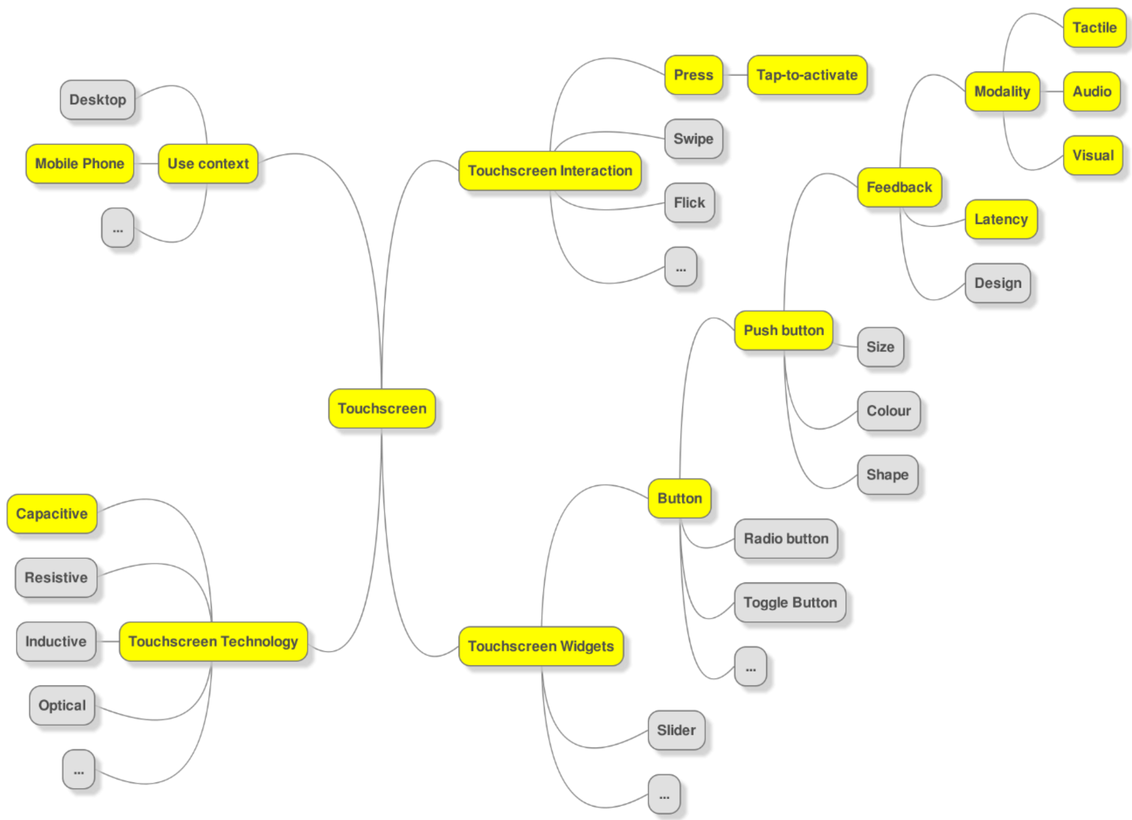


Figure 2-1: The overview of the scope and context of the research in this thesis is latencies in mobile touchscreens. The research focuses on virtual button interaction.

After the fundamentals, latency measurement technology and methods are introduced. There are several techniques and attempts to measure latencies in interactive systems, such as virtual environments, head tracking systems and computer music systems, for example. There are a few commercial timing measurement systems in the market as well.

Following the feedback and latency measurement literature review is a discussion how human perception of simultaneity of two stimuli is measured. This is followed by a review of the research findings in perceived simultaneity of different events; first for exogenous stimuli, meaning stimuli the participants passively experience, and then for the interaction and the feedback.

After the psychophysics literature review, the consequences of latency on usability and user experience in indirect manipulation, such as using a mouse, are introduced. Next, before the conclusions, the latency research in direct manipulation i.e., touchscreen interaction, is introduced, especially in a mobile context. Last section concludes the chapter and addresses all the research questions against the literature reviewed.

2.2 Touchscreens

The British Standard ISO 9241-400 (2007) defines touchscreens (Touch Sensitive Screens – TSS) as an "input device that produces a position and selection input signal from a finger touching, lifting off or moving across a display". In other words, a touchscreen is a computer or mobile device screen which allows the interaction with the device via the graphical user interface (GUI) directly, with a finger or a stylus (although the standard mentions only a finger). A touchscreen is one of the direct input devices, in contrast to indirect input devices such as a mouse or touchpad (Hinckley and Wigdor, 2012). Touchscreens have been shown to have many advantages, such as the directness itself, speed of use, ease of learning and flexibility of the device real estate usage, as stated by Sears *et al.* (1991) more than 20 years ago. Today, touchscreens are very widely used. The vast majority of the touchscreen devices are consumer mobile devices.

The two major technologies used in touchscreen touch sensor in mobile devices are resistive and capacitive (Nichols, 2007). Resistive touch sensors can be used either with a finger or stylus, but are vulnerable to scratches and require some force from the user to activate. Capacitive touchscreens are now more commonly used, especially in high-end mobile devices such as smartphones and tablets. They require only a light touch; in fact, the user does not need to press the screen at all in order to activate it. This possibility of light touch has also enabled other useful interaction techniques, such as flick, pan and multitouch. Since mobile phones with capacitive touchscreens are the mainstream and the research in this thesis was conducted in Nokia – a company making mobile phones until 2014 – this research focuses on the capacitive touchscreen used with a finger.

2.3 Touchscreen Interaction

There are various ways to interact with a touchscreen. Saffer (2009) introduces touchscreen interactions as "touchscreen gestures", referring to touchscreen usage with a finger rather than a stylus. According to Saffer, the main gesture categories are Tap, Drag, Slide, Spin, Flick, Fling and Pinch. A virtual button press, which is a focus of the research in this thesis, is equivalent to “Tap to Activate” as introduced by Saffer.

Sears *et al.* (1991) discuss “land-on” and “lift-off” activation strategies in touchscreen virtual button interaction. The land-on strategy activates the function when the finger touches the button whereas the lift-off strategy activates it when the user releases the finger from the button. It is recommended by Sears *et al.* that the land-on strategy would be used with big buttons and lift-off with small buttons, since the latter makes the correction of the press easier with the more error-prone small buttons. Both strategies can be found in contemporary mobile phones: Land-on in the dialler, which usually features bigger buttons, and lift-off for messaging or other text entry applications, which feature a full QWERTY keypad with small buttons. Activation strategy, however, refers to the action feedback introduced in Section 2.5 and a full description is beyond the remit of this thesis, since it is not directly part of the virtual button’s characteristics. Regardless of the activation strategy, virtual buttons can have feedback either on 1) touch, 2) release or 3) both as explained in Section 2.5. The research in this thesis focuses on touch-related feedback. Before looking closer at the anatomy of a multimodal virtual button press, touchscreen feedback, different feedback modalities, their design, and effect on the usability and user experience are addressed.

2.4 Touchscreen Feedback

Feedback is an essential part of every user interface, for the user to acknowledge that the device is responding to his/her actions, as the British Standard ISO 9241-400 (2007) defines: “An input device shall provide effective feedback, i.e. the user is given an immediately perceptible and easily understandable indication that the device is responding to user actuation”. The aim of the virtual button feedback is to indicate to the user that the button has been pressed correctly and to simulate some of the experience of pressing a real, physical button. Later in this thesis, feedback refers always (if not otherwise stated) to touchscreen virtual button feedback.

Research on touchscreen virtual button interaction almost always includes visual feedback (although some exceptions exist). When combined with audio or tactile feedback, or both, it makes the virtual button feedback multimodal. Thus, multimodal feedback can be visual-audio, visual-tactile, tactile-audio, or visual-audio-tactile. Research that focuses on the addition of audio or tactile feedback to visual feedback is introduced in separate sections for audio and tactile feedback after the section for visual feedback. The research on other feedback combinations is discussed in Section 2.4.4 Multimodal Feedback. Later in this thesis, when there is only one feedback modality involved, it is called as unimodal feedback. If the feedback consists of two modalities, it is called as bimodal.

2.4.1 Visual Feedback

Visual feedback representing a button is part of the graphical user interface (GUI) and is a fundamental part of the *screen* in a touchscreen, since the screen is showing all the graphics and visual elements. In this section, different visual feedback designs are listed as a summary and the research on the effects of different visual feedback designs is introduced in the following section.

2.4.1.1 Visual Feedback Designs

There are many suggestions in the literature about how the visual feedback of a touchscreen virtual button should be implemented. However, they often introduce one design or just mention that visual feedback is important. The British Standard ISO 9241-400 (2007) does not give guidelines for the detailed graphical design of the feedback, either. Different visual feedback designs for both desktop and mobile touchscreen virtual button press have been proposed in the literature and commercial products, however.

Desktop:

- Change in appearance from hollow to solid (Bennion *et al.*, 1981)
- Graphics change (Valk, 1985)
- Invert a button colour (Sears, 1991)
- Visual three dimensional button depression (Deron, 2000)

Mobile:

- Colour change of a button (Kaaresoja *et al.*, 2006; Hoggan *et al.*, 2008b; Tsai and Lee, 2009; Android, 2015; Windows, 2015)
- Fill an unfilled button icon with a colour (Apple, 2015)
- Colour change of the area surrounding a touch area (Herndon, 2008)

- Visual three dimensional button depression (Hoggan *et al.*, 2008b)
- Movement (Tsai and Lee, 2009)
- Magnification (Tsai and Lee, 2009)

These can be used alone or in combination. In addition, visual feedback can consist of a pop-up of a number or key (Figure 2-2 and Figure 2-4), or it can even cause the whole keypad to change from uppercase to lowercase at the beginning of a sentence. Paek *et al.* (2010) introduced sophisticated visual feedback to assist different auto-correction features in text-entry. They include pop-up colour change and colour change of the keypad letters to indicate the most probable letter coming next.

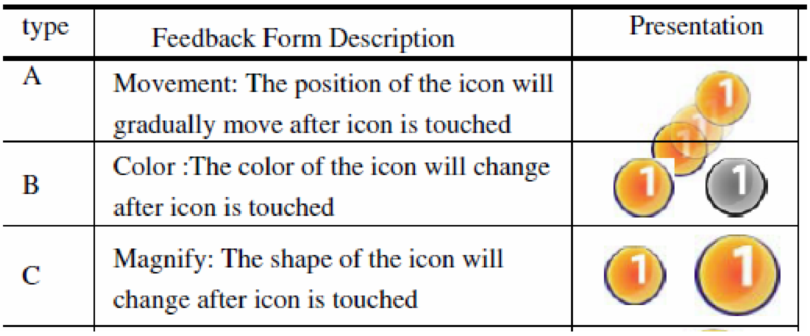
When considering feedback in contemporary mobile phones it can be seen that the most common methods are colour change and pop-up (see Figure 2-4). A pop-up resembles the movement feedback introduced by Tsai and Lee (2009) where the position of a button is changed when the button is pressed (Figure 2-3).

Picture of the Apple iPhone popup has been removed due to Copyright restrictions.

Figure 2-2: Visual feedback for a virtual button, in the form of a popup from the Nokia Lumia 920[10] (left) and the Apple iPhone 5S[11] (right).

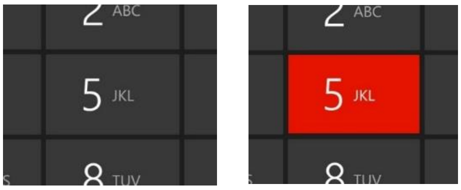
[10] www.microsoft.com (Picture used with permission from Microsoft.)

[11] www.apple.com



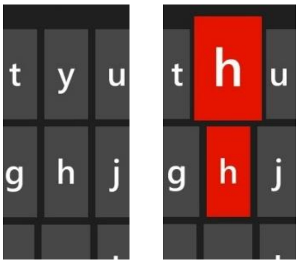
type	Feedback Form Description	Presentation
A	Movement: The position of the icon will gradually move after icon is touched	
B	Color :The color of the icon will change after icon is touched	
C	Magnify: The shape of the icon will change after icon is touched	

Figure 2-3: Movement, Colour change and Magnify feedback designs by Tsai and Lee (2009).

1. 

2. Picture of the visual feedback in Apple iPhone 5S has been removed due to Copyright restrictions.

3. 

4. 

5. Picture of the visual feedback in Apple iPhone 5S has been removed due to Copyright restrictions.

6. 

Figure 2-4: Examples of touchscreen virtual buttons and their visual feedback in contemporary mobile phones (Operating system). From top to bottom: 1.-3. Dialler: Nokia Lumia 920 (Windows Phone)[12], Apple iPhone 5S (iOS), Samsung Galaxy S5 (Android). 4.-6. Messaging: Nokia Lumia 920[13], Apple iPhone 5S, Samsung Galaxy S5.

2.4.1.2 The Effects of Visual Feedback

There are few studies which address the effects of visual feedback on usability or user experience in touchscreen virtual button interaction. Deron (2000) investigated the impact of such feedback on number entry performance in a desktop context. He let the participants type four-digit number series and compared four different visual feedback conditions: no

[12] Picture used with permission from Microsoft.

[13] Picture used with permission from Microsoft.

feedback, text-field feedback above the number keypad, visual three-dimensional button depression and combination of text-field and the depression. The results showed that any of the feedback yielded significantly fewer errors compared to the no-feedback condition and more user satisfaction, but no significant difference was found between the different feedback designs.

Tsai and Lee (Tsai and Lee, 2009) investigated virtual button feedback usability on a small PDA touchscreen with 45 older adults. The users were divided into four groups based on their cognitive skills which were tested with The Cognitore Test ("a general performance test regarding to attention and concentration measurement and analysis" (Tsai and Lee, 2009)). There were 3 different feedback styles and all feedback combinations were tested in the experiment: icon movement, colour change and magnification (see Figure 2-3). The results showed that the older adults with reduced cognitive skills got benefit from the feedback style which included icon movement (resembling a pop-up), whereas, for the users with faster cognitive skills, the usability of the icon feedback styles did not differ significantly from each other.

Despite this research, it seems that there is a gap in the literature in terms of systematic research into the effects of visual feedback design on the usability and user experience. In the research above, the latency of the visual feedback was not measured, reported or taken into account in the results. No research has been conducted on the effects of the visual feedback latency on the usability or user experience, including perceived quality, especially in a mobile phone context.

2.4.2 Audio Feedback

An audible mechanical "click" or "snap" is often an inherent part of a physical button, in addition to the cutaneous sensation from touching it. Since tactile feedback is missing in touchscreens by default, very often an audible "beep", or other short sound, is added as a substitution to virtual buttons, as well as visual feedback. In a mobile phone dialler, DTMF[14]

[14] "DTMF, or "touch-tone". A method used by the telephone system to communicate the keys pressed when dialling. Pressing a key on the phone's keypad generates two simultaneous tones, one for the row and one for the column. These are decoded by the exchange to determine which key was pressed." (http://dictionary.reference.com/browse/dual+tone+multi+frequency)

tones can be played when the buttons are pressed to simulate traditional landline phones. Different audio feedback designs are gone through next as a summary, followed by an introduction of the research on the effects on the audio feedback on usability and user experience.

2.4.2.1 Audio Feedback Designs

In early touchscreen research and design proposals, a “beep” was suggested to accompany the visual feedback in desktop applications (Speckert *et al.*, 1979; Bennion *et al.*, 1981; Valk, 1985; Sears, 1991). A “beep” is a vague design guideline, since a beep can be of any frequency and duration. In addition, it does not resemble the “click” featured in a physical button. In early systems, however, beep was a good choice because the computer systems had primitive audio generation capabilities. It has also been used in more recent studies in a mobile device context (Fukumoto and Sugimura, 2001; Lee and Zhai, 2009). Lately, more advanced audio feedback designs have been introduced, especially in desktop contexts. Altinsoy and Merchel (2009) used six different advanced audio feedback stimuli for button presses, four clicks and two DTMF tones. Lee *et al.* (Lee *et al.*, 2009) used a 150 ms bell sound.

More advanced audio feedback designs have also been proposed for contemporary mobile phone user interfaces, including versatile “click”, “clunk” or “snap” sounds due to the enhanced audio capabilities of contemporary touchscreen devices and the advanced user interface design (e.g. (Paek *et al.*, 2010)). Brewster (2002) used a “standard click” for stylus lift-off, “higher pitched version of the standard click” for indicating that the stylus had successfully tapped the button and an error sound which was “lower pitched version of the standard click” indicating that the stylus had slipped off the button which was pressed. Hoggan *et al.* (2009) continued the work by introducing even more advanced audio feedback for virtual buttons utilising Earcons, abstract synthesized auditory messages to represent parts of an interface (Blattner *et al.*, 1989): an one-beat sharp 30 ms Earcon for button press, 1-note smooth 300ms Earcon for fingertip-over event and 3-beat rough 500 ms Earcon when finger slipped over the edge or a button.

2.4.2.2 The Effects of Audio Feedback

There is little research on audio-only feedback (or with very simple visual feedback) on touchscreens. Much of the research focuses on multimodal feedback, i.e. visual-audio, audio-tactile or visual-tactile-audio, as introduced in Section 2.4.4. Here, audio feedback research is introduced.

Bender (1999) conducted a series of experiments to find out if the duration of audio feedback, or the size of virtual buttons, had an effect on desktop touchscreen number-entry performance. The participants entered series' of four digit numbers on a desktop touchscreen with a finger. A land-on strategy was used and the buttons gave no visual feedback. The duration of the audio feedback was varied between 12,5 and 800 ms and the size of the buttons was either large (30 x 30 mm) or small (10 x 10 mm). It was found that the performance was better with large buttons than small buttons and audio feedback significantly reduced the errors in number entry with the small buttons when the audio feedback duration was between 50 and 400 ms. Movement time was not affected by the audio feedback duration. The audio feedback duration did not affect the errors significantly when the buttons were large.

Brewster (2002) investigated the effect of audio feedback and button size on the usability of mobile touchscreen virtual buttons interacted with a stylus. He let the participants enter 5 digit codes with a touchscreen number keypad with a lift-off strategy. There were three different audio feedback conditions in the experiments: No-feedback; a standard click sound indicating that a button was pressed and released successfully; and enhanced audio feedback. The enhanced audio feedback consisted of an additional higher-pitched click when the button was pressed in addition to the standard click for lift-off and a sound if the user slipped off a button indicating that the button was not successfully pressed. It was found that the sounds increased performance – more 5 digit codes could be entered within the same time – and reduced subjective workload regardless of the button size. When the task was done outdoors while walking the audio feedback assisted the users so much that the performance was not significantly degraded when comparing a small button with sound to a bigger button without sound.

This research is important, but no attention was paid to the feedback latency. The feedback latency was the standard latency offered by the device used in the experiment. It was not measured, reported, controlled nor taken into account to the results.

2.4.3 Tactile Feedback

A tactile "click" or "snap" is also usually an inherent part of a physical button in addition to an audible click. Since tactile feedback is missing from touchscreens by default, several research attempts and implementations have introduced tactile feel back to virtual buttons. This trend can be seen also in commercial products: one third of all smartphones included tactile feedback for more than just vibration alerts in 2012 (Rao, 2012).

Before introducing first the tactile feedback design and then its effects on users, tactile feedback technology is briefly discussed. It is a much more complex challenge than audio feedback, for example, since it requires advanced hardware and mechanical solutions to work properly. In addition, there are many options for implementing tactile feedback in a mobile device and they all have their own latency characteristics.

2.4.3.1 Touchscreen Tactile Feedback Technology

Visual feedback is part of the GUI on the screen, audio feedback can be produced by a simple loudspeaker, but tactile feedback production is not as trivial. It requires more effort, since it requires mechanical movement or vibration of the device. This movement or vibration has been implemented in various ways, and with various technologies, in research and commercial mobile phones[15].

- Eccentric Rotating Motor (ERM[16]) also known as *vibration motor* or *pager motor* is a tiny electric motor attached firmly inside a mobile device and there is an eccentric weight (usually made out of tungsten) attached to the shaft of the motor. When a DC voltage is connected to the motor, it starts to turn and because of the eccentric weight the whole device starts to vibrate (Pesqueux and Rouaud, 2005). The start-up latency

[15] http://www.immersion.com/markets/mobile/solutions/index.html

[16] http://www.precisionmicrodrives.com/vibrating-vibrator-vibration-motors/pager-motors-erm-motors

can be from 50 to 100 ms, which is much higher than with LRA or piezo, below (Rao, 2012).

- Linear Resonant Actuator (LRA[17]) also known as *linear vibrator*, is an electromagnetic vibration actuator which works like a loudspeaker. It can be driven with audio waveforms. The LRA features a resonant frequency, which makes the LRA vibrate at maximum intensity when driven with that frequency. The most commonly used LRA in mobile touchscreen research is the C-2 Tactor[18]. The start-up latency of an LRA varies between 40 to 60 ms, which is better than ERM but still much slower than piezo actuators.
- A piezoelectric actuator (*piezo*) is a flat rectangular or circular component that bends or deforms when high-voltage (50-150 V) is applied across both ends of it. Because they are flat, piezo actuators can be attached in similar manner than ERM or LRA to vibrate the whole body of a mobile device or under a mobile device's touchscreen panel. The latter alternative enables more localized tactile feedback than the whole body vibration. Piezo actuators are used for "high-definition haptics" because they offer faster start-up time, higher bandwidth, lower audible noise and stronger vibration than ERM or LRA (Rao, 2012). The start-up latency can be less than 15ms.

2.4.3.2 Tactile Feedback Design

Since physical buttons most often feature a tactile click, the artificial tactile feedback introduced usually has been built to imitate that single click. Different designs are introduced below, whereas the effects of feedback design on usability and user experience are discussed in the next section.

Desktop

- 5 different designs (50 ms each): half-period sine wave, triangular wave, square wave, sine^2 wave and 50 Hz sine wave (Altinsoy and Merchel, 2009)

Mobile

- 800 ms vibration of 250 Hz sine wave with C2 LRA (Brewster *et al.*, 2007)

---

[17] http://www.precisionmicrodrives.com/vibrating-vibrator-vibration-motors/linear-resonant-actuator-lra-haptic-vibration-motors

[18] http://www.atactech.com/PR_tactors.html (previously http://www.eaiinfo.com)

- One or three periods of sine wave of the LRA's resonance frequency (Fukumoto and Sugimura, 2001)
- One period of 200 Hz sine wave signal (5 ms) with piezo (Poupyrev and Maruyama, 2003)
- Tactile click "designed to simulate real tactile feedback experienced when pressing a physical button" with piezo (Kaaresoja *et al.*, 2006),
- 30 ms 175 Hz square wave with "standard internal vibration actuator in the i718" (Hoggan *et al.*, 2008a)
- "Pop vibration", which is not specified more in detail, with a low latency vibration motor from Immersion tactile mouse (Nashel and Razzaque, 2003).
- 16 different clicks done with piezo technology whose rise time and amplitude were varied (Tikka and Laitinen, 2006)
- 7 different clicks with piezo technology (touch and release feedback) and 6 short vibrations with ERM (Koskinen *et al.*, 2008)
- 4 different designs: Soft short click (piezo), short "clicky" (strong) click (piezo) and long soft click (ERM) and long rough click (ERM) (Hoggan *et al.*, 2008b)
- In some contemporary mobile touchscreen phones a simple tactile click is used, for example in Samsung Galaxy S5[19] and LG Nexus 5[20].

Park *et al.* (2011) played with design parameters of tactile clicks for virtual buttons. They modified the amplitude, duration, carrier signal, envelope function and actuator to create 72 different tactile clicks. Their aim was to simulate a physical button as far as possible.

Hoggan *et al.* (2008a) used advanced tactile feedback in their research: a one-beat Tacton for button press, one-beat smooth 300 ms Tacton for fingertip-over event and 3-beat rough Tacton when finger slipped over the edge or a button. Tactons are tactile counterpart of Earcons and are abstract tactile messages to represent parts of an interface (Brewster and Brown, 2004). There are also attempts to make the users feel the edges of the virtual buttons (Nashel and Razzaque, 2003; Pakkanen *et al.*, 2010), but they are left beyond this thesis since the focus of the research here is the virtual button press feedback, not the feedback assisting the users the find the buttons before press.

---

[19] www.samsung.com

[20] www.lg.com

2.4.3.3 The Effects of Tactile Feedback

Fukumoto and Sugimura (2001) introduced a simple click on a resistive touchscreen with an LRA. They found that tactile feedback improved the performance in a simple calculation task compared to audio feedback, especially in a noisy environment.

Brewster *et al.* (2007) investigated the effect of tactile feedback on text entry usability with mobile touchscreens. The participants entered poems with a stylus on a virtual keyboard with and without tactile feedback both in a laboratory and on an underground train. The tactile feedback was 800 ms 250 Hz sine wave for a successful button press and a “rougher”, amplitude modulated, wave for an error. The results revealed a significant improvement with tactile feedback in the number of lines entered, errors made and errors corrected in the laboratory. On the train, significantly more errors were corrected and the subjective satisfaction was significantly improved with tactile feedback. The tactile feedback design was long compared to the designs in other literature. Any comparison between different designs were not implemented.

Hoggan *et al.* (2008a) compared three mobile touchscreen text entry keyboards in both a laboratory and an underground train. They used a device with a physical keyboard, and a device with touchscreen virtual keyboard with and without tactile feedback. They implemented a one-beat sharp Tacton (Brewster and Brown, 2004) for a button press, one-beat 300 ms Tacton for fingertip-over event and 3-beat rough Tacton when finger slipped over the edge or a button. The tactile feedback was generated by an in-built ERM in the touchscreen device. They found that the accuracy of the text entry was significantly lower with touchscreen virtual keyboard without tactile feedback compared to physical keyboard and touchscreen virtual keyboard with tactile feedback. The virtual keyboard with tactile feedback was comparably accurate with the physical keyboard. The virtual keyboard with tactile feedback also helped the users to enter text significantly quicker than with the virtual keyboard without tactile feedback. They conducted another study with more enhanced tactile feedback implemented with two C2s placed on the back of the PDA. Tactile feedback of the keyboard was divided into two actuators giving spatial information of the button presses in addition to the higher fidelity of the tactile feedback itself. This more advanced tactile feedback helped the users to be even faster and more accurate.

Park *et al.* (2011) deeply investigated the characteristics of touchscreen virtual button tactile feedback. They had an aim to simulate a physical button as far as possible with two designs of LRA with a touchscreen virtual button. They modified the amplitude, duration, carrier signal, envelope function and actuator of the tactile click and created 72 different tactile clicks. They asked the participants to explore virtual buttons on a touchscreen and asked them to scale the perceived similarity with a physical button and their preference. They also asked the participants to scale the perceived quality on adjective scales such as slow-fast and bumpy-smooth. The results revealed a mild correlation between the similarity and preference scores meaning that a tactile feedback resembling a physical button does not necessarily lead to a most preferred tactile feedback. They also found that the tactile feedback design with an LRA needed to have a rise time leading to a delay to the perceived peak of a tactile click in order to gain realism. That is because a physical button needs to comply before the tactile click occurs which takes a bit of time. They also found that the tactile feedback needed to be short and not too strong. The perceived quality adjectives "hard" and "distinct" correlated best with the similarity to a physical button, whereas "clear" and "smooth" correlated with preference. However, while the investigation into the characteristics of virtual button tactile feedback design was extensive, they did not take the effect of latency into account in their results. However, they reported that the delay between the touch and the feedback was so short that they were perceived as simultaneous. It was not described how this validation was conducted in their study.

The research in this section shows the evident benefit of tactile feedback compared to virtual buttons without tactile feedback. With one exception, 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.

2.4.4 Multimodal Feedback

Lee *et al.* (2009) examined older adults' performance in telephone number entry with touchscreen buttons. A mobile phone was simulated with a desktop touchscreen. They compared visual feedback with visual-audio, visual-tactile and visual-audio-tactile feedback. The visual feedback design was not reported. Audio feedback was a 150 ms bell sound and tactile feedback was 50 ms long 50 Hz vibration. They found that the performance was significantly enhanced when the multimodal feedback consisted of audio feedback (i.e. visual-audio or visual-audio-tactile) compared to visual only or visual-tactile feedback.

Altinsoy and Merchel (2009) investigated the effect of tactile and audio feedback designs on desktop touchscreen number entry performance and subjective satisfaction. They found that any kind of tactile feedback was better than no feedback, but tactile feedback implemented with sawtooth-waveform was the most beneficial to the user and was also rated the highest. There was no significant difference in performance nor errors between the audio feedback conditions. Tactile feedback was rated significantly more suitable for confirmation feedback than audio feedback. They also found that ratings were higher when combining tactile feedback with audio feedback.

Tikka and Laitinen (2006) investigated perceived intensity of audio-tactile feedback in a mobile touchscreen. The participants pressed keys of a touchscreen device and the rise time and the displacement of the proprietary piezoelectric tactile actuator underneath the touchscreen (Laitinen and Mäenpää, 2006) were varied. The rise time of the tactile feedback was 3-7 ms, the displacement 3-180 µm and the sound level of the audio component was 28-60 dB. The participants were asked to judge the intensity of the feedback on 1-to-5 scale, where ‘1’ was clearly too weak, ‘3’ moderate, and ‘5’ clearly too strong. The authors found that the intrinsic audio of piezotactile feedback affected the perceived intensity of the feedback: The intensity of the feedback with audio component was rated higher than the intensity of the feedback without audio. They also found a positive correlation between the acceleration (rise time) of the tactile feedback and the perceived intensity; the displacement did not correlate. They did not report the exact parameters of the tactile clicks, the actual designs of tactile and audio feedback nor feedback latency.

Koskinen *et al.* (2008) continued the detailed investigation of the characteristics of tactile feedback for virtual buttons with the same device as Tikka and Laitinen (2006) above, featuring piezoelectric tactile feedback. They selected 7 tactile feedback stimuli, which were rated between ‘2’ and ‘4’ on the 1-to-5 (weak to strong) scale in the experiment by Tikka and Laitinen, thus omitting too weak and too strong feedback. The duration of the feedback was 11-30 ms and the displacement 30-170 µm. The sound level of the audio component was between 42 and 61 dB. In addition, they created 6 additional stimuli implemented with an ERM with a separate but identical device by outside dimensions and the user interface.

The duration of the ERM stimuli was 4-16 ms, displacement 4-35 µm and the sound level less than 38 dB. The aim of the study was to find the most pleasant tactile feedback for mobile touchscreens. The results revealed the most pleasant piezotactile feedback being the one with the parameters of 13 ms, 105 µm and 46 dB; the most pleasant tactile feedback by ERM were the ones with 6-12 ms duration. The results also showed that there were a lot of individual preference differences. The intrinsic audio did not have a significant effect in this study. The most pleasant tactile feedback created with piezo was a bit more pleasant than the most pleasant one with an ERM.

Hoggan *et al.* (2008b) studied the perceived cross-modal congruency between the virtual buttons' graphical design and tactile-audio feedback. Cross-modal feedback means multimodal feedback which presents the same information via different modalities, in this case audio and tactile feedback modalities (Hoggan and Brewster, 2007). The authors describe congruency as "an intuitive match or harmony between the designs of feedback from different modalities". They designed 8 different visual appearances of the virtual button: 2 different sizes, 2 different shapes and two different visual heights. They also designed 4 different types of tactile-audio feedback for the virtual buttons. A short soft click from a piezo was 8 ms long with 22 µm displacement with low frequency and loud (62 dB) audio components. A short "clicky" click from a piezo was even shorter, 5 ms, with high frequency and loud (63 dB) audio components making it more "clicky". A long soft click from an ERM was 30 ms long with 7 µm displacement and a quieter audio component of 45 dB. Finally, a long rough click with the ERM was 45 ms with 12 µm displacement with an audio component of 48 dB.

They conducted an experiment in which the participants were asked to match tactile-audio feedback to the visual appearance of the buttons. It was found that the tactile-audio design matters; there were significant differences in congruence between different tactile-audio feedback and visual design combinations. Overall, the long soft click from an ERM was most frequently voted as the most congruent one. The soft short click done with a piezo was the most congruent match for small raised rectangular and small flat circular buttons. The short "clicky" click was voted the most congruent for large and small flat rectangles. The long soft click appeared to match the best with small raised circular buttons. The feedback done with the ERM matched best with large raised rectangular and circular buttons. The participants were also asked to score the perceived overall quality of the buttons from 1 to

7. The results revealed that the congruence was significantly positively correlated with the perceived quality, meaning that the design of the different modalities and aspects of the virtual button really is important. However, they did not measure or report the feedback latency nor take it into account in the results.

Lee and Zhai (2009) explored virtual button interaction in a mobile context in detail. They compared the effect of virtual and physical buttons, audio and tactile feedback, touch sensor type, input mode (finger and stylus) and virtual button size on a number entry task. They compared the performance and subjective ratings of the number entry done with a device with physical keys and devices featured with a resistive touchscreen used with both a finger and a stylus. The visual feedback of the virtual buttons was not reported. Audio feedback was a 150 ms “system beep” and tactile feedback 50 ms long implemented through the built-in actuator in the device with resistive touchscreen and virtual buttons.

They found that audio, tactile or combined feedback assisted the users to perform better with the resistive touchscreen when they were using a finger. There was no performance effect when used with a stylus. The feedback type (audio, tactile or combined) did not have any effect, either. Finger usage without feedback also caused more Key Presses per Character (KPC) than with feedback. The subjective satisfaction was highest when entering numbers with a stylus and there was audio feedback involved. It was lowest on finger operated virtual buttons with no feedback. The results from the second experiment showed that virtual buttons with audio feedback were as efficient as physical buttons in a number entry task and there were no significant differences in performance or subjective satisfaction between the resistive or capacitive touchscreens.

The results from the third experiment showed that small virtual buttons on a capacitive touchscreen were more error-prone when used with finger than the other conditions, small or large buttons with a resistive or large buttons with capacitive touchscreen. All these findings are important in order to understand virtual button interaction in mobile devices. This work again demonstrated the importance of feedback. However, no attention was paid to the feedback latency, by reporting the latencies in different conditions or measuring its effect on the virtual button interaction performance or user satisfaction.

Paek *et al.* (2010) introduced sophisticated audio-visual feedback to assist different auto-correction features for text-entry in a mobile device. They included pop-up colour change with a "clunk" sound if the user pressed an unexpected letter button. The pop-up did not change colour and the sound was "click" by default if the letter entered was an expected one. Another feedback type was a colour change of the keypad letters to indicate the most probable letter coming next. They conducted a usability study where they compared text entry efficiency of keyboards featured with auto-correction indicator (spacebar framed with red combined with a "swish" sound), pop-up colour change combined with the auto-correction indication and the colour change of the most probable letter combined with the auto-correction indication. Their results showed that the keyboard featuring the colour changing pop-up was significantly more efficient and less error-prone than the auto-correction indicator only. The most probable letter feature combined with the auto-correction was also significantly more efficient than the auto-correction only, but not significantly less error-prone. This work also shows that the pop-up feedback is beneficial for the user. However, latency of the feedback was not measured, controlled, reported or taken into account, although it could have been significant because the auto-correction features need a lot of processing power causing latency.

None of these studies considered latency. The latency between the finger or stylus touch and the associated feedback was not validated, the latency of the feedback was not reported or the effect of the latency was not assessed on the results. Interaction latencies can have negative effect on user experience as can be seen later in Section 2.8. If a system has a lot of latency, the results might be different compared to a low-latency system. Therefore, it is important to take latency into account when investigating button presses.

2.4.5 Conclusions

Touchscreen virtual button feedback can be unimodal visual, sometimes audio or tactile, or bimodal combination of these three. The feedback can also contain all these three modalities. There are several different and diverse design proposals and implementations for touchscreen virtual button feedback in the literature, making the conclusions hard to summarise. However, it can be concluded that feedback helps the users to complete a number or text entry task. Visual feedback alone is already beneficial, and audio or tactile feedback can help the users even more. Another conclusion is that all the feedback modalities can be used in touchscreen mobile phone's user interface: Visual colour change or pop-up, beep, click or other short sound were validated to be beneficial in the literature as well as short tactile clicks.

The feedback implemented in the research in this thesis is based on this earlier work as well as the design on commercial devices. The pop-up was selected for visual feedback, short audible click was adopted from the Apple iPhone and the tactile feedback was designed to be as short as possible but long enough to be perceptible.

With few exceptions, the latency that occurs between the input and the feedback events has been ignored in the earlier research discussed here. It is proposed in this thesis that feedback latency should be part of the virtual button design as it affects the user experience. Thus it should not be ignored but at least validated and reported. In the best case, it should be controlled and taken into account in the research results. Section 2.8.3 covers the research on the effect of latency on touchscreen interaction. Before that, the temporal aspects of a multimodal virtual button press are introduced, followed by an introduction to human time and simultaneity perception.

2.5 Temporal Characteristics of a Multimodal Virtual Button Press

The previous section discussed visual, audio and tactile feedback elements of touchscreen virtual buttons. This section places them in the time domain and introduces the temporal characteristics of a typical button press in detail, including its complexity and especially the temporal challenges included.

Pressing a button seems to be a very simple and trivial task, as it actually is – when interacting with a real button. However, when we have to do everything artificially on a touchscreen device, an itemization of all the stages shows us that a virtual button press is far from trivial. The anatomy of virtual button press is illustrated in Figure 2-5. The touchscreen is touched with a finger or a stylus and feedback is given for the touch after time $t_{touch\_feedback}$ has passed. This time is called *touch feedback latency*. A release of the finger or the stylus happens sooner or later depending on the user and the task, and the release feedback is given after time $t_{release\_feedback}$ has passed. This time is called *release feedback latency*. We can zoom in a little by separating the feedback in different modalities and to simplify, this analysis focuses only on the touch part. After the finger touches the screen, the different feedback elements of the button press start to activate after the individual latency period for each: Visual feedback, audio feedback, tactile feedback and action feedback (see Figure 2-6). As discussed in the previous section, visual, audio and tactile feedback refers to everything that is related to the button itself or the content of the button. For example, as discussed before, visual feedback can be the colour change of the button pressed, audio feedback can be an audible click designed for the button, and tactile feedback can mean a short vibration that can be felt by the finger pressing the button or the hand holding the device. The action feedback means the actions the button press initiates, for example, a number appearing on the screen, an application opening or a piece of music starting to play.

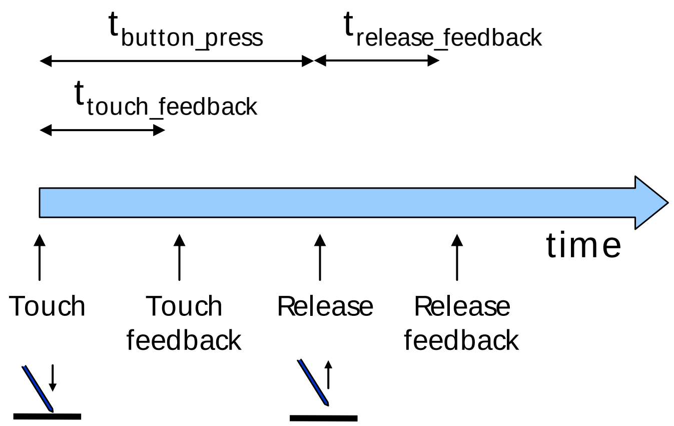


Figure 2-5: A timing diagram for a virtual button press.

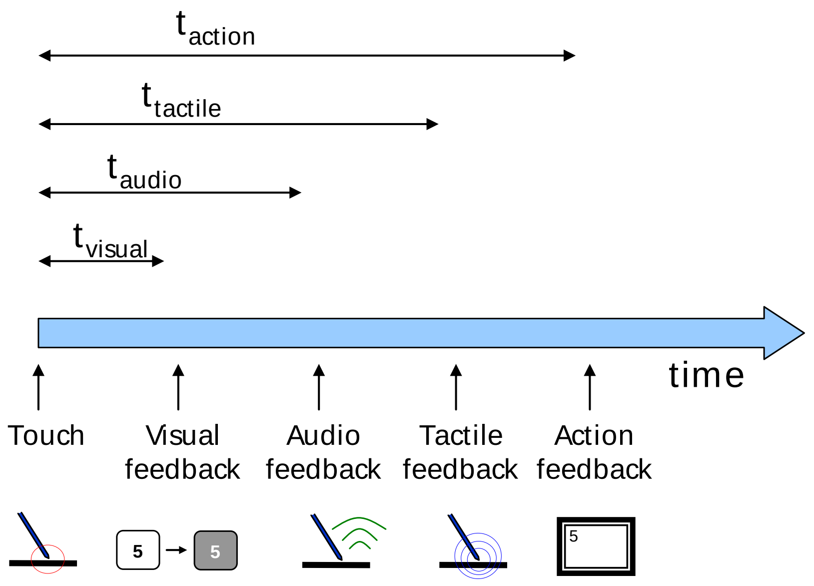


Figure 2-6: Diagram of a touch and the associated feedback modalities. All the feedback can also happen on a button release.

Visual, audio, tactile and action feedback latencies (t_{visual}, t_{audio}, $t_{tactile}$ and t_{action} in Figure 2-6) can differ remarkably from each other due to the nature of the hardware and software used to control them. In addition, they can occur in an arbitrary order. However, to follow causality, it would be wise to have all the button-related feedback (visual, audio and tactile) before the action feedback. In the real-world case, when we press a button we get all of the feedback from the physical action of pressing the button immediately, but the action caused by the press may occur sometime later. With virtual button feedback, this is necessarily not true.

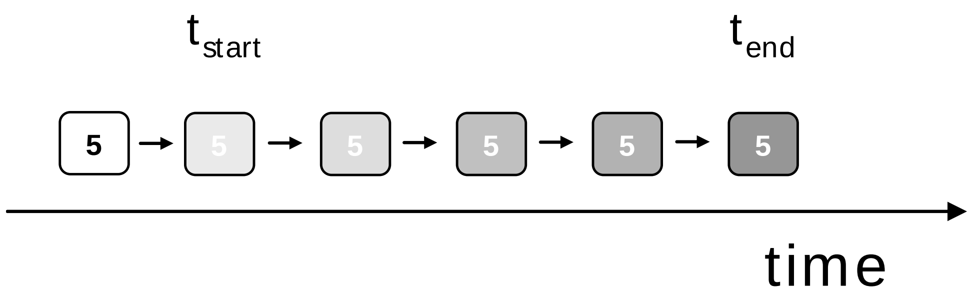


Figure 2-7: Propagation of visual feedback colour change when the button is pressed.

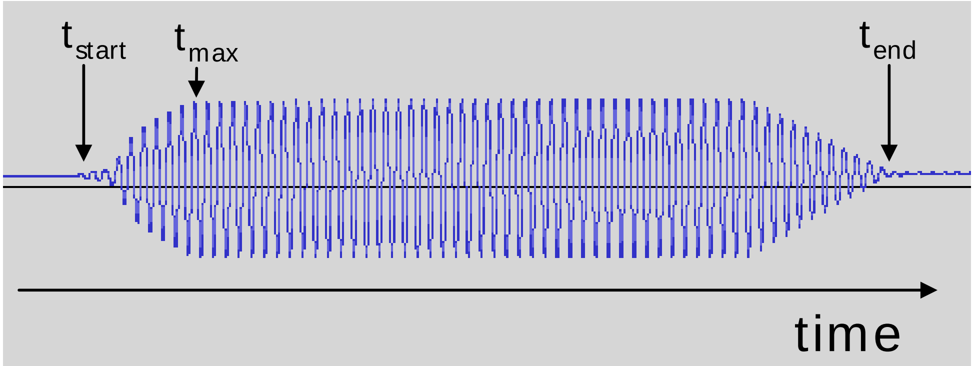


Figure 2-8: The timeline of a 70 ms audio feedback.

Visual feedback is usually shown on the button as long as a user presses the button, whereas audio and tactile feedback are usually designed to be a fixed length. In both cases, we can still consider more deeply the fundamentals of the feedback. The visual feedback needs some time to show. Consider the colour change of a virtual button. The start of the feedback is when the colour starts to change; the end of the feedback is the moment when the colour has fully changed. This takes a certain amount of time depending on the implementation (see Figure 2-7). In this thesis, visual feedback latency is defined to be the time t_{end}, when the exact moment of touch is the start of the time period and t_{visual} is the time difference between the touch and t_{end}.

Audio and tactile feedback, being even a simple beep or click, also need some time to start and reach maximum intensity (see Figure 2-8 for audio feedback). After that they are played and stopped and it takes some time to return to their initial, pre-press state. Throughout this thesis, audio latency t_{audio} and tactile latency $t_{tactile}$ is the time between the touch and t_{max}. Of course, the same is true also for action feedback, although the phases will usually be much more complicated. For simplicity and clarity, we concentrate on the button feedback latencies t_{visual}, t_{audio} and $t_{tactile}$ only in this thesis. Thus the action feedback latency t_{action} is beyond the topic of this thesis and is left for future studies.

This section introduced the temporal characteristics of the multimodal virtual button press on the basis that there are three feedback modalities involved: visual, audio and tactile. The following section introduces how latencies of different modalities can be measured in different contexts.

2.6 Latency Measurement

Latency has different meanings in psychology, telecommunications and computing science. Walker (1995) defines latency in computer science as “Delay, in digital computers, between the initiation of the call for data and the start of the transfer”. MacKenzie and Ware (1993) refine this definition into the HCI field: “The delay between input action and output response”. Hinckley and Widgor (2012) define latency broadly as an “end-to-end measure of the time elapsed between the moment a physical action is performed by the user, versus the moment the system responds to it with feedback that the user can perceive”. Jota *et al.* (2013) define latency as “the lag between a finger touch and the on-screen response”. These definitions are about physical latency (separate from perceived latency) Based on the definitions above, latency is defined in this thesis as follows: Time between the first moment of touch and the first intensity maximum of the feedback**.**

In order to understand if latency is an issue or not and, if so, take corrective actions, it is essential to be able to measure the latencies between user action and device response in the visual, audio and haptic modalities. Physical latency measurement means timekeeping between an action and a response – in this case between the first moment of finger touch and the feedback, as Seow (2008) says: “In the HCI context, system response time has to be measured from the moment a user makes an observable action to the moment the user sees a result”, extended to auditory and tactile feedback.

The following review shows examples how the timekeeping has been done earlier in different contexts and modalities. There are several research prototypes introduced in the literature and also commercial products for latency measurements in different contexts.

2.6.1 Visual Latency Measurement Methods

He *et al.* (2000) introduced a video-based latency measurement system for a Virtual Environment (VE) using a normal speed video camera. The VE system delay was determined by visually inspecting the video tape frame-by-frame. A similar video-based frame-by-frame reading approach was used earlier by Liang *et al.* (1991) when determining latencies in a VE tracking device. Miller and Bishop (2002) introduced a “Latency Meter” for measuring end-to-end latency in VEs. They used two high-speed 1-row CCD (charged coupled device, used commonly in digital cameras) detectors and special algorithms to extract the latency between user movement and VE system display update. The aim of the work was to develop a standalone instrument that would estimate the visual latency without any additional electrical connection or change to the VE software. In these research studies, visual latency between an action and a response has been measured by videotaping the action and response and simply calculating the time difference from the video frames between the action and the response. In the earlier work the frame rate has been restricted to the standard video frame rate of less than 30 frames per second, leading to more than 33 ms temporal resolution. This accuracy would have been enough in complex, highly visual systems with latencies in the magnitude of hundreds of milliseconds. However, it would not have been accurate enough for touchscreen latency measurements for two reasons. First, the human temporal perception is in the same magnitude of tens of milliseconds (see Section 2.7.2 onwards). Second, the initial measurements of touchscreen phones showed that the latency was between 30 and 200 ms. Thus, more accurate latency measurement was needed in order to assess the latency accurately enough. In addition, only visual feedback latency was measured in these earlier measurement systems, so it is not known how the audio or tactile latencies were assessed.

In the more recent work by Steed (2008), the latency of an interactive graphics simulation was investigated with a sine fitting method. A standard frame rate video camera was still used for capturing both the input and the response, but instead of calculating the frames between and action and a response, a mathematical analysis was implemented for the video feed. A position tracker with an LED attached to it was hanging and swinging on a pendulum and the graphics simulation was tracking the swing and drawing its trajectory accordingly. The position of the tracker (the LED) and the on-screen response were video-recorded for the same video stream and a sine wave was fitted for both the tracker and the responses trajectories. The latency was calculated from the phase difference of these sine waves in much higher temporal accuracy than the video frame rate. The method was simple to configure, sensitive and rapid to use and did not require any hardware changes to the system. This kind of mathematical extrapolation of a movement worked fine in continuous interaction but is not suitable for event-based input, such as touch feedback latency measurement. Therefore, the accuracy required has to be gained from a high-speed camera. In addition, only visual feedback latency was measured in the research by Steed.

All the previous work was designed for indirect interaction systems. Ng *et al.* (2012) measured touchscreen inking task latency by placing a ruler on a touchscreen device. A finger was moved on the screen followed by the ink trajectory from the application. A high-speed camera was used to estimate the speed of the finger. The minimum and maximum latency t could be obtained from the individual high-speed video frames by calculating $t = v/s$, where v is the speed of the finger and s is the gap between the finger and the trajectory. Bérard and Blanch (2013) introduced more sophisticated and accurate latency estimation methods for touchscreens also suitable for the touchscreen inking task. Both of these studies required finger movement on the touchscreen and the finger following a trajectory. Thus, the methods are not suitable for estimating or measuring latency of an event-based touchscreen tapping task which also requires the detection of the moment of touch.

Montag *et al.* (2011) aimed to measure latency of a video loop created by a display and a camera. They showed a frame counter window and the camera monitor window on a display and shot the display with a camera. The camera monitor window showed the delayed frame count and the display was photographed. From the photo, the latency was calculated by subtracting the delayed frame count from the frame count.

2.6.2 Audio Latency Measurement Methods

Freed *et al.* (1997) measured operating system latencies with a 2-channel audio recorder. They used a low-latency device to transcode a computer network or MIDI event to a short sound. That way they could use the simple stereo audio recording software to record the input and the output at the same time in order to investigate the latencies. The same 2-channel recorder methodology has been adopted by the other audio latency measurement projects. MacMillan *et al.* (2001) as well as Wright and Brandt (2001) performed an extensive set of audio latency experiments with different audio hardware in different operating systems. They investigated the suitability of general-purpose computer to real-time audio processing. Wright *et al.* (2004) measured system latencies in various computer operating systems from a QWERTY key press to audio out (short “blip”) and Montag *et al.* (2011) measured latency of music performing touchscreen devices and applications. Nelson and Thom (2004) measured MIDI latency under Linux, OS X and Windows with a MIDI-to-audio paradigm.

To summarise, audio latencies have been assessed with 2-channel audio recorders. In practise this uses a standard stereo PC sound card and sound editor software. The input is recorded on one channel and the output sound on another. The latency between the input and output can be investigated easily with the sound editor usually featuring a selection tool and the measure to show the length of the selection either in milliseconds or, more accurately, in samples. Because of the affordability and simplicity, the same approach was taken in the multimodal measurement system introduced in this thesis. However, the previous work for audio latency measurement did not include visual or tactile latency measurements.

2.6.3 Tactile Latency Measurement Method

Lehtosalo (2009) used a force sensor to obtain latencies between finger touch and tactile feedback in touchscreen interaction. He used a setup where the mobile phone equipped with a touchscreen was placed on a force sensor. The force sensor detected the finger press and the tactile feedback provided by a tactile actuator in the phone. Proprietary hardware circuit provided the measurement data from the force sensor to a PC and mathematical software Matlab running in the PC was used for processing and analysing the data to compute the latency. This was a sophisticated method, but did not include visual or audio feedback latency measurements.

2.6.4 Commercial Timing Analysis Products

OptoFidelity[21] sells a product called WatchDog (see Figure 2-9) for automated timing analysis for mobile phone or tablet touchscreen devices. It records the touchscreen with a video camera and automatically detects the changes on the screen. After the measurement, it automatically creates a timing report in HTML. It also has an option for high-speed video and add-on sensors for audio and haptics. It “automatically detects visual, audio and haptics events from the operator interface of the device, and reports the events with accurate timestamps instantaneously” (OptoFidelity, 2014). The touch detection is implemented as a pressure sensitive switch, which is not a good choice if the touch is done with a finger. The pressure on the touchscreen will change from press to press with a human user (comparing

[21] www.optofidelity.com

to a robot finger). If the pressure sensor is not sensitive enough, it can cause extra latency in the measurement and some presses can even go unnoticed. If the pressure sensor is tuned to be too sensitive, unwanted presses can occur in the measurements. Despite this limitation, WatchDog is a tool for deep analysis of a touchscreen user interface consisting of visual, audio and haptic elements for commercial and industrial use. It also has an option to conduct repetitive long-run tests with robot fingers as seen in Figure 2-9. However, it is expensive (starting from 15,000 € without robotics and from 40,000 € with robotics – at the time of writing this thesis) and might not be suitable for occasional small-scale measurements of a single device, which is typically the case in user interface research, for example. It is not portable either, which would be beneficial when used by a team distributed in different sites, for example. The goal of the multimodal measurement tool created in this thesis was to design a tool that can be built from affordable off-the-shelf components in order to be available to everybody and to be able to replicate or move wherever needed.

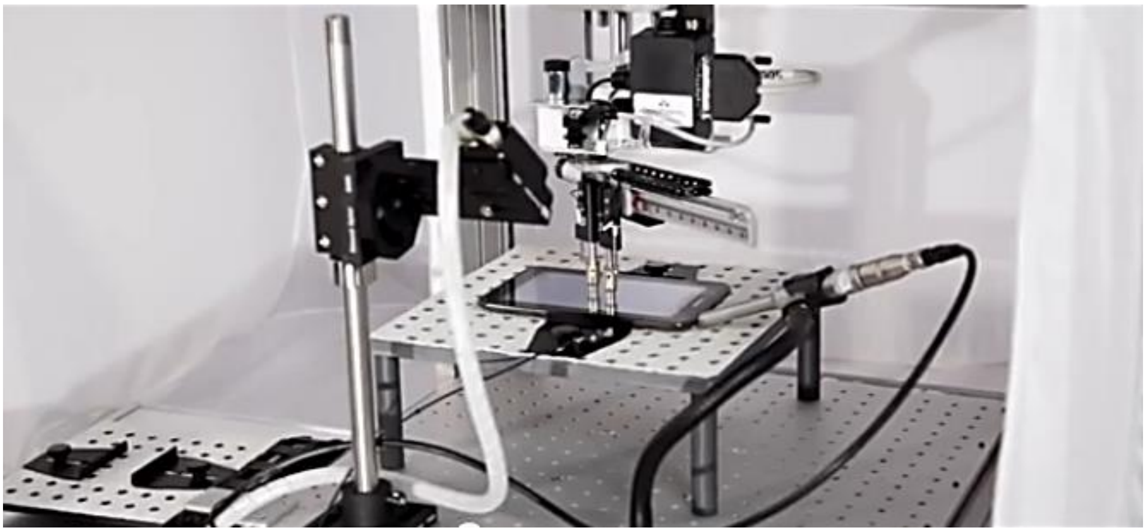


Figure 2-9: OptoFidelity WatchDog with an automated interaction with a robot (captured from a video from www.optofidelity.com)

BlackBox ToolKit[22] is a timing analysis tool for pre-evaluation of a perception study setup or experiment application (see Figure 2-10). It can detect, for example, colour changes on a screen, audio feedback as well as user actions, like button presses. In addition, it can be used for simulating a participant performing a perception experiment by connecting signal cables to mouse switches on the computer running the experiment application. This enables the

[22] www.blackboxtoolkit.com

researcher to analyse the internal latencies of the experiment system before running the real experiment. It is an excellent tool for temporal analysis of events in PC environment, but it is not suitable for the latency measurements of touchscreen devices. It does not feature touch detection or tactile feedback recording, and the visual feedback is recorded with an on-screen detector, as Figure 2-10 shows. Therefore, pressing the virtual button would be impossible. The price[23] of the basic setup was 2150 € at the time of the writing this thesis.

Figure 2-10: The BlackBox ToolKit. An optical visual stimulus sensor is attached to the laptop screen to detect colour or brightness changes. A microphone (white box) is placed on the front of the loudspeaker on the right-hand side of the laptop to detect sounds. (Note: As requested by BlackBox, the figure was updated to ToolKit v2 for the electronic version of this thesis.)

2.6.5 Conclusions

There have been plenty of different attempts to measure latencies in different contexts, in both the literature and in commercial devices. The aim of measuring, or at least estimating, latency in any system is to gain understanding about the minimum, maximum and distribution of latencies in order to make corrective actions on the system.

Visual latency has been mostly assessed by videotaping an input and the system response, calculating the number of frames between them and dividing it with the frame rate. The

[23] http://www.blackboxtoolkit.com/howtobuy.html?tab=1#BBTKv2_pricing

temporal accuracy of this method is restricted by frame rate of the video. A straightforward way to improve this has been to increase the frame rate and use a high-speed camera. Further way to enhance the accuracy has been the use of high-speed cameras together with mathematical methods.

All the earlier research on visual latency assessment has been designed for a dynamic input and an associated response either in virtual environment, graphical simulation or touchscreen inking. Therefore, they are not suitable for latency measurement of a virtual button press. There are also more sophisticated methods introduced for estimating the visual latency in dynamic interaction, but the problem remains; they are not suitable for the latency measurement of event-based input, such as virtual button presses. None of the earlier research included visual solution for touch detection either, which is an essential feature in virtual button latency measurement.

Audio latency has most commonly been measured with a 2-channel audio recorder consisting of a stereo sound card and sound editor software; one channel recording the input, the other the response. The latency has been assessed manually in audio recorder software capable of showing and modifying audio files. The latency assessment has also been done automatically, but the basic principle has been the same as in the manual assessment: calculate the time difference between an input and a response in different channels in a stereo audio stream. Because of its simplicity and affordability, this method was adopted into use in the research in this thesis.

Excluding the commercial latency assessment systems, there is only one attempt to measure tactile feedback latency in the literature. It was implemented by using a sensitive force sensor underneath the device under measurement. The use of force sensor was reasonable since it picked up both the moment of touch and the tactile feedback. The latency was calculated with Matlab software. This method was not used in the research in this thesis since it was considered easier to implement the tactile feedback recording and touch detection with the same hardware and at the same time as audio latency measurement making both the measurement and analysis phases simpler.

The commercial device by OptoFidelity is available for deep temporal analysis of touchscreen events. However, it relies on proprietary and complex hardware and software tools making it expensive and unaffordable for many. It is also big making it difficult to move to another location if needed. The commercial device called Black Box Toolkit is available for timing analysis of multimodal interactive systems. It is not as expensive as OptoFidelity's system, but it does not feature touch detection or tactile feedback recording. In addition, the visual feedback detection requires a cumbersome hardware on top of a display. Therefore, it is not suitable for virtual button feedback latency measurements. One goal of the research in this thesis was to create an inexpensive latency measurement tool using off-the-shelf components as far as possible without compromising the accuracy, making it affordable for universities and other research institutes and affordable to copy when needed.

2.7 Latency Perception

The previous section explained latency measurements methods. In this section, human perception of latency, including simultaneity perception is introduced. There are many definitions and meanings for simultaneity starting from Antiquity, all the way into the Theory of Relativity (Jammer, 2006). In this research in this thesis, simultaneity means that two or more events happen at exactly the same time within the space that a person can see, hear and touch. Perceived simultaneity, in turn, means that the person perceives two stimuli to happen at the same time. "A and B appear to be simultaneous if they seem to happen at the same time", as Power (2011) indicates and continues: "So, there will be an illusion of simultaneity if two perceived events seem to be happening at the same time but are not happening at the same time". That means that one perceives two events happening at the same time but there is a temporal gap between the events. When the gap between the events gets larger, at some point the events are no longer perceived as simultaneous. That point, which is usually defined statistically, is the simultaneity perception threshold, i.e. the asynchrony detection threshold (Coren *et al.*, 2003). For the user, the physical simultaneity and the perceived simultaneity seem to be the same thing (Seow, 2008; Ng *et al.*, 2014).

Seow (2008) has written a comprehensive book about time perception in Human Computer Interaction (HCI). Although interesting and important, it focuses on perceived time and waiting times longer than 100 ms and techniques to make the waiting time feel shorter. The research in this thesis concentrates on investigating the effects of latency in simultaneity perception and perceived quality and the thresholds are mostly less than 100 ms as can be seen in the following sections and chapters.

Finding the touch-feedback simultaneity perception thresholds in touchscreen interaction are important in order to set practical guidelines for hardware, software and interaction designers. It is not reasonable to optimize the touchscreen system latencies below the simultaneity thresholds since the user would not perceive the improvements anyway. On the other hand, when measuring latencies from existing devices, the simultaneity perception threshold will tell if the users will notice the latency or not.

The fundamentals of simultaneity perception assessment, i.e. how to measure human ability to perceive simultaneity, are introduced next, followed by the introduction of the simultaneity perception research done earlier in psychophysical research with different modality pairs. The simultaneity of the visual, auditory and tactile modalities is introduced here, as well as the simultaneity of a tap (with a hammer, mallet or finger) and one of three feedback modalities. They are closely related to the research in this thesis, since all these modalities are included, in addition to tapping virtual buttons. The earlier work on simultaneity perception of more complex stimuli such as video and speech by e.g. Dixon, *et al.* (1980), van Wassenhove *et al.* (2007) and Carter *et al.* (Carter *et al.*, 2010) is out of the scope of this thesis.

2.7.1 Perceived Simultaneity Assessment

The perceived simultaneity of two stimuli has been studied a great deal in psychophysics. It is usually assessed with two methods: Simultaneity Judgments (SJ) and Temporal Order Judgments (TOJ). 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 SJ provides a *detection* threshold and TOJ provides a *differentiation* threshold (Vogels, 2004; Harris *et al.*, 2010). 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 (see Figure 2-11). 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”. SJ method was selected for the research in this thesis, since TOJ would have been inappropriate: It would have not been reasonable to ask participants to judge the temporal order of touch and feedback (which one came first) since the feedback always came after touch when the buttons were pressed in the experiments described later in this thesis.

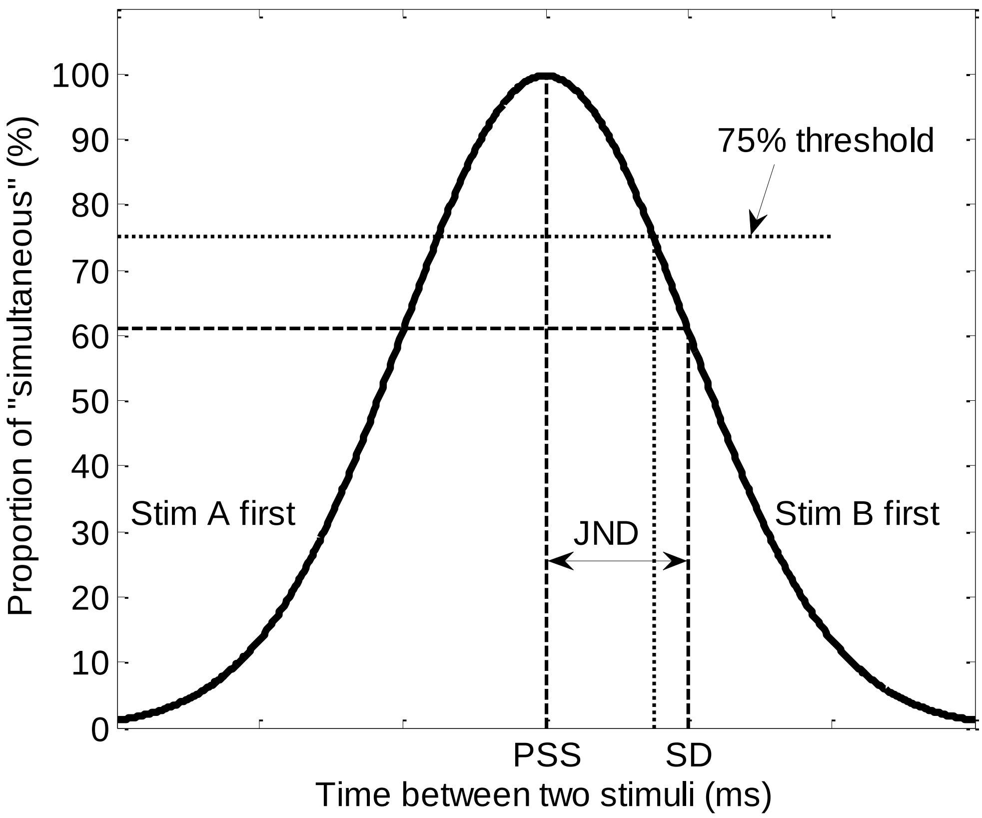


Figure 2-11: A Gaussian curve fitted to Simultaneity Judgment data as a function of time between two stimuli. The Point of Subjective Simultaneity (PSS) is the maximum of the fitted Gaussian function and it states the time between two stimuli at which the participants most probably judged the two stimuli 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 100% in this figure. Adopted from Vogels (2004) and Harris *et al.* (2010).

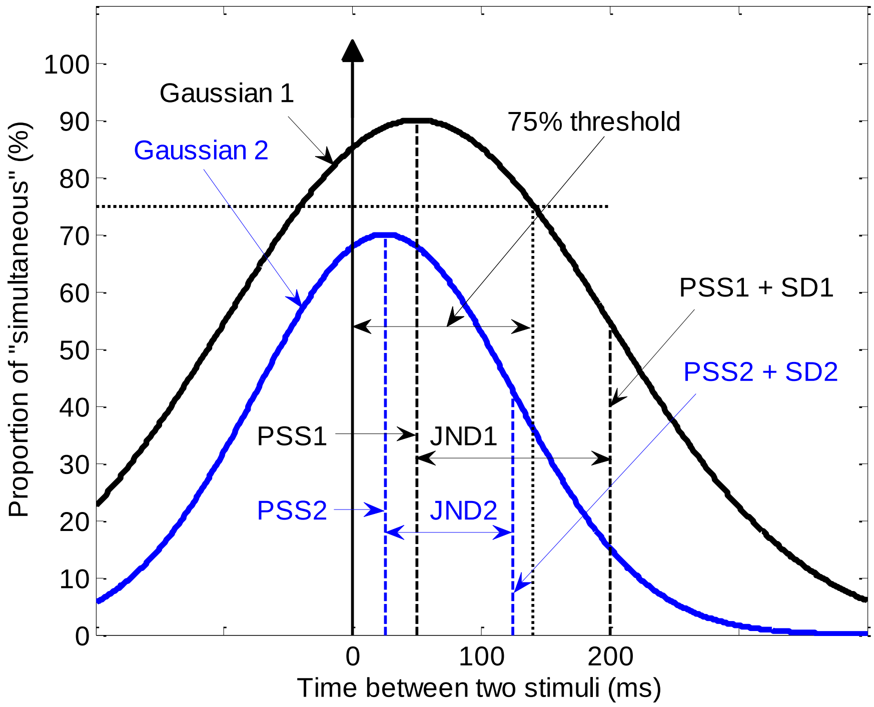


Figure 2-12: The illustration of two different Gaussian curves showing the importance of the 75% threshold versus the traditional JND.

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, i.e. 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. The same is true when the JND is defined to be "half width at the half height" of the Gaussian bell (Vogels, 2004; Fujisaki and Nishida, 2009). However, the JND defined either 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 2-12 illustrates this with two hypothetical frequency distributions of simultaneity perception modelled by Gaussian functions (JND = standard deviation). 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 it was chosen to be used in this thesis research; it has also been used in by Levitin *et al.* (1999) and Jota *et al.* (2013). In addition, the 75% threshold is always more conservative than the JND based on standard deviation σ (≤ 0.759 × σ, if PSS ≥ 0 ms and height of the Gaussian ≤ 100%) making it a stricter rule for the design guidelines (Figure 2-12).

SJ research can be conducted in two ways. In the first, a single stimulus pair is presented to the participants and they are asked if the pair was simultaneous or not. Another way is to present two successive pairs of stimuli of which one, called the "probe", is always truly simultaneous (or the minimum latency set by the system baseline). The participant then has to choose which pair was more simultaneous. In fact, the first one is truly a simultaneity *detection* method, whereas the latter measures the participants' ability to *differentiate* two latencies. According to Harris *et al.* (Harris *et al.*, 2010) the first method is "highly subjective, psychophysically uncontrolled, and subject to criterion shifts in the JND". This sounds undesirable from a psychophysical point of view, but actually that is exactly what was looked for in the research in this thesis. The aim was to establish a threshold for a subjective simultaneity of touch and feedback. The comparison method, which measures the latency differentiation, is not ecologically valid, since the users most often use one device at a time instead of constantly comparing two devices let alone one being "truly simultaneous". In addition, the comparison method is incapable of providing the PSS since the "probe" *is* the PSS (Harris *et al.*, 2010). Thus, it was decided to conduct the simultaneity perception experiments with a single stimulus pair.

2.7.2 Intramodal Asynchrony Detection

To set the foundation of human temporal perception and find the requirements for the latency measurement tool and experiments, it is essential to understand the research on asynchrony (successiveness) detection between two stimuli of same modality. Human temporal perception has been studied for more than a century in psychology. As early as 1875, Exner (1875) found the thresholds for simultaneity perception of two intramodal (same modality) stimuli to be as low as 2 ms for two auditory clicks and 44 ms for two brief flashes of light. Wundt found very similar figures: 2 ms for audio, 27 ms for tactile and 43 ms for visual (Boring, 1923; Levitin *et al.*, 1999). These values have set the baseline for human temporal perception.

2.7.3 Simultaneity Perception of Exogenous Multimodal Stimuli

Since the Simultaneity Judgment (SJ) method was selected for the research in this thesis, only SJ research is introduced here and Temporal Order Judgment (TOJ) research is mentioned only if SJ research is not available for a certain modality or event pair.

The fundamental research into simultaneity perception has measured the perceived simultaneity of two exogenous stimuli that means that they are played to the passively observing participant and there is no interaction involved. The modality pairs studied have been visual-audio, visual-tactile and tactile-audio, and the stimuli in the research introduced below have been brief momentary events such as flashes, sounds, beeps or tactile clicks resembling feedback usually implemented in touchscreen interaction. A lot of other simultaneity perception research exists, such as audio-visual speech synchronisation, but it is beyond the remit of this thesis. It is remarkable that even though asynchrony detection research started in the 19th century, simultaneity perception research has evolved only in the past two decades.

Fujisaki and Nishida (2009) explored the simultaneity perception of multimodal visual, audio and tactile stimulus pairs with same seven participants. The visual stimulus was a blob presented on a computer screen, the auditory a pulse of white noise presented on headphones and the tactile a pulse of vibration presented on the both index fingers. The stimuli were all 6.25 ms long and the time between the stimulus pair was varied from -300 to +300 ms. They defined the threshold (JND) to be half width at the half height of the Gaussian curve. Their results revealed that the audio-visual simultaneity perception threshold was 75 ms, visual-tactile 55 ms and audio-tactile 35 ms.

2.7.3.1 Visual-Audio Simultaneity

Stone *et al.* (2001) varied the time between audio and visual stimuli from -250 ms (sound first) to +250 ms (light first). Their results showed that the average PSS was 51 ms and JND was also 51 ms. 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 19 ms and the average JND was 114 ms when the stimuli came from the same location. The PSS was 32 ms and the JND 91 ms 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) Stone and Zampini found were of the same magnitude being 102 ms and 123 ms. 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. The study by Fujisaki and Nishida (2009) revealed the PSS -5 ms (audio first) and the JND 75 ms.

In an experiment by Levitin *et al.* (1999), participants judged the simultaneity of a mallet hit and a percussive sound. One blindfolded participant hit the mallet and felt the hit haptically while another observed visually 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 -250 ms (sound first) to +250 ms (visual/haptic hit first). The results revealed that the observer's audio-visual PSS was 0 ms and the 75% threshold was approximately 43 ms on average and symmetrical. Levitin's results showed smaller figures than in the research by Stone, Zampini and Fujisaki, since the test setup enabled participants to anticipate the event, thus making the judgment easier.

2.7.3.2 Visual-Tactile Simultaneity

The study by Fujisaki and Nishida (2009) is the only one exploring the simultaneity perception of visual-tactile simultaneity in a similar manner than the other stimulus pairs. Their finding was that the PSS was -20 ms (tactile first) and JND 55 ms. In an earlier study by Vogels (2004) 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 creating 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 59 ms when force came first and 44 ms when the cursor hit the horizontal line first. The PSS was nearly 0 ms. Although their test setup and application were different from the ones used in the research in this thesis, the findings were taken as a reference.

2.7.3.3 Tactile-Audio Simultaneity

Altinsoy (2003) conducted a simultaneity judgment task where a 25 ms tactile stimulus was presented to participants' index finger along with an audible stimulus (noise) of equal length to a set of headphones. The results showed that the average PSS was 8 ms and the threshold was asymmetrical: 24 ms when audio was presented first (audio lead) and 50 ms when audio was presented after (audio late). A similar pattern, showing more sensitivity to audio lead, was found by Begault *et al.* (2005) and also Fujisaki and Nishida (2009). The figures found by Begault *et al.* were much larger: 100 ms for audio first and 200 ms for audio late whereas the figures found by Fujisaki and Nishida were similar to Altinsoy: 23 ms for audio lead and 46 ms for audio late.

2.7.4 Simultaneity Perception of Momentary Actions and Feedback

Little research has been conducted on the simultaneity perception, or the effect of feedback latency, on a momentary action such as tap, hit or press and its associated feedback. This is highly relevant since a button press is central to this thesis research. It is important to find out the methodology and the results in order to have a reference to the research. However, as can be seen, none of the research introduced below involved more than one feedback modality at a time. In addition, the experimental setups were constructed to understand human perception. A setup which was more focused on the practical application domain, based around a mobile phone prototype, was constructed for the research in this thesis.

2.7.4.1 Tap-Visual Simultaneity

There is no SJ research on tap-visual simultaneity, but TOJ experiments conducted with tap and visual feedback exist in the literature. Rohde and Ernst (2013) experimented with a virtual button press implemented with a Phantom[24] force feedback device and associated visual feedback and the temporal order of the press and the feedback. The visual feedback could also precede the press. The threshold for the "key press comes first" condition was 70 ms, on average. Since there are no SJ experiments available, the research in this thesis will be an important contribution to both psychophysics and HCI, showing the simultaneity

[24] Currently called Geomagic Touch (http://geomagic.com/en/products/phantom-omni/overview/)

perception threshold for virtual button press and visual feedback. The found TOJ threshold has to be taken as a reference of threshold of tap-visual simultaneity perception, though.

2.7.4.2 Tap-Audio Simultaneity

In the experiment by Levitin *et al.* (1999) introduced in Section 2.7.3.1, the blindfolded participant who hit the mallet and felt the hit haptically heard an associated percussive sound from headphones. The results revealed that the tap-audio simultaneity threshold was -25 ms (sound first) and 42 ms (hit first) on average.

Adelstein *et al.* (2003a) investigated the perceived asynchrony of a hammer tap and a related percussive sound. They did a comparative study where participants tapped a tile with a hammer and were given a delayed sound over headphones. They used the comparison method and the participants had to judge which of the two hit-sound pairs had less delay. They found that the average 75% threshold was 24 ms ranging from 5 to 70 ms within participants. Although the experiment was conducted with an ecologically invalid comparison method, these simultaneity perception threshold figures set a baseline for the hypotheses.

In a recent study, Van Vugt and Tillmann (2014) investigated the simultaneity perception of a computer keyboard key press with a finger and a percussive sound given on headphones. They found a simultaneity perception threshold of 180 ms for non-musicians and 102 ms for musicians. These numbers are, for some reason, much higher than the ones found by Adelstein *et al.* and Levitin *et al.* In their paper, they did not report the baseline latency, which would be essential to know when comparing the results with other research. In addition, as in the research by Winter *et al.* (2008), introduced later, the keyboard button needs to go down and it takes time before the electrical contact. This might explain the higher threshold.

A tap with a mallet, hammer or a physical button with an associated but delayed sound strongly relates to the practical approach to the simultaneity perception of touch and audio feedback in the research of this thesis. These simultaneity perception threshold figures set a baseline for the hypotheses. In the studies discussed above the sound was provided by headphones. It is important to investigate the simultaneity when audio feedback is given from the same location of the tap, because in a mobile phone that is the case (if the user do not wear headphones). In addition, no other modality than audio was involved in the tap interaction in the previous research.

2.7.4.3 Tap-Tactile Simultaneity

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 (Stone *et al.*, 2001), here the results showed that the simultaneity perception followed a Gaussian function. They also showed that the average PSS was -29 ms (tactile feedback first), although it was not significantly different from 0 ms. This means that the point of perceived simultaneity could have been equal to physical simultaneity, which would have been 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. Although the Morse key is different from a touchscreen virtual button, this research was informative, and a psychophysical approach was applied in this thesis in order to understand the simultaneity perception of a button press and its associated feedback. The JND was defined to be one standard deviation of the Gaussian function and was found to be 105 ms on average in Winter *et al.*'s research, yielding the estimated threshold of 76 ms (PSS + JND). This also gave a reference for simultaneity perception between a touch and tactile feedback in touchscreen virtual button interaction.

2.7.5 Latency Perception in Touchscreen Interaction

Latency perception in touchscreen interaction has been researched only recently (from 2012 onwards) because achieving near 0 ms latency has demanded advances in hardware. It still requires a special hardware setup to achieve near-zero latencies and no commercial device is capable of doing it. This kind of hardware has been set up in Microsoft Research (MSR) and the research, introduced in this section, has only been conducted by that institute. They have also investigated the effect of latency in usability and user experience during touchscreen interaction, and that work will be discussed in Section 2.8.3.2. However, the research has been focused on tablet-sized devices and visual feedback only.

Ng *et al.* (2012) investigated latency perception in a dragging task on a touchscreen. They constructed a proprietary system capable of producing visual response with very low baseline latency (1 ms) for touchscreen gestures. They let participants 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 (the 1 ms baseline latency) or the probe (1-65 ms latency), was faster. They found that the 75% threshold for latency perception varied from 2.4 to 11.4 ms, being 6.0 ms on average, far below the latency in current commercial devices. In addition, their paper focused on the technical details of touchscreen visual latencies and solutions to overcome the challenges of reducing touch-to-display latency.

Jota *et al.* (2013) continued to investigate latency in direct-touch input on a touchscreen with a finger. They explored the effect of latency on the performance of the dragging task with similar hardware setup to Ng *et al.* (2012). The participants dragged an object from one position to another with their finger on a touchscreen. The latency of the cursor movement was varied between 1 and 50 ms, in addition to varying target width and location, while task speed and accuracy were measured. They found that performance degraded as latency increased, width decreased and target distance increased. The results showed that there was no significant difference in performance between touch and feedback latencies 1 and 10 ms, 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 touch-feedback latency detection with comparison method in a tapping task. Their results showed that the 75% latency detection threshold varied from 20 to 100 ms depending on the participant, the average being 64 ms. 85% of the participants could not discriminate 40 ms from 1 ms.

Later, Ng *et al.* (2014) modified their low latency touchscreen to also work with a stylus and they investigated the perception of latency in dragging and scribbling tasks. The participants were given three different tasks: large box dragging, small box dragging and scribbling. The results showed different thresholds for the latency perception depending on the task. When dragging a large rectangle with a stylus, the participants could discriminate latency of 6 ms from 1 ms, whereas when dragging the small one, the discrimination threshold was 2 ms latency. With simple scribbling task in which electronic ink appeared on the screen after the curvilinear stylus movement, the participants could discriminate 7 ms – which was the baseline in this task – from 40 ms, but not lower than 40 ms.

The authors claimed that the reason for the very low threshold when dragging and for the difference between dragging, tapping and scribbling is the latency judging strategy. When dragging with a stylus (or finger) the latency is perceived by visually detecting the movement of the dragged object in relation to the stylus tip (or fingertip), not being truly a latency judgment. Whereas, when tapping the screen, the visual feedback is compared with the inherent tactile sensation caused by the tap, making it truly a latency judgment. When scribbling, the latency judgment strategy was more varied between the users, but was still based on visual attention making the judgment again not truly based on latency. Based on the results, latency judgment of scribbling was more difficult than in dragging.

Annett *et al.* (2014) continued the work with different inking tasks with the same hardware as used by Ng *et al.* (2014). They found that latency perception threshold in an inking task was approximately 50 ms and did not differ significantly between different tasks (simple line, writing a word and drawing a star). Instead they found a significant difference in latency perception if the inking hand was visible or not.

In conclusion, MSR researchers have investigated latency perception in touchscreen interaction extensively, focusing on visual feedback on tablet-sized device. They tested touchscreen dragging, scribbling and tapping tasks with a finger. In addition, they explored dragging, scribbling and inking tasks with a stylus. They found that the latency perception threshold was the lowest, 2 ms, when dragging a small rectangle with a stylus. Scribbling or inking with a stylus led to latency perception threshold as high as 40-50 ms. The latency perception threshold of the tapping task was 40 ms on average. They suggested that latency perception appears to be dependent on the task. Theoretically, according to their Latency Perception Model, latency perception is also dependent on other factors, such as referents available and contextual demands. This all means that the results from experiments with one task cannot necessarily be transferred directly to another, but have to be investigated separately.

The limitation of their work, from the perspective of the research in this thesis, was the use of a comparison method that is not ecologically valid, meaning that the users usually do not compare devices with different latencies in everyday use. In addition, the MSR researchers focused only on visual feedback, which means that the latency perception thresholds of other modality feedback in touchscreen interaction remained uninvestigated. However, the visual latency perception threshold of the tapping task, 40 ms, will serve as the best baseline for the research in this thesis, focusing on touchscreen virtual button interaction in a mobile phone context.

2.7.6 Conclusions

The human ability to perceive asynchrony between same modality stimuli starts from 2 ms, with two audio stimuli. However, when a simultaneity perception of two event-based stimuli of different modalities is investigated, the threshold are around 10x higher, at about 25 ms minimum (Levitin *et al.*, 1999; Adelstein *et al.*, 2003a; Altinsoy, 2003). These numbers set the strictest requirements for latency measurement tools, meaning that the latencies to be measured are on the magnitude of milliseconds and tens of milliseconds. Thus, for example microsecond accuracy is not required.

Overall, the simultaneity perception threshold has been found in earlier research to be between approximately 25 and 200 ms (most being between 25 and 100 ms). Touchscreen virtual button latency perception has been investigated with visual feedback, with the threshold found to be 64 ms, which sets the baseline for this modality. The perception of tactile and audio feedback latencies in touchscreen interaction remains uninvestigated.

2.8 The Effects of Latency on Usability and User Experience

The previous sections have discussed the temporal aspects of psychophysics and simultaneity perception for different events and in various contexts. This section will focus on the Human-Computer Interaction (HCI) aspects of timing, such as the effect of latency on usability and user experience.

2.8.1 Current Latency Recommendations

As mentioned at the beginning of Section 2.4, the British Standard ISO 9241-400 (2007) states: “An input device shall provide effective feedback, i.e. the user is given immediately perceptible and easily understandable indication that the device is responding to user actuation”. However, it does not define what “immediately” means in real world devices. It refers to “responsiveness” and the same standard (ISO, 2007) defines responsiveness as follows: “An input device shall be responsive, i.e. the feedback following its actuation shall be consistent, timely and accurate”. Another document, MIL-STD 1472 (Cohen, 1995; Defence, 2012) gives recommendations for responsiveness. Feedback for a key or touchscreen press should happen within 100 ms and the action feedback within 200 ms. Seow (2008) gives similar guidelines referring to the previous document: for a simple input such as a key press the response time should be less or equal to 100 ms and more complex actions, such as a drop-down menu, up to 200 ms. Although these numbers have not been backed up by scientific research, they give a reference and baseline to the hypotheses in this thesis and they seem to be in the same magnitude as the simultaneity perception figures seen in the previous sections.

However, less than or equal to 100 ms is still a vague measure and, as seen in latency perception research, the threshold varies between tasks and modalities. Less than or equal to 100 ms means something between 0 and 100 ms (if response comes after action according to the causality). A focus in the research of this thesis is this: where exactly – between 0 and 100 ms – is the threshold when the usability and user experience start to degrade, in addition to the threshold of simultaneity perception between touch and different modality feedback? It might take a great deal of time and effort for engineers to reduce touch feedback latency from 100 ms to 50 ms, so it is very important to know the limits of the human and optimize the latency just right and not too much.

Perceived quality is one aspect of overall user experience. Although broad and subjective in nature, it will provide one measure of user experience in virtual button interaction. It has also been successfully used before by Hoggan *et al.* (2008b) and Park *et al.* (2011).

Although beyond the focus of this thesis, it is valuable to know for background and reference that the effect of latency has been investigated in Virtual Environments (VE) with different tracking systems (Ellis *et al.*, 1997; Ellis *et al.*, 1999a; Ellis *et al.*, 1999c; Ellis *et al.*, 1999b; Allison *et al.*, 2001; Adelstein *et al.*, 2003b; Meehan *et al.*, 2003; Ellis *et al.*, 2004; So and Chung, 2005; Teather *et al.*, 2009). The common result has been that latency has a significant negative effect in VEs, although the effect thresholds vary because of the different VE setups and tasks. The thresholds are typically below 100 ms.

The effect of latency on speech has also been explored in detail starting from the 1950's (e.g. (Fairbanks and Guttman, 1958; Chase *et al.*, 1961; Yates, 1963), but it is also beyond this thesis since the research here focuses on manual interaction. Some research also exists on the effect of latency on musical performance, especially with piano players (Finney, 1997; Dahl and Bresin, 2001).

The effect of latency on user experience and usability (performance, speed, errors, for example) has been studied relatively little in indirect manipulation such as using a mouse in interaction. The effect of latency in direct manipulation, which basically means touchscreen interaction with a finger or stylus, has been investigated even less. The next sections will introduce this research.

2.8.2 The Effect of Latency on Indirect Manipulation

The pioneering research by MacKenzie and Ware (1993) investigated the effect of cursor movement latency on a visual targeting task with a mouse. For the first time, they showed that latency has a negative effect on usability in mouse interaction. They found that with a latency of 225 ms, the movement time increased 64% and error rates 214% compared to the minimum latency of 8.3 ms. Based on their findings they created a mathematical model between the latency and the task completion time based on Fitts' Law.

Pavlovych and Stuerzlinger (2009) further investigated the effect of latency on performance during a targeting task with a mouse on a computer screen. They found that, with the small targets, the movement time started to increase when latency was above the system baseline latency of 33 ms whereas, with middle and large sized targets, the movement time increased after 58 and 83 ms, respectively. The effect of latency on the error rate was not as strong: a latency increase from 33 to 133 ms caused 10-15 % more errors.

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 performance, but haptic feedback latency had much less effect. Movement time went up significantly with visual and visual-haptic delays above 69 ms, whereas, with haptic feedback, only delays above 187 ms had an impact. There were no more errors with the haptic feedback delay, nor did the users rate the use as more difficult with haptic feedback delay. In contrast, both of these were significantly affected by 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.

2.8.3 The Effect of Latency on Direct Manipulation

Little research has been conducted on the impact of latency on direct manipulation of touchscreens with finger or a stylus. Even less research is available on the effect of latency on virtual button interaction on mobile devices. The research will be introduced next, preceded by a brief discussion about the reasons for the touchscreen interaction latencies.

2.8.3.1 What Causes Latency in a Capacitive Touchscreen

The sources of latency include a capacitive touch sensor, software which processes the interaction and output to the display, and the visual display itself (Hinckley and Wigdor, 2012; Ng *et al.*, 2012). The capacitive touch sensor causes latency because of its function. The location of a finger is scanned through the sensor with a certain sampling rate, which takes time, and often several scanning cycles are needed in order to reliably recognize the finger position. The software latencies include the time needed to recognize the interaction technique the user intended to use, processing time to interpret the input and calculate the response (Anderson *et al.*, 2011). As shown in Section 2.5, the feedback production also takes time. It takes time for the visual display to, for example, change from one colour to another for visual feedback. The audio production pipeline usually includes buffers that again can cause latency, if the audio data used for the feedback is not stored but generated again every time for every button press. The tactile actuators, especially ERMs, suffer from slow starting time because of the inertia of the spinning eccentric weight. An LRA features quicker start time than ERM, but can suffer from the same buffer challenges as audio feedback, since it works like a loudspeaker and can be driven with an audio signal.

2.8.3.2 The Effect of Latency on Touchscreen Interaction

Miller (1968) suggested in his early work that 100 ms would be acceptable for a direct manipulation interaction with a light pen. However, he stated that the strokes should be careful and slow. However, this work does not give much evidence or details about the derivation of this result.

Kaaresoja *et al.* (2011a; 2011b) studied the effects of latency on performance, error rate, and user preference in text entry with mobile touchscreen virtual buttons used with fingers. The device was similar to the device used by Tikka and Laitinen (2006) and Koskinen (2008) and was equipped with piezoelectric tactile actuator underneath the touchscreen (Laitinen and Mäenpää, 2006). They found that the text entry and error rates were not affected when the latency between touch and tactile feedback was constant and in the range of 18 to 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. These studies were the first attempt to understand the effect of latency on mobile touchscreen virtual button interaction. However, the latency range used was narrow compared to other literature. For example Winter *et al.* (2008) and Stone *et al.* (2001) used latency range up to 200 ms. In addition, the device featured a resistive touchscreen, which is not the technology utilised in most contemporary mobile phones. Capacitive touchscreens differ 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 the research in this thesis, a capacitive device was used to give data useful for current mobile phone designs. The research above only controlled tactile feedback latency and did not report, control or take visual feedback latency into account. They also completely ignored the audio component. These are all the focus of the research in this thesis, as they all are common forms of feedback in mobile devices.

Anderson *et al.* (2011) investigated the subjective effect of touchscreen latency in common touchscreen tablet tasks with a finger: web browsing, photo viewing and e-book reading. They modified latency in all these tasks from 80 to 780 ms. They particularly asked the participants to score the usability of the task when in different latency conditions. Their results revealed that the participants rated the usability of the conditions relatively highly in large latency levels: 580 ms was scored higher than 4 in their 1-5 scale ('1' corresponding 'bad' and '5' 'excellent'). They used a commercially available device, which could not provide latencies low enough to test the high latency conditions against the near-zero latency conditions. So the participants were not offered a really responsive system as comparison. In fact, the users in the work by Ng *et al.* (2012) explicitly noted this limitation: after being exposed to low latencies, they found the latencies of the commercial devices unacceptable. Anderson's work included only subjective data and did not include latency perception or a performance study.

Although Hinckley and Widgor (2012) claim that latency is always a problem, in their latency studies, Kaaresoja *et al.* (2011b) shed light on an alternative effect of latency in touchscreen interaction: latency may actually have some benefits, if used in a controlled way, as it can be used as an interaction design parameter. They again used a device featuring piezotactile feedback similarly as the device used by Tikka and Laitinen (2006) and Koskinen (2008). It was shown that virtual buttons could be made to feel heavier when tactile feedback latency was increased in virtual button finger interaction. Participants were asked to estimate the weight of a button in relation to a reference featuring the minimum latency of the system. A positive significant correlation was found between latency and perceived weight, and 78 ms tactile feedback latency was rated significantly heavier than the reference, and 118 ms latency was rated significantly heavier than 78 ms. A resistive touchscreen was again used and visual feedback latency was not controlled nor reported. In addition, there was no audio feedback involved in the interaction. However, these results show that the effects of latency have to be better understood in touchscreen virtual button interaction.

2.8.4 Conclusions

According to the recommendations, maximum latency between a button press and the responses should be 100 ms (Seow, 2008; Defence, 2012). The same figure comes out from the early study on touchscreen interaction (Miller, 1968). However, the contemporary research results are diverse and show that the effect of latency on usability and user experience is dependent on the task, modality and the research method. Thus, the effect has to be investigated case by case. The lowest latency threshold for the usability degradation in mouse interaction was found to be 33 ms (Pavlovych and Stuerzlinger, 2009) for visual feedback. When the haptic feedback was delayed in a targeting task, the latency affected the performance only when the latency was 187 ms (Jay and Hubbold, 2005). In touchscreen interaction it seems there is no floor effect for the performance enhancement until the latency goes down to 0 ms in dragging task (Jota *et al.*, 2013), whereas the highest latency for the subjective usability degradation was 580 ms (Anderson *et al.*, 2011).

2.9 Overall Conclusions

This chapter has presented an overview of touchscreen virtual button interaction. It has also discussed the earlier work on latency measurements in different contexts, as well as latency and simultaneity perception. Lastly, the effect of latency in indirect and direct manipulation has been introduced. This section goes through the research questions set in Introduction (Chapter 2) in the light of the earlier research. Before that the feedback designs selected for the research in this thesis are presented.

2.9.1 Feedback Designs

The overall conclusion is that any feedback helps the user in touchscreen interaction, visual feedback being the essential one and audio and tactile feedback helping the user even more, especially in the mobile interaction context. The feedback in this research is based on best practises from previous touchscreen feedback designs.

Visual feedback: pop-up feedback is popular in contemporary mobile phones and it has also been shown to be beneficial to users. Thus, a visual pop-up was selected as the visual feedback for the research in this thesis.

Audio feedback: according to Bender (1999) the audio feedback duration did not affect the number entry performance when the buttons were large (30 x 30 mm). According to

Adelstein *et al.* (2003a), duration (1, 50 or 200 ms) of the audio feedback did not have significant effect on asynchrony detection either. Therefore, a short “standard click” adopted from the popular Apple iPhone was selected as the audio feedback for the research in this thesis.

Tactile feedback: the research by Park *et al.* (Park *et al.*, 2011) suggests that a short rising time of tactile feedback is important as well as the short duration when creating realistic button clicks. Therefore, both short rise time and short duration were used as a tactile feedback in the research in this thesis. A short rise time also functions to reduce latency from the tactile actuator.

2.9.2 Research Questions

Research Question 1 asks:

> **RQ1:** Can an affordable and accurate touchscreen latency measurement tool be built?

In big mobile phone industry, it would be essential to be able to use a latency measurement tool in different places since research and development for multiple products is conducted typically on multiple of desks, buildings, cities and countries. Therefore, the aim in the research in this thesis was to build the latency measurement tool either inexpensive or portable. As affordable, the copying costs would be low benefiting the business. As portable (even not low-cost), it could be carried or posted to another location where it is currently needed. The review in Section 2.6 showed that a commercial multimodal touchscreen latency measurement tool exists, but it is complex and expensive (OptoFidelity) and not portable. The earlier research introduced shows that latency measurements for different modalities can be done with simple methods and equipment. The multimodal latency measurement tool designed and implemented based on these methods is introduced in Chapter 3.

Research Question 2 asks:

> **RQ2:** What are the touch-feedback simultaneity perception thresholds in virtual button interaction?

The latency perception in touchscreen interaction has been researched mostly in dynamic gesture-like interaction techniques, such as dragging, scribbling and inking. However, one study exists on the latency perception of visual feedback on touchscreens, as introduced in Section 2.7.5. The latency perception threshold was found to be 64 ms. However, this research was established with a less ecologically valid comparison method and on a tablet-sized device placed on a table. The simultaneity perception of touch and visual feedback in handheld mobile touchscreen virtual button interaction has not been studied before. Also, unknown are the simultaneity perception thresholds for touch and audio, as well as touch and tactile feedback. The simultaneity perception threshold for two events of different modality has been found in earlier research to be between approximately 25 and 200 ms (most of the being between 25 and 100 ms). Thus, the hypothetical answer for RQ2 is between 25 and 200 ms. The research in Chapter 4 investigates the simultaneity perception threshold for touch and unimodal visual, audio or tactile feedback.

The simultaneity perception threshold for more than two events has never been investigated. In a standard touchscreen mobile phone there are usually more than one modality involved in the virtual button interaction. Therefore, a simultaneity perception of touch with a bimodal feedback is investigated and the perception thresholds derived for all three modality pairs: visual-audio, visual-tactile and tactile-audio. This research is introduced in Chapter 5.

Research Question 3 asks:

> **RQ3:** How the perceived quality of a virtual button changes when latency between touch and feedback changes?

Perceived quality was used as a measure by Hoggan *et al.* (2008b) and Park *et al.* (2011) when assessing user experience during touchscreen virtual button interaction. However, these studies focus on the feedback design. The latency threshold for the usability degradation has shown to be from 1 to 580 ms, a huge deviation. Therefore, the research in Chapter 4 investigates the effect of latency on the perceived quality when the feedback is a unimodal event: visual, audio or tactile.

As stated above, the feedback in a touchscreen mobile phone commonly features more than one modality, so it is important to investigate the effect of perceived quality on bimodal feedback. This research is introduced in Chapter 5 for all three modality pairs: visual-audio, visual-tactile and audio-tactile.

Research Question 4 asks:

> **RQ 4:** What are the latency guidelines recommended for visual, audio and tactile feedback for a virtual button press?

The guideline for each modality will be derived based on the simultaneity perception threshold and the latency values when the perceived quality starts to degrade. This guideline is not necessarily the same for unimodal and bimodal cases. The guidelines will be established based on the research results in Chapter 4 and Chapter 5.

Chapter 3 Multimodal Latency Measurement Tool

3.1 Introduction

According to the literature, latency may be harmful for interaction. To understand if latency is an issue in a device, or to take some corrective actions, it is necessary to be able to measure the latencies between user action and device response in the visual, audio and haptic modalities. There are several research prototypes and methods for measuring visual latencies (He *et al.*, 2000; Miller and Bishop, 2002; Steed, 2008; Di Luca, 2010). However, they are based on indirect input in Virtual Environments or dynamic interaction in touchscreen instead of a virtual button press. In addition, none of them detect the moment of touch. The measurement method has been based on normal or high-speed video camera and the latter was used also in the measurement tool described in this thesis.

Audio latencies have been investigated with 2-channel audio recorder setup (Freed *et al.*, 1997; MacMillan *et al.*, 2001; Wright and Brandt, 2001; Wright *et al.*, 2004). In practise it means using the standard stereo sound card in every Personal Computer and a sound editor software. For example, the sound of finger tap onto a button is recorded to one channel and the sound of the button feedback to another and the latency has been easy to derive with a sound editor. Although this method has not been applied to touchscreen use, it has been shown to be good practice. In addition, and because of the affordability and the simplicity, this approach was adapted into use in the multimodal measurement tool introduced in this chapter.

Tactile feedback latency in touchscreens has been measured with a force sensor which also detected the moment of touch (Lehtosalo, 2009). The latency was derived with a mathematical software. Even though effective, this approach was not adopted because of the proprietary hardware needed for both recording and saving the data into a PC. In addition, using the same 2-channel recording setup for tactile feedback latency measurement as for audio feedback latency measurement would shorten the measurement, make the tool simpler and the analysis easier.

The previous research methods have all been unimodal meaning that only visual, audio or tactile latencies were measured at a time. Commercial products[25] exist for multimodal latency measurements in different contexts, but they are big, expensive or clumsy to use as seen in Chapter 2. In product development of mobile phones, many times multiple products are developed in the same time in different sites. In addition, many different developers work in the same time on the same product. It would be beneficial to have either inexpensive multimodal latency measurement tool in order to copy it at low-cost for all developers who need it, or portable in which case it could be carried from desk to desk or site to site if needed. Therefore, Research Question 1 asks:

RQ1: Can an affordable and accurate touchscreen latency measurement tool be built?

This chapter describes the multimodal latency measurement tool responding to the needs and shortcomings in the earlier work. Section 3.2 introduces the detailed design drivers for the latency measurement tool. General latency measurement methodology theory is introduced in Section 3.3. Section 3.4 describes the latency measurement method between touch and visual feedback in this tool and Section 3.4.1 explains how the moment of touch was detected with the visual feedback latency assessment. Audio and tactile latency measurement method is described in Section 3.5. The different components of the tool, general system setup and the latency extraction of all the modalities are discussed in Sections 3.6 and 3.7. Section 3.8 describes the calibration of the measurement tool. Section 3.9 describe sample measurements of five touchscreen phones and introduce the results of the measurements. Section 3.10 discusses the limitations of the tool before the discussion and conclusions are given in Section 3.11.

3.2 Design Drivers

The aim was to build a multimodal latency measurement tool which would be affordable, easy to build and accurate, making it beneficial or even a necessity for both researchers conducting user experiments and product designers struggling with delays in hardware, software and user interfaces. The following detailed drivers were derived for the tool design:

[25] www.blackboxtoolkit.com, www.optofidelity.com

1. The tool should be as inexpensive as possible so as to be for purhcase by every researcher and designer;
2. The tool should be easy to build without deep electrical or mechanical expertise. That means the use of off-the-shelf equipment and components as much as possible;
3. There should be absolutely no hardware or software modifications to the devices to be measured. That would make the measurement quicker and easier and the results more reliable because the device would be measured 'as is';
4. The tool should be capable of measuring latencies between the moment of touch and the visual, audio and tactile feedback;
5. The tool should be capable of measuring devices both with resistive and capacitive touchscreens;
6. The tool would be for laboratory use, but should be still as portable as possible in order to change the place of the measurement to different R&D sites of a big company, for example.
7. Based on the human perceptual capabilities, measurement resolution and accuracy of the tool should be a minimum 1 ms in audio and tactile, and 10 ms for the visual modality.

3.3 General Methodology

In any measurement there is a stage of data capture and a stage of data analysis. The measurement of latency actually means timekeeping between an action and a response. The data can be captured by recording and then finding the moment of the start and end of the clock to extract the latency. Sometimes, however, special capture methods are needed to make either the recording or the analysis easier. Based on earlier work in latency measurements, a simple classification of these advanced capture methods is introduced: Simplification, oversampling and transcoding.

Simplification means that the data acquisition compresses and filters the data so that changes in the time domain are more easily observable. The time domain can be also stretched with oversampling in order to zoom in time and simplify the analysis. Both of them will make the investigation of the relevant information easier. Transcoding means that the data is transformed from one modality to another to be more easily captured or analysed. In the implementation of the multimodal latency measurement tool, oversampling and transcoding methods were used.

Data analysis can be done either manually by a human from the captured data, or it can be automated with signal processing methods. The tool introduced in this chapter is the initial version, therefore, the data was analysed manually.

An example of simplification as well as oversampling is the Latency Meter (Miller and Bishop, 2002). Instead of using normal speed video of a user and the VE, they used high-speed CCD sensors to simplify the "picture" by calculating two single numbers from the brightness distribution from the CCD sensors. These quickly updated numbers were used to calculate the latency between the user action and the VE response. Another example of simplification, but also transcoding is the work of Freed *et al.* (1997) in which they measured operating systems latencies. They transcoded network events into audible beeps, but also simplified by transcoding only the start and the end of the events.

As stated before, one another common practice used in latency measurements has been the use of an audio recorder. It can be a two or multiple channel soundcard found in a standard PC or it can be a separate piece of hardware. This method requires the events to be transcoded into audio first if needed. Events are recorded in the different channels in order to easily extract the latencies between them using an audio editor. This methodology has been in use in several research projects (Freed *et al.*, 1997; MacMillan *et al.*, 2001; Wright *et al.*, 2004).

3.4 Measurement of Latency Between Touch and Visual Feedback

A high-speed camera was selected to find out what exactly happens when a touchscreen button is pressed and the graphical UI changes. There are plenty of industrial high-speed cameras available, but they are big and expensive. Fortunately, there are consumer-grade digital cameras at reasonable prices that can record high-speed video, for example the Casio Exilim EX-F1[26] seen in Figure 3-1. It is capable of recording 300, 600, and 1200 frames per

[26] exilim.casio.com

second with 512x384, 432x192, 336x96 pixel resolutions, respectively. The 300 fps with the almost VGA resolution showed good data (see Figure 3-3), so that was selected. The high-speed video can be viewed with the freely available MiDas player[27] by Xcitex. However, the EX-F1 produces high-speed video only in MOV video format, which is not supported by the MiDas player, which only supports AVI format. Fortunately, there is a MOV to AVI video converter[28] freely available to solve this problem. With these tools it was possible to transfer the MOV high-speed video from the EX-F1 to the MiDas player for analysis in AVI format. One of the advantages of the EX-F1 is that it can also be used as normal digital camera, which is also sometimes needed in product development, saving money from buying another.

Figure 3-1: Casio Exilim EX-F1 used as high-speed camera for visual feedback latency measurements.

3.4.1 Detecting the Moment of Touch

There were multiple candidates for detecting the touch of the fingertip or stylus on the surface of the touchscreen. One was to use a programmable robot arm with a force gauge. That would have provided a controllable stimulus, but would have been challenging for capacitive touch screens and a robot would have been expensive and far from portable. A force sensor is used in OptoFidelity's Watchdog (OptoFidelity, 2014) and was used also by Lehtosalo (2009) in his tactile feedback latency measurement system. A force sensor would

[27] www.xcitex.com

[28] www.pazera-software.com

have needed a synchronisation mechanism between the force signal and the visual feedback, which was not available off-the-shelf.

An alternative was to create a transcoder from touch to light, by building a stylus-like device with a sensitive switch on one end and an LED on the other. This way the moment of touch would have been seen in the high-speed video as an LED light. However, this would have again caused issues with capacitive touchscreens and required building of new hardware, which not all users of the tool would be capable.

After some investigation, a simple solution was found: an inexpensive make-up mirror with an adjustable support (see Figure 3-2). It was placed next to the touchscreen device to be measured under the high-speed camera and adjusted so that the stylus or finger tap could be seen in the video stream. In this way, the user interface recording was inherently synchronized with the touch detection. The method enabled also a normal interaction with the device, with a finger or stylus. In addition, this methodology made it possible to measure both resistive and capacitive touchscreens. Figure 3-3 shows an example picture sequence of stylus approaching the touchscreen. The moment of touch is easily seen in the pictures viewed by the mirror.

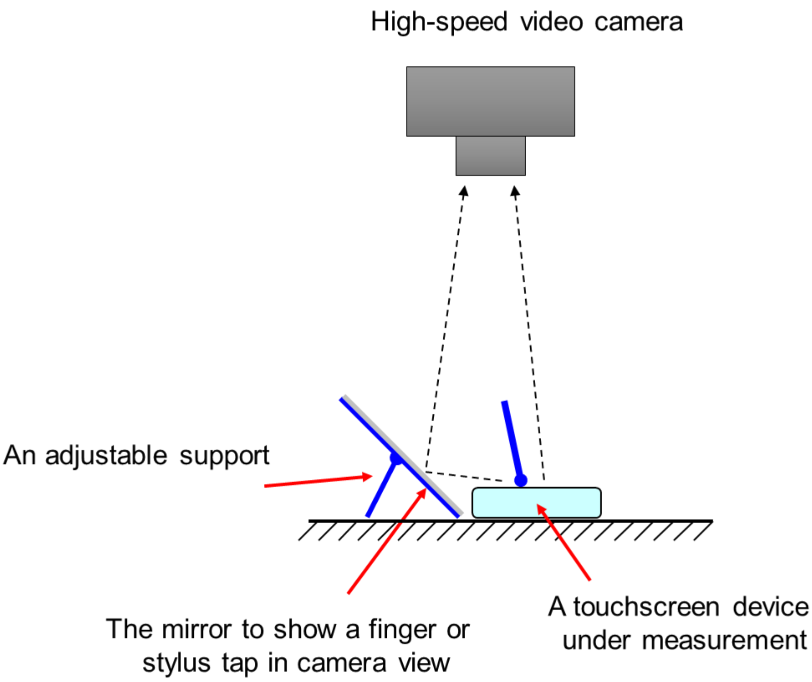


Figure 3-2: The mirror arrangement for recording a finger or stylus tap on high-speed video together with its visual feedback.

1.
2.
3.

Figure 3-3: A picture sequence showing the stylus tip (inside red circles) approaching and finally touching the touchscreen surface of a mobile phone. The mirror can be seen on the left and gives a clear view of when the stylus hits the screen. The picture is of the Samsung Omnia i900[29] dialler user interface.

---

[29] www.samsung.com

3.5 Measurement of Latency Between Touch and Audio-Tactile Feedback

Recording the latency between touch and audio feedback is a simple task, since both of the events can be recorded with a small microphone. An inexpensive Vivanco EM216[30] (later known as EM35[31] as shown in Figure 3-4) lavalier microphone was used, since the microphone element was located on the side (instead on the top) of the microphone capsule enabling an easy setup on a mobile phone under measurement. The microphone was connected to the left channel of the Terratec Aureon 5.1 USB MKII[32] soundcard via line-in input.

Figure 3-4: The Vivanco EM 35 lavalier microphone used for audio feedback latency measurement. The microphone was used without the tie clip. The box labelled as EM 35 is a preamplifier enabling the connection to soundcard line-in input.

Using a microphone also would work for some tactile feedback since the vibration generally causes audible noise that can be recorded by a microphone. However, initial tests showed that the microphone did not pick up the sound of the tactile feedback accurately enough, especially if the tactile feedback intensity was low and the duration short. Therefore, a small accelerometer was added. It was desirable to make use of the sound card for recording also the accelerometer data, using the 2-channel recorder principle, because that would have made the both setup and the analysis simpler. Thus, an analogue accelerometer circuit was

[30] discontinued, www.vivanco.de

[31] discontinued as well

[32] Currently sold as Aureon 7.1 USB, www.terratec.com

arranged to enable the connection to the same line-in input as the microphone. The analogue output of the accelerometer was transcoded into audio. It turned out to work well with the line-in input of the soundcard, so no extra data acquisition hardware was needed. The accelerometer circuit board was the only piece of hardware that was not off-the-shelf, but it was a simple one. The accelerometer component was Kionix KXPS5[33]. The accelerometer board weighted 0.9 grams and was 11 x 13 mm in size (Figure 3-5). Figure 3-6 shows the principal circuit of the accelerometer used. In addition to tactile feedback, the accelerometer picked up the touch event much better than the microphone (see Figure 3-8).

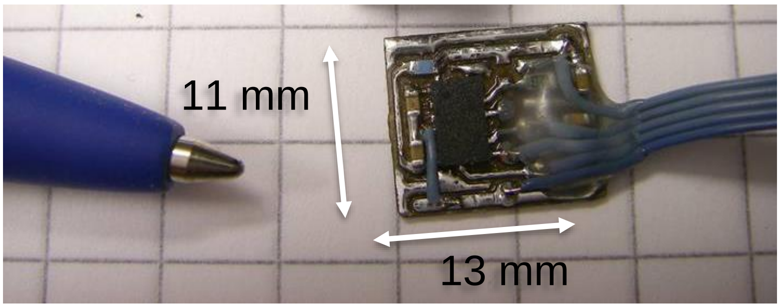


Figure 3-5: The accelerometer board built specifically for measuring the tactile feedback. The board was designed and implemented by Tom Ahola. Picture by Tom Ahola (Nokia Research Center).

At the time of the measurement tool development, there were no small off-the-shelf analogue accelerometer boards available. Today, there are several, for example from Adafruit[34], Seeed[35] and Sparkfun[36]. They are all based on Analog Devices' ADXL335[37] accelerometer component. If one of these boards will be used, one needs to provide voltage to the board and connect Z-axis output with the ground to the right channel of the line-in input of the soundcard. Figure 3-7 shows an example how the microphone and accelerometer could be attached to a phone under measurement. For recording and analysis, the freely available

[33] http://www.kionix.com/accelerometers/kxps5

[34] http://www.adafruit.com/products/163

[35] http://www.seeedstudio.com/depot/Grove-3Axis-Analog-Accelerometer-p-1086.html

[36] https://www.sparkfun.com/products/9269

[37] http://www.analog.com/en/products/mems/mems-accelerometers/adxl335.html

Audacity[38] v.1.3.6 open source sound editor software was used, capable of showing timing information in millisecond resolution (see Figure 3-8).

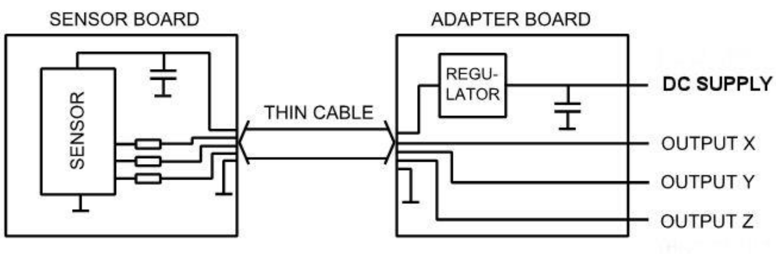


Figure 3-6: The accelerometer circuit. Only OUTPUT Z was used in the measurements.

Figure 3-7: The microphone and accelerometer attached to a phone under measurement.

[38] audacity.sourceforge.net

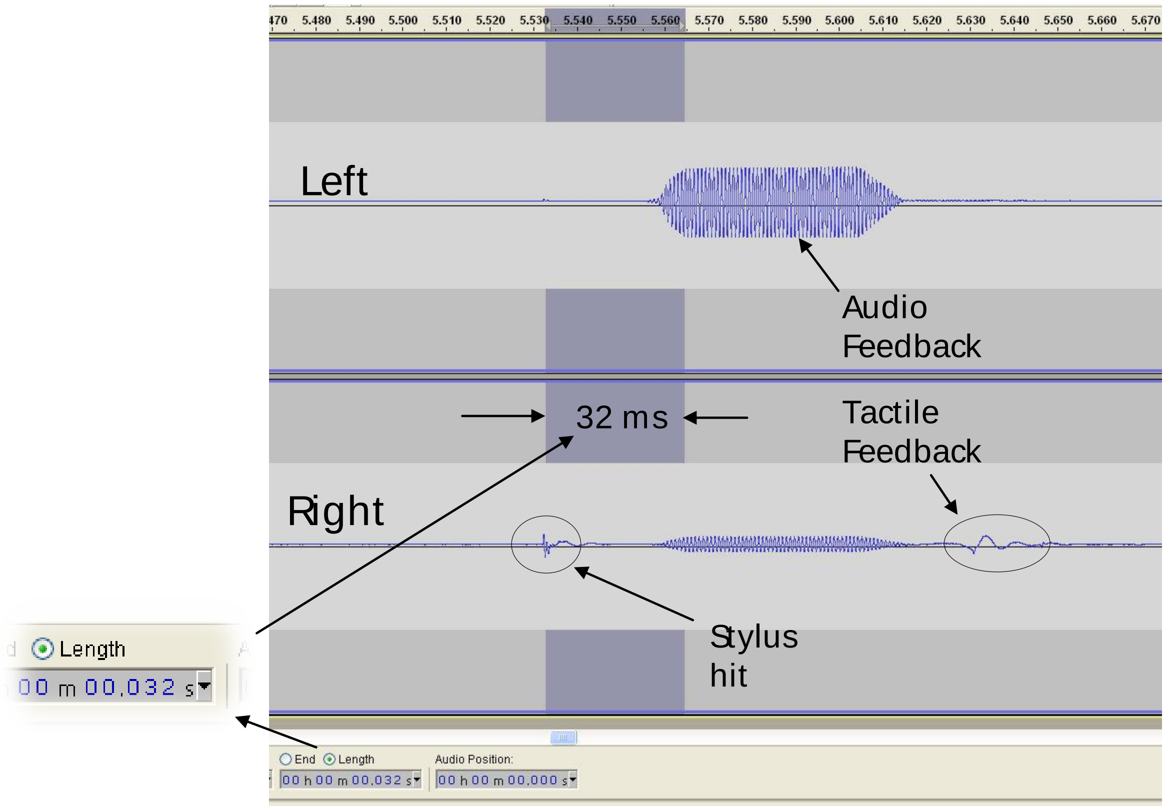


Figure 3-8: Example of a record of touch, audio and tactile feedback of a virtual button press in Audacity, a free sound recorder software.

3.6 Integration

3.6.1 Overall Setup

Figure 3-9 shows the overall setup and Figure 3-10 a close-up of the latency measurement tool. The centre of the tool was the camera attached to the table with a clamp and a 6 Degree-of-Freedom adjustable arm. With the arm the camera was easy to set to the right position to record both the phone UI and the view from the mirror. The mirror was placed next to the measured device and opposite to it was placed a white background to make a clearer image (a folded sheet of paper). Two inexpensive LED lamps as light sources can be seen on both sides of the camera, as well as the microphone and accelerometer (for clarity, not attached to the phone in this picture).

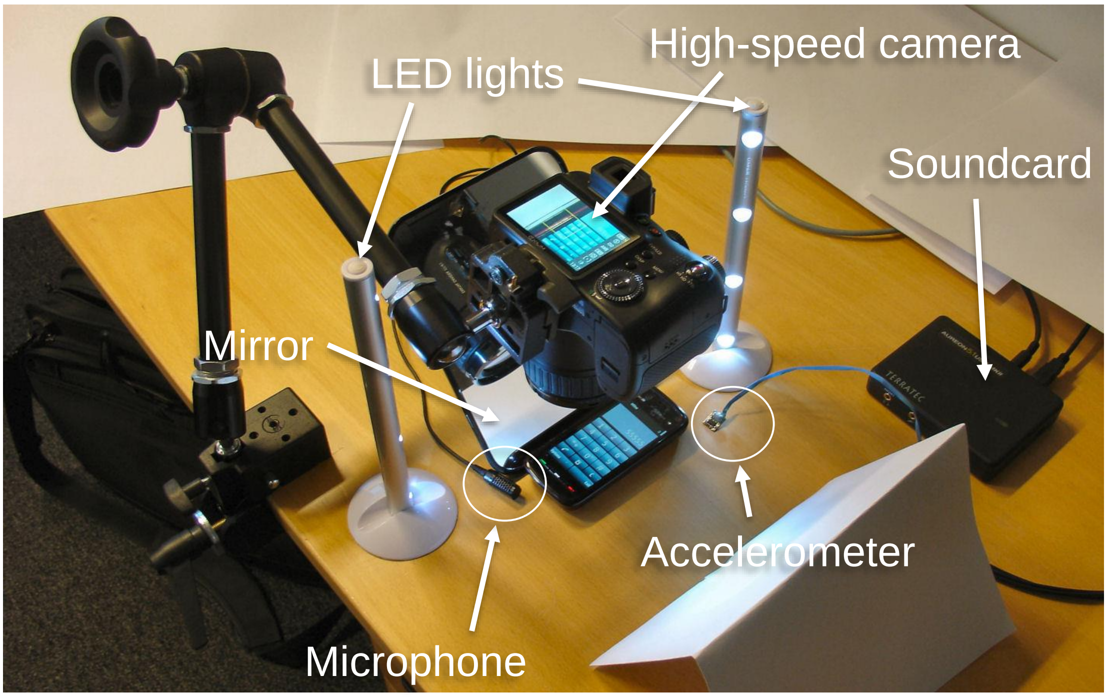


Figure 3-9: Overall setup of the multimodal latency measurement tool. For clarity, the microphone and the accelerometer are not attached to the measured phone in this picture.

Figure 3-10: Close-up of the mirror arrangement. The microphone (front) and the accelerometer (back) are not attached to the phone in this figure.

For transporting the tool, the components were packed in a cushioned carrying case by HPRC[39] and the camera with accessories in a standard camera carrying bag. The carrying case size was 40.5 x 33.0 x 16.5 cm (l x h x d) and weight 4.3 kg. The camera bag by Lowepro[40] size was 21.0 x 17.0 x 22.0 cm and weight 1.5 kg. Since the overall weight of the tool was 4.3 + 1.5 = 5.9 kg and the size of the carrying case and the bag were of the size of a briefcase, it could be carried to different locations to conduct measurements.

3.6.2 Block Diagram of the Whole System

Figure 3-11 shows the overall block diagram of the tool containing the key elements of the system: the high-speed camera, microphone, accelerometer, mirror, MOV to AVI converter, slow-motion video player, soundcard and sound editor.

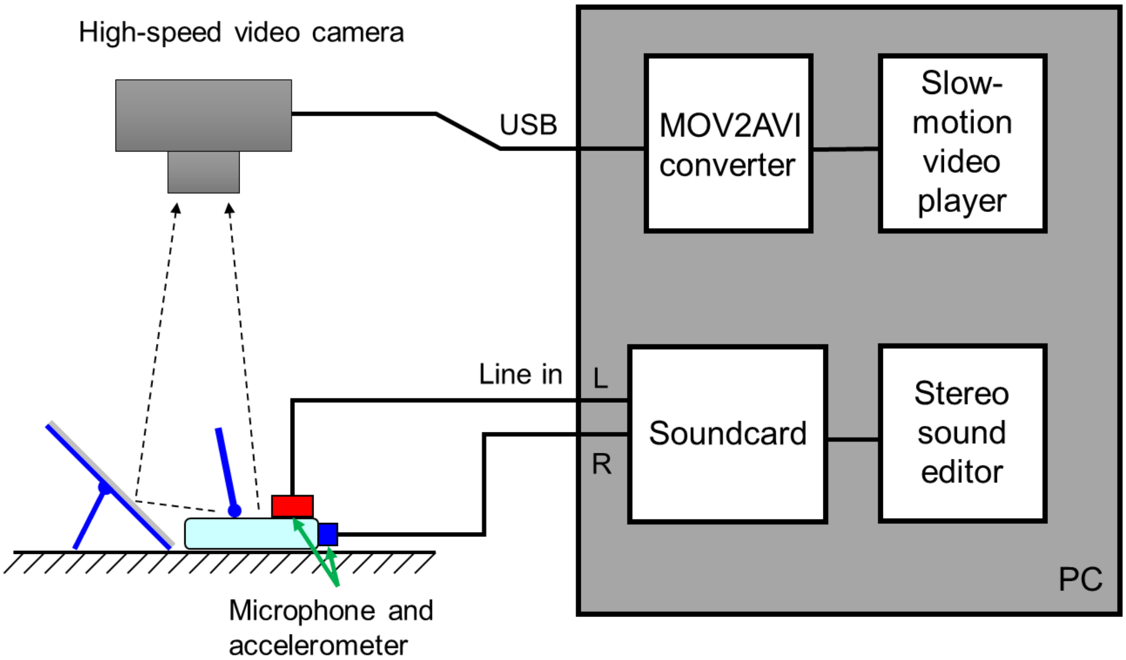


Figure 3-11: Overall block diagram of the multimodal latency measurement tool.

3.6.3 Bill of Materials

---

[39] http://www.hprccases.com.au/hprc/2400.htm

[40] www.lowepro.com;
http://www.amazon.co.uk/Lowepro-Nova-140-AW-Shoulder/dp/B0016JA2YS

Table 3-1 shows the list of components and their prices. The total cost of the components was approximately 1000 €. The most expensive component was the camera, which might be available cheaper today as the technology has developed. The off-the-shelf alternatives for accelerometer boards introduced earlier cost 10 – 15 € today.

Table 3-1: List of components and Bill of Materials of the multimodal latency measurement tool.

Hardware	Brand	Reference	Price
High-speed camera	Casio Exilim EX-F1	www.exilim.casio.com	600.00 €
Adjustable arm	Manfrotto 244rc	www.manfrotto.com	200.00 €
Make-up mirror	Any with adjustable support	Supermarket	10.00 €
LED lamps	OSRAM Ledstixx	www.osram.com	40.00 €
Lavalier Microphone	Vivanco EM-35	www.vivanco.com	25.00 €
Soundcard	Terratec	www.verkkokauppa.com	50.00 €
Accelerometer	Kionix KXPS5	www.kionix.com	100.00 €
Stereo Breakout Cable	Hosa YMM-261	www.amazon.co.uk	5.00 €
Double sided tape	Any		5.00 €
Cellotape	Any		5.00 €
White background	Folded A4	Any copyroom	0.01 €
Software			
MOV2AVI converter	Pacera	www.pazera-software.com	free
Audio editor	Audacity	www.audacity.sourceforge.net	free
Frame-by-frame video player	MiDas player	www.xcitex.com	free
Together			**1 040.01 €**

3.7 Extracting the Latency between Touch, Visual, Audio and Tactile Modalities

The high-speed video and audio-tactile streams were inherently synchronized by the stylus or finger tap seen in both streams. The visual feedback latency was extracted by playing the high-speed video with MiDas player frame-by-frame and finding the frame where the stylus or finger first touched the screen surface and the frame when the visual feedback was fully given. The latency in seconds was then measured as the frame number difference divided by 300 fps, the frame rate of the high-speed video. For convenience the number was multiplied by 1000 in order to get milliseconds. See Figure 3-12.

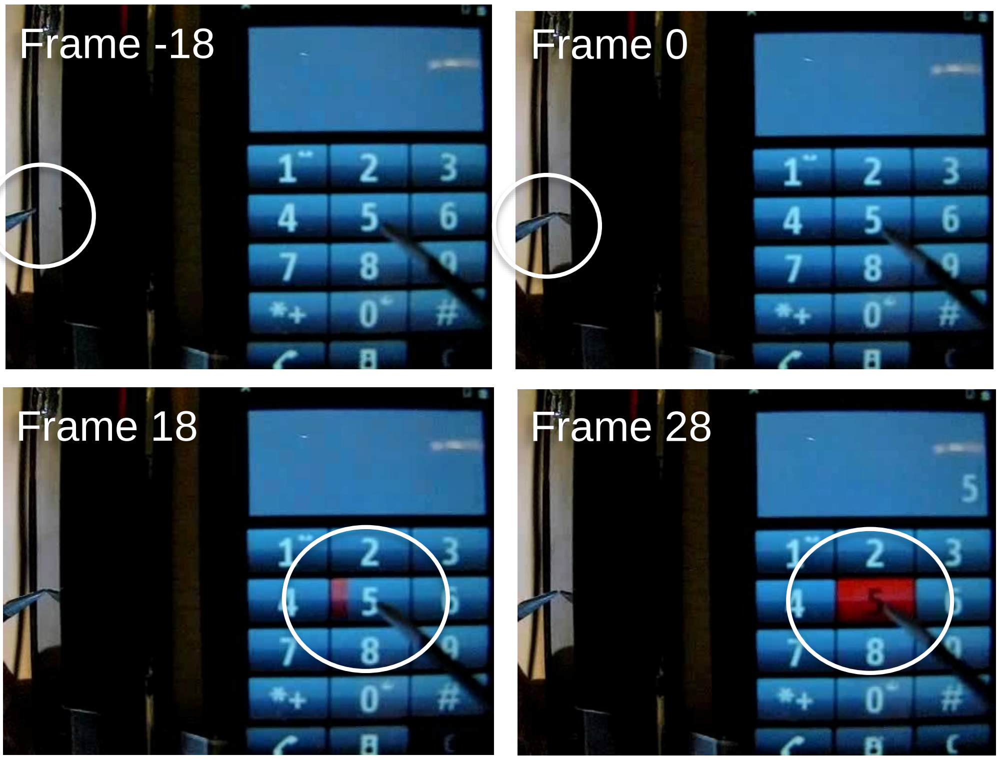


Figure 3-12: Example of visual feedback latency assessment. The stylus and the moment of touch can be seen on the left in every picture thanks to the mirror arrangement. Top left: The stylus is approaching. Top right: Stylus touches the screen for the first time. This is the Frame 0 and the frame calculation starts. Bottom left: After 18 frames (= 60 ms) the visual feedback of the button pressed starts to emerge. Bottom right: After 28 frames (= 93 ms) the visual feedback is fully given. This is the visual feedback latency in this single measurement. Figures used with permission from Microsoft.

Figure 3-8 shows an example recording of touch, audio and tactile feedback in Audacity. The audio feedback can be seen above on the left channel, and the touch (stylus hit) can be seen on the right channel in addition to the tactile feedback (and an attenuated trace of audio feedback). The latency between the touch could be measured with Audacity’s selection tool. Latencies between different modalities can be extracted simply by subtracting the latencies of them.

Before conducting any measurements, a calibration procedure was needed to validate if the tool was accurate and well synchronized. This procedure is described in the next section.

3.8 Calibration of the Measurement Tool

Although the internal clocks of the video camera and soundcard should have been accurate enough, a calibration test was arranged for the tool to validate the accuracy of the all the recording channels. An LED and a small loudspeaker were connected to the output of a calibrated Agilent[41] 33120A[42] Arbitrary Waveform Generator (AWG) (Figure 3-13). The high-speed camera was recording the LED, the microphone was picking up the audio above (2 mm) the loudspeaker and the accelerometer was attached to the bottom of the loudspeaker. The calibration method was to time two different durations: 100 ms and 1000 ms.

The AWG generated bursts of 1 kHz to create both visible and audible signal. The AWG was programmed to play a burst sequence of ten 10 ms bursts per second for measuring 100 ms (the time between two bursts). For timing 1000 ms the AWG was programmed to play a burst sequence of one 100 ms burst per second. The measurement was repeated 10 times for each time length.

The calibration results with standard deviations for the measurement resolution (σ_r) and the measurements (σ_m) are shown in Table 3-2. The mean measurement error remained under 0.1 % for all cases except for the visual measurement of 100 ms which gave a 0.37 % mean error (0.37 ms). These errors, as well as the standard deviations, were small relative to human temporal perception (> 2 ms), therefore it could be concluded that the tool had the resolution and accuracy needed to measure the latency of different feedback modalities.

---

[41] www.agilent.com, http://www.keysight.com

[42] http://mntl.illinois.edu/Equipment/docs/Agilent33120Auserguide.pdf

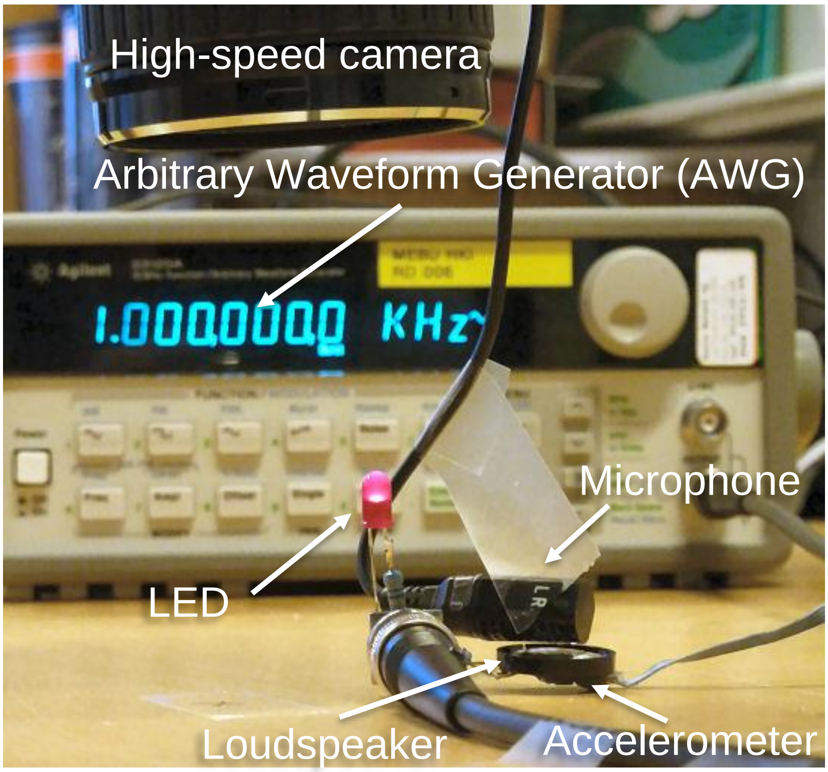


Figure 3-13: The calibration setup consists of an Arbitrary Waveform Generator AWG), LED and loudspeaker. The microphone picked up the sound and accelerometer the mechanical vibration from the loudspeaker.

Table 3-2: Measurement resolutions and calibration results with standard deviations. All units are milliseconds.

Record channel	**Reference**	**Measurement resolution**	**Mean result**	**Difference**	**Relative Error**	**Stdev of measurement resolution (σ_r)**	**Stdev of measurements (σ_m)**
Audio	100	0.0227	99.93	0.07	0.07 %	0.0065	0.027
Audio	1000	0.0227	999.4	0.6	0.06 %	0.0065	0.026
Tactile	100	0.0227	99.92	0.08	0.08 %	0.0065	0.019
Tactile	1000	0.0227	999.4	0.6	0.06 %	0.0065	0.023
Visual	100	3.4	99.67	0.37	0.37 %	0.96	1.05
Visual	1000	3.4	999.3	0.7	0.07 %	0.96	1.41

3.9 Sample Measurements

To evaluate the functionality of the latency measurement equipment, a study was conducted to measure virtual button latencies in some commercial mobile phones. All the phones featured touchscreens and all of them had audio and tactile feedback for the buttons in addition to visual. Four touchscreen mobile phones from different manufacturers were chosen. In addition, half of them should feature resistive and half of them capacitive touchscreen. Using both, it could be shown that the latency measurement tool was capable of measuring devices with both technologies. Figure 3-14 shows the phones which fulfilled the criteria were: HTC[43] Desire, LG[44] Chocolate BL40, Nokia[45] 5800 XpressMusic, and Samsung Omnia[46] i900. Four was considered a reasonable number without measurement sessions taking an unnecessary long time, but still delivering good amount of data for the analysis. Figure 3-15 shows the audio-tactile feedback latency measurement setup for all the phones.

Picture of HTC Desire has been removed due to Copyright restrictions.

Figure 3-14: Four touchscreen phones were measured: HTC Desire, LG Chocolate BL-40, Nokia 5800 XpressMusic[47] and Samsung Omnia i900.

[43] www.htc.com

[44] www.lg.com

[45] www.nokia.com

[46] omnia.samsungmobile.com

[47] Picture used with permission from Microsoft.

Figure 3-15: Audio and tactile feedback (with the microphone and accelerometer) measurement setups for the mobile phones. Top: LG, Samsung and Nokia. The loudspeaker was located on the left side of the Nokia phone. Bottom: The loudspeaker was located underneath of the HTC phone. Therefore, the microphone was placed under the phone and the measurement happened on top of a book. The accelerometer was placed on the bottom of the phone.

For consistency, similar applications were tested in each phone so a dialler and text editor were chosen, because these applications can be found in any mobile phone. The dialler is the application with a number keypad for making a phone call and the text editor with a full QWERTY keypad is used for creating messages. The text editor was used in number mode to make it easier to compare to the dialler.

Number ‘5’ virtual button in both application keypads was pressed 10 times. All the four phones were tested with a finger and the two phones with resistive keypads (Nokia and Samsung) were also tested with a stylus. The visual, audio and tactile feedback were recorded all in the same time with the tool and the latencies extracted as described in the previous sections.

3.9.1 Hypotheses

The hypotheses were:

(H1) There will be **observable virtual button latencies** in the devices measured by the tool;

(H2) The latencies will be **shorter with stylus rather than finger** interaction since the finger needs time to deform as it makes contact with the surface of the device.

3.9.2 Results

3.9.2.1 Finger usage

The figures in the next page show the recordings of visual feedback and the touch detection for all the measured phones (HTC: Figure 3-16, LG: Figure 3-17, Nokia: Figure 3-18, Samsung: Figure 3-19). The figures clearly show the frame during a finger is touching the screen and the frame when the feedback has been fully given. In LG the latency is so long that finger has already released from the screen when the feedback occurs (Figure 3-17). Next figures (HTC: Figure 3-20, LG: Figure 3-21, Nokia: Figure 3-22, Samsung: Figure 3-23) show audio and tactile feedback recording in addition to the touch of one button press during the measurements. The feedback could be clearly seen in the Audacity sound editor screen and the feedback delays could be assessed with the selection tool by Audacity. The average latencies with standard deviations of both dialler and text editor with all the feedback modalities in all measured phones are shown in Figure 3-24. These figures show that (H1) is supported. Latencies vary from 30 ms audio feedback latency in Nokia 5800 text editor to 367 ms audio feedback latency in LG BL40 dialler application.

Figure 3-16: Finger touch and visual feedback (grey) in HTC dialler, which can be seen despite of the occluding finger.

Figure 3-17: Finger touch and visual feedback (red) in LG dialler. Note that the feedback comes so late that finger has already been released.

Figure 3-18: Finger touch and visual feedback (red) in Nokia dialler.

Figure 3-19: Finger touch and visual feedback (blue) in Samsung dialler.

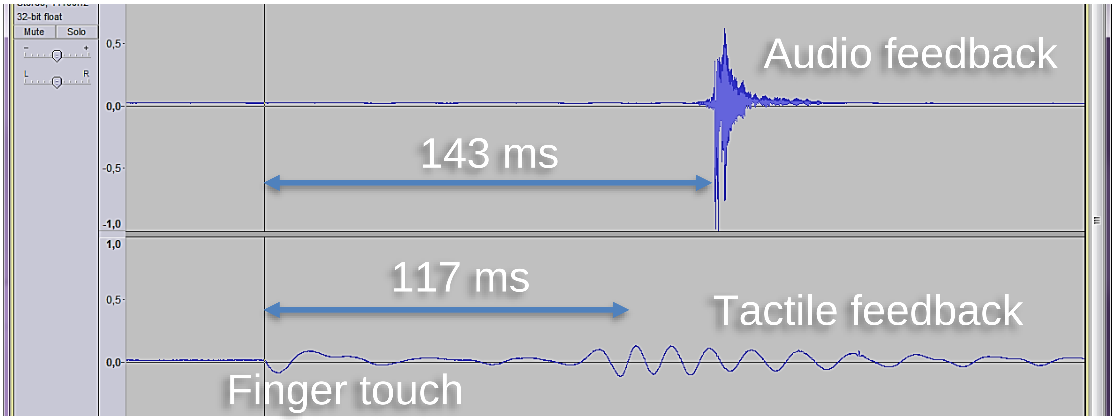


Figure 3-20: One measurement of finger touch and audio and tactile feedback in HTC text entry application recorded with Audacity.

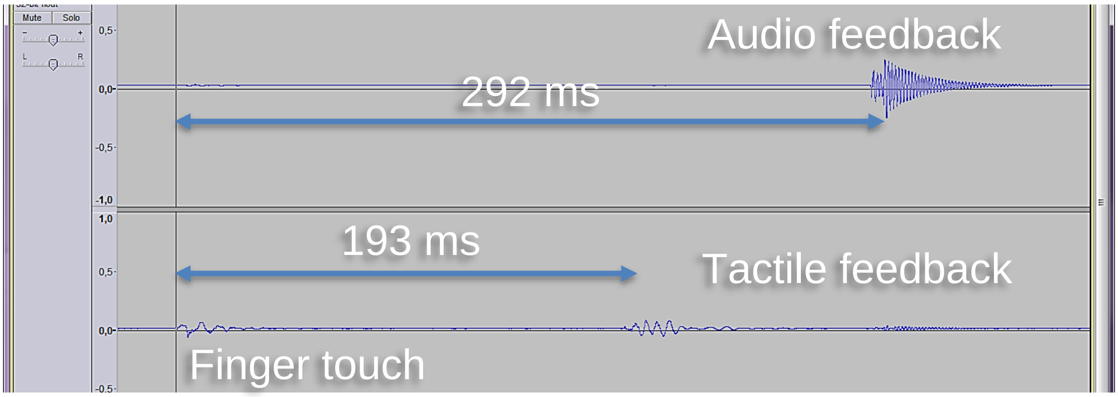


Figure 3-21: One measurement of finger touch and audio and tactile feedback in LG text entry application recorded with Audacity.

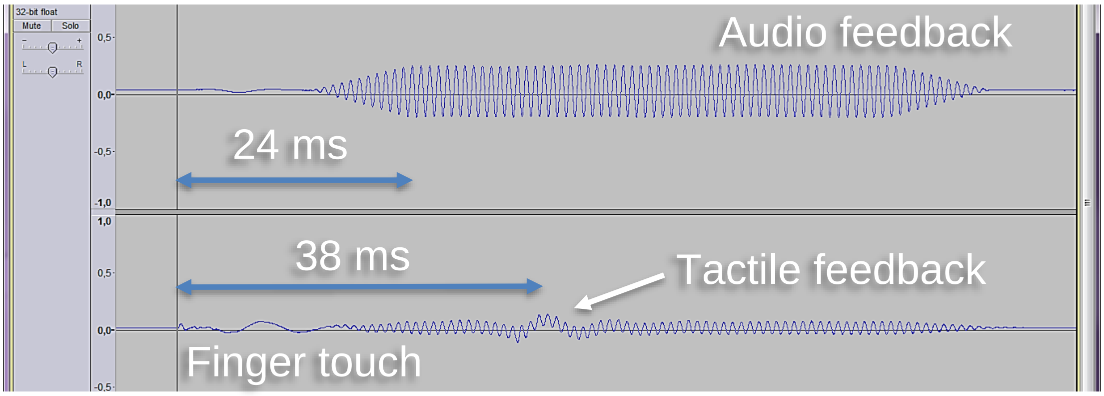


Figure 3-22: One measurement of finger touch and audio and tactile feedback in Nokia text entry application recorded with Audacity. Note that the tactile feedback is partially masked by the audio feedback since tactile feedback occurred during audio feedback was still playing. It is still clearly visible.

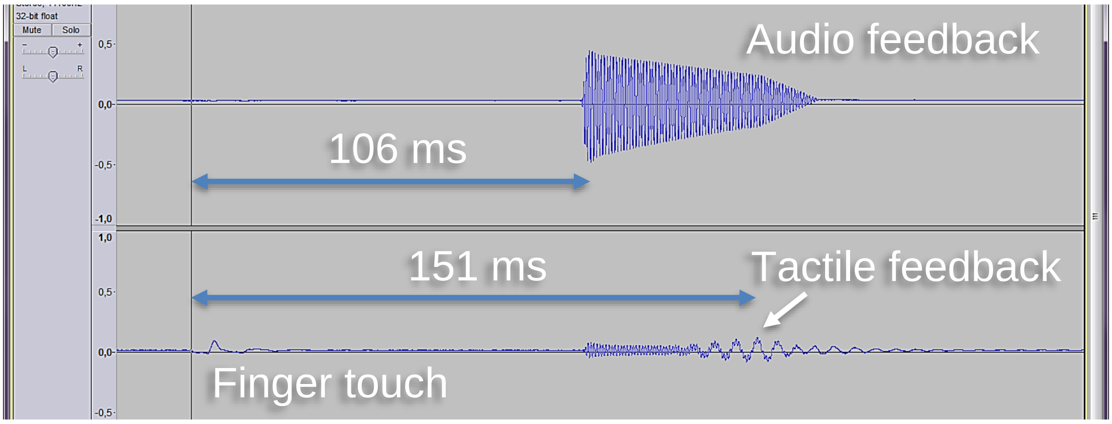


Figure 3-23: One measurement of finger touch and audio and tactile feedback in Samsung text entry application recorded with Audacity. Note that the tactile feedback is partially masked by the audio feedback since tactile feedback occurred during audio feedback was still playing. It is still clearly visible.

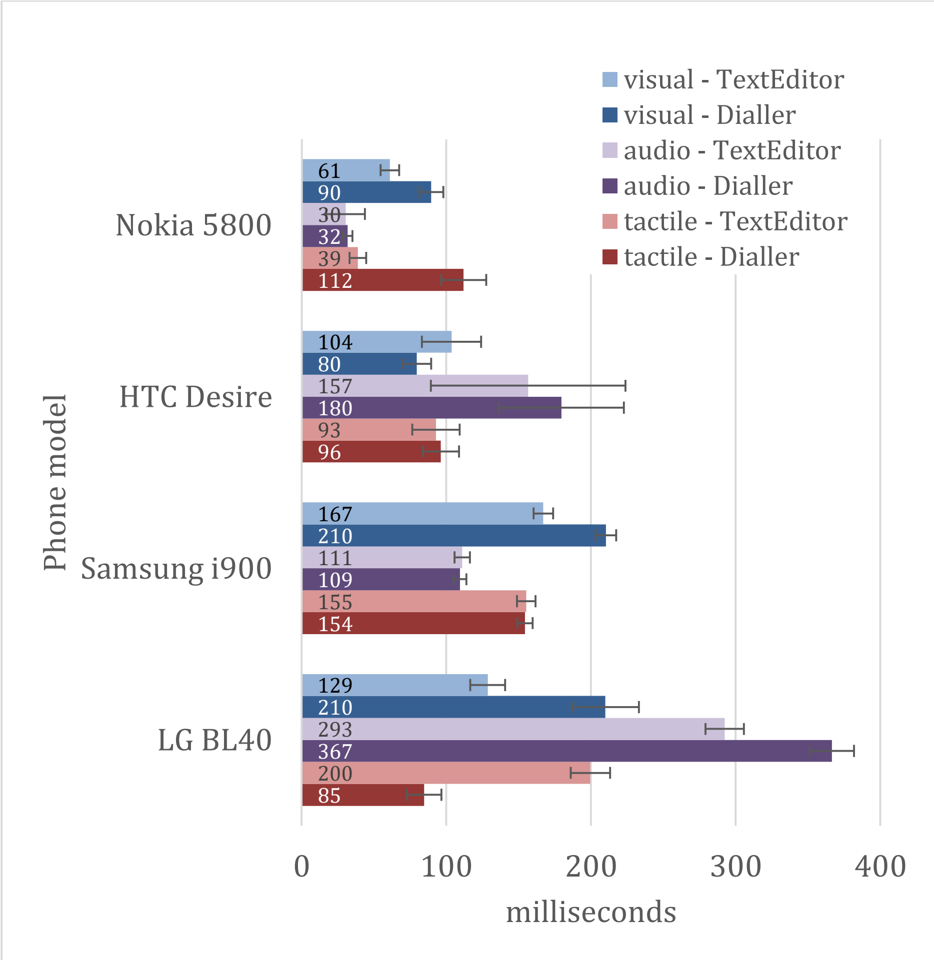


Figure 3-24: Average virtual button latency in the applications with different feedback modalities and all the phones. The error bars are standard deviations.

3.9.2.2 Finger vs. Stylus

Latencies were also measured with a stylus in order to find out if the tool is capable of measuring latencies with both input methods. In addition, it was considered useful to find out if there is a latency difference between finger and stylus usage. If they were the same, it can be assumed that the latency will stay constant independent of the input method.

The Nokia and Samsung phones featured resistive touchscreen[48] and that is why only they were measured. It was assumed that different applications would not behave differently within a phone when considering the latency difference between finger and stylus usage.

[48] Currently there are styli also for capacitive touchscreen, but they were not available at the time of this research.

The same assumption was made for the feedback modality. The audio modality was selected since the calibration procedure showed that the tool is capable of measuring it the most accurately. There was no specific criterion for the application selection and the dialler was selected for the test.

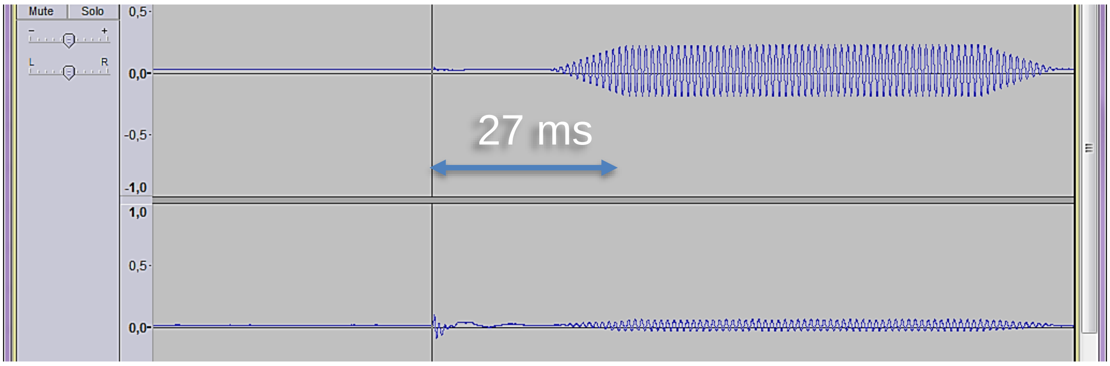


Figure 3-25: One measurement of stylus touch and audio feedback in Nokia dialler application recorded with Audacity.

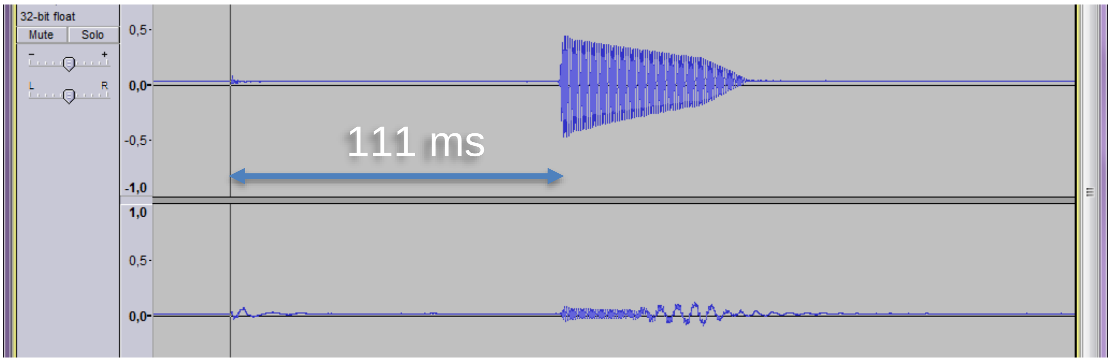


Figure 3-26: One measurement of stylus touch and audio feedback in Samsung dialler application recorded with Audacity.

Figure 3-25 and Figure 3-26 show the stylus touch event and the audio feedback of one virtual button press in Nokia and Samsung. The finger touch event and the audio feedback were presented earlier in Figure 3-22 and Figure 3-23. The measurement results indicated that in the Samsung phone the latency was 109.30 ms with a stylus and 109.40 ms with a finger meaning that the latency was 0.1 % smaller with a stylus. In the Nokia phone the latency was 30.67 ms with a stylus and 31.78 ms with a finger, latency being 3.6 % smaller with stylus. These differences were not significant (ANOVA: $F = 0.16, p > .05, df = 1{,}36$). Therefore, (H2), hypothesizing that the latency would be smaller when interacting with a stylus, was not supported. Since the means were close to each other in this measurement, it might mean that latency does not change depending on the input style and for measuring latency of virtual button either stylus or finger can be used. However, finger is preferable since it is the most used input style in today's mobile phones and, it can be used also with capacitive touchscreens.

3.10 Limitations

The main limitations of the tool were that the assessment of latency can be difficult if 1) audio or tactile feedback latency is near zero, i.e. the recording of the touch itself and the feedback are very close to each other and 2) audio and tactile feedback latencies are approximately the same.

3.10.1 Tactile Feedback Latency Is Near Zero

The assessment of audio and tactile feedback is easy when the latencies are some tens of milliseconds or more like shown for example in Figure 3-25, which is usually the case in commercial mobile phones. However, when tactile feedback latency is near zero, the touch detection recording and the tactile feedback recording will be very close if not overlapping. Therefore, differentiating touch and tactile feedback can be challenging. These cases require learning and education to make correct measurements. Fortunately, these situations are rare, however, with current phones as seen from the measurements in this chapter and to be seen in the following chapters. In the case of near zero latency the trick is to measure the button press first without any feedback and examine the waveform and approximate length of the touch itself. In the second measurement round, the enabled tactile feedback can be recognized from the recording by comparing second measurement round to the first measurement round.

3.10.2 Audio and Tactile Feedback Latencies Are the Same

As mentioned before, in addition to the tactile feedback, the accelerometer picked up some traces of audio feedback. This effect is not harmful if audio and tactile feedback latencies are different as, for example, in Figure 3-21. However, if the audio and tactile feedback latencies are approximately the same, the trace of audio feedback will overlap the tactile feedback recording, as Figure 3-22 shows. In this kind of case, finding tactile feedback might be challenging. The workaround is to disable audio for the first measurement session and learn the magnitude of the tactile feedback latency and the shape of the tactile feedback pulse. During the second measurement round the tactile feedback latency can be assessed although the audio feedback trace overlaps the tactile feedback recording.

As a solution for this limitation, one could imagine arranging a measurement from an audio output connector used for connecting headphones or an auxiliary amplifier. In principle, it would solve the problem, since audio would be silent for the accelerometer and thus separated from tactile feedback latency measurement. However, the goal of the measurement tool was to measure the end-to-end latency a user experiences from the user's first touch to the output of the sound. Keeping this in mind, the audio connector measurement would not be accurate, since the connector is not the final end. Since it cannot produce sound for the user, the final end would be the headphones or the loudspeaker of the auxiliary amplifier. Since these devices can possibly insert additional latency to the system, they have to be taken into account when measuring the latencies. Also internal audio buffers and internal amplifiers of the phone under measurement can cause latencies different from the internal amplifier of the in-built loudspeaker. Therefore, the measurement would not be accurate, and that is why – if the goal is not explicitly measure the audio latency of output connector – this method is not recommended.

3.11 Discussion and Conclusions

This chapter introduced a multimodal latency measurement tool for assessing touchscreen feedback latencies. Affordable off-the-shelf components and freeware software were used to make the tool inexpensive and easy for all to use whilst still being capable of measuring latencies accurately between different touch and visual, auditory and tactile feedback. The tool featured a high-speed camera, a mirror, a microphone and an accelerometer to measure them. The microphone and accelerometer were both interfaced with a standard soundcard that made the measurement and analysis simple. The latencies were extracted visually using a slow-motion video player and an audio editor. The focus was in mobile touchscreen devices, but with minor modifications the tool could be used also in other domains. To validate the tool, it was first calibrated and then, four commercial mobile phones were measured. The results showed that the tool was capable of measuring visual, audio and tactile feedback latencies in all the phones. They also showed that the latency varied between phones and within phones and between applications and feedback modalities. In addition, the results showed that the latencies did not differ significantly from each other regardless of the use of stylus or finger.

Research Question 1 asked:

> **RQ1:** Can an affordable and accurate touchscreen latency measurement tool be built?

This chapter shows that the tool was designed and implemented and the price tag was approximately 1000 € at the time the tool was first made. The most expensive parts were the high-speed camera and the accelerometer. Today, four years later at the time of the writing this thesis, consumer-grade cameras and the analogue accelerometer boards can be found lower prices making the tool even more affordable. That makes it possible to build multiple of the tools in low-cost, making it good choice for universities and companies with strict budgets. In addition, the weight of the tool, when packed in a carrying case and camera bag, was less than 6 kg, which made it easily portable to different desks and sites if needed. The calibration procedure validated accuracy and the sample measurements the reliability.

Although the multimodal latency measurement tool worked as planned, there is room for improvement. For example, the manual analysis of the slow-motion video and the audio files containing auditory and tactile feedback is time consuming. It also is a potential source of errors. Automating the analysis of the videos to find the moment of touch and the visual events would improve the speed and accuracy of the analysis, and also the reliability of the results. Automating the analysis of the audio files would again speed things up. This could be done by using different pattern recognition algorithms for both visual and audio files, for example.

It would be interesting to expand the measurements to other touchscreen widgets and interaction patterns, such as sliders, scrollbars, and drag-and-drop to see latency changes in more continuous interactions. With slight modifications to the setup, the tool could also be also used for latency measurements of whole device gestures and their responses.

The creation of this tool was the first step in the research in this thesis. As the measurement results show, the virtual button feedback latencies vary between phones. The latency values were in the magnitude of human perception as seen in Literature Review in Chapter 2. However, these measurements do not tell us anything about human perception. According to the literature, latency matters, but it is not known how much in virtual button press. Therefore, the research conducted in next chapters aimed to find out the human perception threshold for latencies in virtual button press and the effect of latency in one important aspect of user experience, perceived quality.

3.11.1 Where This Tool Has Been Used

The tool introduced in this chapter has been used in the real mobile phone Research and Development projects in different sites inside Nokia to validate the latencies in mobile phones in the development phase. In addition, this tool was used to assess the latencies in a research project in the University of Glasgow (McAdam and Brewster, 2011). In that project, distal tactile feedback with a mobile phone was introduced for tabletop computing environment. The tactile feedback latencies were validated using this tool.

Chapter 4 Latency Guidelines for Unimodal Feedback

4.1 Introduction

Latency affects a device's responsiveness and the perceived ability of the device to react to the user's input (Anderson *et al.*, 2011; Ng *et al.*, 2014). That is why latency can be harmful in interaction. It has been stated that latency is one of the major issues limiting the quality, interactivity, and effectiveness of virtual and augmented reality (Allison *et al.*, 2001; 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; Pavlovych and Stuerzlinger, 2009) or 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). As Hinckley and Widgor (Hinckley and Wigdor, 2012) state, latency can be especially harmful in direct input devices such as touchscreens used with a finger or stylus. Latency has been shown to degrade subjective satisfaction in touchscreen interaction (Kaaresoja *et al.*, 2011a; Kaaresoja *et al.*, 2011b) as well as user performance (Jota *et al.*, 2013).

The previous chapter introduced an inexpensive, but accurate latency measurement tool for touchscreen phones. It is capable of measuring visual, audio and tactile feedback latencies in touchscreen virtual button interaction. The measurements conducted showed substantial latency differences between and within phones. However, no guidelines for virtual button feedback latencies exist.

As latency causes a system to be slower, degrading the user experience, it is natural to conclude that simultaneity, where there is no latency, would enable an improved user experience through responsiveness. As simultaneity perception has been widely studied in psychophysics, an applied psychophysical approach was taken to find the simultaneity perception threshold of touch and feedback. In addition, to further understand how user experience changes as a function of latency, one qualitative dimension of virtual button latency was examined: perceived quality.

Therefore, Research Questions from 2 to 4 ask:

> **RQ2:** "What are the touch-feedback simultaneity perception thresholds in virtual button interaction?"
>
> **RQ3:** "How does the perceived quality of a virtual button change when latency between touch and feedback changes?
>
> **RQ 4:** What are the latency guidelines recommended for visual, audio and tactile feedback for a virtual button press?

This chapter introduces Experiment 1, investigating simultaneity perception of touch and feedback and the effect of feedback latency on perceived quality of the virtual button press. In Experiment 1 the feedback consisted of single visual, audio or tactile event. The results from these two different perspectives were combined as a latency guideline. Sections from 4.2.1 to 4.2.9 describe the method, including the description of Virtual Button Simulator (4.2.4), the research device used in both experiments in this thesis. Section 4.2.10 introduces the results. Section 4.2.11 discusses the results and Section 4.3 introduces the latency guidelines. Section 4.4 reflects the guidelines against the latencies measured from commercial mobile phones. Finally, Section 4.5 concludes the chapter.

4.2 Experiment 1

4.2.1 Design

A within-subjects design was selected for both perceived simultaneity and quality sessions. Every subject judged the simultaneity of every touch-feedback pair and scored the quality of the buttons. The method of constant stimuli [Coren et al. 2003] was chosen with a forced-choice Simultaneity Judgment 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 (a forced-choice SJ task). In perceived quality session they responded on 1-to-7 scale, “1” meaning low quality and “7” high quality.

4.2.2 Participants

Twenty-four (12 female) volunteer participants aged 26-50 (mean 36.4, std 6.3) took part in the experiment. Three were left-handed. 23 of the participants were employees of Nokia Research Center in Helsinki, Finland and one was recruited from outside of the company. 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.

4.2.3 Equipment

The goal of the research in this thesis was to investigate the perception of touch-feedback simultaneity and the effects on it in the perceived quality. To achieve this goal a device with programmable latencies of wide range was needed. Especially the device needed to provide low-latency tactile, audio and visual feedback with low variance. As seen in Chapter 3, current commercial mobile phones cannot provide feedback latencies near zero with low variance. Therefore, a proprietary research device was built. It would resemble a mobile phone as much as possible. The research device was called the *Virtual Button Simulator*. Another reason to use a simulator was to rule out the possible effect of mobile phone design on perceived quality.

4.2.4 Virtual Button Simulator

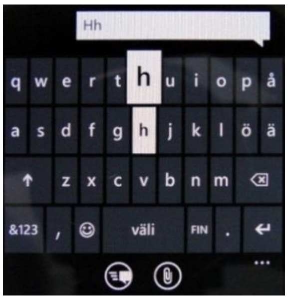
The size and weight of the Virtual Button Simulator were similar to a small mobile phone: 54 x 112 x 21 mm (max width x height x thickness) and 83 g (see Figure 4-1). In order to feature capacitive sensing, but to keep the sensing latency as low as possible, two metallic capacitive buttons at bottom on the front of the device were used (see Figure 4-2) instead of installing a full touch sensor which would have caused extra latency. One button would have caused still less latency, but it would have been difficult to set up a reasonable task for the participants.

Figure 4-1: Left: The Virtual Button Simulator (white) with the response pad for the experiment (black). The Virtual Button Simulator enclosure was designed and manufactured by Antti Rönkkö.

Figure 4-2: The Virtual Button Simulator and the USB cable used for connecting to a PC. 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.

4.2.4.1 Feedback Hardware

Visual feedback was provided by two rectangular green LEDs (HLMP-0504, light wavelength 565 nm, size 2.5 x 7.6 mm) placed just above the buttons for giving visual feedback to imitate a button popup (see Section 1.3.1 in Chapter 2 and Figure 4-3). Audio feedback was played through a miniature loudspeaker (9 x 9 x 3 mm) located inside the cover on top of the device like in a real mobile phone (see Figure 4-4). The loudspeaker used in this research was originated from an old mobile phone, but similar ones are available for example from Puiaudio[49] Tactile feedback was provided by a C2 Tactor by Engineering Acoustics (currently ATAC Technology)[50], which has been used in several mobile experiments before (e.g. (Brewster *et al.*, 2007; Hoggan *et al.*, 2008a)) and was located inside the device in its own covered cavity (see Figure 4-4) . Two red rectangular LEDs (HLMP-0301, light wavelength 635 nm, size 2.5 x 7.6 mm) were located on top of the device to give cueing information (see Figure 4-2).



Picture of the Apple iPhone popup has been removed due to Copyright restrictions.

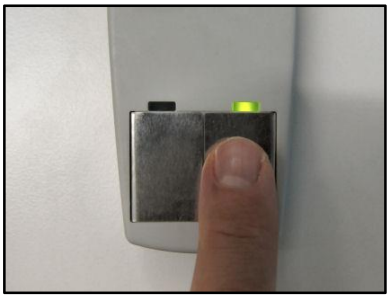


Figure 4-3: A text entry popup in Nokia Lumia[51], Apple iPhone and the simulated one in the Virtual Button Simulator.

To minimize latencies, all the processing of button presses and feedback generation happened in an Arduino Nano[52] 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[53]. Since the Arduino was not capable of driving strong enough signals to the loudspeaker or the tactile actuator C2, a Texas Instruments L293DN (Instruments, 2004) 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 1 ms latency to the circuit.

---

[49] http://www.puiaudio.com/pdf/SMT-0540-S-R.pdf, http://www.digikey.com/product-detail/en/SMT-0540-S-R/668-1060-1-ND/1464927

[50] http://www.atactech.com/PR_tactors.html (was: www.eaiinfo.com)

[51] Picture used with permission from Microsoft.

[52] http://arduino.cc

[53] playground.arduino.cc/Code/CapacitiveSensor

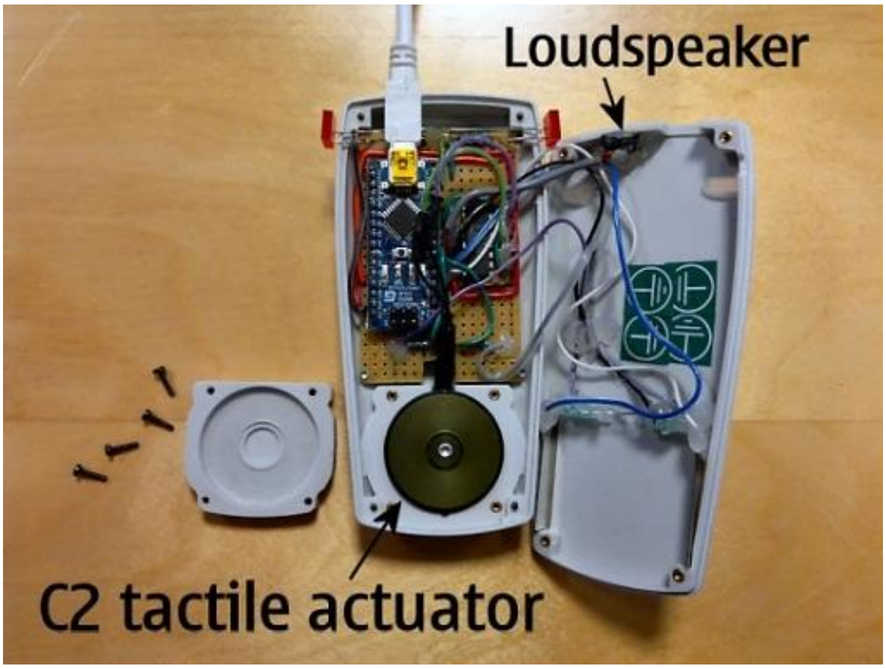


Figure 4-4: The opened enclosure of the Virtual Button Simulator. The USB cable was connected to Arduino Nano, the tactile and audio driver was located next to Arduino. C2 tactile actuator was located in its own enclosed cavity on the bottom of the device (cover open). The miniature loudspeaker was attached inside the cover on the top of the device.

The LEDs were connected directly to the Arduino's output pins via serial resistors. 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 green LEDs, loudspeaker and C2 tactile actuator, the Virtual Button Simulator was able to provide visual, audio and tactile feedback with less than 4 ms 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 multimodal latency measurement tool introduced in Chapter 3. Each feedback modality and latency condition was measured seven times which equals to the number of button presses per condition in the experiment. The average baseline latency was 3.92 ms for visual, 0.65 ms for audio and 2.81 ms for tactile feedback, and the mean standard deviation was 1.6, 0.46 and 0.41 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. Therefore, the baseline latency consists of both software delay in Arduino and the feedback raise time. The measurements proved that with Virtual Button Simulator it was possible to control latencies across the modalities at levels below human perception. Figure 4-5 shows the block diagram of Virtual Button Simulator.

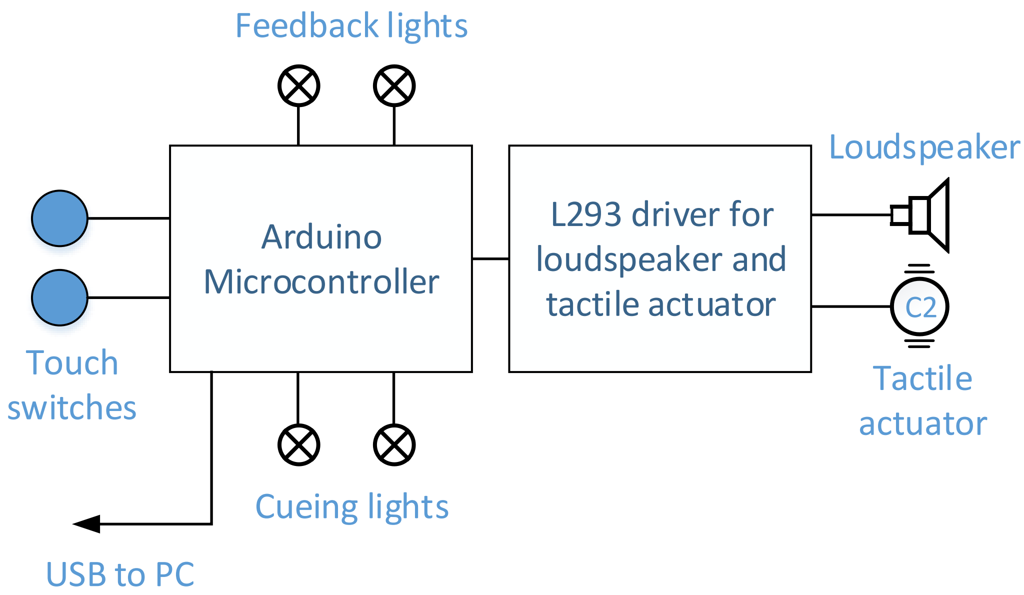


Figure 4-5: Block diagram of Virtual Button Simulator.

4.2.5 PC Software

The experiment software ran on a laptop PC and was programmed with Presentation®[54], 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 and save them on a hard disk. The Virtual Button Simulator and the Presentation® application communicated via a serial communication protocol through USB. Virtual Button Simulator took care of the time-critical processing such as the touch detection and timing of the feedback.

4.2.6 Feedback Design and Latency

[54] www.neurobs.com

There were two independent variables in the experiment: Feedback Modality and Feedback Latency (later Modality and Latency). Modality had three types: tactile, audio and visual. Latency consisted of nine (9) latency levels (based on earlier work): 0, 10, 20, 30, 50, 70, 100, 150 and 300 ms. This led to 27 different conditions and every condition was repeated 4 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 Modality and Latency condition without training leading to 27 additional stimuli.

The feedback were designed to be simple, pleasant and meaningful. The effect of the feedback design was beyond the scope of the research in this thesis. There was no attempt to equalize the intensity of the feedback of different modality for the experiment. However, they were all clearly over the perception thresholds.

4.2.6.1 Visual Feedback

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, green LEDs were placed just above the finger position. They highlighted simulating button popups shown in Figure 4-3 on page 111. It was not possible to use a proper LCD display as it would not have had a low enough latency for the study design. The green feedback LED glowed as long as the button was pressed. However, to tackle bouncing effects an 8 ms dead period was added after the release, which meant that the LED actually glowed 8 ms after the button was released. This did not cause any problems since 8 ms is a short time compared to the time the user presses the button and the LED is on. The earlier research on asynchrony perception of tap and audio feedback (Adelstein *et al.*, 2003a) showed that the asynchrony perception is not dependent on the duration of the feedback but was based on the attack (beginning) of the feedback. Therefore, it is assumed that the duration of the stimulus does not affect the simultaneity perception of touch and visual feedback, either.

4.2.6.2 Audio Feedback

The short audible click used in Apple iPhone virtual buttons was used as the basis for the audio feedback design, because of popularity of the phone and since it was considered pleasant. Figure 4-7 shows the recorded waveform from the Virtual Button Simulator. It was an audible click with a duration of 10 ms and a frequency of 2033 Hz. The sound level of the audio feedback was 60 dB (A) measured at a 30 cm distance from the Virtual Button Simulator.

4.2.6.3 Tactile Feedback

The tactile feedback was designed to be a short tactile click (Figure 4-6) mimicking a tactile feedback of a physical button. It was produced by sending a 1 ms pulse of 5 V the C2 resulting in a click with 1.5 ms rise time and 13 ms fall time (50%). The acceleration level of the tactile click was 2.2 g peak-to-peak. The sound level of the tactile feedback was 40 dB (A) measured at a 30 cm distance from the Virtual Button Simulator.

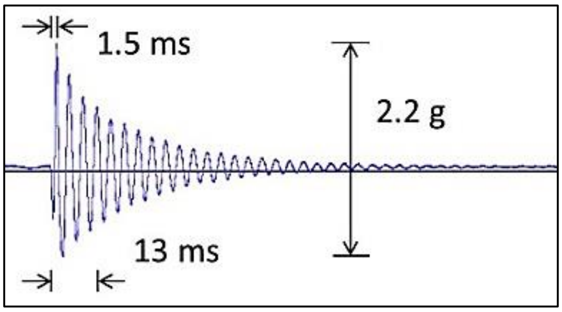


Figure 4-6: The acceleration and timing of the tactile feedback in Experiment 1. The rise time of the feedback was 1.5 ms, and the fall time to 50% level was 13 ms. The acceleration level was 2.2 g.

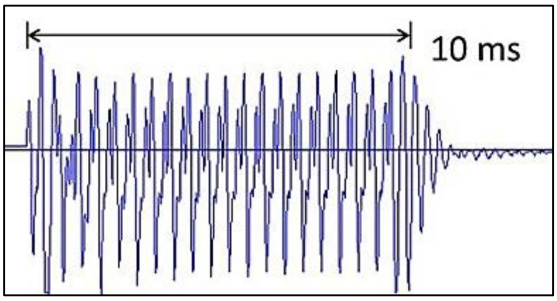


Figure 4-7. The recorded waveform and the timing of the audio click used as the audio feedback. The length of audio feedback was 10 ms and frequency 2033 Hz.

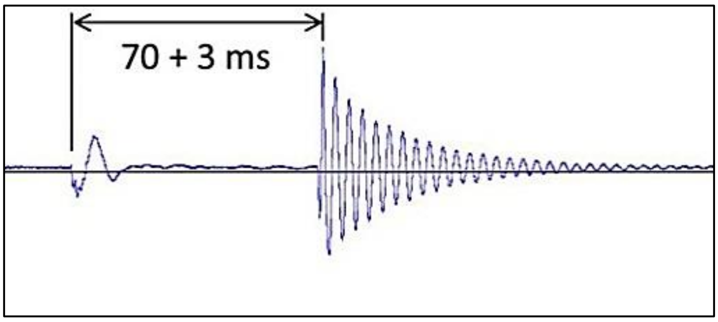


Figure 4-8: The 70 ms latency for tactile feedback. The 70 ms latency was added to the 3 ms system baseline (measured 2.81 ms on average for the tactile feedback).

4.2.6.4 Latency Conditions

The nine Latency levels were added to the Virtual Button Simulator's measured baseline (see Section 4.2.4.1) for each of the modalities (an example is shown in Figure 4-8). The selection of the latency values was based on earlier work introduced in Chapter 2. The baseline latency is usually added to the latency conditions (e.g. (Adelstein *et al.*, 2003a)) since it makes the analysis simpler and the latency conditions can be selected evenly.

4.2.7 Hypotheses

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

4.2.7.1 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 0 ms, e.g. (Levitin *et al.*, 1999; Winter *et al.*, 2008);

(H3) The 75% simultaneity perception threshold of touch and visual feedback will be 64 ms (Jota *et al.*, 2013), audio feedback 42 ms (Levitin *et al.*, 1999) and tactile feedback 58 ms ($(PSS + JND) \times 0.759 = 58$ (Winter *et al.*, 2008)).

4.2.7.2 Perceived Quality

(H4) The perceived quality score for the buttons will drop when latency is higher than 118 ms (Kaaresoja *et al.*, 2011a);

(H5) The participants would perceive a drop in quality earlier than simultaneity perception threshold (Kaaresoja *et al.*, 2011a).

Figure 4-9: Experiment setup. Participants held the Virtual Button Simulator in their non-dominant hand and pressed the buttons with their dominant hand. They responded with a modified keypad connected to a PC.

4.2.8 Procedure

Participants sat at a desk in a quiet office room and first read the experiment instructions and filled in a background questionnaire and consent form. They were instructed to hold the Virtual Button Simulator in their non-dominant hand and asked to press the capacitive buttons with the index finger of their dominant hand (Figure 4-9).

The Modality conditions were counterbalanced and the Latency conditions were randomized during both parts of the experiment. The experiment took approximately 1 hour. Figure 4-10 presents the overall experiment procedure.

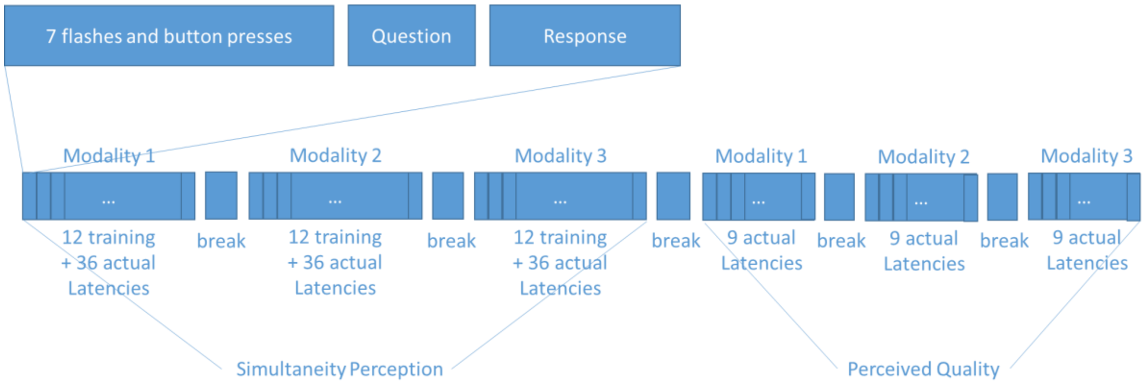


Figure 4-10. Experiment procedure.

4.2.8.1 Task Design

The task was designed 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 the length of the experiment session could not be controlled. It would have been possible to ask participants to write text with just two buttons. However, it was considered useful to 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, affected to the simultaneity or the perceived quality judgment and the reliability of the 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 two cueing LEDs were added at the top of the device, one as each side as described in Section 4.2.4.1, page 110. 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.

4.2.8.2 Task

The participants' task was to follow the flashing red cueing LEDs by pressing the buttons according to the side of the flash: if the right red LED flashed participants were to press the right capacitive button 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 button and wait for the maximum touch-feedback latency before the next cue. After some iteration rounds the length of the cueing flash was chosen to be 50 ms and a flash interval of 1 s. 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.

4.2.8.3 Perceived Simultaneity Assessment

Feedback was given depending on the Modality and Latency condition for each button press. One stimulus set consisted of seven cueing flash and button press pairs, within which the Modality 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” (“Y” for “yes” and “N” for “no”) on the response pad according to her/his perception. The response pad was a modified number keypad connected to the experiment PC containing only three keys, “Y”, “N” and “Enter” (see Figure 4-1, page 110 and Figure 4-9, page 117). After the response, the participant pressed “Enter” key as a confirmation to continue, and another stimulus set was presented to the participant. Background noise was played from two external active loudspeakers (Genelec 2029AL Digital[55]) 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 64 dB (A) measured 60 cm from the midpoint of the loudspeakers. The room background noise level was 39 dB (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 task properly. All nine Latency conditions were repeated four times in one Modality condition, meaning that there were 36 flash-press-response sequences in the real experiment

[55] www.genelec.com

for each of the three modalities. There were $3 \times (12 + 36) = 144$ flash-press-response sequences for simultaneity judgment altogether for one participant.

4.2.8.4 Perceived Quality Assessment

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 buttons 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.

4.2.9 Simultaneity Perception Analysis Methods

There were $n = 9 \times 4 \times 24 = 864$ binary simultaneity perception 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 finger touch feedback latency equal to LAG_i ms is

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

Equation 4-1

where

- μ: touch-feedback latency at which the "simultaneous" answer is most likely to happen,
- a: maximum probability of a "simultaneous" answer at the touch-feedback latency $LAG = \mu$, and
- σ: standard deviation 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 $(1 - p_1)$:

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

Equation 4-2

The probabilities p_1 and p_0 defined above were fitted jointly to all the observed responses, i.e. to all “simultaneous” and “not simultaneous” responses by all the participant 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). It was assumed that the responses were made independently from each other thus the likelihood function $L(\mu, \sigma, a)$ was of a product form

$$L(\mu, \sigma, a) = \prod_{i=1}^{n_1} ae^{-\frac{1}{2}\left(\frac{LAG_i - \mu}{\sigma}\right)^2} \times \prod_{i=1}^{n_0} \left(1 - ae^{-\frac{1}{2}\left(\frac{LAG_i - \mu}{\sigma}\right)^2}\right)$$
$$= \prod_{i=1}^{n} \left(\boldsymbol{ae^{-\frac{1}{2}\left(\frac{LAG_i - \mu}{\sigma}\right)^2}}\right)^{r_i} \times \left(\mathbf{1} - \boldsymbol{ae^{-\frac{1}{2}(\frac{LAG_i - \mu}{\sigma})^2}}\right)^{(\mathbf{1} - r_i)}$$

Equation 4-3

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.* (Stone *et al.*, 2001). However, in this experiment only positive (touch-feedback) latencies were observed, in other words, the feedback always came after the touch. For a realistic buttons 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[56] function *fminsearch* which is based on Nelder-Mead simplex algorithm.

[56] www.mathworks.se

Function *fminsearch* needs an initial starting point set for the parameter optimization and it was obtained by fitting curves with 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 the modality conditions and there were no constraints involved in the minimization procedure.

4.2.10 Results

This section presents all the results from this experiment. First, the simultaneity perception results are introduced, including the Gaussian model, PSSs and 75 % thresholds for each Modality. Second, the results from the perceived quality part are introduced with the help of significance maps developed in the research in this thesis. After the discussion, latency guidelines are established in the next section.

4.2.10.1 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 (LRT) 95% confidence intervals of the parameters are summarized in Table 4-1. The LRTs of all three parameters of all the feedback specific Gaussian models were implemented against $\chi_3^2(0.95)$. Figure 4-11 shows the three-dimensional confidence body with its two-dimensional projections of the MLE of the Gaussian model parameters for visual feedback. It can be seen that the projections are not ellipsoids and the MLE is in the middle of them. This indicates that the distribution of the parameter estimates was not normal. This was the case also when considering the Gaussian models for audio and tactile feedback conditions and their confidence bodies (Figure 4-12 and Figure 4-13). Stone *et al.* (2001) used Wald's test to determine the uncertainty of the maximum likelihood estimated 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, the restricted LRT was implemented 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 the feedback modality conditions were calculated by going through the parameter triplets within the whole 3-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 (1 ms resolution).

Table 4-1: The Gaussian curve fitting results for the probability p_1. $\hat{\mu}$ 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. The 95% confidence intervals are asymmetric around MLE estimates due to non-normal distribution of the parameters.

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

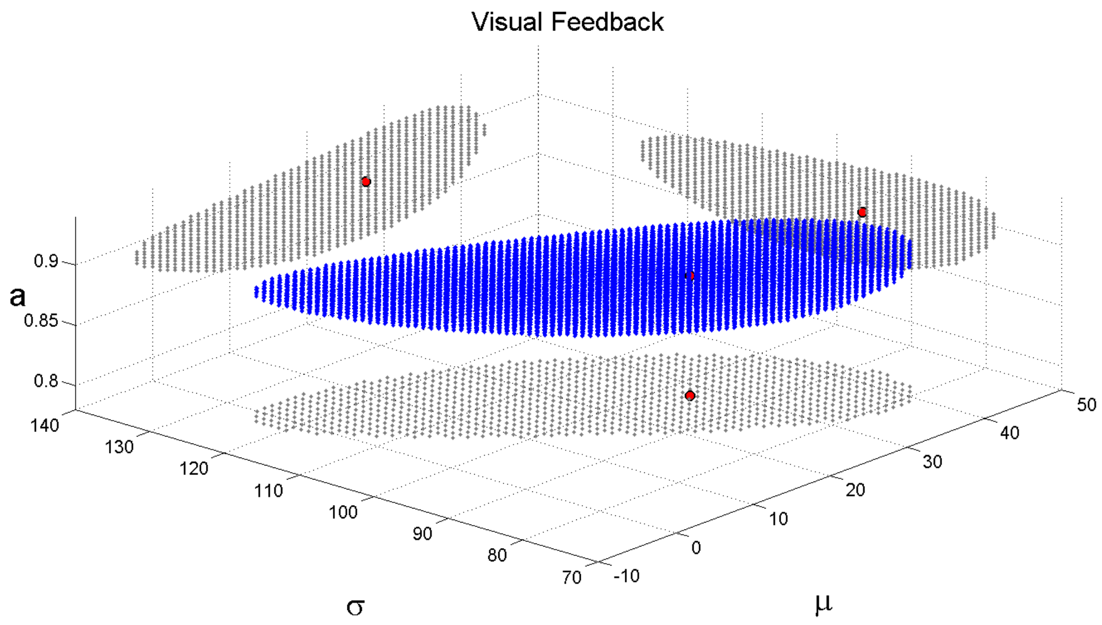


Figure 4-11: The 3D confidence body (blue), and its 2D projections (grey), of the MLE of Gaussian function parameter estimates $\hat{\mu}$, $\hat{\sigma}$ and $\hat{a}$ for the simultaneity perception in touch and *visual feedback* condition. The MLE points are marked as red dots. This confidence body was used to calculate the 95% confidence interval for the Gaussian model.

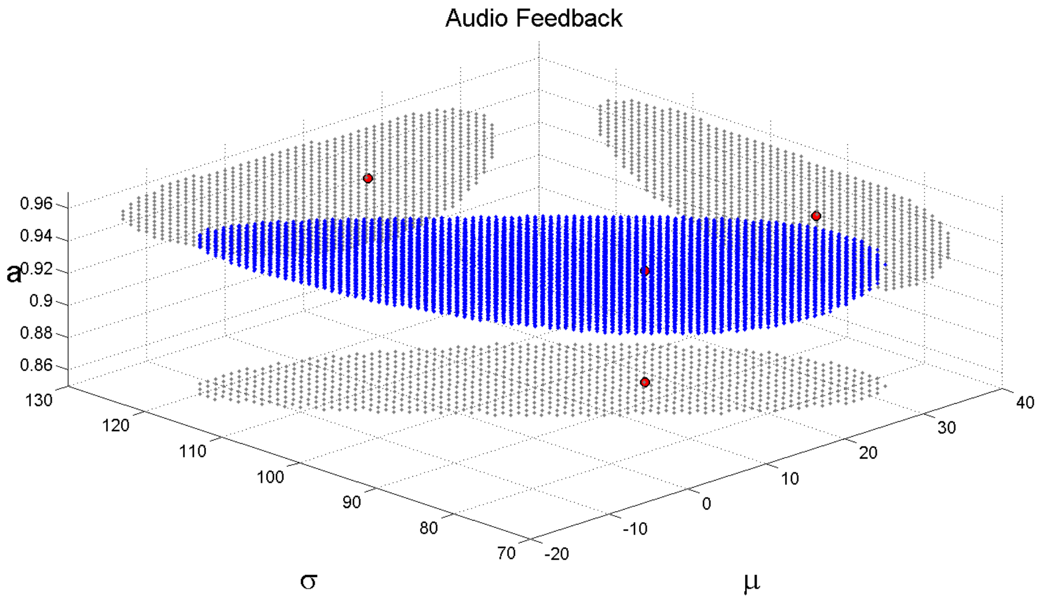


Figure 4-12: The 3D confidence body (blue), and its 2D projections (grey), of the MLE of Gaussian function parameter estimates $\hat{\mu}$, $\hat{\sigma}$ and $\hat{a}$ for the simultaneity perception in touch and *audio feedback* condition. The MLE points are marked as red dots. This confidence body was used to calculate the 95% confidence interval for the Gaussian model.

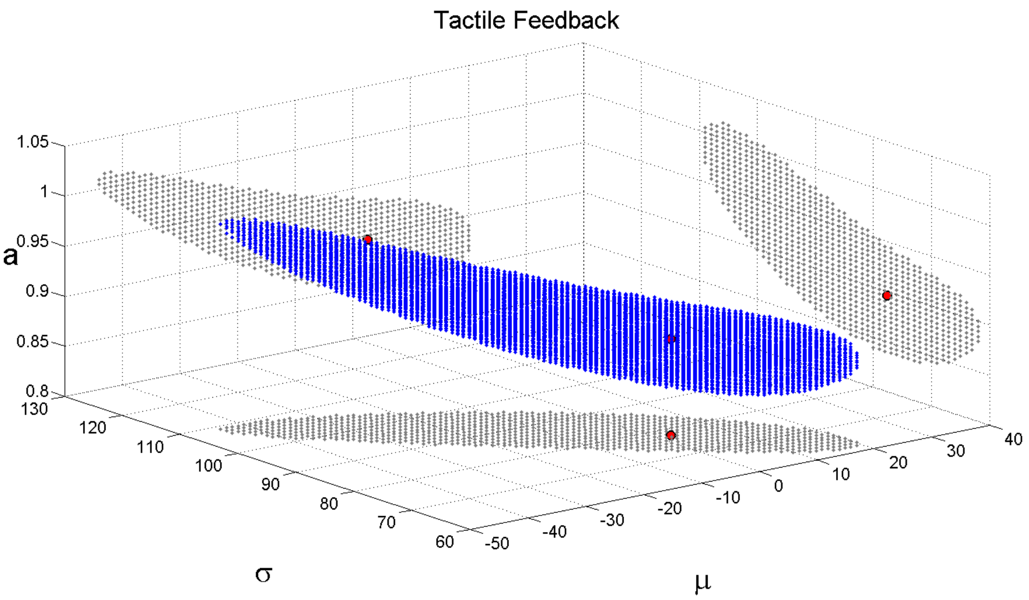


Figure 4-13: The 3D confidence body (blue), and its 2D projections (grey), of the MLE of Gaussian function parameter estimates $\hat{\mu}$, $\hat{\sigma}$ and $\hat{a}$ for the simultaneity perception in touch and *tactile feedback* condition. The MLE points are marked as red dots. This confidence body was used to calculate the 95% confidence interval for the Gaussian model.

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 Point of Subjective Simultaneity (PSS) was calculated as $\hat{\mu}$ + system baseline latency for each modality. For simultaneity perception of touch and visual feedback the PSS was 32 ms with 95% confidence interval being 20 – 43 ms, touch and audio feedback 19 ms with 95% confidence interval 8.2 – 30 ms and touch and tactile feedback 5 ms with 95% confidence interval -3.1 – 14 ms. The PSS of touch and visual as well as touch and audio feedback were significantly different from physical simultaneity since 0 ms was not within the 95% confidence intervals. However, the PSS of touch and tactile feedback did not differ statistically significantly from physical simultaneity since 0 ms was within the 95% confidence interval. Thus, (H2) – the PSS will not be significantly different from 0 ms – was partially supported. This means that in order to touch and feedback to be perceived as simultaneous as possible, audio feedback needs to have 19 ms latency and visual feedback 32 ms latency, on average.

A pair-wise 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 visual feedback was not significantly different when the latency condition was 0, 10, 20, 30, 50 or 70 ms, but dropped significantly between 70 and 100 ms ($\chi_1^2 = 9.9187, p < 0.0016$). 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 ($\chi_1^2 = 9.8091, p < 0.0017$). The proportion of the simultaneity perception of touch and tactile feedback was not significantly different when the latency condition was 0, 10, 20 or 30 ms, but significantly higher at the latency condition 20 ms than at 50 ms ($\chi_1^2 = 10.074, p < 0.0015$) meaning a significant drop between 20 and 50 ms.

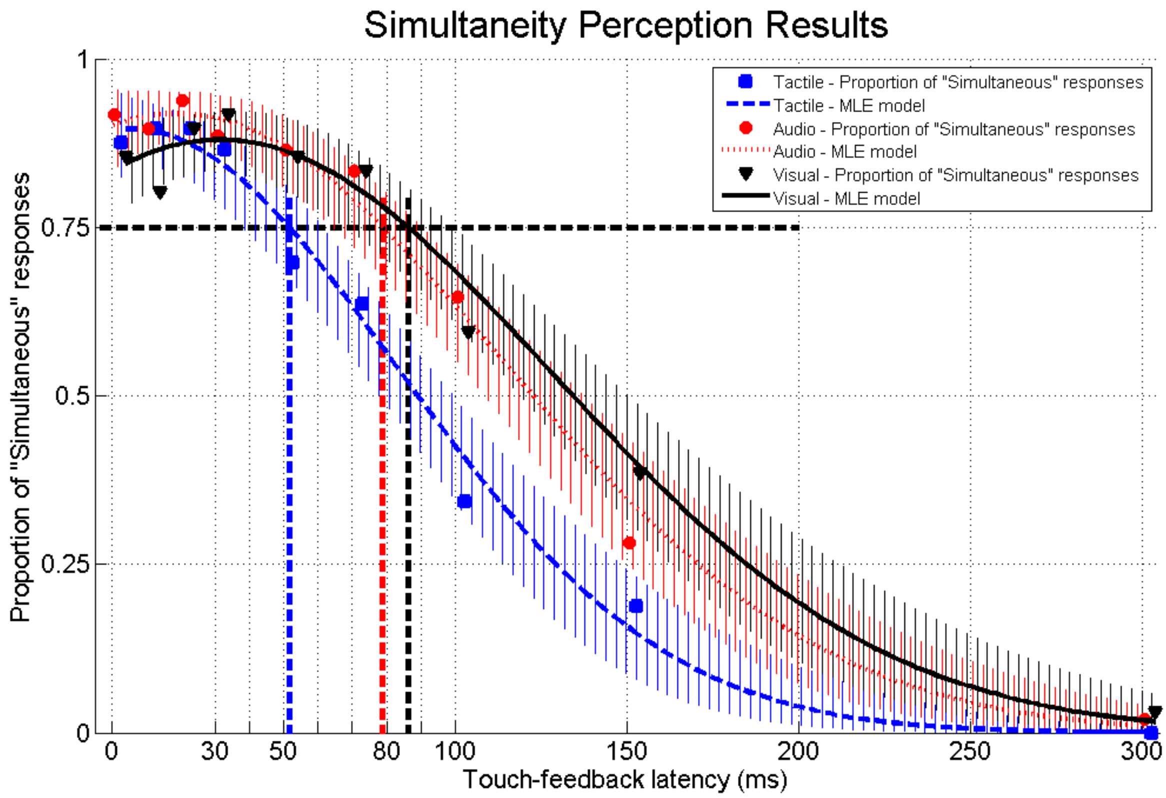


Figure 4-14: Proportion of "simultaneous" responses and the corresponding MLE Gaussian functions with the 95% confidence intervals (the line clouds around the Gaussian functions). Vertical dashed lines show the 75% simultaneity perception thresholds. The system baseline latencies have been added to all the latency values.

The proportions of "simultaneous" responses and the MLE probability p_1 models with 95% confidence intervals are plotted in Figure 4-14. The figure also shows also the uncertainty (95% confidence intervals) of the values of the Gaussian models. The 75% simultaneity perception threshold for touch and visual feedback was 85 ms with 95% confidence interval 70 – 100 ms. For touch and audio feedback the threshold was 80 ms with 95% confidence interval 65 – 90 ms. For touch and tactile feedback was 52 ms with 95% confidence interval being 40 – 62 ms.

Therefore, the hypothesis about the 75% simultaneity perception threshold (H3) – visual 64 ms, audio 42 ms and tactile 58 ms was partially supported: The obtained 75% threshold for visual and audio was higher than hypothesised and did not even fall within the 95% confidence intervals. However, the hypothesized 75% threshold for tactile feedback was within the confidence interval of the threshold obtained from the Gaussian model. In addition, any of the hypothesised values did not fall within the time windows found in the Chi-Square statistical inference of the observations above. The resulted 75% simultaneity perception threshold for tactile feedback was lower than hypothesised and higher for audio and visual. This all means that hypothesised values derived from earlier studies were not supported by the results from the experiment. Therefore, it can be concluded that when a user is tapping capacitive touchscreen buttons with a finger the simultaneity perception of touch and feedback differs from the simultaneity perception of tapping of different kind and feedback or two exogenous stimulus.

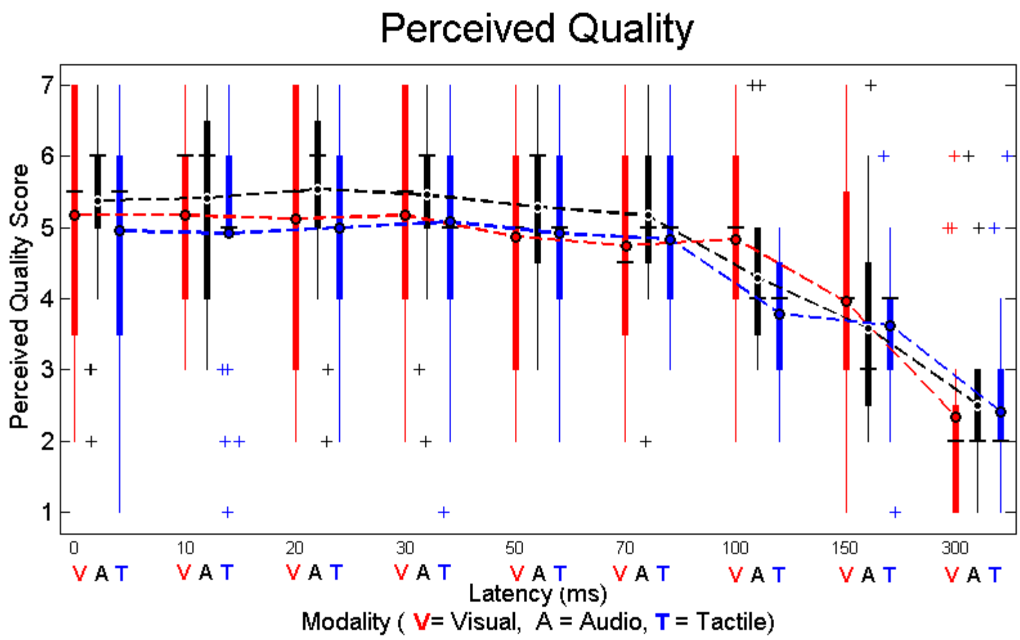


Figure 4-15: A boxplot showing medians and the distribution of the scores from the perceived quality questionnaire. The horizontal black lines inside or on the edge of the boxes show medians for each latency and modality condition. The edges of the boxes show the 25th and 75th percentiles of the data, 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. 'o' markers show the means of the data for each latency and modality condition and the dashed lines show the trendlines. Note that the outliers are shown only for this visualization, not considered in the data analysis.

4.2.10.2 Perceived Quality

A boxplot with the medians and means with trendlines of the scores from the perceived quality questionnaire are shown in Figure 4-15. 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 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 *significance maps* shown in Figure 4-17. Significance maps are a new way to visualize a complex set of condition comparisons. An example of a significance map is shown in Figure 4-16. The black square shows 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 at a 5% level than another condition, the other condition is marked green and with a “+”. A significance level of 0.1% is marked with dark green and an “X”. If the average quality score of the current combination is statistically significantly lower at a 5% level than of another condition, the other condition is marked red and with an “o”. A significance level 0.1% is marked with dark red and an “O”. Non-significant differences are marked with 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. The standard significance levels introduced here were mapped on the Bonferroni corrected significance levels.

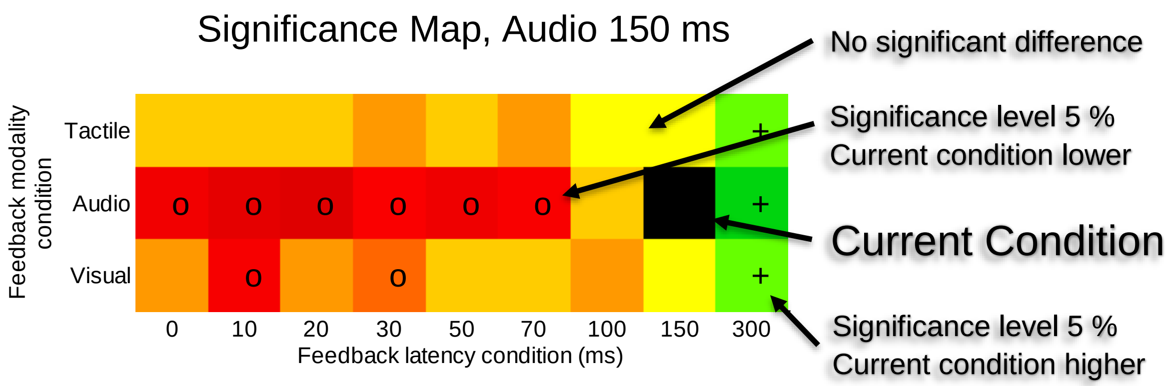


Figure 4-16: An example of significance map used to illustrate the statistical significance differences in perceived quality scores. This figure shows the audio feedback modality and 150 ms feedback latency conditions visualizing quality in relation to the other condition combinations. The black square marks the current condition combination (audio, 150 ms). 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 more detailed description.

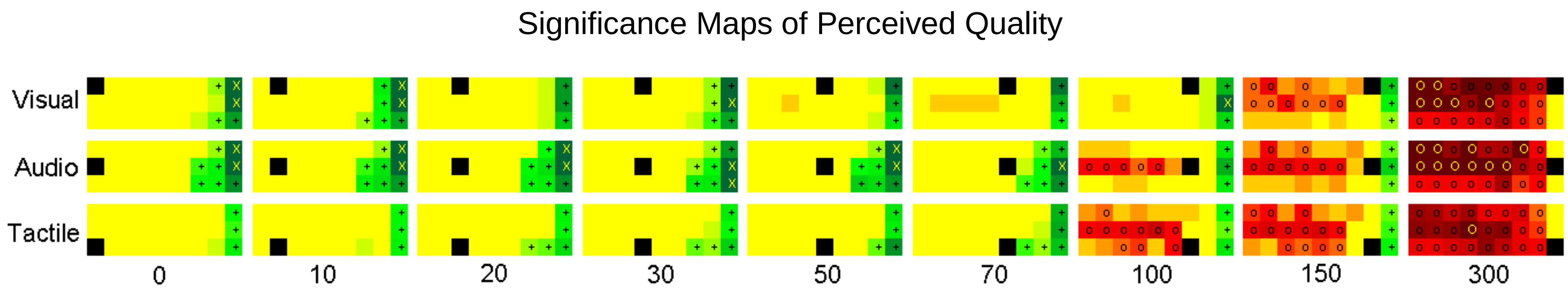


Figure 4-17: Significance maps of all the feedback modality and latency condition. The black square means the current condition combination labelled on horizontal and vertical axes. Each map follows the scheme introduced in Figure 4-16. See the text for more detailed description.

From the maps it is easy to see that there was a significant drop in perceived quality between 100 and 150 ms in visual feedback condition. The audio and tactile conditions differed from visual; the perceived quality dropped significantly between 70 and 100 ms. Therefore, (H4) – The perceived quality score for the buttons will drop when latency is higher than 118 ms – was not supported, since the quality dropped earlier. The buttons with any feedback with a latency of 300 ms were rated significantly lower than the buttons with any feedback with latency 0-150 ms. It also can be seen that the modality conditions did not 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 (Figure 4-15). Figure 4-18 shows the proportion of each score level as a function of latency conditions for all Modalities. It can be seen that the proportion of favourable ratings (scores from 5 to 7) is at least 50 % until the perceived quality is degraded for visual (150 ms) and more than 50 % for tactile and audio (100 ms). This means that the most of the participants liked the buttons when the latency was less than the perceived quality drop. It can be seen also that, for all the modalities, the majority of the scores were non-favourable (80 % for tactile and visual feedback and 90 % for audio feedback) when the latency was 300 ms. This means that the participants clearly did not like the buttons with 300 ms latency.

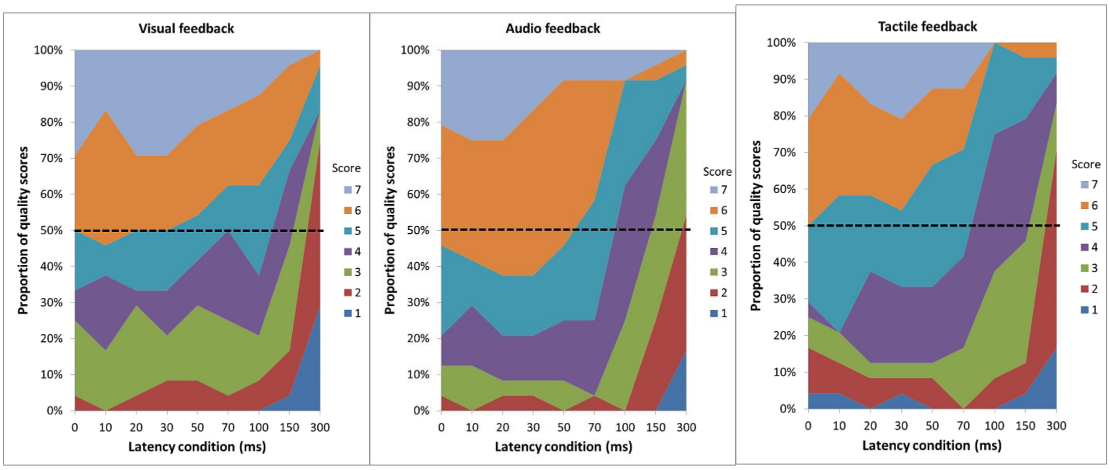


Figure 4-18: Proportion of quality scores as a function latency conditions for all Modalities. The dashed black lines show a 50% threshold. It can be seen that the proportion of favourable ratings (scores from 5 to 7) is more than 50% until the perceived quality is degraded (visual 150 ms, audio and tactile 100 ms).

4.2.10.3 Simultaneity Perception Threshold and Perceived Quality Drop

The last hypothesis (H5) – the participants would perceive a drop in quality earlier than simultaneity perception threshold – was not supported for visual or tactile feedback conditions. The significant drop in the proportion of “simultaneous” responses was before the significant drop in the perceived quality scores. For audio feedback 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, but not earlier. Therefore, it can be concluded that the users perceive the non-simultaneity before the drop in the button quality when latency increases.

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.

4.2.11 Discussion

4.2.11.1 Gaussian model

The aim of this study was to achieve a general model of touch-feedback simultaneity perception in order to derive practical design guidelines for tactile, audio and visual feedback. It was hypothesized (H1) that the distribution of “simultaneous” responses would follow a Gaussian function. The experimental data and statistical analysis showed that this hypothesis was a feasible choice for that purpose. The results confirmed that touch-feedback simultaneity perception behaved in similar manner to the simultaneity perception of exogenously applied stimuli in earlier work (e.g. (Stone *et al.*, 2001; Winter *et al.*, 2008; Fujisaki and Nishida, 2009)). In these earlier studies the model fitting was implemented for individual participants’ data. In the current study, a practical choice was made to keep the duration of the test reasonable since the aim was to inspect the touch-feedback simultaneity, in addition to the perceived quality assessment, with all the feedback modalities in the same experiment. Collecting more data points needed for individual modelling would have increased the experiment duration beyond reasonable levels, considering participants fatigue causing potentially unreliable data. More importantly, the objective was to define general design guidelines for the feedback latencies. Thus, the general model of touch-feedback simultaneity was under interest, instead of accurately modelling simultaneity perceptions of individual participants and understanding the differences between them.

4.2.11.2 Point of Perceived Simultaneity (PSS)

It was hypothesized (H2) that the PSS would not differ significantly from physical simultaneity (i.e. when feedback comes exactly at the same time as the touch). The results only partially supported this. The PSS of touch and visual feedback was 32 ms and physical simultaneity was not within the 95% confidence interval, meaning that the PSS was significantly different from 0 ms. The PSS of touch and audio feedback latency was 19 ms and significantly different from 0 ms as well. The PSS of touch and tactile feedback was 5 ms, but not differ significantly from 0 ms. The PSS shift from 0 ms was supported by an additional anecdotal finding; participants verbally reported in 26% (19/72) of all the 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 explicitly asked about it. This might be broadly an adaptation issue. It has been shown by Rohde and Ernst (2013) that when participants were exposed to visual feedback delays, the PSS shifted towards the delay. Sugano *et al.* (2010) showed the PSS shift also happened with audio feedback. In the other hand, after the exposure, if the tap and feedback happen physically simultaneously, the feedback is perceived to come before touch – even though it is against the causality (Heron *et al.*, 2009). This adaptation has been shown to carry over time (Rohde and Ernst, 2013). These findings can be applied to the current use case: The participants have most probably been exposed to the latencies of their own mobile devices and they have accustomed to virtual buttons with certain latency. When they pressed buttons with shorter latencies, especially near 0 ms, it may have felt unnatural and could even cause the feeling that the feedback came earlier than the touch. PSS shift means that the participants perceived touch and visual feedback as simultaneous if visual feedback was 32 ms late, on average. The same was true for audio feedback when it was 19 ms late. This is good news for the hardware and software engineers aiming to minimize the touchscreen device latencies; 32 ms is enough for visual feedback and 19 ms for audio. Therefore, it is not necessary to reach zero latency which might be technically challenging, expensive and time consuming.

4.2.11.3 Simultaneity Thresholds

The practical simultaneity perception thresholds were obtained both by examining the 75% level in the Gaussian models and validated by conducting statistical significance analysis of the observations. It was hypothesized (H3) that the 75% threshold of touch-feedback simultaneity perception will be 64 ms for visual, 42 ms for audio and 58 ms for tactile feedback. The derived thresholds were significantly larger (although the same magnitude) than the hypothesised ones when the feedback was visual (85 ms, with CI 95% 70 – 100 ms) or audio (80 ms, with CI 95% 65 – 90 ms). However, although smaller, the derived threshold did not differ significantly from the hypothesized one when the feedback was tactile (52 ms, with CI 95% 40 – 62 ms). Thus, (H3) was only partially supported.

The 75 % threshold of simultaneity perception of touch and visual feedback was larger than the latency perception threshold in Jota *et al.* (2013) which was the basis for the hypothesis in the current experiment. Two explanations can be found: The judging method and metrics difference and the difference on the cognitive load caused by the task. In the current experiment participants were asked to follow flashing lights and press the buttons accordingly. The visual feedback was given for their button presses above the fingertip. After seven presses they were asked to judge the simultaneity. In Jota's experiment, participants were asked to press a solid target on a screen which changed to a rectangle around their fingertip. After another press they were asked to judge which one of the two was quicker (one being a probe with 1 ms latency). The feedback itself was very similar in nature, but the judgment was different. The current study was detection task whereas Jota's were differentiation task, which might cause the lower threshold in Jota's results. In the current study the flashing lights caused extra visual sensory load (ecologically valid mimicking a use of a mobile phone) compared to Jota's were participants could freely find the target and press it. As Ng *et al.* (2014) stated, the cognitive load and attention required to complete a task seem to decrease the ability to perceive latencies. This further explains the higher threshold derived from the data in the current experiment. However, this is good news for the engineers and designers: Because mobile phone users highly concentrate their task rather than evaluating latencies and have a lot of visual attention demands, they tolerate more latencies.

The hypothesised 75% threshold for simultaneity of touch and audio feedback was based on the study by Levitin *et al.* (1999). In their study, the participant was hitting a surface with a mallet and asked to judge the simultaneity with audio feedback which was given on headphones. Comparing the experiment described in this chapter to Levitin's, two main differences can be found. Firstly, participants used a finger instead of a mallet. The finger needs to comply before the actual sensation happens which causes extra latency to the touch perception. Whereas with a tool, the sensation might happen earlier because of the tight grip from the mallet rod. Secondly, the audio feedback was given on the same device, meaning the same location, as the tap happened, whereas headphones were far away from the mallet in Levitin's experiment. Zampini *et al.* (2005) found that the (visual-audio) simultaneity perception threshold was lower when the stimuli were coming from different location. This might give an additional explanation, although the effect of location of the feedback to the simultaneity perception of touch and feedback has not been studied formally. However, the results from this thesis would support this hypothesis. In any case, the results again work as a favour of engineers and designers: it is not necessary to optimize the audio feedback latency so much when the feedback intrinsically comes from the same device than where the user is tapping the buttons.

4.2.11.4 Perceived Quality

There were no significant peaks in the perceived quality scores. The perceived quality score dropped significantly for visual feedback latencies between 100 and 150 ms and audio and tactile feedback latencies between 70 and 100 ms. The hypothesis (H4) – the perceived quality score for the buttons would drop when latency is larger than 118 ms – was not supported since the quality dropped earlier than 118 ms. The hypothesis was based on the earlier results by Kaaresoja *et al.* (2011a). In these studies, the participants entered numbers with virtual buttons and the tactile feedback latency was varied. Their results revealed that the performance did not drop within the latency range of 18 ms and 118 ms. The subjective satisfaction score was higher for all metrics when latency was 18 ms compared to the others from 38 ms to 118 ms. It was hypothesised that the latency range was just too narrow to show the satisfaction drop. However, the results from the current study showed the significant drop within this range. This underlines the substantial value of the research in this experiment. The quality drop definitely needs to be taken into account when defining latency requirements for virtual buttons.

4.2.11.5 Simultaneity Perception Threshold and Perceived Quality Drop

The last hypothesis (H5) – the participants would perceive a drop in quality earlier than simultaneity perception threshold – was based on an initial finding by Kaaresoja *et al.* (2011a). The results in their study showed that the subjective satisfaction was higher when the tactile feedback latency was only 18 ms compared to the other conditions with more latency. Still, five out of twelve participants did not notice any latency when it was 118 ms in some conditions. However, neither of these findings was validated by the results of the current experiment and therefore, the (H4) was not supported. The explanation might lie in the different input technology: Resistive touchscreen was used in the earlier study by Kaaresoja *et al.*, whereas a highly sensitive capacitive touch sensor was used in the current experiment. The participants needed to press the resistive touchscreen before the feedback was given. Thus the drop of the subjective satisfaction when the latency was only 18 ms could be comparable the drop between 70 ms and 100 ms in the current experiment. Another difference was that there was visual feedback (latency not validated) accompanying the tactile feedback in the earlier study. The interference between visual and tactile feedback modalities might explain the different behaviour of the subjective satisfaction in the earlier study and perceived quality score in the current experiment where only one modality at the time was studied. This is a topic which will actually be tackled in the next chapter where the simultaneity and quality perception with bimodal feedback is investigated.

4.3 Latency Guideline

The results presented above are all summarized in the Table 4-2. These results were synthesized as a guideline as follows. The recommended minimum latency was selected to be the PSS of the touch and feedback as explained above. The maximum recommended latency was selected both from the Gaussian models and the significant drop in the perceived quality score: the smaller of either the 75% simultaneity perception threshold or the latency condition when the perceived quality started to drop. For tactile and visual feedback the 75% threshold was smaller and for audio feedback the latency when the perceived quality started to drop was smaller. As the guideline (results rounded to the nearest 5 ms),

visual feedback latency should be 30 – 85 ms,
audio feedback latency 20 – 70 ms and
tactile feedback latency 5 – 50 ms.

Referring to the adaptation discussion earlier, 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 as discussed also by Ng. *et al.* (Ng *et al.*, 2012). Ng *et al.* found that after their participants were exposed to very low latencies (1-2 ms) they found the latencies in the current commercial devices totally unacceptable. However, as can been seen in the next section, the virtual button feedback latencies in the mobile phones are still mostly larger than the guidelines just established, making them highly valuable guiding the engineers and designers to optimize latencies just right.

Table 4-2: Summary of the simultaneity perception thresholds and drops in the perceived quality scores.

	PSS	Significant drop in the proportion of "simultaneous" responses	75% threshold of the Gaussian model	Significant drop in the perceived quality scores	GUIDELINE
Visual	32 ms	70-100 ms	85 ms	100-150 ms	30 – 85 ms
Audio	19 ms	50-100 ms	80 ms	70-100 ms	20 – 70 ms
Tactile	5 ms	20-50 ms	52 ms	70-100 ms	5 – 50 ms

4.4 Reflection of Latencies in Mobile Phones

In order to show how the latency guideline can be put in the practice, the latencies of the touchscreen phones measured in Chapter 3 were reflected against the guidelines. These phones were HTC Desire, LG Chocolate BL40, Nokia 5800 XpressMusic, and Samsung Omnia i900. The text messaging application results were considered here. In addition, the latencies of five other newer generation mobile phones were measured with the latency measurement tool introduced in Chapter 3. These phones are introduced in Figure 4-19:

HTC[57] Wildfire S running Android, iPhone[58] 4S running iOS, Nokia[59] Lumia 800 running Windows Phone 7, Nokia N9 running MeeGo and Samsung[60] Galaxy Note running Android. All the wireless functions were switched off in the phones during the measurement in order to avoid extra variance in latencies. The default text message application was opened and for the measurement the "g" button 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 pop-up of the button was fully drawn on the screen.

Picture of HTC Wildfire S has been removed due to Copyright restrictions.	Picture of Apple iPhone 4S has been removed due to Copyright restrictions.

Figure 4-19: Five new generation touchscreen phones measured for latency guideline reflection: HTC Wildfire S, Apple iPhone 4S, Nokia Lumia 800[61], Nokia N9[62] and Samsung Galaxy Note.

The results can be seen in Table 4-3. The latencies highlighted with green fulfilled the guidelines. Some of the phones performed very well according to the guidelines. Some phones had 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. From the first four phones, only the latencies of Nokia 5800 XpressMusic were within the guideline. From the newer phones, Nokia Lumia 800 had audio and visual feedback latencies within the

[57] www.htc.com

[58] www.apple.com/iphone

[59] www.microsoft.com (was www.nokia.com)

[60] www.samsung.com

[61] Picture used with permission from Microsoft.

[62] Picture used with permission from Microsoft.

guideline and Nokia N9 tactile and audio feedback latencies. The visual feedback latency in Apple iPhone 4S was also within the guideline. The rest of the feedback had longer latencies than recommended in the guidelines. Only the older Nokia 5800 XpressMusic provided all three forms of feedback within the latency guidelines.

Table 4-3: Average touch-feedback latencies in milliseconds for virtual buttons in the default messaging application in nine touchscreen mobile phones. The first four are the phones measured in Chapter 3. The following five are newer generation phones. The table is sorted according to the average latency of all the feedback within the two groups. The green highlight shows that the latency was within the guideline set in this study.

Mobile Phone (Operating System)	Visual feedback latency (guideline 30 – 85 ms)	Audio feedback latency (guideline: 20 - 70 ms)	Tactile feedback latency (guideline: 5 - 50 ms)
Nokia 5800 XpressMusic	61 ms	30 ms	39 ms
HTC Desire	104 ms	157 ms	93 ms
Samsung Omnia i900	167 ms	111 ms	155 ms
LG Chocolate BL-40	129 ms	293 ms	200 ms
Nokia Lumia 800 (Windows Phone)	53 ms	37 ms	Not supported
Nokia N9 (MeeGo)	110 ms	38 ms	35 ms
Apple iPhone 4S (iOS)	83 ms	102 ms	Not supported
HTC Wildfire S (Android)	140 ms	149 ms	74 ms
Samsung Galaxy Note (Android)	197 ms	172 ms	123 ms

4.5 Conclusions and Research Questions 2, 3 and 4

For the first time, latency guidelines for unimodal visual, audio and tactile feedback were established in this chapter. This was done by combining the novel results of touch and feedback simultaneity perception and of the effect of latency on perceived quality of button press. The guidelines are important for manufacturers of touchscreen devices, because now they can optimize the feedback latencies for individual modalities just right, enough for most of the users to perceive touch and feedback simultaneous and not feel bad quality buttons, but not too low which can be time consuming and expensive.

In the research in this thesis, the aim was 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 the results would be usable for touchscreen mobile device designers. The participants pressed capacitive buttons and the associated feedback was provided from the same device as in a real mobile phone and they were asked to judge if the feedback was simultaneous with the touch. The results showed for the first time that the perception of simultaneity of touch and visual, touch and audio and touch and audio feedback in a realistic setup can all be modelled with a Gaussian function. This confirms the earlier 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 (Stone *et al.*, 2001; Fujisaki and Nishida, 2009). The Gaussian models were convenient tools for finding parameters for applicable guidelines. It was found that the Points of Subjective Simultaneity (PSSs) according to the Gaussian models were not the same as physical simultaneity; the PSS of touch and tactile feedback was 5 ms, touch and audio feedback 19 ms and touch and visual feedback 32 ms. In order to further understand the effect of latency to the user experience, the participants were asked to score the perceived quality of the buttons.

Research Questions 2 asked:

> **RQ2:** What are the touch-feedback simultaneity perception thresholds in virtual button interaction?

The 75% thresholds were obtained from the Gaussian models: 85 ms for visual, 80 ms for audio and 52 ms for tactile feedback.

Research Question 3 asked:

> **RQ3:** How the perceived quality of a virtual button changes when latency between touch and feedback changes?

The perceived quality scores dropped significantly between latency conditions 100 and 150 ms when feedback was visual and between 70 and 100 ms when the feedback was tactile or audio.

Although any correlation statistics were not performed, the results suggested 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 non-simultaneous. Thus, the quality perception assessment reinforced the simultaneity perception findings in this study and *vice versa*.

Finally, the Research Question 4 asked:

> **RQ 4:** What are the latency guidelines recommended for visual, audio and tactile feedback for a virtual button press?

The guidelines for interaction designers were established as follows. The minimum latency was selected to be the PSS of the touch and feedback and the maximum both from the Gaussian models and the significant drop in the perceived quality score: the smaller of either the 75% simultaneity perception threshold or the latency condition when the perceived quality started to drop. The guidelines established this way recommend that (rounded to the nearest 5 ms):

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.

As already mentioned, these new guidelines have a two-fold importance to the field. First, 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. Second, hardware and software engineers do not need to optimize the latency between touch and feedback towards 0 ms.

Chapter 5 Latency Guidelines for Bimodal Feedback

5.1 Introduction

As stated in previous chapters, most often latency in interaction is undesirable, but the perception and user experience thresholds depends on many factors such as task and modality, for example. In the previous chapter, the guideline for unimodal virtual button feedback latency in touchscreen interaction was established for the first time. The feedback involved in the virtual button press was single modality, i.e. only visual, audio or tactile feedback was given as a response for a virtual button press at a time. The unimodal feedback study was the first step towards understanding the effects of latency in touchscreen virtual button interaction and it set the baseline for the future research and guidelines.

Often in mobile phones, there is accompanying audio or tactile feedback in addition to inherent visual feedback for a virtual button. Therefore, it is essential to derive guideline also when there is bimodal feedback involved. In addition, it is important to find out if this guideline differs from the unimodal feedback one so that a designer can select the right latency requirements for engineers depending on the current modality combination. In addition, there are known perceptual consequences caused by an interaction between modalities, for example a Colavita effect: a visual perception tends to dominate over audio (Spence, 2009).

This chapter introduces an experiment which investigated the touch-feedback simultaneity perception as well as the effect of latency on perceived quality when the feedback was bimodal, i.e. there was an additional feedback modality involved in addition to visual feedback. Thus, the feedback modality pairs were visual-audio and visual-tactile and, in the matter of completeness, tactile-audio.

In the previous chapter, answers for Research Questions 2, 3 and 4 were found for unimodal feedback case. In this chapter answers are retrieved for the same questions for bimodal feedback.

Research Questions 2, 3 and 4 again ask:

> **RQ2:** "What are the touch-feedback simultaneity perception thresholds in virtual button interaction?"
>
> **RQ3:** "How does the perceived quality of a virtual button change when latency between touch and feedback changes?
>
> **RQ 4:** What are the latency guidelines recommended for visual, audio and tactile feedback for a virtual button press?

The research with bimodal feedback was fundamentally more demanding than with unimodal, since the latency space was two dimensional. This is because the latency of both two feedback modalities could be varied. Novel data analysis methods were developed to better retrieve the result. Since the simultaneity perception of unimodal feedback could be modelled with univariate Gaussian, it was hypothesised that it will follow the bivariate Gaussian distribution when the feedback was bimodal, although it was never been investigated before. Also for perceived quality assessment, Significance Maps, introduced already in Section 4.2.10.2 in Chapter 4, and further developed in this chapter, were a great help for deriving the results for perceived quality.

Section 5.2 introduces the Experiment 2 including analysis methods (Section 5.2.7) and results (Section 5.2.8). The guidelines, which were established based on the results, are introduced in Section 5.3. The results are discussed in Section 5.4. The guidelines are further reflected on latencies with the new set of commercial mobile phones in Section 5.5 in similar manner than in the previous chapter (Section 4.4). Finally, Section 5.6 gives conclusions and introduces the answers for the research questions above.

5.2 Experiment 2

Experiment 2 setup was very similar to the one in Experiment 1. The task was the same, to press buttons with Virtual Button Simulator and to judge the simultaneity and score the perceived quality. However, there are some fundamental differences because of the two-dimensional latency space and different amount of conditions. That is why the experiment setup is introduced briefly again.

5.2.1 Design

A within-subjects design with the method of constant stimuli (Coren *et al.*, 2003) was chosen for all three different feedback modality pairs and 49 or 64 latency combinations (see Section 5.2.4 for explanation). The experiment was divided in two sessions: simultaneity judgment and perceived quality assessment. In both sessions each participant went through all the feedback latency combinations. In simultaneity perception session they were instructed to respond either "yes" ("simultaneous") or "no" ("not simultaneous") for each (a forced-choice SJ task). In quality perception session the participants were instructed to judge the quality of the buttons on 1-to-7 scale, "1" meaning low quality and "7" high quality.

5.2.2 Participants

Twenty-four (5 female) volunteer participants aged 28-62 (mean 40.0, std 7.5) took part in the experiment. One was left-handed. All of them were employees of Nokia Research Center and they were experienced mobile phone users: 23 of them had used a mobile phone more than 10 years, one reported 5-10 years. All of them used a touchscreen mobile phone at the moment of the experiment. All filled in a consent form at the start of the experiment and were given two cinema tickets as a reward for their participation.

5.2.3 Equipment

Virtual Button Simulator (Figure 5-1), introduced in Section 4.2.4 in Chapter 4, was also used in this experiment with reprogrammed internal embedded software. It featured two metallic capacitive buttons at bottom on the front of the device, two green LEDs placed just above the button area for giving visual feedback to imitate a button popup. Audio feedback was played through a miniature loudspeaker and tactile feedback was provided by a C2 Tactor. Two red LEDs were located on top of the device to give cueing information.

With the LEDs, loudspeaker and C2 tactile actuator, the Virtual Button Simulator was able to provide visual, audio and tactile feedback with less than 5 ms average baseline latency between finger touch and feedback. The baseline latency was a bit higher than in Experiment 1, because two parallel feedback modalities caused extra processing time. The system baseline latency of the Virtual Button Simulator was again measured with the latency measurement tool introduced in Chapter 3. Each feedback modality and latency condition was measured seven times which was equal to the repetitions of one condition in the experiment. The average baseline latency was 2.5 ms for visual, 1.7 ms for audio and 4.7 ms for tactile feedback. Because of the two parallel feedback generation processes in Virtual Button Simulator, the average latency also varied across the latency pair conditions. However, the latency was never above 4.8 ms for visual, 4.6 ms for audio and 5.8 ms for tactile feedback. Individual latency for each latency pair condition was added to the conditions before the analysis.

As shown above, the measurements after reprogramming proved that the performance of Virtual Button Simulator again was able to control latencies across the modalities and the modality pairs at levels below human perception also when bimodal feedback was given for each button press.

The experiment software ran on a laptop PC and was again programmed with Presentation®. 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 via wired USB.

Figure 5-1: The Virtual Button Simulator (white, front) with the response pads for the experiment (black, back). The left, small one was used in the simultaneity perception and the right, bigger one in the perceived quality assessment session.

The response pad in the simultaneity judgment session was a modified number keypad connected to the experiment PC containing only three keys, “Y” for “yes” and “N” for “no” responses and the Enter key (see Figure 5-1). In the perceived quality assessment session the responses were given with a modified PC keyboard. Number keys from 1 to 7 were moved to the second lowest row and keys around them were removed in order to make the scoring between 1 and 7 easier. It also contained an Enter key (see Figure 5-1).

5.2.4 Feedback Design and Latency

There were two independent variables in the experiment: Feedback Modality Pair (later Modality Pair) and Feedback Latency Pair (later Latency Pair). There were three Modality Pairs in the experiment: visual-audio, visual-tactile and tactile-audio and eight latency levels for each feedback modality: 0, 20, 30, 50, 70, 100, 150 and 300 ms. This equalled 64 Latency Pairs for each Modality Pair. Figure 5-2 illustrates the Latency Pairs formed from the latency levels from 0 ms to 300 ms. In fact, the first three participants were exposed latency levels from 0 ms to 200 ms only, thus $7 \times 7 = 49$ latency conditions. However, after preliminary analysis for the data for these 3 participants, it was noticed that there were not too big differences for the quality scores. Thus 300 ms latency level was added to the latency levels leading to $8 \times 8 = 64$ Latency Pairs in order to gain hypothesised quality drop. That means that there were 24 responses for each Latency Pairs from 0 to 200 ms and 21 scores for the Latency Pairs containing 300 ms. This led to 147 or 192 different conditions altogether for both simultaneity perception and perceived quality parts. In the beginning of each Modality Pair block, there were 9 training stimuli for each modality pair which equals 27 training stimuli for both parts. Thus, there were 174 or 219 stimuli for both simultaneity perception and perceived quality parts for three first participants and the rest, respectively. Altogether the experiment included $2 \times (3 \times 174 + 21 \times 219) = 10{,}242$ conditions tested that equals $2 \times (3 \times 174 + 21 \times 219) \times 7 = 71{,}694$ button presses (7 presses per Modality Pair and Latency Pair combination). The different number of Latency Pairs per participant did not cause problems in the analysis, however (see Section 5.2.7.1 and Section 5.2.7.2).

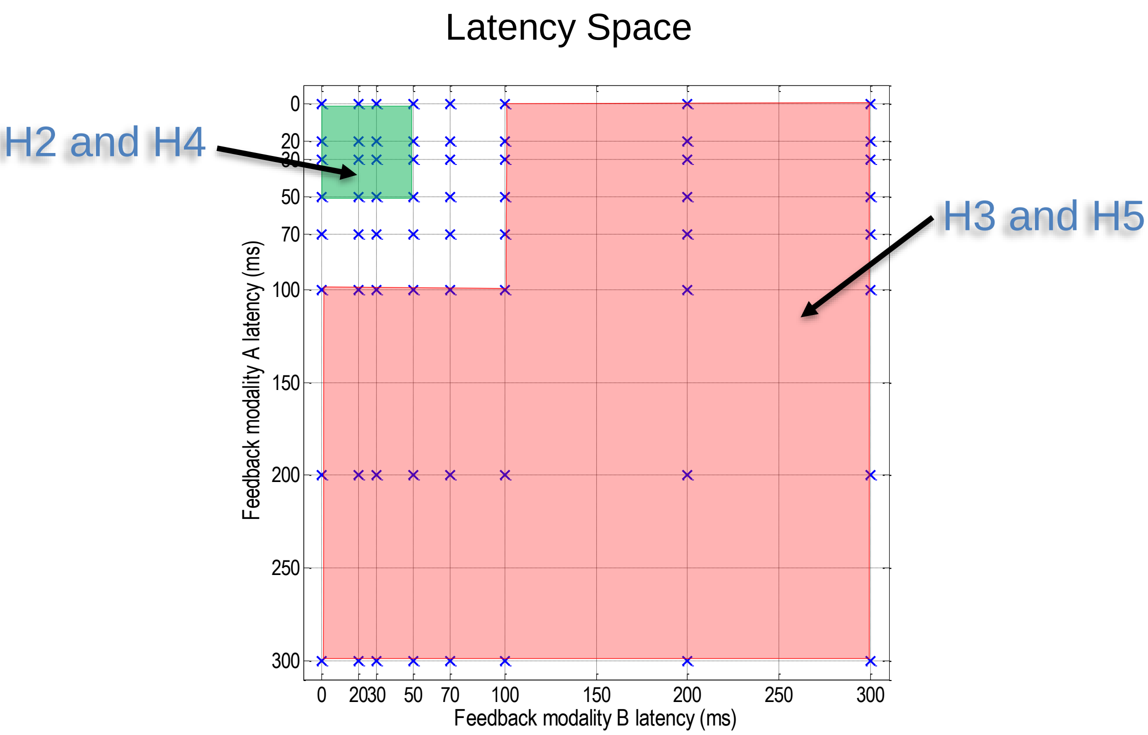


Figure 5-2: Feedback Latency Pairs used in this experiment (marked with blue crosses). The green area illustrates the Hypotheses H2 and H4 and the red are the Hypotheses H3 and H5 (see Section 5.2.5)

All individual feedback was identical to the ones in Experiment 1 described in detail in previous chapter (Section 4.2.6 in Chapter 4). The visual feedback was the green feedback LED glowing as long as the button was pressed. The audio feedback was a short click with a duration of 10 ms. The tactile feedback was designed to be a short tactile click mimicking a tactile feedback of a physical button. The latency was varied between the first moment of finger touch and both feedback events in all Modality Pair conditions in addition to the system baseline latency. Baseline latency was measured individually for each Latency Pair as explained above and it was added individually to latencies in each Latency Pair. Thus, the latency values used in the data analysis were the latency levels with baseline latencies. The selection of the latency values was based on the experiment described in the previous chapter.

5.2.5 Hypotheses

The experiment hypotheses for each modality were based on the previous experiment as follows (see Figure 5-2):

5.2.5.1 Perceived Simultaneity

(H1) Simultaneity perception of touch and bimodal feedback can be modelled with bivariate Gaussian for all Modality Pairs.

(H2) The touch and feedback will be perceived as simultaneous when the latency values for both feedback events are small (0 - 50 ms) for all the modality pairs. Simultaneity means that the proportion of modelled "simultaneous" responses is equal or greater than 75%.

(H3) The touch and feedback will be perceived as non-simultaneous when the latency value for at least the another feedback event is large (100 - 300 ms) for all the modality pairs. Non-simultaneity means that the proportion of modelled "simultaneous" responses is less than 75%.

In the other words, if H1 - H3 were all supported it would mean that it would be possible to create a model for touch-feedback simultaneity and the 75% simultaneity threshold would be between 50 ms and 100 ms for all the modalities and modality pairs (the white area between the green and red areas in Figure 5-1).

5.2.5.2 Perceived Quality

(H4) The quality of the buttons will be perceived as higher when the latency values for both feedback modalities are small (0 - 50 ms) for all the modality pairs.

(H5) The quality of the buttons will be perceived as lower when the latency value for at least another feedback event is large (100 - 300 ms) for all the modality pairs.

In the other words, if H4 and H5 were supported it would mean that the perceived quality would drop significantly between 50 ms and 100 ms for all the modalities and modality pairs (the white area in Figure 5-1).

5.2.6 Procedure

The experiment was divided in two one hour sessions: Simultaneity judgment and perceived quality assessment. Participants sat at a desk in a sound proof music listening room and in the beginning of the first session they read the experiment instructions and filled in a consent form. The background questionnaire was conducted by the experiment moderator. The participants were instructed to hold the Virtual Button Simulator in their non-dominant hand and asked to press the capacitive buttons with the index finger of their dominant hand (Figure 5-3). Modality Pair conditions were counterbalanced and Latency Pair conditions were randomized during both parts of the experiment.

5.2.6.1 Task

The task was identical to the task in Experiment 1 described in Chapter 4. In the both sessions, the participants' task was to follow the flashing red cueing LEDs by pressing the buttons according to the side of the flash: if the right red LED flashed participants were to press the right capacitive button and *vice versa*. If they made a mistake they were instructed to continue the task without interruption.

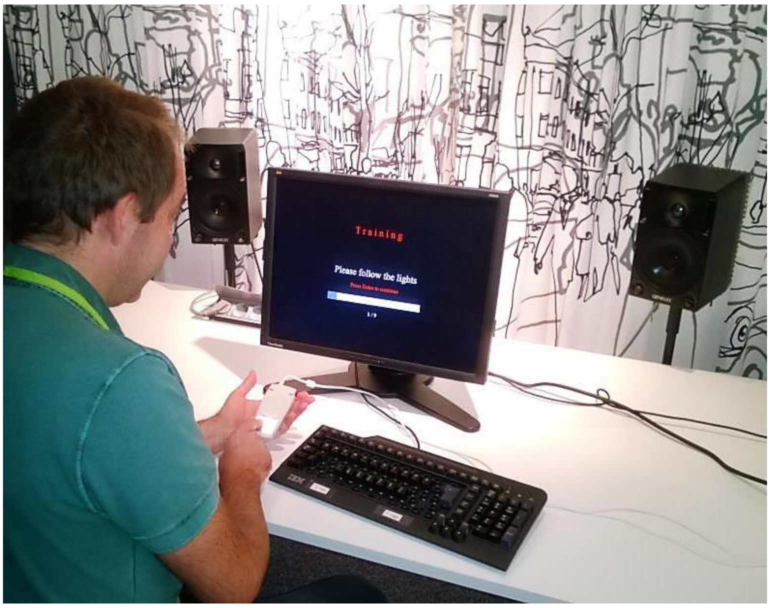


Figure 5-3: Experiment setup. Participants held the Virtual Button Simulator in their non-dominant hand and pressed the buttons with their dominant hand. They responded with a modified keypad connected to a PC. Brown noise was played from the loudspeakers.

5.2.6.2 Feedback, Questions and Responding

Figure 5-4 presents the experiment procedure for both sessions. Feedback was given depending on the Modality Pair and Latency Pair conditions for each button press. One stimulus set consisted of seven cueing flash and button press pairs, within which the Modality Pair and the Latency Pair were kept constant. The Modality Pair was kept constant until all the Latency Pairs were gone through. After the seven flash-press pairs the participant was asked a question. The question was different in the simultaneity judgment and the perceived quality assessment sessions. In the simultaneity judgment session, the question was: “Was the feedback simultaneous with your touch?” The participant responded “Y” or “N” on the response pad according to her/his perception, and pressed the Enter key to confirm that they were ready to continue to the next flash-press sequence. In the quality perception session, the question was: “How would you rate the quality of the keys?”. The participants responded on 1-to-7 scale with the buttons from “1” to “7” on another response pad, and pressed the Enter key for confirmation. After the Enter key press, another stimulus set was presented to the participant. This flash-press-response procedure continued until all the Latency Pairs were gone through. The participant had a training period of 9 flash-press stimulus sets in the beginning of each Modality Pair using all the latency combinations of 0, 200 and 400 ms. These latency conditions were selected for the training period to ensure that the participant understood the tasks properly.

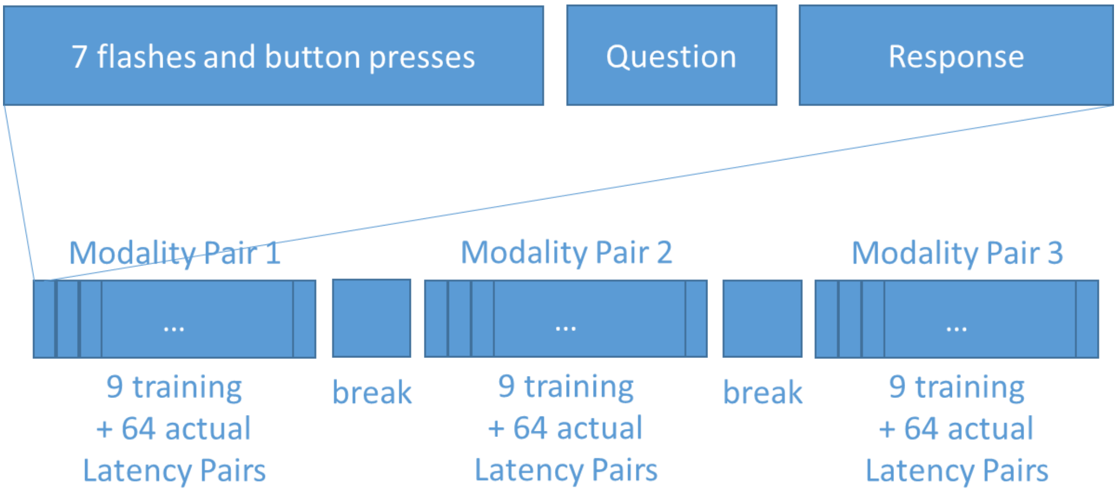


Figure 5-4. Experiment procedure.

Background noise was played from two external active loudspeakers (Genelec 2029AL Digital[63]) 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 visual-audio feedback condition. 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 60 dB (A) measured 80 cm from the midpoint of the loudspeakers. The room background noise level was 28 dB (A).

5.2.7 Analysis Methods

5.2.7.1 Simultaneity Perception

There were $n = 8 \times 8 \times 23 = 1536$ binary responses[64] altogether for each modality pair condition. Earlier work shows that the probability of simultaneity perception of two exogenous events can be modelled with a univariate Gaussian function (Stone *et al.*, 2001;

[63] www.genelec.com

[64] As stated in the Section 5.2.4, the the first three participants experienced Latency Pairs containing latencies from 0 ms to 200 ms only. Because the responses were assumed as independent, the responses for the Latency Pairs with latency value 300 ms were set as “not simultaneous” leading to full number of binary responses for all participants for the analysis.

Zampini *et al.*, 2005; Fujisaki and Nishida, 2009) and the results from Experiment 1 (Chapter 4) proved that this also was true for touch and feedback. In Experiment 2 there were two feedback events involved and it was hypothesized that the simultaneity perception of touch and feedback consisting of two modalities with different latencies will follow *bivariate* Gaussian function.

Bivariate Gaussian is the special case of Multivariate Gaussian (Ash, 2013)

$$f_x\left(x_{1,\cdots,}x_k\right) = \frac{1}{\sqrt{(2\pi)^k|\boldsymbol{\Sigma}|}}e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

Equation 5-1

where in the Bivariate Gaussian case

$$\boldsymbol{\mu} = \binom{\mu_x}{\mu_y}, \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix} \Rightarrow \boldsymbol{\Sigma^{-1}} = \frac{1}{\sigma_x^2\sigma_y^2-(\rho\sigma_x\sigma_y)^2}\begin{bmatrix} \sigma_y^2 & -\rho\sigma_x\sigma_y \\ -\rho\sigma_x\sigma_y & \sigma_x^2 \end{bmatrix}$$

Equation 5-2

Therefore, the probability distribution of bivariate Gaussian (normal) is

$$P(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}-\frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\right]}$$

Equation 5-3

where

$\mu_{1,2}$: the means, defining the location of the top of the Gaussian function.

$\sigma_{1,2}$: standard deviation determining the width of the Gaussian function.

ρ: the correlation of x_1 and x_2.

Figure 5-5 shows an example of bivariate Gaussian surface and its contours. In the experiment in this chapter, x1 and x2 will be only positive since they are latencies between touch and a feedback and feedback came always after touch.

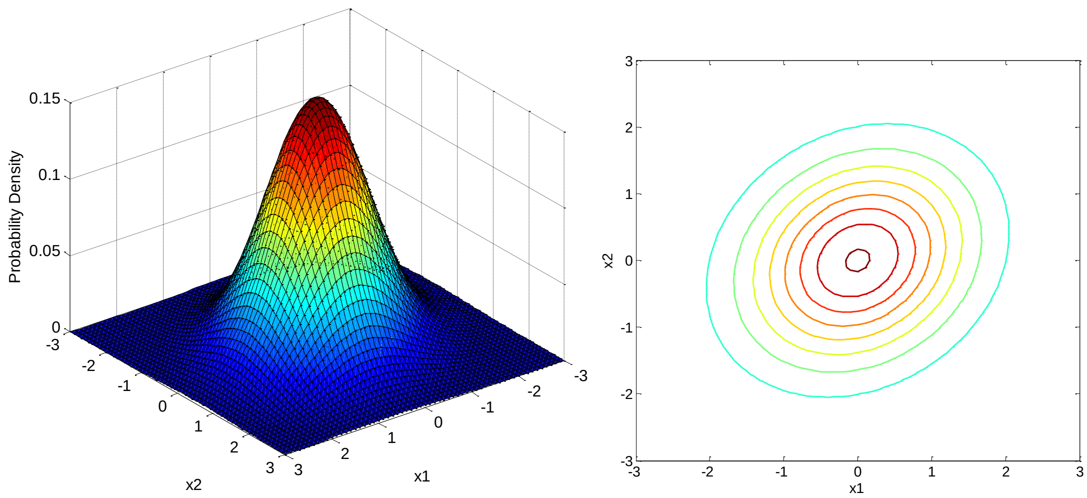


Figure 5-5: Example of the bivariate Gaussian model and its contours.

Similarly to Experiment 1, the probability $p_1$of observing a "simultaneous" response $r_i = 1$ $(i = [1, n])$ at first feedback latency of Latency Pair equal to $LAGx_i$ ms and second feedback latency equal to $LAGy_i$ ms is

$$p_1(r_i = 1|LAGx_i, \mu_x, \sigma_x, LAGy_i, \mu_y, \sigma_y, \rho, a)$$
$$= ae^{-\frac{1}{2(1-\rho^2)}\left[\frac{(LAGx_i-\mu_x)^2}{\sigma_x^2}+\frac{(LAGy_i-\mu_y)^2}{\sigma_y^2}-\frac{2\rho(LAGx_i-\mu_x)(LAGy_i-\mu_y)}{\sigma_x\sigma_y}\right]}$$

Equation 5-4

where μ_x and μ_y define the coordinates of the top of the bivariate Gaussian surface, a is the maximum probability of a "simultaneous" responses at the latencies $LAGx = \mu_x$ and $LAGy = \mu_y$, σ_x and σ_y are the standard deviations associated with responses determining the width of the Gaussian function in $LAGx$ and $LAGy$ dimensions and ρ is the correlation between $\boldsymbol{LAGx}$ and $\boldsymbol{LAGy}$. Probability p_0 of a "not simultaneous" response $r_i = 0$ at a latency values equal to $LAGx_i$ ms and $LAGy_i$ ms is $(1 - p_1)$

$$p_0\left(r_i = 0|LAGx_i, \mu_x, \sigma_x, LAGy_i, \mu_y, \sigma_y, \rho, a\right)$$
$$= 1 - ae^{-\frac{1}{2(1-\rho^2)}\left[\frac{(LAGx_i-\mu_x)^2}{\sigma_x^2}+\frac{\left(LAGy_i-\mu_y\right)^2}{\sigma_y^2}-\frac{2\rho(LAGx_i-\mu_x)\left(LAGy_i-\mu_y\right)}{\sigma_x\sigma_y}\right]}$$

Equation 5-5

The probabilities p_1 and p_0 defined above were fitted jointly to all the observed responses, i.e. to all "simultaneous" and "not simultaneous" responses by all the participant in each and every latency combination condition. The fitting was implemented separately for each Modality Pair using the maximum-likelihood estimation (MLE) method. The MLE procedure was adapted from the ones from Stone *et al.* (Stone *et al.*, 2001) and Experiment 1 (Chapter 4), in which similar procedure to Stone's was used. It was assumed that the responses were made independently from each other. Thus the likelihood function $L\left(\mu_x, \sigma_x, \mu_y, \sigma_y, \rho, a\right)$ became of a product form

$$L\left(\mu_x, \sigma_x, \mu_y, \sigma_y, \rho, a\right) =$$
$$= \prod_{i=1}^{n_1} ae^{-\frac{1}{2(1-\rho^2)}\left[\frac{(LAGx_i-\mu_x)^2}{\sigma_x^2}+\frac{\left(LAGy_i-\mu_y\right)^2}{\sigma_y^2}-\frac{2\rho(LAGx_i-\mu_x)\left(LAGy_i-\mu_y\right)}{\sigma_x\sigma_y}\right]}$$
$$\times \prod_{i=1}^{n_0}\left(1 - ae^{-\frac{1}{2(1-\rho^2)}\left[\frac{(LAGx_i-\mu_x)^2}{\sigma_x^2}+\frac{\left(LAGy_i-\mu_y\right)^2}{\sigma_y^2}-\frac{2\rho(LAGx_i-\mu_x)\left(LAGy_i-\mu_y\right)}{\sigma_x\sigma_y}\right]}\right)$$
$$= \prod_{i=1}^{n}\left((ae^h)^{r_i} \times (1 - ae^h)^{(1-r_i)}\right)$$
$$h = -\frac{1}{2(1-\rho^2)}\left[\frac{(LAGx_i-\mu_x)^2}{\sigma_x^2} + \frac{\left(LAGy_i-\mu_y\right)^2}{\sigma_y^2} - \frac{2\rho(LAGx_i-\mu_x)\left(LAGy_i-\mu_y\right)}{\sigma_x\sigma_y}\right]$$

Equation 5-6

where $n = (n_1 + n_0)$ (n_1 "simultaneous" and n_0 "not simultaneous" responses).

The MLE estimates $\hat{\mu}_x$, $\hat{\mu}_y$, $\hat{\sigma}_x$, $\hat{\sigma}_y$, $\hat{\rho}$ and $\hat{a}$ of the parameters μ_x, μ_y, σ_x, σ_y, ρ and a were obtained for each Modality Pair by minimizing the negative log-likelihood function. This minimization was done with Matlab[65] function *fmincon* which attempts to find the minimum of constrained nonlinear multivariable function using Interior Point algorithm[66]. An initial starting point for the parameter optimization was 0 for all the parameters. The constraints involved in the minimization procedure were ([-50 50], [-50 50], [50 200], [50 200], [0.3 0.5], [0.9 0.99]) for visual-audio, ([-50 50], [-100 50], [50 200], [50 200], [-0.5 0.5], [0.9 1.0]) for visual-tactile and ([-100 50], [-50 50], [50 200], [50 200], [0.3 0.5], [0.9 1.0]) for tactile-audio Modality Pairs (the brackets correspond to the parameters).

A pair-wise Chi-square test of proportion was conducted between the observations to validate when the proportion of simultaneity drops significantly. A Bonferroni correction was applied, resulting in new significance levels set at $p < 7.8 \times 10^{-4}$ and $p < 1.6 \times 10^{-5}$ (corresponding the significance levels 5 % and 0.1 %).

5.2.7.2 Perceived Quality

Each and every Latency Pair were analysed individually. A Skillings-Mack test was conducted in order to find general effect in significance. Skillings-Mack test is equivalent to Friedman but it takes into account missing data for some data points (Hollander *et al.*, 2013). As a *post-hoc* test the Wilcoxon signed-rank test was applied with Bonferroni correction resulting in new significance levels set at $p < 7.8 \times 10^{-4}$ and $p < 1.6 \times 10^{-5}$ (corresponding the significance levels 5 % and 0.1 %).

5.2.7.3 Significance Maps

As the *post-hoc* tests for both simultaneity perception (Chi-Square) and perceived quality (Wilcoxon signed-rank) data involved $64 \times 64 - 64 = 4032$ comparisons (all the Latency Pair combinations minus comparisons to itself), the introduction and interpretation of the results would have been very challenging and exhausting for both the writer and the reader with traditional tables or lists. That is why the significance maps, introduced briefly already

[65] se.mathworks.com

[66] http://se.mathworks.com/help/optim/ug/fmincon.html

in Chapter 4 (Section 4.2.10.2), were developed in this thesis to make the results more understandable.

The significance maps are a novel way to visualize a complex set of condition comparison results. An example of a significance map is illustrated in Figure 5-6. The black square means the Current Feedback Condition (Modality Pair and Latency Pair) – the condition under comparison with the other conditions. If the average value (proportion of "simultaneous" responses or perceived quality score) of the Current Condition 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 value of the Current Condition 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 statistically not-significant is coloured with gradients either between yellow and green or yellow or red depending on whether the average value of the Current Condition is higher or lower than of another condition. Gradients are also implemented between the significance levels 5 % and 0.1 % with '+' or 'o' mark. This colouring scheme highlights the relative proportion of the simultaneous responses or the relative quality of the current condition. The standard significance levels introduced here were mapped on the Bonferroni corrected significance levels when the significance maps were implemented.

Significance maps are useful, since they provide colour coded overview of the full set of data analysis results. In addition, if needed, one can go into details by searching the significance map under interest and zoom in (this is easy with electronic version of this thesis, for example). This procedure is in line with Shneiderman's (1996) instructions for visual information seeking: Overview first, zoom and filter, then details-on-demand.

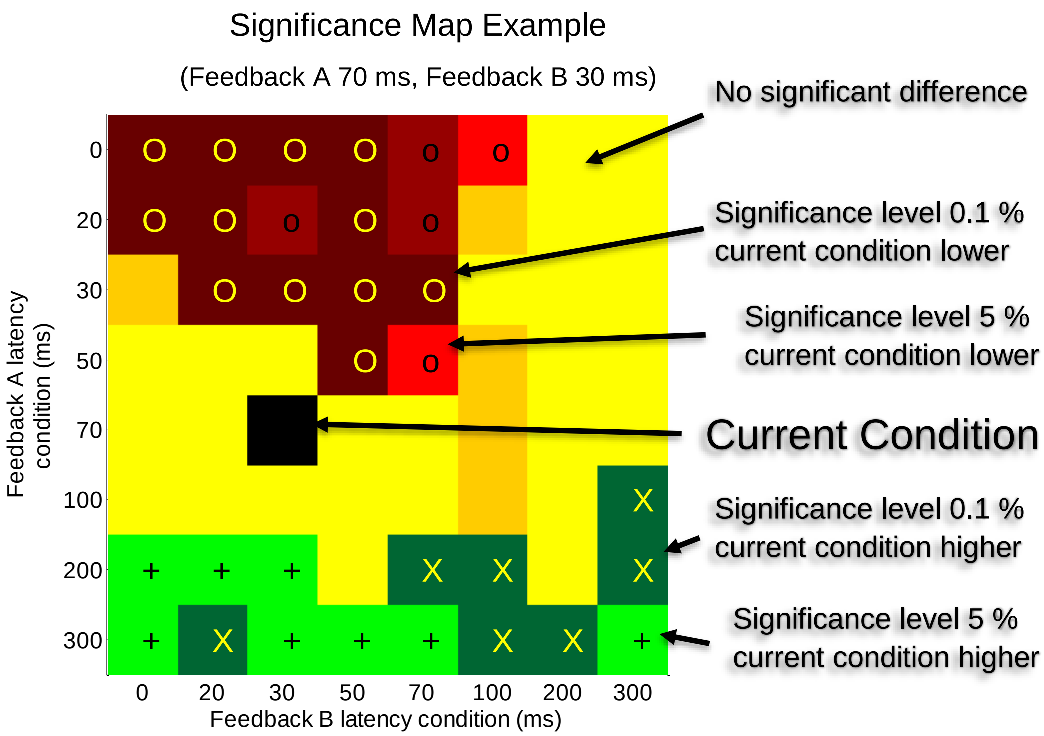


Figure 5-6: An example of one significance map which illustrates statistical differences between large set of conditions. This figure shows the Latency Pair of 70 ms and 30 ms (black square = Current Condition) visualizing the statistical significance of the Current Condition in relation to the other Latency Pairs. See the text for more detailed description.

5.2.8 Results

This section presents all the results from this experiment. First, the simultaneity perception results are introduced, including the bivariate Gaussian models and 75 % thresholds for each modality in each Modality Pair. Second, the results from the perceived quality part are introduced with the help of significance maps. After the discussion, the latency guidelines for bimodal feedback are established in the next section.

5.2.8.1 Simultaneity Perception

Parameters and Confidence Bodies

The results of the bivariate Gaussian model fitting for the probability p_1 are introduced in Table 5-1. The results consist of the model parameter MLE estimates and their 95%

confidence intervals which were derived with the restricted Likelihood Ratio Test (LRT). The restricted LRTs of all six parameters of all the feedback modality pair specific Gaussian models were implemented against $\chi_5^2(0.95)$. The confidence body for the parameters was derived against $\chi_6^2(0.95)$ distribution. Because in the bivariate Gaussian model includes six parameters, the confidence body of the MLE was also 6-dimensional (6D). Since a 6D object is very difficult to visualize here, the two-dimensional (2D) projections of it were investigated. Figure 5-7, Figure 5-8 and Figure 5-9 show these 2D projections of the confidence body when the Modality Pair was visual-audio, visual-tactile and tactile-audio respectively. Visual investigation of the figures reveals that the projections are not ellipsoids indicating that the distributions of the parameter estimates were not normal. Therefore LRT was used instead of Wald's test for determining the confidence intervals of the parameters ((Millar, 2011) and Section 4.2.10.1 in Chapter 4).

Table 5-1: The Gaussian model fitting results for the proportion of "simultaneous" responses (probability p_1, see Section 5.2.7.1). All the values are in milliseconds (ms) and the quantities are MLE parameter estimates and their 95% confidence intervals. Note that the 95% confidence intervals are asymmetric around MLE estimates due to non-normal distribution of the parameters. "r" after the estimated value means that the parameter was restricted during the estimation ($\hat{a}$).

Feedback Modality pair	$\hat{\mu}_x$ 95% CI$_{\hat{\mu}_x}$	$\hat{\mu}_y$ 95% CI$_{\hat{\mu}_y}$	$\hat{\sigma}_x$ 95% CI$_{\hat{\sigma}_x}$	$\hat{\sigma}_y$ 95% CI$_{\hat{\sigma}_y}$	$\hat{\rho}$ 95% CI$_{\hat{\rho}}$	$\hat{a}$ 95% CI$_{\hat{a}}$
Visual-audio	-13 [-29 4.7]	-8.0 [-29 13]	154 [139 172]	174 [155 197]	0.43 [0.062 0.61]	0.95 [0.90 – 0.99]
Visual-tactile	37 [15 58]	-48 [-61-36]	131 [114 152]	143 [131 157]	-0.012 [-0.21 0.25]	1.0 r [0.95 – 1.1]
Tactile-audio	-97 [-111-83]	4.8 [-22 31]	164 [152 177]	155 [136 181]	0.36 [0.11 0.54]	1.0 r [0.94 – 1.1]

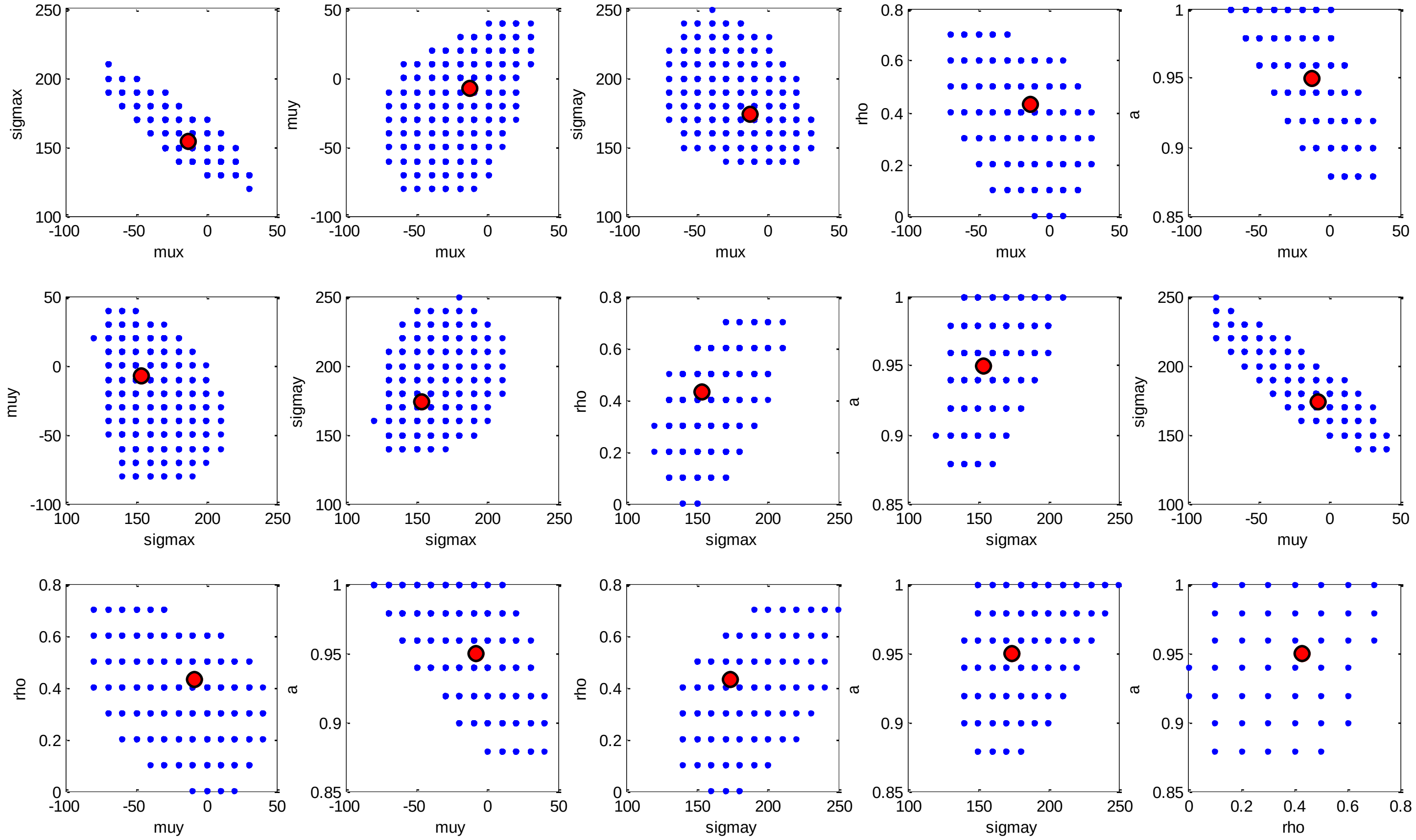


Figure 5-7: The 2D projections of the 6D confidence body of the MLE of Gaussian function parameter estimates for the simultaneity perception in touch and *visual-audio* Modality Pair. The MLE points are marked as red dots. The 6D confidence body was used to calculate the 95% confidence intervals for the Gaussian model values.

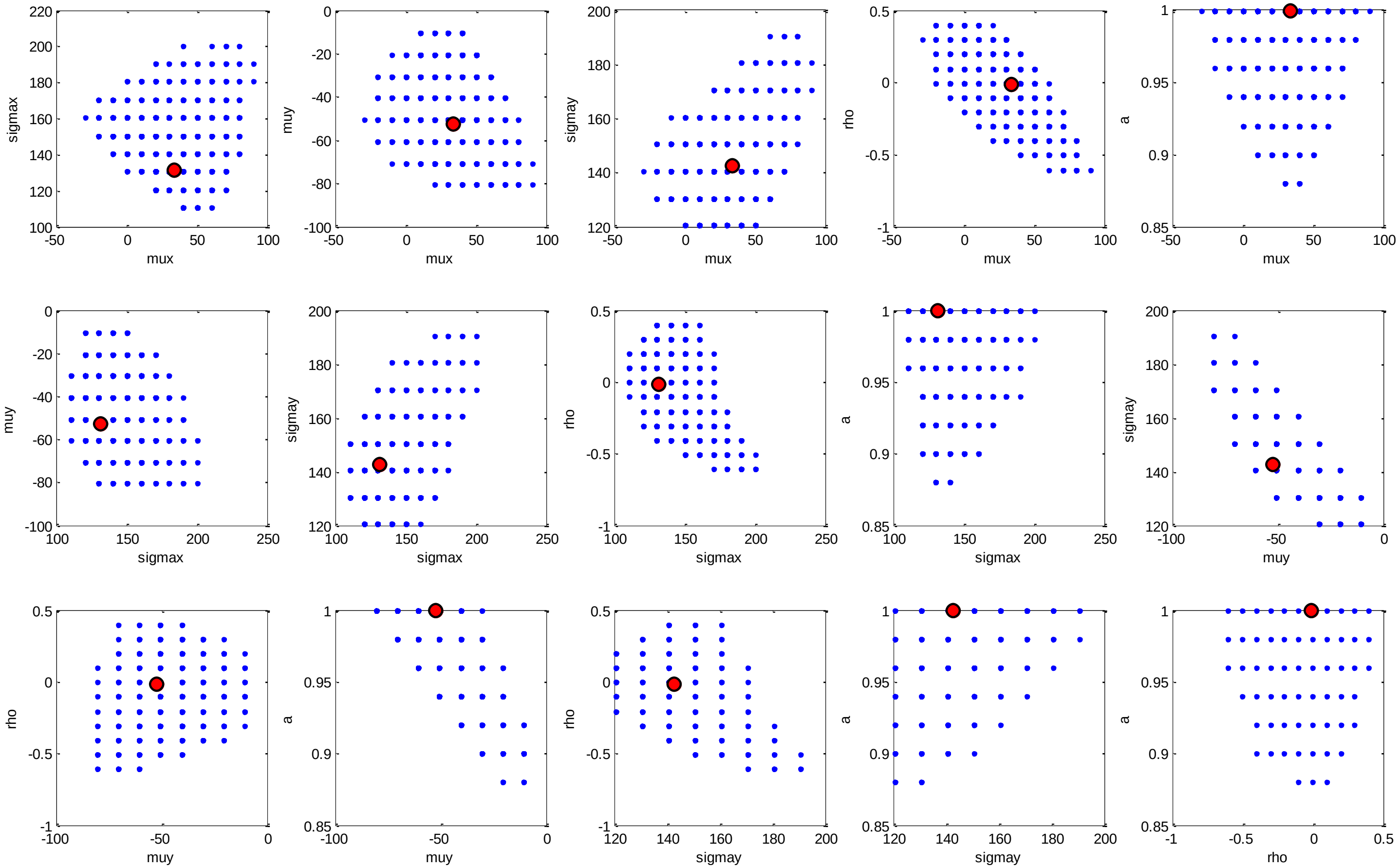


Figure 5-8: The 2D projections of the 6D confidence body of the MLE of Gaussian function parameter estimates for the simultaneity perception in touch and *visual-tactile* Modality Pair. The MLE points are marked as red dots. The 6D confidence body was used to calculate the 95% confidence intervals for the Gaussian model values.

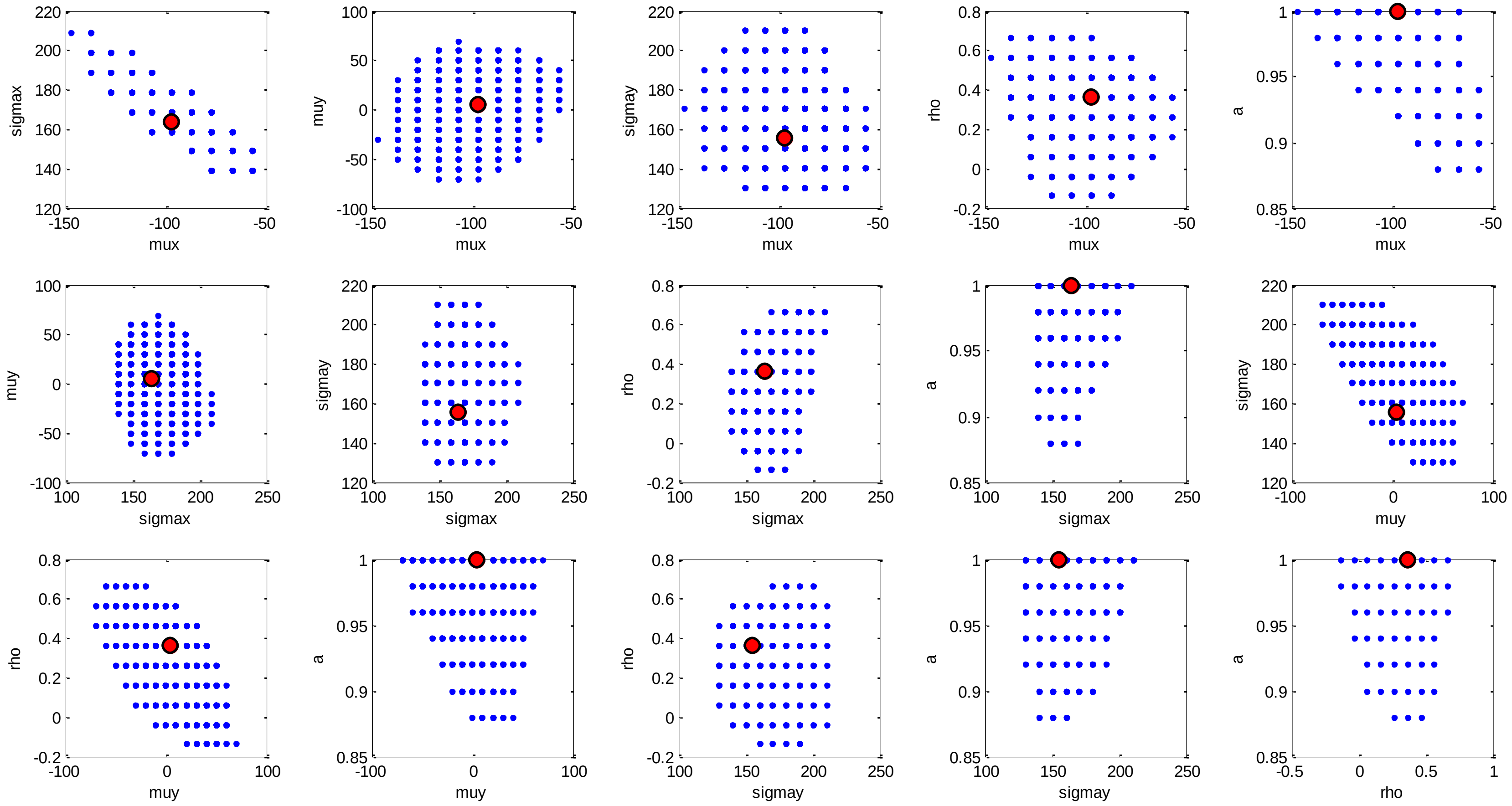


Figure 5-9: The 2D projections of the 6D confidence body of the MLE of Gaussian function parameter estimates for the simultaneity perception in touch and *tactile-audio* Modality Pair. The MLE points are marked as red dots. The 6D confidence body was used to calculate the 95% confidence intervals for the Gaussian model values.

Surfaces of the Gaussian Models

The 95% confidence surfaces for the probability p_1 for all the feedback modality conditions were calculated by going through the parameter sextets within the whole 6D confidence body and finding the minimum and the maximum values of the probability p_1 at each combination of $LAGx$ and $LAGy$ running from 0 to 300 ms (1 ms resolution).

Figure 5-10 shows the bivariate Gaussian model fitted to the proportions of "simultaneous" responses with MLE when the Modality Pair was visual-audio. The solid surface (light to dark blue gradient) is the Gaussian MLE model and the surfaces consisted of red and blue dots form the lower and upper 95% confidence surfaces for it. The magenta circles show the observed touch-feedback simultaneity proportions in each latency combination condition. The set of circles underneath the model show the differences between the observations and the model. The green and red circles indicate if the individual data point is significantly above or below the bivariate Gaussian surface. If the dot is yellow, there is no significant difference. The bivariate Gaussian models for visual-tactile and tactile-audio Modality Pairs are shown in Figure 5-11 and Figure 5-12.

The goodness of a bivariate Gaussian fit was tested with Chi-square goodness-of-fit test. The proportion of "simultaneous" responses was compared with the modelled proportions at all the latency conditions. All the fits passed the test (visual-audio: $p = 0.994$, visual-tactile: $p = 0.807$ and tactile-audio: $p = 0.0503$). This proves that the experimental data support (H1) – the distribution of "simultaneous" responses will follow a bivariate Gaussian distribution.

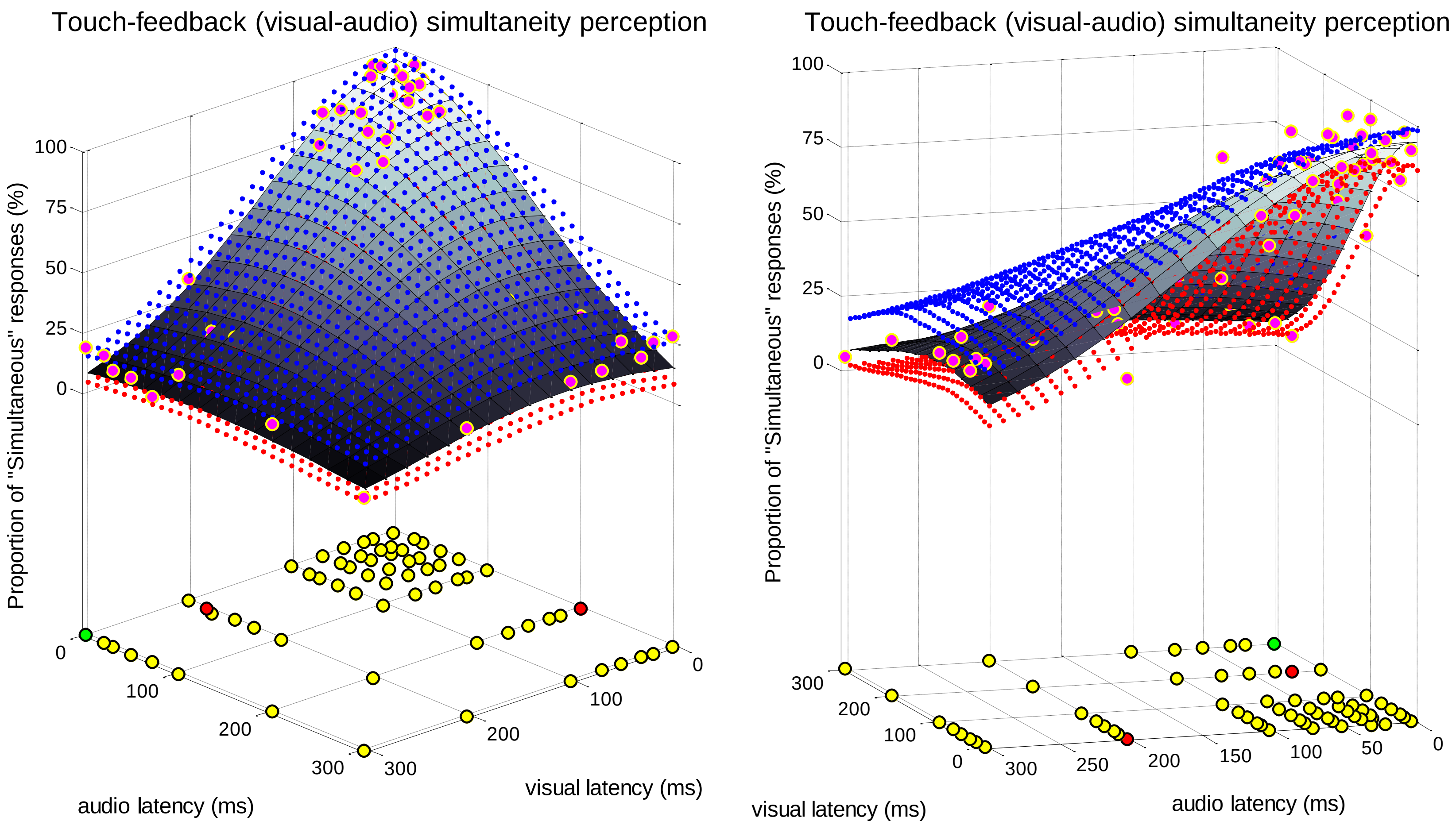


Figure 5-10: Fitted bivariate Gaussian model of simultaneity perception of touch and *visual-audio* feedback from two different angles. See text for details.

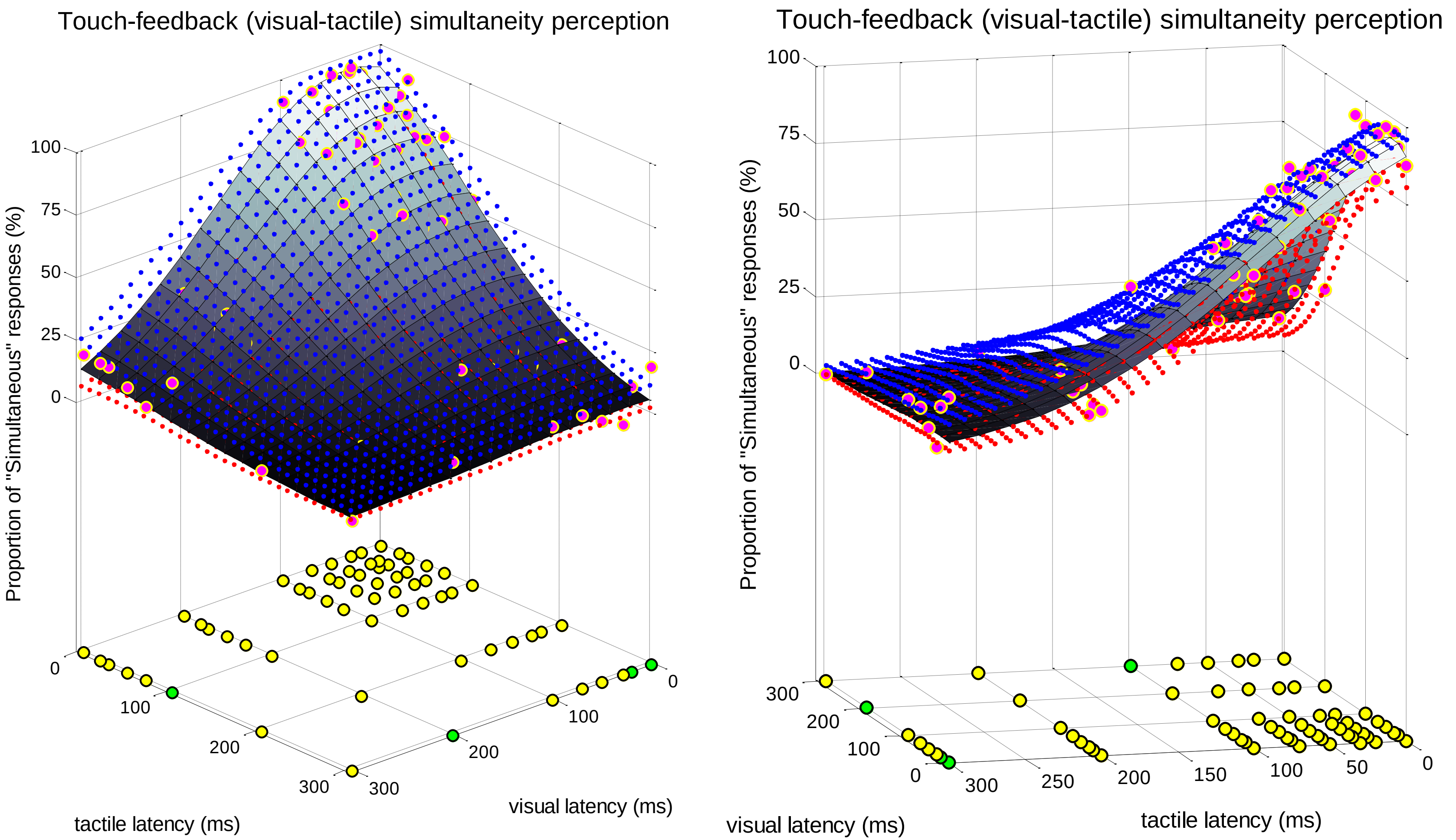


Figure 5-11: Fitted bivariate Gaussian model of simultaneity perception of touch and *visual-tactile* feedback from two different angles. See text for details.

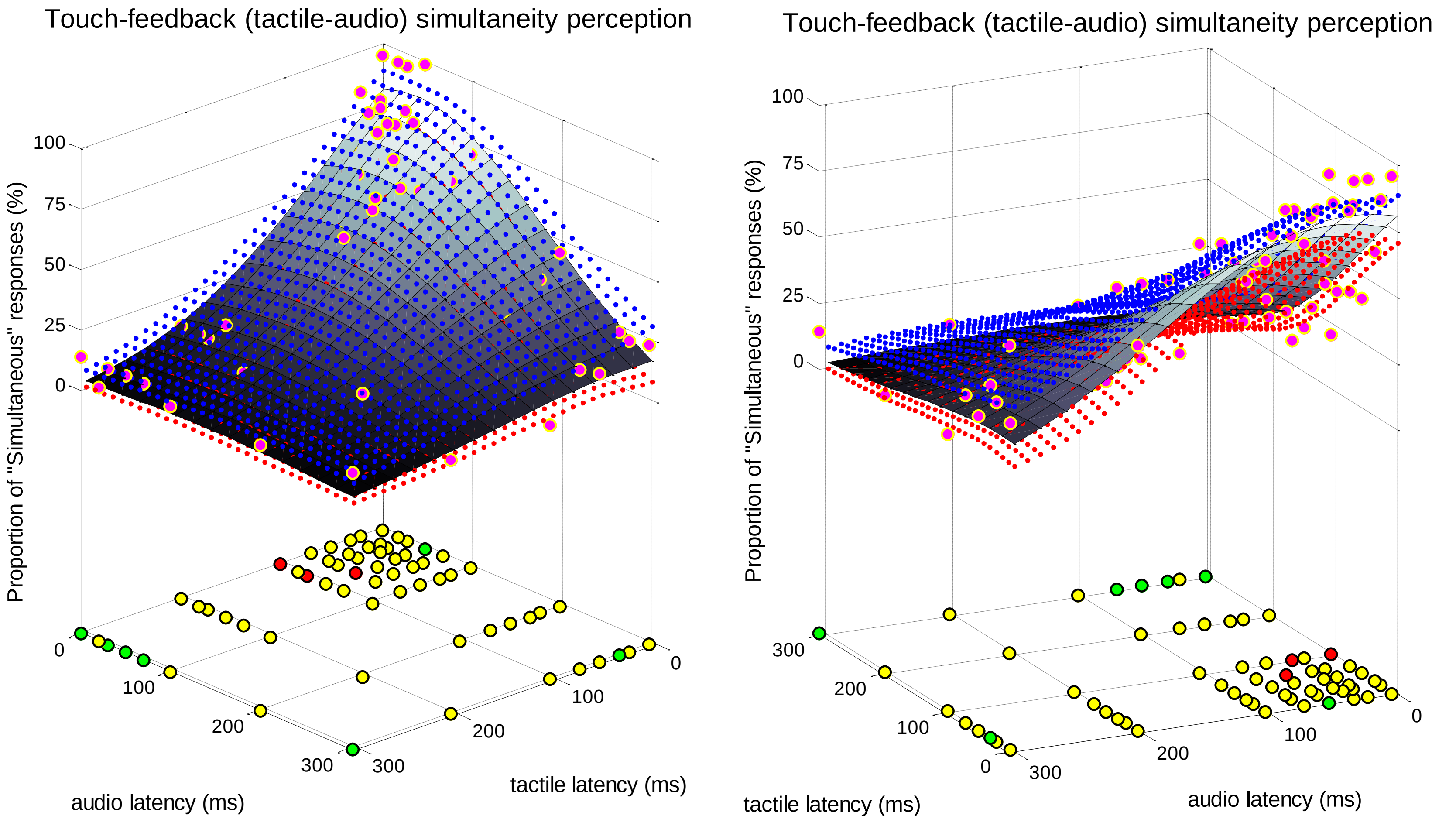


Figure 5-12: Fitted bivariate Gaussian model of simultaneity perception of touch and *tactile-audio* feedback from two different angles. See text for details.

Contours of the Gaussian Models

Figure 5-13 shows the contour plots of the bivariate Gaussian model with 75% threshold contour and the 95% confidence regions when the Modality Pair was visual-audio. The dotted dark blue line shows the 75% threshold and the blue and red dotted lines show the 95% upper and lower confidence regions for that corresponding the 95% blue and red confidence surfaces shown in Figure 5-10. The numbers on the contours mean the modelled proportion of simultaneous responses. The green and red dots indicate that the observed proportion of "simultaneous" responses was significantly higher or lower than 75% threshold (Wilcoxon signed-rank test $p < 0.00078$ with Bonferroni correction). The yellow dots indicate that there was no statistically significant difference. The right-hand side of the figure shows the zoomed-in view for the contour plot on the left (latency for both of the feedback is between 0 ms and 150 ms). The shaded area shows the region within which the touch and feedback were perceived as simultaneous according to the model (proportion of the "simultaneous" responses was ≥ 75 %). Figure 5-14 and Figure 5-15 show the corresponding contour figures for visual-tactile and tactile-audio, respectively.

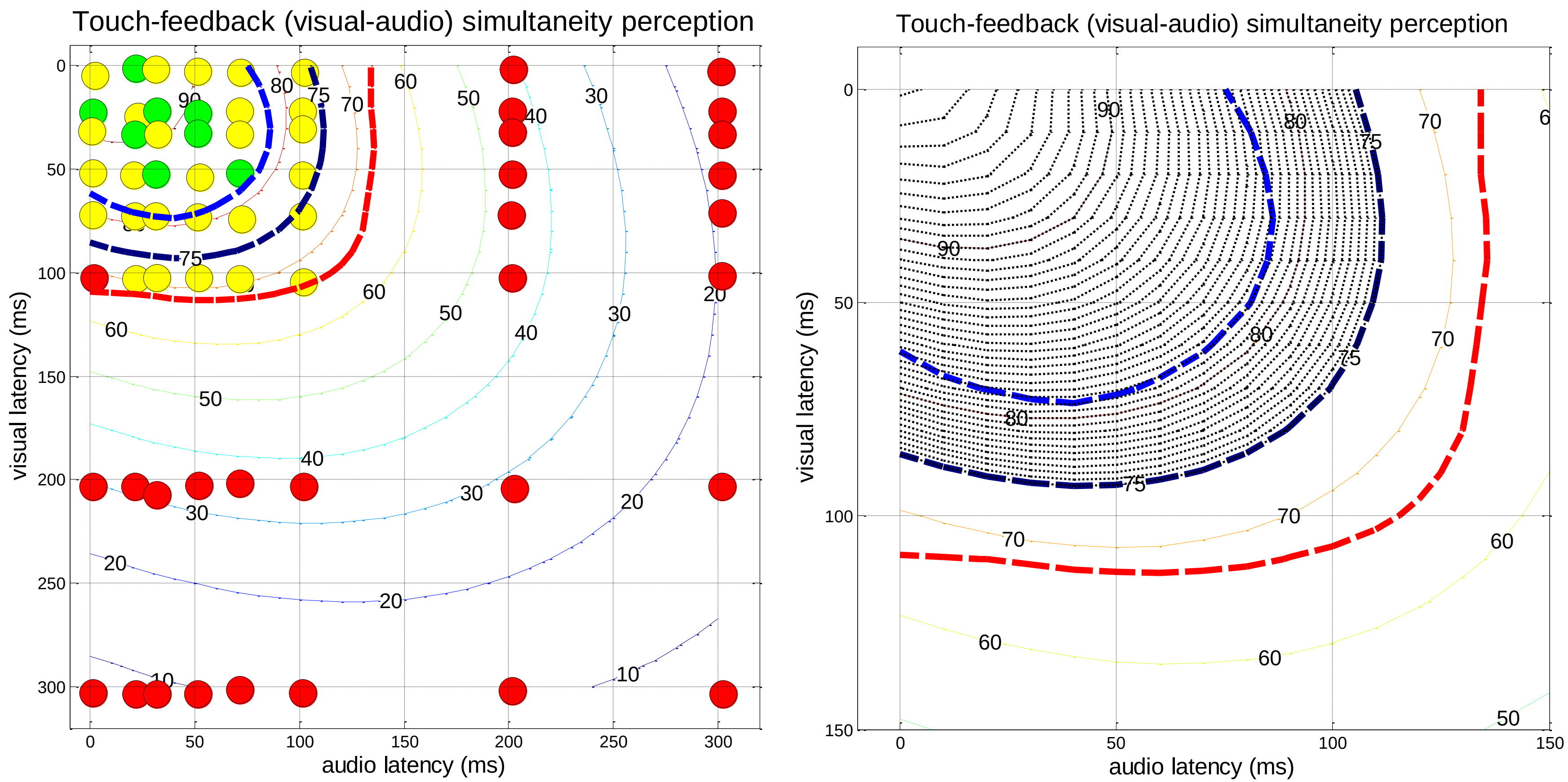


Figure 5-13: Left: Contour plots of the bivariate Gaussian model of simultaneity perception of touch and *visual-audio* feedback. Right: Zoomed-in view of the contour plot. The shaded area shows the region within which the touch and feedback were perceived as simultaneous according to the model. See text for details.

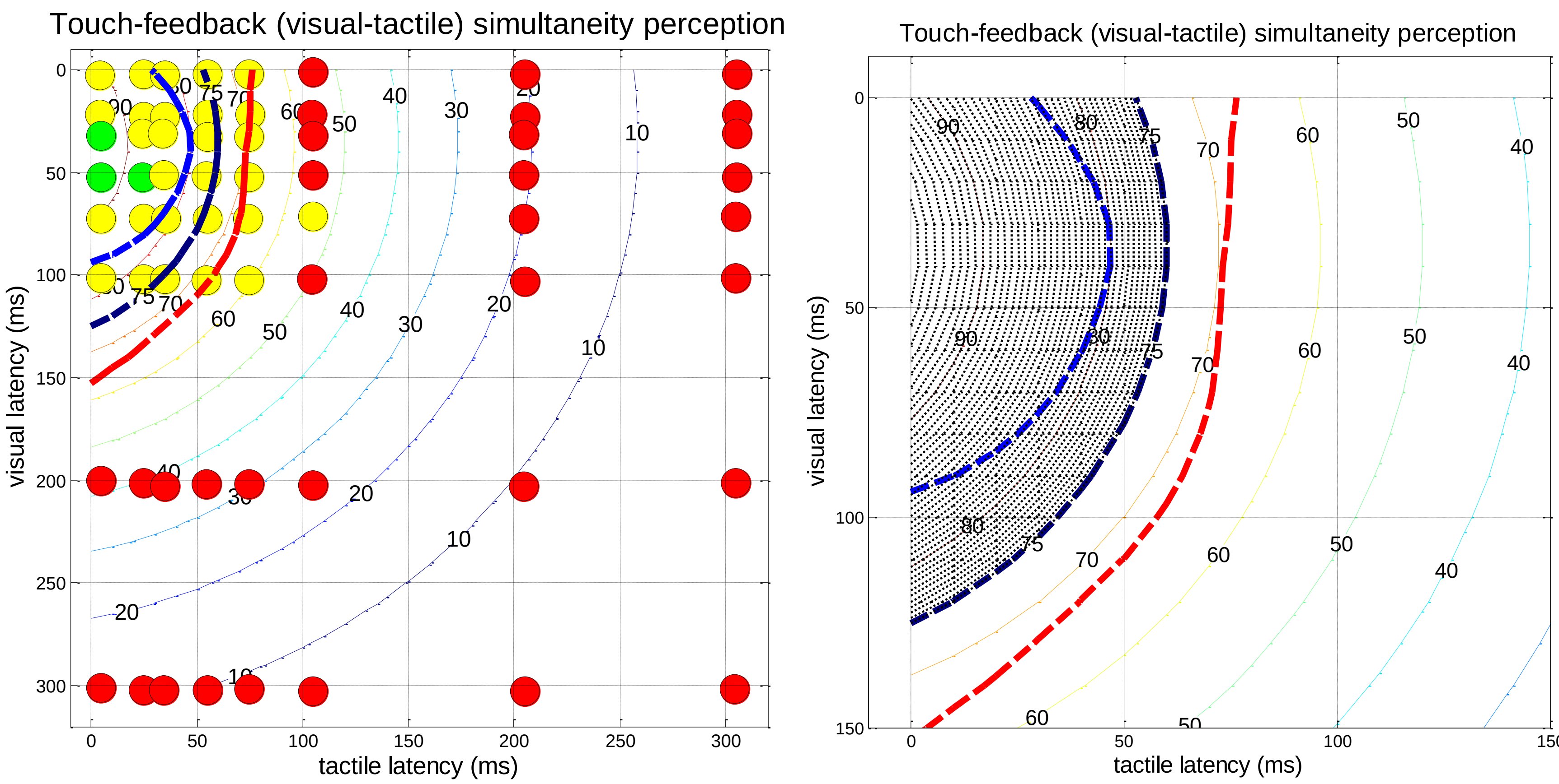


Figure 5-14: Left: Contour plots of the bivariate Gaussian model of simultaneity perception of touch and *visual-tactile* feedback. Right: Zoomed-in view of the contour plot where the shaded area shows the region within which the touch and feedback were perceived as simultaneous according to the model. See text for details.

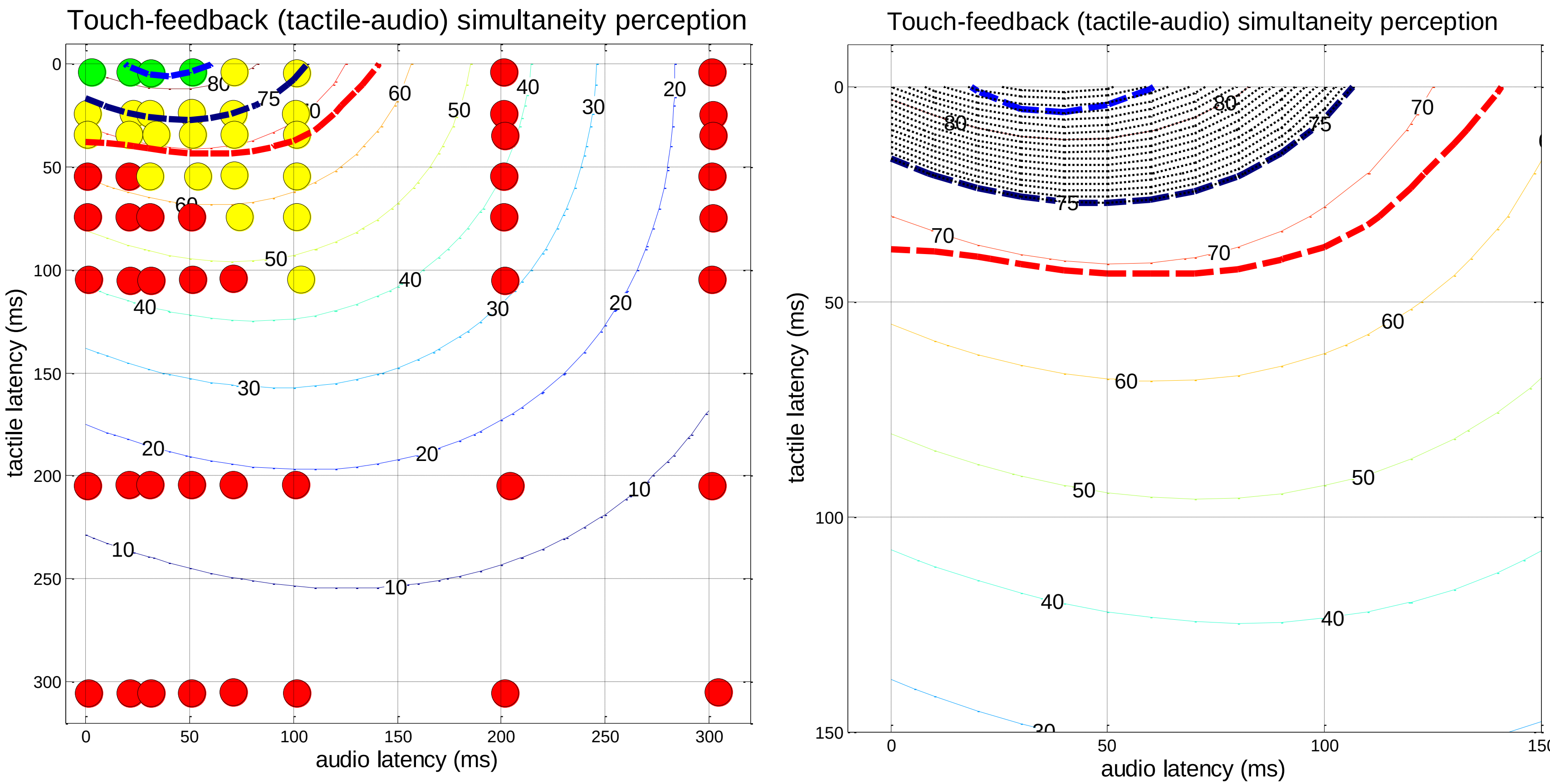


Figure 5-15: Left: Contour plots of the bivariate Gaussian model of simultaneity perception of touch and *tactile-audio* feedback. Right: Zoomed-in view of the contour plot where the shaded area shows the region within which the touch and feedback were perceived as simultaneous according to the model. See text for details.

Visual-Audio Feedback

The bivariate Gaussian model for visual-audio Modality Pair shows that the simultaneity perception was almost symmetric between the modalities (Figure 5-10 and Figure 5-13). The model behaved nearly as hypothesized: The simultaneity perception threshold settled down between 50 ms and 115 ms approximately (hypothesis: between 50 ms and 100 ms). This suggests that (H2) – The touch and feedback will be perceived as simultaneous when the latency values are small (0-50 ms) – was fully supported. (H3) – the touch and feedback will be perceived as non-simultaneous when the latency value for at least the another feedback event is large (100-300 ms) – was partially supported since the threshold reached as far as 115 ms for audio feedback latency. However, it can be seen from the red circles in Figure 5-13 that when the one of the latencies in Latency Pair was 200 ms or 300 ms the touch and feedback were perceived as non-simultaneous (significantly below the 75 % threshold).

Visual-Tactile Feedback

The bivariate Gaussian model for visual-tactile Modality Pair shows that the simultaneity perception was more asymmetric between the modalities (Figure 5-11 and Figure 5-14) than when Latency Pair was visual-audio. Compared to visual feedback latency, the proportion of "simultaneous" responses dropped more steeply when the tactile feedback latency increased. The 75% threshold was not higher than 60 ms for tactile modality, but as much as maximum 125 ms approximately for visual modality (when tactile feedback latency is 0 ms). Still the model worked closely as hypothesised: The simultaneity perception threshold was between 60 ms and 125 ms approximately (hypothesis: between 50 ms and 100 ms). This suggests that (H2) – The touch and feedback will be perceived as simultaneous when the latency values are small (0-50 ms) – was fully supported. The circles in Figure 5-14 show that the results partially supported (H3) – The touch and feedback will be perceived as non-simultaneous when the latency value for at least the another feedback event is large (100-300 ms): Touch and feedback was perceived as non-simultaneous when tactile feedback was between 100 and 300 ms (except tactile feedback latency 100 ms and visual 70 ms), but the visual feedback latency could be 200 ms before the feedback was perceived significantly as non-simultaneous (tactile feedback latency between 0 ms and 70 ms. All the Latency Pairs when one or both feedback latency was 200 ms or 300 ms the touch and feedback were perceived as non-simultaneous.

Tactile-Audio Feedback

The bivariate Gaussian model for tactile-audio Modality Pair shows that the simultaneity perception of touch and tactile-audio feedback was even more asymmetric between the modalities (Figure 5-12 and Figure 5-15). The proportion of “simultaneous” responses dropped much sooner when the tactile feedback latency increased compared to the audio feedback modality. The 75% simultaneity perception threshold was not higher than 25 ms for tactile modality (at audio feedback latency 50 ms). In turn, the audio latency could be 100 ms and the touch and the feedback were perceived still as simultaneous (at tactile feedback latency 10 ms, approximately).

The bivariate Gaussian model suggests that (H2) – The touch and feedback will be perceived as simultaneous when the latency values are small (0-50 ms) – was partially supported. The simultaneity perception concentrated to the low latencies of tactile feedback, but the threshold was as low as 25 ms, maximum. In contrast, the audio feedback latency as large as 100 ms could be perceived as simultaneous. This also indicates that (H3) – The touch and feedback will be perceived as non-simultaneous when the latency value for at least the another feedback event is large (100-300 ms) – also was partially supported. However, it can be seen from the circles in Figure 5-15 that when the one of the latencies in Latency Pair was 200 ms or 300 ms the touch and feedback were perceived always as non-simultaneous.

5.2.8.2 Quality perception

Boxplots with Trendlines

A boxplot with the medians and means with trendlines of the perceived quality scores when the Modality Pair was visual-audio is shown in Figure 5-16, visual-tactile in Figure 5-17 and tactile-audio in Figure 5-18. The Skillings-Mack test showed significant differences in perceived quality depending on Latency Pair when Modality Pair was visual-audio ($T = 1357.2,\ p < 0.001,\ df = 63$) and visual-tactile ($T = 1295.7, p < 0.001, df = 63$) and tactile-audio ($T = 1020.7,\ p < 0.001,\ df = 63$).

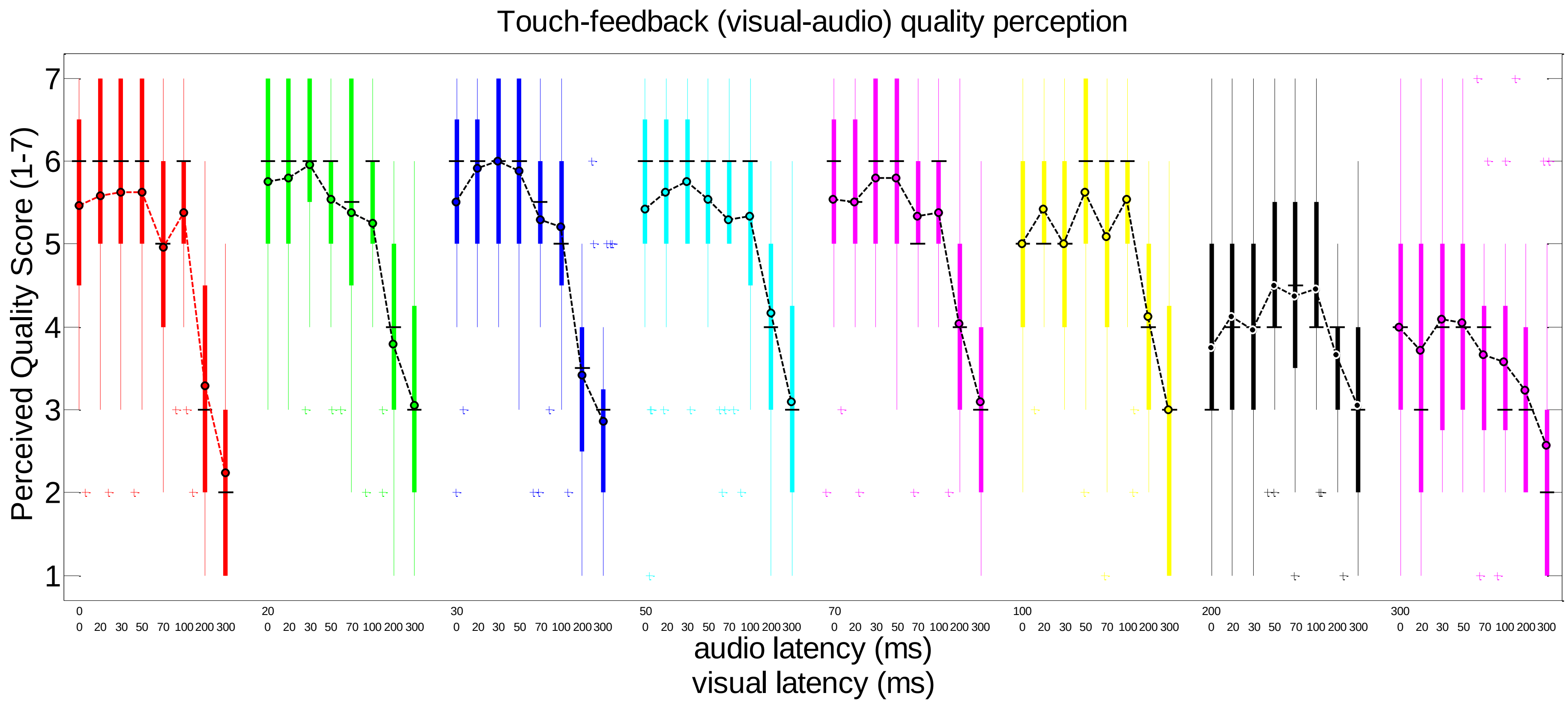


Figure 5-16: Boxplot of the perceived quality scores when the Modality Pair was *visual-audio*. The horizontal black line inside or on the edge of a box show the median of the scores. The edges of the boxes show the 25th and 75th percentiles of the data, and the whiskers show the most extreme data points not considered outliers (Tukey, 1977). Outliers are presented as "+" marks and are for visualization only (not considered in data analysis). 'o' markers show the means of the data for each condition and the dashed lines show the trendlines.

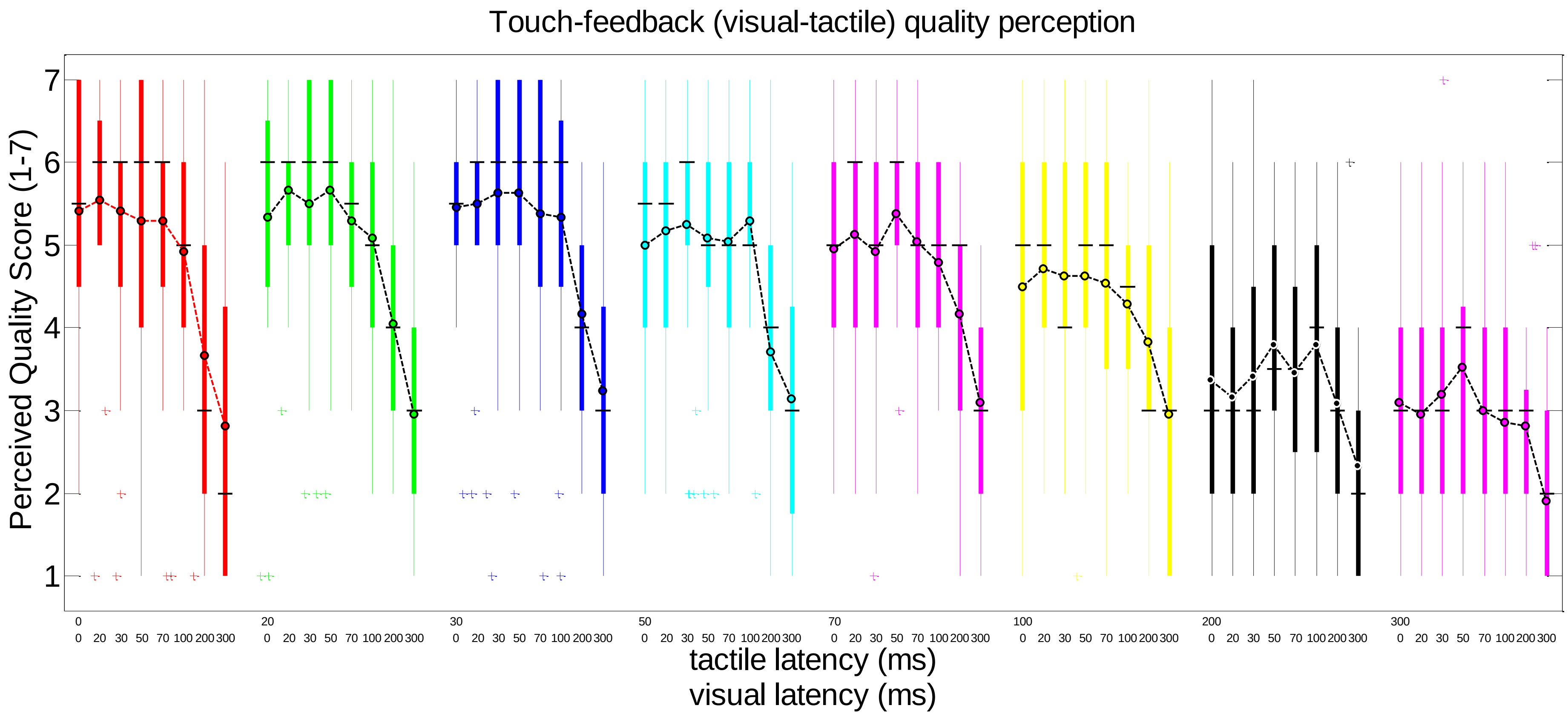


Figure 5-17: Boxplot of the perceived quality scores when the Modality Pair was *visual-tactile*. The horizontal black line inside or on the edge of a box show the median of the scores. The edges of the boxes show the 25th and 75th percentiles of the data, and the whiskers show the most extreme data points not considered outliers (Tukey, 1977). Outliers are presented as "+" marks and are for visualization only (not considered in data analysis). 'o' markers show the means of the data for each condition and the dashed lines show the trendlines.

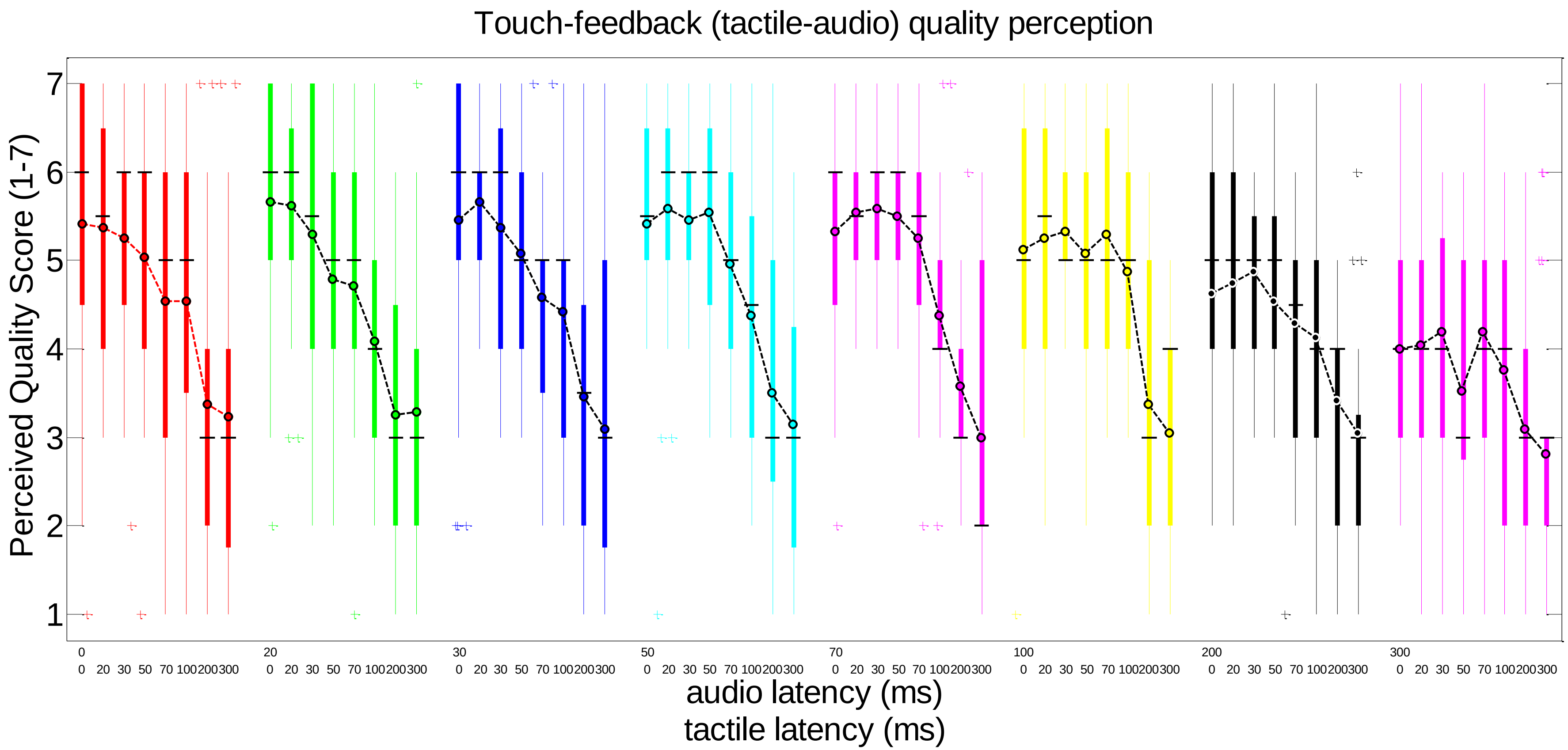


Figure 5-18: Boxplot of the perceived quality scores when the Modality Pair was *tactile-audio*. The horizontal black line inside or on the edge of a box show the median of the scores. The edges of the boxes show the 25th and 75th percentiles of the data, and the whiskers show the most extreme data points not considered outliers (Tukey, 1977). Outliers are presented as "+" marks and are for visualization only (not considered in data analysis). 'o' markers show the means of the data for each condition and the dashed lines show the trendlines.

Post Hoc *Analysis and Significance Maps*

The *post hoc* analysis with Wilcoxon Signed-Rank tests was conducted and the results are introduced in significance maps (see Section 5.2.7.3) shown in Figure 5-19 (visual-audio) Figure 5-20 (visual-tactile) and Figure 5-21 (tactile-audio). The Latency Pairs inside the area highlighted with green were always significantly higher than some others and in addition did not differ significantly from each other. That means that the quality of the buttons was perceived high with all these latency conditions. The boundaries of the green highlight were defined to be the perceived quality drop threshold. The conditions inside the area highlighted with red were always significantly lower than some others.

Visual-Audio Feedback

As the green highlight in Figure 5-19 shows the quality of the buttons was perceived high symmetrically between the modalities when the Modality Pair was visual-audio and when the latency was between 0 and 100 ms for both modalities (with three exceptions: Latency Pairs 70 ms and 0 ms, 0 ms and 100 ms and 30 ms and 100 ms). Therefore, the results support (H4) – The quality of the buttons will be perceived as higher when the latency values for both feedback modalities are small (0-50 ms). (H5) – The quality of the buttons will be perceived as lower when the latency value for at least another feedback event is large (100-300 ms) – was partially supported since also 100 ms was within the high quality area for some Latency Pairs.

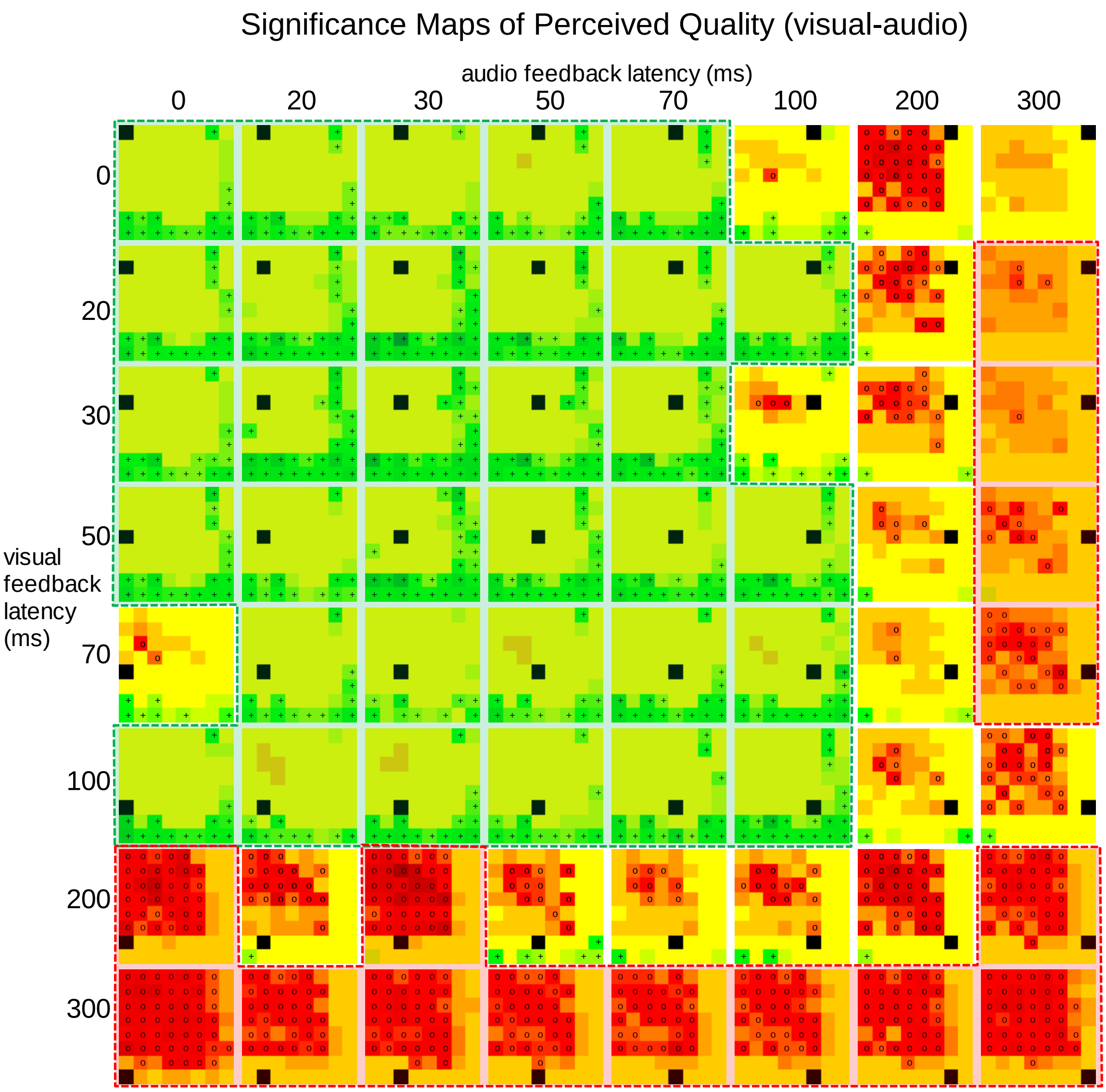


Figure 5-19: Significance maps of perceived quality scores for all Latency Pairs when Modality Pair was *visual-audio*. See text for further details.

Visual-Tactile Feedback

As the green highlight in Figure 5-20 shows the quality of the buttons was perceived high also symmetrically between the modalities when the Modality Pair was visual-tactile and when the latency was between 0 and 100 ms for both modalities (with two exceptions: Latency Pairs 70 ms and 100 ms, 100 ms and 100 ms). Thus, (H4) – high quality at low latencies – was supported. (H5) – Low quality at high latencies – was partially supported since also latency of 100 ms was within the high quality area for some Latency Pairs.

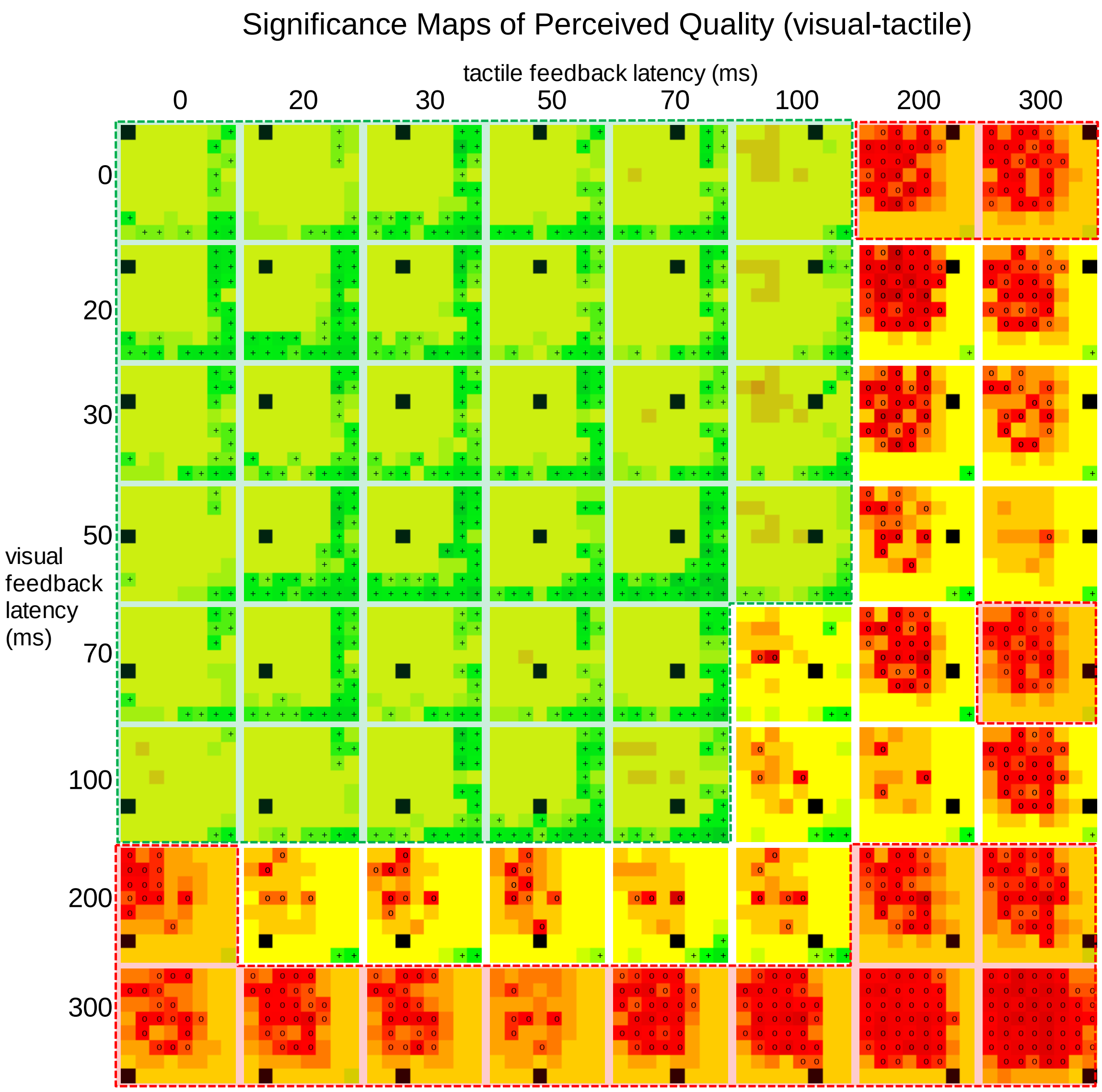


Figure 5-20: Significance maps of perceived quality scores for all Latency Pairs when Modality Pair was *visual-tactile*. See text for further details.

Tactile-Audio Feedback

As simultaneity perception results, the quality perception also showed asymmetry when Latency Pair was tactile-audio: The perceived quality score drop happened earlier when tactile feedback latency increased compared to audio. The green highlight in Figure 5-20 shows that the quality of the buttons was perceived high with higher latencies (even 200 ms) when the feedback modality was tactile compared to audio. Still the results support (H4). (H5) was partially supported since also latencies of 200 ms and 100 ms were within the high quality area for some Latency Pairs.

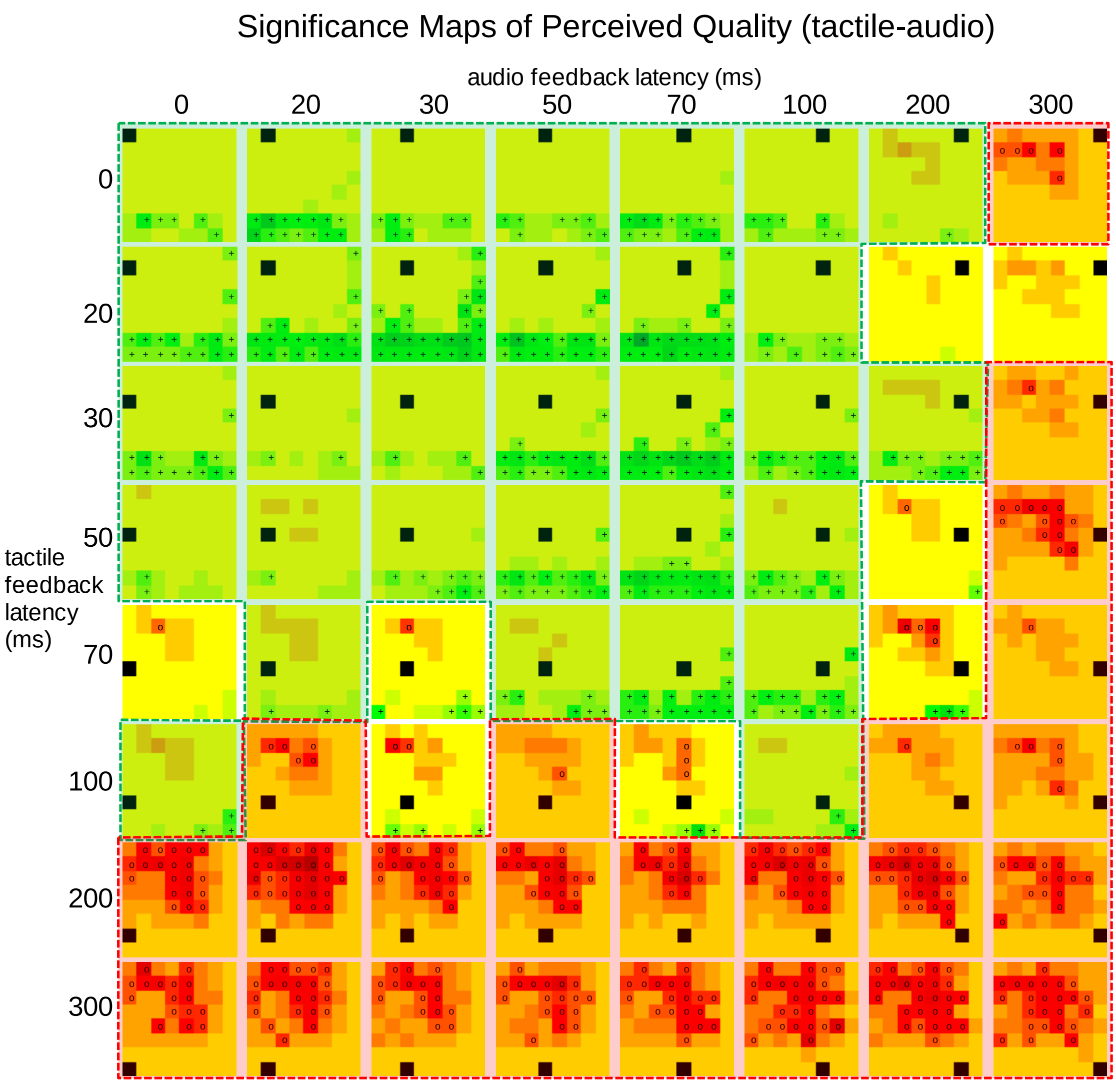


Figure 5-21: Significance maps of perceived quality scores for all Latency Pairs when Modality Pair was *tactile-audio*. See text for further details.

5.3 Latency Guidelines

To conclude the simultaneity and quality perception results above, the touch and feedback were perceived approximately as simultaneous when the latency was small and not simultaneous when the latency was large, on average. The same was found on quality perception: the quality of the buttons were perceived as high when latency was small and low when latency was large. The thresholds were not as unambiguous as hypothesised, though. Based on the findings above, latency recommendations were established as a form of simplified latency guideline. The guideline was established as conservative to fulfil the strictest requirements of high quality products based on the combination of the fitted bivariate Gaussian models and the perceived quality assessment results. Figure 5-22 shows the combined results from the simultaneity perception and perceived quality for all three Modality Pairs. In each subfigure the contours of the bivariate Gaussian model, the 75 % threshold and the 95 % confidence intervals for that are shown (explained in Section 5.2.8.1). In addition, the green area with dashed green outline shows the high quality region defined in Section 5.2.8.2 . The intersection of these two is presented as yellow area with solid yellow outline.

In big companies, with a lot of employees and fast pace of product projects, it is crucial to be able to communicate recommendations effectively. This can be done if the guidelines are clear and simple. That is why the combination of the simultaneity and quality perception results were further simplified. The approximation was conducted carefully, without compromising the statistical significance and not making the guidelines more liberal than the results, as follows: The guideline for each modality pair is a rectangle (black thick outline in Figure 5-22) which follows the intersection as closely as possible, but straightens the lines within the lower 95% confidence interval of the 75% simultaneity threshold (the red dashed line in the figures). The guideline formed this way (rounded to the nearest 5 ms) are meant to be easy to remember and use. The guideline is: When the Modality Pair is

***visual-audio*, the visual feedback latency should not exceed**
90 ms and audio feedback latency should be between 20 and 70 ms,

***visual-tactile*, the visual feedback latency should not exceed**
100 ms and tactile feedback latency should not exceed 55 ms and

***tactile-audio*, the tactile feedback latency should not exceed**
25 ms and audio feedback latency should not exceed 100 ms.

Combined Results and the Guidelines for All the Modality Pairs

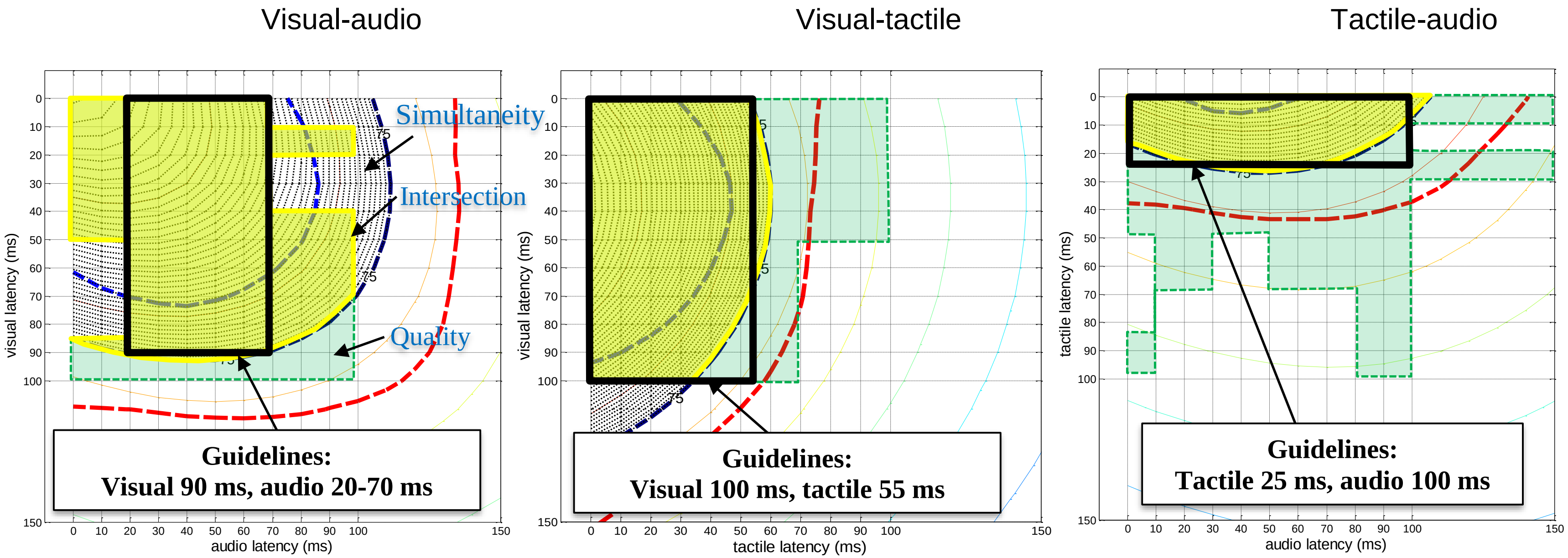


Figure 5-22: Latency guidelines derived from the combination of the simultaneity perception and perceived quality results when the Modality Pair was visual-audio (left) and visual-tactile (middle) and tactile-audio (right). The shaded area shows the area where the simultaneity perception was > 75 % according to the Gaussian model. The light green area with dashed outline shows the high quality area. The yellow area with thick outline shows the cross-section of these two. The rectangle with black outline shows the guideline derived from the cross-section.

5.4 Discussion

5.4.1 Bivariate Gaussian Model

One aim of the experiment described in this chapter was to derive a model for simultaneity perception of touch and bimodal feedback which would help to establish practical guidelines. It was hypothesised (H1) that the results would follow a bivariate Gaussian model. The choice of the model was reasonable since it was proved in previous chapter that the simultaneity perception of touch and unimodal feedback followed a univariate Gaussian model. The experimental data and the data analysis showed that bivariate Gaussian was a right choice. The Chi-Square showed that none of the models did not differ significantly from the observations. The fit was especially good when the Modality Pair was visual-audio ($p = 0.994$), meaning that the bivariate Gaussian function modelled the simultaneity perception of touch and visual-audio feedback very well. When the Modality Pair was visual-tactile, the fit was also good ($p = 0.807$), meaning that the bivariate Gaussian modelled the simultaneity perception of touch and visual-tactile feedback well.

When the Modality Pair was tactile-audio, the Chi-Square statistics showed nearly statistical significant difference between the observations and the bivariate Gaussian model ($p = 0.0503$), meaning that although the fit was successful, the model could not defer to the set of observations well, and there are remarkable differences between the observations and the model. Therefore, it could have been possible that some other model would have fitted better. However, this work is the new opening and it is the first time when simultaneity perception of bimodal feedback was studied. The results set a baseline and it is the best knowledge so far.

5.4.2 Simultaneity Perception

It was hypothesised that the simultaneity perception threshold of touch and bimodal feedback would be between 50 ms and 100 ms ((H2) and (H3) combined). This was based on the results from the previous experiment which showed that the simultaneity perception thresholds were approximately between 50 ms and 100 ms (85 ms, 80 ms and 52 ms for visual, audio and tactile feedback, respectively).

When the Modality Pair was visual-audio, the threshold was close to the unimodal case: the maximum of the 75 % threshold contour in visual feedback dimension was 90 ms vs. 85 ms in unimodal visual feedback case. However, the audio feedback latency threshold was larger when feedback featured visual modality in addition to audio: maximum 115 ms vs. 80 ms. 80 ms is inside the 95 % confidence region, though.

When the Modality Pair was visual-tactile, the maximum of the threshold was 125 ms in visual dimension, which was clearly larger than in unimodal case 85 ms. It was partly outside of the 95 % confidence region as well (tactile feedback 0 ms). However, the maximum of the threshold in tactile dimension was close to single modality case: 60 ms vs. 52 ms.

When the Modality Pair was tactile-audio, the difference between bimodal feedback and unimodal feedback was remarkable. The maximum of the threshold was 25 ms in tactile dimension, whereas the simultaneity perception threshold for touch and tactile feedback was 52 ms in unimodal case, which was clearly outside of the 95 % confidence region (see Figure 5-15). In turn, the maximum threshold for visual feedback in bimodal feedback case was higher than in unimodal case: 110 ms vs. 85 ms. It was within the 95 % confidence region, though.

It can be concluded that the simultaneity of touch and bimodal feedback approximately in the same way than in unimodal feedback case, when the Modality Pair was visual-audio. When tactile feedback was involved, in visual-tactile and tactile-audio Modality Pairs, there were differences. This highlights the value of the second experiment: The simultaneity perception can be different when the feedback features two modalities compared to one. Therefore, it is important to find the thresholds for both cases.

5.4.3 Perceived Quality

It also was hypothesised that the threshold for perceived quality would fall between 50 ms and 100 ms. This was based on the results from the previous experiment. The results from the current experiment showed that the perceived quality was constantly high until 100 ms latency when the Modality Pair was visual-audio for both feedback modalities (with few exceptions). This was the case also when the Modality Pair was visual-tactile. When the Modality Pair was tactile-audio, the results were again a bit peculiar, as in simultaneity perception. The quality drop threshold varied between tactile feedback latency 50 ms and 100 ms when the audio feedback latency varied between 0 ms and 100 ms. Also when tactile feedback was 0 ms or 30 ms, the quality drop threshold was as large as 200 ms for audio feedback latency.

It is also interesting to notice that for each tactile feedback latency, the quality scores do not differ significantly from each other when the audio feedback latency increases (with only one exception, tactile 50 ms and audio 300 ms). When audio feedback latency changed, it did not affect significantly to the perceived quality score. This means that the quality perception of tactile-audio feedback was dominated only by tactile feedback latency.

This again emphasises the importance of this experiment. The results from the unimodal feedback case cannot be transferred as such to bimodal feedback case.

5.4.4 Latency guidelines

When comparing latency guidelines for both unimodal and bimodal feedback and bimodal feedback only, it can be seen that there are some differences (see Table 5-2). The upper limits of visual guidelines were stretched up a bit when another modality was involved. Another modality seems to make the requirements more liberal, working in favour of designers and engineers.

In case of audio feedback, the bimodal guideline was exactly the same when the feedback was visual-audio, whereas when combined with tactile feedback, the guideline was more liberal, working again in favour of designers and engineers. Tactile feedback turned to be more challenging when combined with audio, the guideline being remarkable stricter than for unimodal and bimodal visual-tactile, which were approximately the same.

As stated before when discussing simultaneity perception and perceived quality results, these differences stress the importance of this experiment. The guidelines were different depending on the number of modalities and the modality combination. This is most probably because of different human perception mechanism when there are two modalities included in feedback instead only one. This is unexplored area in psychophysics and human-computer interaction and future research is needed to understand the reasons behind the differences in guidelines.

Table 5-2: Comparison of Latency Guidelines

	Unimodal	Visual-audio	Visual-tactile	Tactile-audio
Visual	30 - 85 ms	≤ 90 ms	≤ 100 ms	
Audio	20 - 70 ms	20 – 70 ms		≤ 100 ms
Tactile	5 - 50 ms		≤ 55 ms	≤ 25 ms

5.5 Reflection of Guidelines in Mobile Phones

To show how the latency guideline for bimodal feedback can be put in practice, the latencies of the touchscreen phones measured in Chapter 3 and Chapter 4 were reflected against the guideline. In addition, the latencies of four other newer generation mobile phones were measured with the latency measurement tool introduced in Chapter 3.

The phones measured in Chapter 3 were:

- HTC Desire
- LG Chocolate BL40
- Nokia 5800 XpressMusic
- Samsung Omnia i900

The phones measured and also reflected against the latency guideline for unimodal feedback in Chapter 4 were:

- HTC Wildfire S running Android
- iPhone 4S running iOS
- Nokia Lumia 800 running Windows Phone 7
- Nokia N9 running MeeGo
- Samsung Galaxy Note running Android.

The newer generation phones are introduced in Figure 4-19:

- Apple[67] iPhone 5S running iOS 7.1
- LG [68] Nexus 5 running Android 4.4.4
- Nokia [69] Lumia 930 running Windows Phone 8.1
- Samsung[70] Galaxy S5 running Android 4.4.2

The latency measurement process was exactly the same than in the previous section (Section 4.4 in Chapter 4). A letter “g” in standard message application was pressed 20 times and calculated the average and standard deviation of tactile, audio and visual feedback latency.

Picture of Apple iPhone 5S has been removed due to Copyright restrictions.

Figure 5-23: Four new generation touchscreen phones measured for bimodal feedback latency guideline reflection: Apple iPhone 5S, LG Nexus 5, Nokia Lumia 930[71] and Samsung Galaxy S5.

The results can be seen in Table 5-3. The latencies highlighted with green fulfilled the guidelines. None of the phones fulfilled all the guidelines. However, the oldest Nokia phone (5800 XpressMusic) nearly fulfilled the guidelines. Only with tactile-audio feedback, the tactile feedback latency guideline was not fulfilled since with audio feedback the guideline

[67] https://www.apple.com/iphone

[68] http://www.google.fi/nexus/5

[69] http://www.microsoft.com/en/mobile/phones/lumia

[70] http://www.samsung.com/uk/consumer/mobile-devices/smartphones

[71] Used with permission from Microsoft.

was very strict. From the phones introduced in Chapter 4, the Nokia phones did also a good job fulfilling the guidelines. From the newest phones, Samsung fulfilled half of the guideline and LG's audio feedback latency was below the guideline when the feedback modality pair was tactile-audio. All the rest of the feedback latencies were higher than recommended by 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. There is a lot of room for improvement in the virtual keyboards even in the current high end phones.

Table 5-3: Average touch-feedback latencies in milliseconds for virtual buttons in the default messaging application in 13 touchscreen mobile phones. The first four are the phones measured in Chapter 3 and the next five in Chapter 4. The last four are newer generation phones. The table is sorted according to the average latency of all the feedback within the groups. The green highlight shows that the latency was within the guideline set in this study.

	Mobile Phone (Operating System)	Visual-audio feedback latency (guideline: visual 90 ms, audio 20-70 ms)		Visual-tactile feedback latency (guideline: visual 100 ms, tactile 55 ms)		Tactile-audio feedback latency (guideline: tactile 25 ms, audio 100 ms)	
		visual	audio	visual	tactile	tactile	audio
Chapter 3	Nokia 5800 XpressMusic	61	30	61	39	39	30
	HTC Desire	104	157	104	93	93	157
	Samsung Omnia i900	167	111	167	155	155	111
	LG Chocolate BL-40	129	293	129	200	200	293
Chapter 4	Nokia Lumia 800 (Windows Phone)	53	37	Not supported		Not supported	
	Nokia N9 (MeeGo)	110	38	110	35	35	38
	Apple iPhone 4S (iOS)	102	83	Not supported		Not supported	
	HTC Wildfire S (Android)	140	149	140	74	74	149
	Samsung Galaxy Note (Android)	197	172	197	123	123	172
Chapter 5	Samsung Galaxy S5 (Android 4.4.2)	84	129	84	50	50	129
	LG Nexus 5 (Android 4.4.4)	118	77	118	83	83	77
	Apple iPhone 5S (iOS 7.1)	98	101	Not supported		Not supported	
	Nokia Lumia 930 (Windows Phone 8.1)	140	180	Not supported		Not supported	

5.6 Conclusions and Research Questions 2, 3 and 4

In this chapter, latency guidelines for bimodal feedback in touchscreen interaction were established for the first time. Encouraged by the results from Experiment 1, the guidelines were formed by combining the simultaneity perception and perceived quality results. These guidelines are important extension to the guidelines established for unimodal feedback. The research showed that both the simultaneity and quality perception of bimodal feedback differed from the unimodal ones, affecting further to the guidelines. Therefore, the bimodal latency guidelines would have been partly incorrect, if they were formed from unimodal ones.

As in the previous chapter, the aim of the experiment in this chapter was 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 the results would be usable for touchscreen mobile device designers. The participants pressed capacitive buttons and the bimodal feedback were provided from the same device (as usually is the case in a real mobile phone). The participants were asked to judge if the feedback was simultaneous with the touch. The bivariate Gaussian models were usable tools for finding parameters for applicable guidelines. The research in this chapter showed for the first time that the perception of simultaneity of touch and bimodal feedback (visual-audio, visual-tactile and tactile-audio) in a realistic setup could be modelled with a bivariate Gaussian function. This expanded the earlier, already important finding from Chapter 4 that the simultaneity perception of touch a unimodal feedback can be modelled with a univariate Gaussian function. In order to establish practical guidelines, the 75% thresholds were obtained from the Gaussian models. In order to further understand the effect of latency to the user experience, the participants were asked to score the perceived quality of the buttons when pressed and the latency was varied for both of the bimodal feedback.

Research Questions 2 asked:

> **RQ2:** “What are the touch-feedback simultaneity perception thresholds in virtual button interaction?”

This question was already answered in Chapter 4 when the feedback was unimodal. The experiment described in this chapter yielded further answers for this question when the feedback was bimodal: the 75 % threshold was the 75 % contour from the bivariate Gaussian model for each Modality Pair. Figure 5-24 illustrates this in detail.

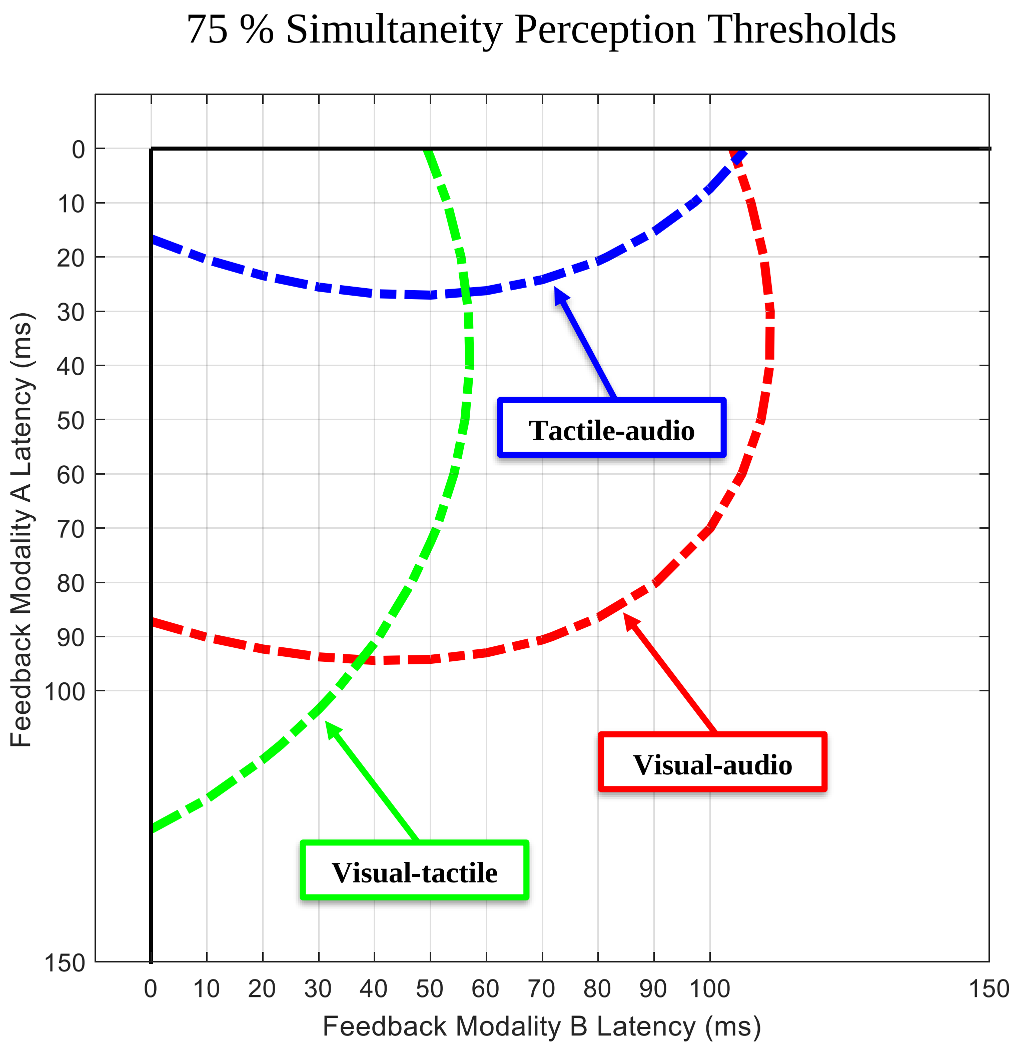


Figure 5-24: 75 % thresholds for simultaneity perception of touch and bimodal feedback.

Research Question 3 asked:

> **RQ3:** "How does the perceived quality of a virtual button change when latency between touch and feedback changes?

This questions also were answered in Chapter 4 when the feedback was unimodal. The experiment described in this chapter yielded further answers for this question when the feedback was bimodal. The answers were derived from the significance maps which showed the conditions that got statistically significantly higher quality scores than some others and did not differ significantly from each other. The area which was formed by these conditions was defined as the high quality region in the two dimensional latency space (Figure 5-25).

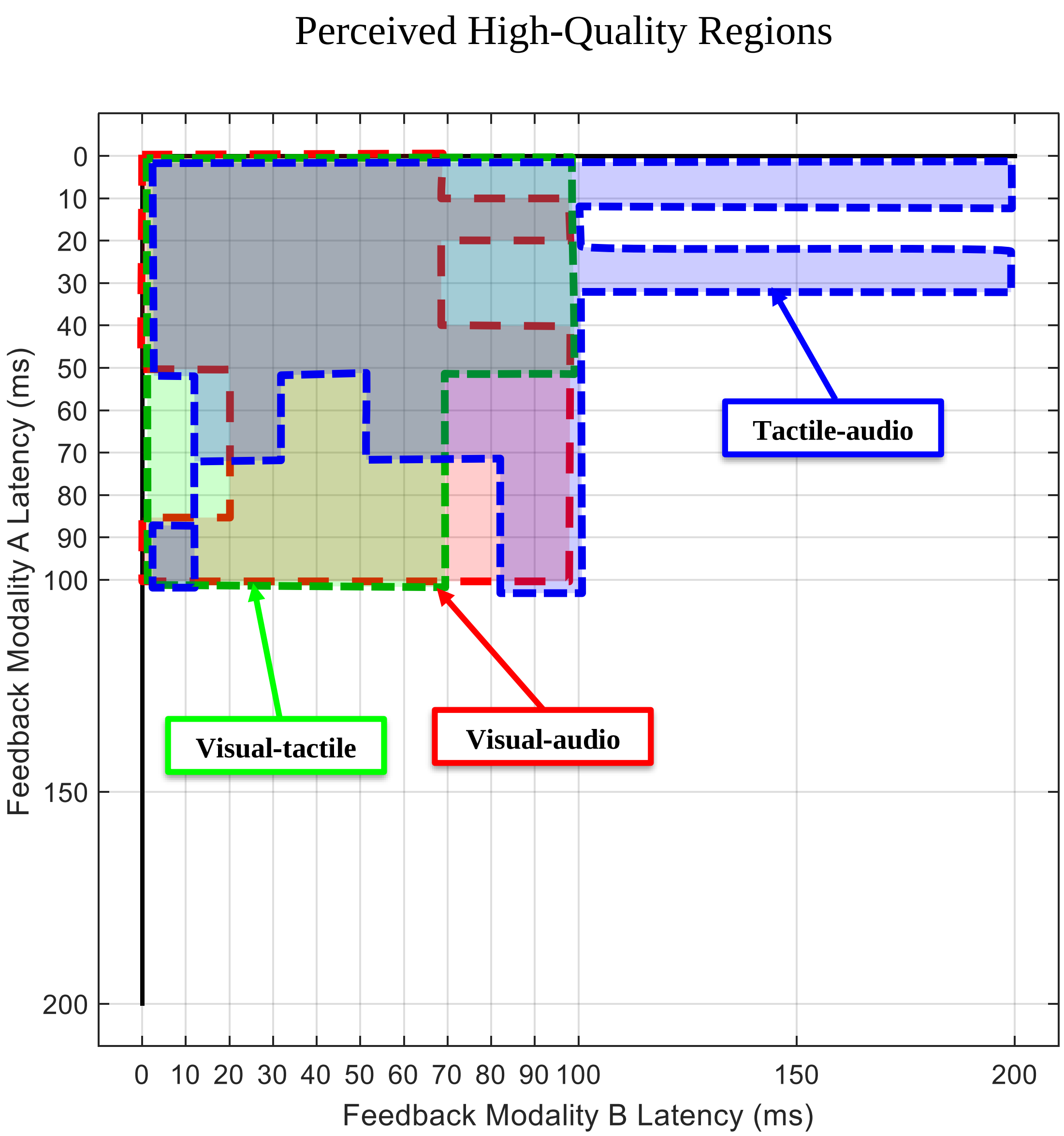


Figure 5-25: Perceived high-quality regions as a function of bimodal latency for different Modality Pairs. For clarity, the regions have been shifted from the origin (0 ms, 0 ms).

Although no correlation statistics were performed, these results suggested 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 non-simultaneous. Thus, the quality perception assessment results reinforced the simultaneity perception findings in this study.

Finally, the Research Question 4 asked:

> **RQ 4:** What are the latency guidelines recommended for visual, audio and tactile feedback for a virtual button press?

Guidelines for interaction designers to set the right requirements for touch and bimodal feedback were established for the first time. The guideline (rounded to the nearest 5 ms) derived was a simplified, easy-to-use combination of the both the simultaneity perception and quality perception results:

visual-audio, the visual feedback latency should not exceed 90 ms and audio feedback latency should be between 20 and 70 ms,

visual-tactile, the visual feedback latency should not exceed 100 ms and tactile feedback latency should not exceed 55 ms and

tactile-audio, the tactile feedback latency should not exceed 25 ms and audio feedback latency should not exceed 100 ms.

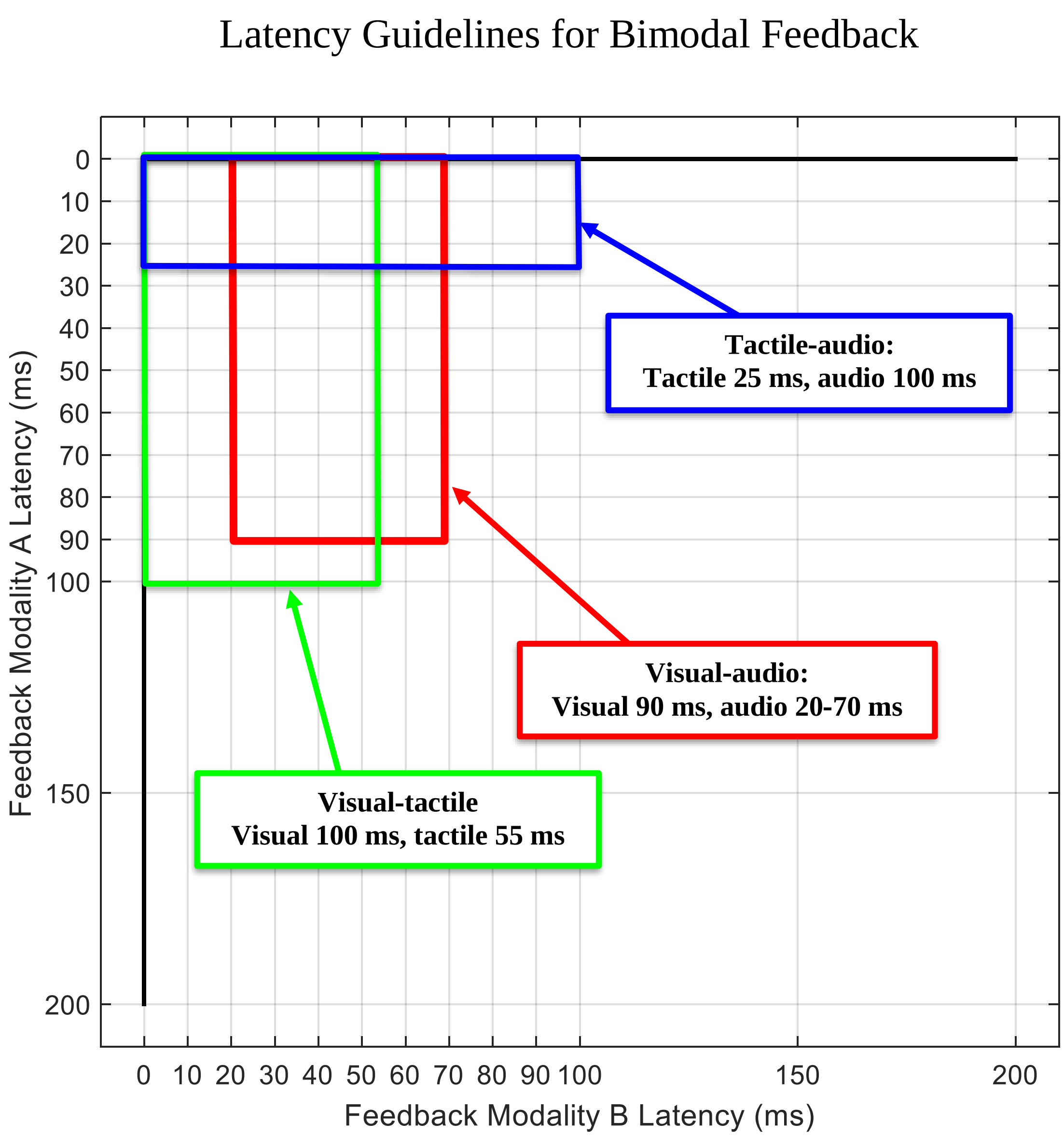


Figure 5-26: Latency Guidelines for bimodal feedback for different Modality Pairs.

These guidelines have a three-fold importance to the field. First, the guidelines established here take into account two feedback modalities experienced together instead of only one, which was the case for the guideline before this experiment. Second, hardware and software engineers do not need to optimize the latency between touch and feedback all the way to 0 ms since the users are tolerant to latencies. Third, 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. Based on the guidelines just established, the measurements showed that there is a lot of room for improvement in virtual keyboard latencies in contemporary mobile touchscreen phones.

Chapter 6 Discussion and Conclusions

6.1 Introduction

There are billions of touchscreen mobile phones that are used every day in the world. The most common interaction method in them is a virtual button press. Earlier research shows that latency in interaction is mainly harmful, but recommended latency between a touch and the virtual button feedback has been unknown. The aim of this research was to find touch-feedback simultaneity and quality perception thresholds for the visual, audio and tactile feedback modalities, and combine the results into guidelines for designers and engineers. The thesis statement read as follows:

It is not known what the recommended latency for virtual button feedback in touchscreens should be. An affordable latency measurement tool was built and two extensive experiments conducted to find out simultaneity and quality perception thresholds. From the thresholds, latency guidelines were derived and commercial touchscreen products were measured with the tool and validated against the guidelines.

The following Research Questions (RQ) have been addressed by the research in this thesis:

RQ1: Can an affordable and accurate touchscreen latency measurement tool be built?

RQ2: What are the touch-feedback simultaneity perception thresholds in virtual button interaction?

RQ3: How does the perceived quality of a virtual button change when latency between touch and feedback changes?

RQ 4: What are the latency guidelines recommended for visual, audio and tactile feedback for a virtual button press?

These Research Questions have been addressed by the literature review in Chapter 2, design and implementation of a measurement device in Chapter 3 and two extensive experiments in Chapter 4 and 5. A multidisciplinary approach was taken during the research to ensure the best possible results from different angles. The experiments were conducted using both psychophysical and human-computer interaction research paradigms using novel analysis methods. This chapter summarises the research in this thesis and discusses the findings in light of the research questions. Limitations of the research as well as the potential future work are also discussed. The chapter ends with a final conclusion. The answers for these Research Questions are discussed in the following section.

6.2 Research Question 1

RQ1: Can an affordable and accurate touchscreen latency measurement tool be built?

Research Question 1 was answered in Chapter 3. For the first time, a single and affordable tool for measuring touchscreen visual, audio and tactile feedback was built. Because it can be built entirely using inexpensive off-the-shelf components, universities and other research institutes can build it easily and at low-cost. It would enable the touchscreen interaction researchers to validate the latencies in their experiments, report them and take them into account in the results. Mobile phone manufacturers will benefit from the tool, because engineers will be able to validate the latencies when developing hardware and software for touchscreen phones. Because the tool is inexpensive, it can be copied wherever needed and, because it is also portable, it can be moved between locations easily.

Best practises from early work were adopted for the design of the multimodal latency measurement tool. A high-speed camera and a simple, but novel mirror construction were used for detecting a touch and recording its visual feedback to the same video stream. Having them both in the same stream made the latency measurement easy. The latency was measured conveniently by counting the high-speed video frames. A microphone and an analogue accelerometer connected to a 2-channel soundcard were used for touch detection, audio and tactile feedback recording. As with visual feedback, recording all these into the same audio stream made the measurement easier than separate streams. Connecting an analogue accelerometer straight to a soundcard was also novel, simple and useful innovation enabling inherent movement-sound transcoding. The audio and tactile latencies were easy to measure in a sound editor by selecting the area between the touch and the feedback.

The bill-of-materials was approximately 1000 € which can be considered as affordable compared to approximately 40,000 € for the OptoFidelity system. The tool was calibrated to validate it and to be accurate enough for human latency measurement. Sample measurements were conducted to show the tool in action and showed that it was possible to build an affordable and accurate touchscreen latency measurement tool.

6.3 Research Question 2

> **RQ2:** What are the touch-feedback simultaneity perception thresholds in virtual button interaction?

The experiments in both Chapter 4 and Chapter 5 answered this question. The experiment in Chapter 4 investigated the simultaneity perception of touch and unimodal feedback. For the first time, simultaneity perception of touch and unimodal visual, audio or tactile feedback was modelled with a univariate Gaussian function. This finding extended the earlier simultaneity perception research to practical application – finger touch and feedback. Based on the models, the 75 % thresholds of the simultaneity perception of touch and unimodal feedback were established for the first time: 85 ms when the feedback was visual, 80 ms when audio and 52 ms when tactile.

The experiment in Chapter 5 extended the simultaneity perception of touch and feedback to unexplored fields. For the first time, simultaneity perception of touch and *bimodal* feedback was investigated. A novel approach was taken into the analysis: the simultaneity perception of touch and bimodal feedback was modelled with a *bivariate* Gaussian function for all the feedback modality pairs. A successful modelling was a new finding as well. The 75 % thresholds were established for the first time: for each modality pair, they were the 75 % contours of each Gaussian model surface. Two-dimensional simultaneity thresholds have not been established before. The detailed contours were introduced as figures that are easy to interpret. The maximum values of the thresholds were

- for visual-audio feedback: 90 ms for visual and 115 ms for audio,
- for visual-tactile feedback: 125 ms for visual and 60 ms for tactile and

- for tactile-audio feedback: 25 ms for tactile and 110 ms for audio.

RQ2 was successfully answered by both experiments showing the 75 % thresholds for simultaneity perception for both unimodal and bimodal feedback.

6.4 Research Question 3

> **RQ3:** How does the perceived quality of a virtual button change when latency between touch and feedback changes?

The experiments in Chapter 4 and 5 also answered this question. In addition to simultaneity perception of touch and unimodal feedback, the experiment in Chapter 4 investigated the effect of latency on perceived quality when the feedback was unimodal, which has not been investigated before. The participants were asked to score the perceived quality of the buttons with different feedback latencies. The data was analysed with novel *significance maps*, developed in the research in this thesis. It was found for the first time that the latency affects the perceived quality of a virtual button: quality scores dropped significantly between latencies of 100 and 150 ms when the feedback was visual, and between 70 ms and 100 ms when the feedback was audio or tactile.

The effects of bimodal feedback latency on perceived quality has also been unexplored until now. They were investigated in the experiment described in Chapter 5. Participants were asked to score the perceived quality of the buttons with different latency combinations for feedback pairs. The significance maps invented during the analysis of Experiment 1 were developed further to make the analysis easier and the interpretation simpler. The results revealed for the first time that the perceived quality of a virtual button was affected by latency also when the feedback was bimodal. The new findings were: the quality was perceived significantly higher when the latency was between 0 ms and 100 ms for both modalities (with few exceptions) when the bimodal feedback pair was visual-audio or visual-tactile compared to the latencies above 100 ms. When the bimodal feedback pair was tactile-audio, the high quality region was less regular. However, when the latency was between 0 ms and 50 ms for tactile feedback and 0 ms and 100 ms for audio feedback, the buttons were rated as high quality.

These findings are remarkable for touchscreen phone manufacturers; it means that it is important to consider latency in design and engineering and the quality drop means that below these thresholds latency does not matter in the light of perceived quality working in favour of the engineers.

6.5 Research Question 4

RQ 4: What are the latency guidelines recommended for visual, audio and tactile feedback for a virtual button press?

The results from the experiments introduced in Chapter 4 and 5 were combined as a latency guideline for both unimodal and bimodal feedback. These guidelines were established for the first time. For unimodal feedback, the Point of Subjective Simultaneity (PSS) was selected as the minimum recommended latency. The maximum of the recommended latency was the 75 % simultaneity perception threshold or the latency condition where the perceived quality score had not yet dropped. The latency guideline for unimodal feedback established this way was:

visual feedback latency 30 – 85 ms,
audio feedback latency 20 – 70 ms and
tactile feedback latency should be 5 – 50 ms.

For bimodal feedback, the latency guideline, which combined the simultaneity perception and perceived quality results, was established for the first time. The guideline was the intersection of the area inside the 75% threshold contour and the high quality area with the carefully selected approximation:

For *visual-audio*, the visual feedback latency should not exceed
90 ms and the audio feedback latency should be between 20 and 70 ms,

For *visual-tactile*, the visual feedback latency should not exceed 100 ms and the tactile feedback latency should not exceed 55 ms and

For *tactile-audio*, the tactile feedback latency should not exceed 25 ms and the audio feedback latency should not exceed 100 ms.

Using these guidelines, designers and engineers will ensure that the majority of users perceive the touch and feedback as simultaneous and feel a good quality button under their finger, enhancing the user experience. Better user experience, in turn, means potentially more customers and sold devices benefiting the business of a mobile phone manufacturer.

6.6 Limitations and Future Work

The research in this thesis has been the first attempt to enable affordable latency measurements and understand simultaneity and quality perception in touchscreen virtual button interaction, and it sets the baseline for future research. Although the latency measurement tool worked as planned there are some limitations and room for improvements. The measurements could also be expanded into other contexts. In addition, the simultaneity and perceived quality research approaches were limited, although carefully considered. The limitations of the research in this thesis and also potential future work for the research are discussed next.

6.6.1 Multimodal Latency Measurement Tool

The latency measurement tool has two limitations in measurement involving tactile feedback. The first one can occur if tactile feedback latency is close to zero milliseconds, and the second when tactile and audio feedback latencies are close to each other. If tactile feedback latency is near zero, the recording of touch and tactile feedback might overlap which would make the measurement of tactile feedback difficult. Fortunately, these cases are rare, since the latencies in current mobile phones are not less than 30 ms (as seen in the measurements). If it happens, the workaround is to first record the touch only without tactile feedback and examine the length and waveform of the touch. In the second measurement round, the enabled tactile feedback can be recognized from the recording by comparing second measurement round to the first measurement round.

Because the accelerometer also picks up traces of audio feedback in addition to tactile, measuring tactile feedback can be challenging if the latency of audio and tactile feedback are close to each other. The workaround is similar to above: tactile feedback can be recorded first without audio and the waveform examined, then the measurement can be done with both feedback modalities and the examined tactile feedback can be found from the audio stream. Of course the latency of both feedback modalities could be recorded and measured separately, but to get the most accurate results, they must be played at the same time, because their latency can be different compared to the case when only one is played at a time. Measurement out of audio output connector is not recommended as discussed in Section 3.10.2 in Chapter 3.

The manual inspection of latencies was time consuming and a potential source of errors. Automatic recognition of touch and feedback would make the process quicker which, in turn, would benefit the business of a company. A semi-automatic process would make the measurement quicker and more accurate. The system could recognize the events and the user of the tool would only need to confirm them (and correct them if not properly recognized), for example. This approach would save the user searching for the start of the touch and feedback events. After confirmation of all the events, the system could automatically calculate mean latencies and standard deviations, which would again save time.

6.6.1.1 Measuring Latency in Other Touchscreen Interaction

As Saffer (2009) suggested, there are more touchscreen interaction methods in addition to a virtual button press. In touchscreen mobile phones, flick, pan, drag, pinch and spread are often used. As future work, it would be beneficial to expand the measurements to these interaction techniques to validate their latencies. The latencies would most probably differ from the virtual button press within the same device, since the software processes are different. The same applies on other touchscreen widgets, such as sliders and scrollbars, to see latency changes in more continuous interactions. With slight modifications to the setup, the tool could also be also used for latency measurements of whole device gestures and their responses.

6.6.1.2 Measuring Latency in Other Touchscreen Devices

As already shown by McAdam and Brewster (2011), latency measurements have already been conducted in a different context (a tabletop computer and a mobile phone). The tool could also be used in other contexts, such as bigger touchscreens found in tablets. The measurement process is almost identical to the mobile phone. The camera only needs to be raised to see the whole screen. Touchscreens in a vertical orientation could also be measured, for example in cars. The mirror should be then attached to either the side or bottom of the screen.

6.6.1.3 Measuring Latency in Other Devices

The high-speed camera can be effectively used to measure latency on many other manual input and visual output. For example, a key press on a normal PC keyboard and the screen output can be measured, if the camera is arranged in such manner that a finger and the computer screen are in the same picture.

6.6.2 Simultaneity perception

6.6.2.1 Individual Simultaneity Perception Modelling

In the experiments in this thesis, simultaneity perception was not modelled for individual participants as done usually in pure psychophysical experiments (as the focus was on a practical application). Future work in psychophysics should include experiments collecting more data per touch-feedback modality so that the simultaneity perception of each participant can be modelled, the threshold derived and statistics done.

Experiment 1 took one hour to complete and it consisted of 144 + 27 = 171 flash-press-response sequences (simultaneity + quality). If these were all the same unimodal feedback, the sample size should be sufficient for MLE modelling: 50 samples per model parameter is stated to be enough for MLE modelling (Hart and Clark, 1999); the univariate Gaussian consists of three parameters, equalling 150 samples. Increasing the number of samples reduces the confidence intervals though, making the model more reliable. If more samples were needed, the amount of button presses could be reduced from 7 to 5, for example. This would be needed if the modelling were done individually for bimodal feedback. The number of samples needed would be 300 (6 parameters x 50). It would be interesting to see the differences between different modality pairs and the distribution of thresholds in this kind of ecologically valid, but unexplored context.

6.6.3 Perceived quality

Perceived quality consists of many aspects, such as price, market share or brand name, for example (Wankhade and Dabade, 2010). According to Aaker (2009), in the case of a product, perceived quality factors are, for example, performance, features, reliability and fit and finish. These are important factors and should be taken in to account when the perceived quality of a product is considered. In this thesis, however, for the first time, the effect of latency on perceived quality was investigated and, for the first time, latency was found to be one of the aspects affecting on the perceived quality. Therefore, it has to be taken into account, alongside the other factors.

6.6.3.1 Lack of qualitative data

The judgment of perceived quality gave basic subjective opinion of the effect of latency on perceived quality. No other form of qualitative data was collected. Spontaneous comments were made by participants, however, which may indicate that more specific questions, or even interviews would give other information about the quality of the buttons when latency varies. The effect of latency could have been divided into subcategories as Kaaresoja *et al.* (2011a; 2011b) did (e.g. pleasantness, comfort). However, for the thesis, this would have led to an unacceptable duration for the experiments. Now that baseline measures have been established, further experiments could be undertaken to measure more of these qualitative issues.

6.6.4 Simultaneity and Quality Perception

6.6.4.1 Missing Mobile Context

In addition to home and work environments, mobile phones are often used in mobile contexts, such as when walking, sitting on a bus or even cycling. Headphones are also often used to listen to music. To set the baseline for the simultaneity and quality perception and further the guidelines, the experiments in this thesis were solely conducted in a laboratory context without headphones. Therefore, the effects of mobile context on simultaneity and perceived quality when latency changes remain unexplored. When mobile, the simultaneity perception threshold might go higher, since it is a detection threshold and detection needs attention. When the user is mobile, the attention is divided between several environmental aspects and might be shifted from the simultaneity detection (Spence *et al.*, 2001), causing the simultaneity threshold to increase. The same might happen to the perceived quality, shifting the quality drop for higher latencies. It would be important to know these effects, since the mobile setting is a very common use case for mobile phones. The future study might adopt methods and ideas e.g. from Brewster (2002), Hoggan *et al.* (2008a) and Koskinen (2008) who tested different virtual button feedback designs and modalities in mobile contexts.

6.6.4.2 Use of Simulator Instead of Real Mobile Phone

Simultaneity and quality perception were assessed with a simulator to allow the precise control of latency. The results were not validated with commercial touchscreen mobile phones. It is not yet known how the results will generalize to real touchscreen phones. However, the Virtual Button Simulator was designed to be as mobile-phone-like as possible as was the button-pressing task. There are several challenges when considering the validation of the results with a real touchscreen phone, which is why the Virtual Button Simulator was built. Firstly, latencies in phones are mostly above the perception thresholds and with few exceptions, current commercial phones cannot provide feedback with less than 50 ms latencies. Secondly, perceived quality might be affected by factors other than latency. One option would be to implement a special button application, which looked similar in all the phones and hid the design of the phones by covering them, for example.

Although conducted in a laboratory environment with a simulator, the research in this thesis sets the important baseline for the understanding of latency. These results could be compared to the real mobile phones when near zero latencies have been implemented in them.

6.6.4.3 Only one set of feedback designs

The feedback used in the experiments in this thesis was designed by using the best knowledge and practises from earlier work and current commercial virtual button design in touchscreen mobile phones. However, it remains unknown how the design of different feedback would affect simultaneity and quality perception when latency changes. There might be different effects depending on the feedback variable. Based on earlier research, feedback duration might not have an effect (Efron, 1970; Adelstein *et al.*, 2003a). Intensity, in turn, has been found to have an effect: when intensity is reduced, the stimulus is delayed (Efron, 1963), meaning that the simultaneity of touch and low intensity feedback would be perceived as non-simultaneous earlier when latency is varied. In the research in this thesis, the intensity was selected to be clearly perceivable, which is practical, and the feedback with lower intensity could be investigated in future work.

6.6.4.4 Only one type of cueing

Visual cueing guided the participants in the experiments. It is not known how the visual cueing affected simultaneity and quality perception. There might be issues with the visual feedback, for example, as that modality was used for both cuing and feedback. Therefore, different cueing methods could also be tested. In unimodal and bimodal feedback cases, the modality which is not used as feedback could be used for cueing. Another approach could be to name the two buttons of the Virtual Button Simulator as A and B and cue the participants by showing letter combinations, such as ABABBAB, on a computer screen. The participants should then memorize the letter sequence and press the buttons accordingly. However, although it would address the cueing problem, it would set another challenge, to memorize, which in turn could affect the simultaneity perception by shifting the attention.

The cueing used in the experiments in this thesis was carefully designed to be the most realistic possible, given the latency requirements, to simulate cognitive load and keep the pace so that the experiment duration would be reasonable and approximately the same for all the participants.

6.6.4.5 Trimodal feedback

The natural continuation of the research in this thesis is to experiment with trimodal feedback (visual, audio, tactile). The results would further extend the latency guidelines by finding the thresholds for simultaneity perception and perceived quality. Testing trimodal combinations of feedback would be valuable since virtual buttons in mobile phones often include all three modalities. The simultaneity perception or the effects of latency on perceived quality of a touch and trimodal feedback has never been explored before. This would allow designers to see all of the possible trade-offs between the different modalities when designing touchscreen button interactions. It would be a simple task to modify the Virtual Button Simulator software to provide three feedback modalities and vary their latencies. A simultaneity perception study with trimodal feedback would also be a new opening in psychophysics.

It could be hypothesised that the simultaneity perception of touch and trimodal feedback would follow a trivariate Gaussian distribution. The probability density function P would be a function of the scaling factor a, the means μ, standard deviations σ and correlations ρ (Stuart and Ord, 1994; Rose and Smith, 1996; Rose and Smith, 2002; Weisstein, 2015):

$$P(LAGx, LAGy, LAGz) = f(a, \mu_x, \mu_y, \mu_z, \sigma_x, \sigma_y, \sigma_z, \rho_{xy}, \rho_{xz}, \rho_{yz})$$

As we can see, the trivariate Gaussian has 10 variables. Therefore, the number of samples needed for successful MLE modelling would be: 10 x 50 = 500. To keep the experiment duration reasonable for one participant, six latency levels could be used (e.g. 0, 20, 70, 100, 200 and 300 ms, leading to 216 Latency Triplets. If the number of flashes and presses would be reduced from 7 to 5, the experiment would take approximately 1 hour including a short training period. According to the literature, only 3 participants would be enough to achieve a big enough sample size (3 x 216 = 648 > 500), but to achieve sound results and a reliable model, at least 12 participants should be used. This number would also be needed for the perceived quality assessment.

The trivariate Gaussian surface is four-dimensional, meaning that it would be challenging to visualise. Fortunately, the contours of equal probability would be three-dimensional ellipsoids (see example in Figure 6-1) (Shea, 2015). As before, feedback would always come after the touch, meaning that all the latencies would be positive ($LAGx > 0, LAGy > 0, LAGz > 0$). Therefore, the contours would be $\frac{1}{8}$ of the ellipsoid (as in the bimodal case, where the bivariate Gaussian models were $\frac{1}{4}$ of the full bivariate Gaussian surface.

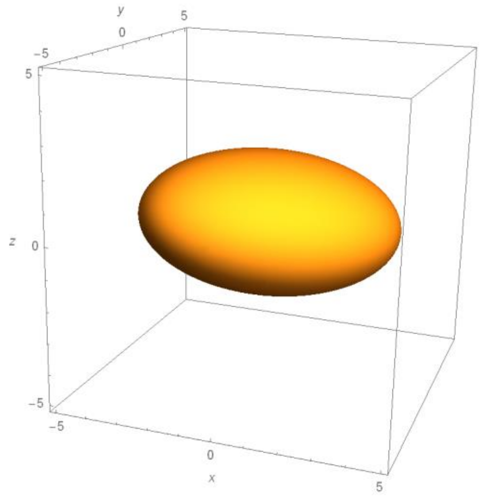


Figure 6-1: A contour of a trivariate Gaussian. It is a three-dimensional ellipsoid (Shea, 2015).

The effect of latency on perceived quality when the feedback is trimodal could be explored in a similar manner to the bimodal case. The significance maps would be three-dimensional and would need to be modified to be visualised reasonably. For example, the high-quality area could be shown as green to have an idea of the high-quality threshold. The details of the significance maps could be shown on-demand.

6.6.4.6 Touch and release feedback

The audio and tactile feedback were provided on touch down in the experiments in this thesis, rather than both on touch down and release. However, physical buttons also give release feedback (after being pressed). If virtual buttons also had release feedback, the button might feel more like a real button. The results by Kaaresoja *et al.* (2011a) hint that if both feedback types are given, the users might make less errors and score the user experience high even in higher latencies, meaning they better tolerate latencies. However, in mobile phones, which were a focus in this thesis, the feedback is given on touch down only. Therefore, the touch down was selected as a feedback type in the studies in this thesis.

6.7 Conclusions

The research in this thesis has investigated multimodal latency in mobile touchscreen virtual button interaction. For the first time, an affordable, but accurate tool was built to measure all three feedback latencies in touchscreens. For the first time, simultaneity perception of touch and feedback as well as the effect of latency on virtual button perceived quality has been studied and thresholds found for both unimodal and bimodal feedback. The results from these studies were combined as latency guidelines for the first time. These guidelines enable interaction designers to establish requirements for mobile phone engineers to optimise the latencies on the right level.

The measurement tool built was capable of accurately measuring latencies of all three feedback modalities. The novel contribution of this thesis is the price tag, which was approximately 1000 €. It can be considered as affordable for a multi-functioning measurement device. In addition, the tool was made mostly with off-the-shelf components (today only off-the-shelf components could be used) and, as a bonus, it was portable. Therefore, it could be copied at low cost or moved wherever needed. The tool enables touchscreen interaction designers to validate latencies in their experiments, making their results more valuable and accurate. The tool could benefit the touchscreen phone manufacturers, since it enables engineers to validate latencies during development of mobile phones. That gives them valuable information about the quality of the product, which of course should be the best possible in the end.

For the first time, latency guidelines for touchscreen unimodal and bimodal feedback were established. It is important to have separate guidelines for both cases, since they differ from each other. Applied psychophysical and human-computer interaction methods were used to obtain simultaneity and quality perception thresholds for the first time. Perceived quality results were presented in a novel way as significance maps. The guidelines will help engineers and interaction designers to select and optimise latencies to be low enough, but not too low. Designers using these guidelines will make sure that most of the users will both perceive the feedback as simultaneous with their touch and experience high quality virtual buttons. This in turn will enhance user experience, which will reflect on the quality of the product. A better product will potentially mean more customers which of course works in favour of the business.

The results from this thesis show that latency has a remarkable effect on touchscreen virtual buttons, and it is a key part of virtual button feedback design. The novel results enable researchers, designers and engineers to master the effect of latencies in research and development. This will lead to more accurate and reliable research results and help mobile phone manufacturers make better products.

Appendices

To save on paper, the Appendices are available at:
http://dx.doi.org/10.5525/gla.researchdata.220

A. Experimental files for Experiment 1

B. Experimental files for Experiment 2

References

Aaker, D. A. (2009). Managing brand equity, Simon and Schuster.

Adelstein, B. D., Begault, D. R., Anderson, M. R. and Wenzel, E. M. (2003a). Sensitivity to haptic-audio asynchrony. *5th International Conference on Multimodal Interfaces*. Vancouver, Canada, ACM**:** 73-76.

Adelstein, B. D., Lee, T. G. and Ellis, S. R. (2003b). Head tracking latency in virtual environments: psychophysics and a model. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, SAGE Publications.

Allison, R. S., Harris, L. R., Jenkin, M., Jasiobedzka, U. and Zacher, J. E. (2001). Tolerance of temporal delay in virtual environments. *Virtual Reality, 2001. Proceedings. IEEE*, IEEE.

Altinsoy, M. E. (2003). Perceptual aspects of auditory-tactile asynchrony. *Proceedings of the Tenth International Congress on Sound and Vibration*, Citeseer.

Altinsoy, M. E. and Merchel, S. (2009). Audiotactile feedback design for touch screens. *Haptic and audio interaction design*, Springer**:** 136-144.

Anderson, G., Doherty, R. and Ganapathy, S. (2011). User Perception of Touch Screen Latency. *Design, User Experience, and Usability. Theory, Methods, Tools and Practice*, Springer**:** 195-202.

Android (2015). Touch Feedback. *Android Design in Android 5.0, http://developer.android.com/design/style/touch-feedback.html*.

Annett, M., Ng, A., Dietz, P., Bischof, W. F. and Gupta, A. (2014). How low should we go?: understanding the perception of latency while inking. *Proceedings of the 2014 Graphics Interface Conference*, Canadian Information Processing Society.

Apple (2015). Bar Button Icons. *iOS Human Interface Guidelines, https://developer.apple.com/library/ios/documentation/UserExperience/Conceptual/MobileHIG/BarIcons.html*.

Ash, R. B. (2013). Lectures on Statistics (Lecture 21). *http://www.math.uiuc.edu/~r-ash/Stat/StatLec21-25.pdf*. Dept. of Mathematics, University of Illinois, IL, USA.

Begault, D. R., Adelstein, B. D., McClain, B. U. and Anderson, M. R. (2005). Thresholds for auditory-tactile asynchrony. *Audio Engineering Society Convention*. Barcelona, Spain, Audio Engineering Society.

Bender, G. T. (1999). Touch screen performance as a function of the duration of auditory feedback and target size, Wichita State University.

Bennion, S. I., Creager, J. D. and VanHouten, R. D. (1981). "Touch sensitive graphics terminal applied to process control." *SIGGRAPH Comput. Graph.* 15(4): 342-350.

Bérard, F. and Blanch, R. (2013). Two touch system latency estimators: high accuracy and low overhead. *Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces*, ACM.

Blattner, M. M., Sumikawa, D. A. and Greenberg, R. M. (1989). "Earcons and icons: Their structure and common design principles." *Human–Computer Interaction* 4(1): 11-44.

Boring, E. (1923). A History of Experimental Psychology. New York, Pendragon.

Brewster, S. (2002). "Overcoming the Lack of Screen Space on Mobile Computers." *Personal and Ubiquitous Computing* 6(3): 188-205.

Brewster, S. and Brown, L. M. (2004). Tactons: structured tactile messages for non-visual information display. *Proceedings of the fifth conference on Australasian user interface-Volume 28*, Australian Computer Society, Inc.

Brewster, S., Chohan, F. and Brown, L. (2007). Tactile Feedback for Mobile Interactions. *CHI2007*, ACM Press.

Carter, E. J., Sharan, L., Trutoiu, L., Matthews, I. and Hodgins, J. K. (2010). "Perceptually motivated guidelines for voice synchronization in film." *ACM Transactions on Applied Perception (TAP)* 7(4): 23.

Carvalho, P. V. R., Gomes, J. O. and Borges, M. R. S. (2011). Human centered design for nuclear power plant control room modernization. *CEUR Proceedings 4th Workshop HCP Human Centered Processes*. Genoa, Italy.

Charwat, H. J. (1992). Lexikon der Mensch-Maschine-Kommunikation. Oldenbourg.

Chase, R., Harvey, S., Standfast, S., Rapin, I. and Sutton, S. (1961). "Studies on sensory feedback I: Effect of delayed auditory feedback on speech and keytapping. ." *Quarterly Journal of Experimental Psychology* 13: 141-152.

Cohen, A. (1995). "A description of MIL-STK-1472." *ACM SIGCHI Bulletin* 27(2): 48-49.

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

Dahl, S. and Bresin, R. (2001). Is the player more influenced by the auditory than the tactile feedback from the instrument. *Proc. of the COST-G6 Workshop on Digital Audio Effects (DAFx-01), Limerick*, Citeseer.

Defence, D. o. (2012). Military Standard: Human Engineering Design Criteria for Military Systems, Equipment and Facilities. *MIL-STD 1472G*. Washington, Department of Defence.

Deron, M. (2000). "How important is visual feedback when using a touch screen?" *Usability news* 2(1).

Di Luca, M. (2010). "New method to measure end-to-end delay of virtual reality." *Presence* 19(6): 569-584.

Dixon, N. F. and Spitz, L. (1980). "The detection of auditory visual desynchrony." *Perception* 9(6): 719-721.

Efron, R. (1963). "The Effect of Stimulus Intensity on the Perception of Simultaneity in Right- and Left-Handed Subjects." *Brain* 86(2): 285-294.

Efron, R. (1970). "Effect of stimulus duration on perceptual onset and offset latencies." *Perception & Psychophysics* 8(4): 231-234.

Ellis, S. R., Adelstein, B. D., Baumeler, S., Jense, G. and Jacoby, R. H. (1999a). Sensor spatial distortion, visual latency, and update rate effects on 3D tracking in virtual environments. *Virtual Reality, 1999. Proceedings., IEEE*, IEEE.

Ellis, S. R., Breant, F., Manges, B., Jacoby, R. and Adelstein, B. D. (1997). Factors influencing operator interaction with virtual objects viewed via head-mounted see-through displays: viewing conditions and rendering latency. *Virtual Reality Annual International Symposium, 1997., IEEE 1997*, IEEE.

Ellis, S. R., Mania, K., Adelstein, B. D. and Hill, M. I. (2004). Generalizeability of latency detection in a variety of virtual environments. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, SAGE Publications.

Ellis, S. R., Young, M. J., Adelstein, B. D. and Ehrlich, S. M. (1999b). Discrimination of changes in latency during head movement. *HCI (2)*.

Ellis, S. R., Young, M. J., Adelstein, B. D. and Ehrlich, S. M. (1999c). Discrimination of changes of latency during voluntary hand movement of virtual objects. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, SAGE Publications.

Exner, S. (1875). "Experimentelle Untersuchung der einfachsten psychischen Processe. ." *Archiv für die gesamte Physiologie des Menschen und der Tiere* 11(1): 403-432.

Fairbanks, G. and Guttman, N. (1958). "Effects of delayed auditory feedback upon articulation." *Journal of Speech & Hearing Research*.

Finney, S. A. (1997). "Auditory feedback and musical keyboard performance." *Music Perception*: 153-174.

Freed, A., Chaudhary, A. and Davila, B. (1997). Operating Systems Latency Measurement and Analysis for Sound Synthesis and Processing Applications. *International Computer Music Conference*, Thessaloniki, Creece, ICMA (International Computer Music Association).

Fujisaki, W. and Nishida, S. y. (2009). "Audio–tactile superiority over visuo–tactile and audio–visual combinations in the temporal resolution of synchrony perception." *Experimental brain research* 198(2-3): 245-259.

Fukumoto, M. and Sugimura, T. (2001). Active Click: Tactile Feedback for Touch Panels. *CHI2001*, ACM.

Harris, L. R., Harrar, V., Jaekl, P. and Kopinska, A. (2010). Mechanisms of simultaneity constancy. *Space and Time in Perception and Action*. R. Nijhawan. Cambridge, UK, Cambridge University Press**:** 232-253.

Hart, R. A. and Clark, D. H. (1999). Does size matter? Exploring the small sample properties of maximum likelihood estimation. *Annual Meeting of the Midwest Political Science Association*.

He, D., Liu, F., Pape, D., Dawe, G. and Sandin, D. (2000). Video-Based Measurement of System Latency. *IPT2000, International Immersive Projection Technology Workshop*. Ames IA, USA.

Herndon, N. (2008). ATTITUDE: A Tidy Touchscreen Interface to a UML Development Environment, ProQuest.

Heron, J., Hanson, J. V. and Whitaker, D. (2009). "Effect before cause: supramodal recalibration of sensorimotor timing." *PLoS One* 4(11): e7681.

Hinckley, K. and Wigdor, D. (2012). Input Technologies and Techniques. *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, Third Edition*. J. A. Jacko, CRC Press, Taylor & Francis Group.

Hoggan, E. and Brewster, S. (2007). Designing Audio and Tactile Crossmodal Icons for Mobile Devices. *ICMI2007*, Nagoya, Japan, ACM.

Hoggan, E., Brewster, S. A. and Johnston, J. (2008a). Investigating the Effectiveness of Tactile Feedback for Mobile Touchscreens. *CHI 2008*. Florence, Italy.

Hoggan, E., Crossan, A., Brewster, S. and Kaaresoja, T. (2009). Audio or Tactile Feedback: Which Modality When? *CHI'09*, Boston, USA, ACM.

Hoggan, E., Kaaresoja, T., Laitinen, P. and Brewster, S. (2008b). Cross-Modal Congruence:The Look, Feel and Sound of Touchscreen Widgets. *ICMI2008*, Crete, Greece.

Hollander, M., Wolfe, D. A. and Chicken, E. (2013). Nonparametric statistical methods, John Wiley & Sons.

Instruments, T. (2004). "L293, L293D QUADRUPLE HALF-H DRIVERS." *Datasheet*.

ISO (2007). 9241 Ergonomics of human-system interaction. *Part 400: Principles and requirements for physical input devices*, British Standard BS EN ISO 9241-400:2007.

Jammer, M. (2006). Concepts of simultaneity: From antiquity to Einstein and beyond, JHU Press.

Jay, C. and Hubbold, R. (2005). Delayed Visual and Haptic Feedback in a Reciprocal Tapping Task. *World Haptics*, Pisa, IEEE.

Jota, R., Ng, A., Dietz, P. and Widgor, D. (2013). How fast is fast enough?: a study of the effects of latency in direct-touch pointing tasks. *CHI'13*. Paris, France, ACM**:** 2291-2300.

Kaaresoja, T., Anttila, E. and Hoggan, E. (2011a). The Effect of Tactile Feedback Latency in Touchscreen Interaction. *World Haptics 2011*. Istanbul, Turkey, IEEE.

Kaaresoja, T. and Brewster, S. (2010). Feedback is… Late: Measuring Multimodal Delays in Mobile Device Touchscreen Interaction. *ICMI'2010*, Beijing, China, ACM.

Kaaresoja, T., Brewster, S. and Lantz, V. (2014). "Towards the Temporally Perfect Virtual Button: Touch-Feedback Simultaneity and Perceived Quality in Mobile Touchscreen Press Interactions." *ACM Trans. Appl. Percept.* 11(2): 1-25.

Kaaresoja, T., Brown, L. M. and Linjama, J. (2006). Snap-Crackle-Pop: Tactile Feedback for Mobile Touch Screens. *Eurohaptics 2006*. Paris, France**:** 565-566.

Kaaresoja, T., Hoggan, E. and Anttila, E. (2011b). Playing with Tactile Feedback Latency in Touchscreen Interaction: Two Approaches. *Interact 2011*, Lisbon, Portugal, Springer-Verlag.

Koskinen, E. (2008). Optimizing Tactile Feedback for Virtual Buttons in Mobile Devices Master's Thesis, Helsinki University of Technology.

Koskinen, E., Kaaresoja, T. and Laitinen, P. (2008). Feel-Good Touch: Finding the Most Pleasant Tactile Feedback for a Mobile Touch Screen Button. *ICMI2008*, Crete, Greece, ACM.

Laitinen, P. and Mäenpää, J. (2006). Enabling mobile haptic design: Piezoelectric actuator technology properties in hand held devices. *HAVE'06 (Haptic Audio Visual Environments)*. Ottawa, Canada.

Lee, J.-H., Poliakoff, E. and Spence, C. (2009). The effect of multimodal feedback presented via a touch screen on the performance of older adults. *Haptic and Audio Interaction Design*, Springer**:** 128-135.

Lee, S. C. and Zhai, S. (2009). The Performance of Touch Screen Soft Buttons. *CHI'09*, Boston, USA, ACM.

Lehtosalo, A. (2009). Haptics Technologies and Electromechanical Latency Measurements in Tactile Feedback System. Master of Science Master's Thesis, Tampere University of Technology.

Levitin, D. J., MacLean, K., Mathews, M. and Chu, L. (1999). The Perception of Cross-Modal Simultaneity. Invited paper. . *3rd International Conference on Computing Anticipatory Systems*, Liège, Belgium.

Liang, J., Shaw, C. and Green, M. (1991). On Temporal-Spatial Realism in the Virtual Reality Environment. *UIST'91, Symposium of User Interface Software and Technology*, Hilton Head, South Carolina, US.

MacKenzie, S. and Ware, C. (1993). Lag as a Determinant of Human Performance In Interactive Systems. *CHI'93*. Amsterdam, the Netherlands, ACM.

MacMillan, K., Droettboom, M. and Fujinaga, I. (2001). Audio Latency Measurements of Desktop Operating Systems. *International Computer Music Conference*, Habana, Cuba, ICMA (International Computer Music Association).

McAdam, C. and Brewster, S. (2011). Using mobile phones to interact with tabletop computers. *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*. Kobe, Japan, ACM**:** 232-241.

Meehan, M., Razzaque, S., Whitton, M. C. and Brooks Jr, F. P. (2003). Effect of latency on presence in stressful virtual environments. *Virtual Reality, 2003. Proceedings. IEEE*, IEEE.

Merriam-Webster (2015). Touch. *http://www.merriam-webster.com/dictionary/touch*.

Miall, R. C. and Jackson, J. K. (2006). "Adaptation to visual feedback delays in manual tracking: evidence against the Smith Predictor model of human visually guided action." *Experimental Brain Research*.

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

Miller, D. and Bishop, G. (2002). Latency meter: a device for easily monitoring VE delay. *Stereoscopic Displays and Virtual Reality Systems IX*. San Jose, CA, USA.

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

Miller, R. B. (1968). Response time in man-computer conversational transactions. *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, ACM.

Montag, M., Sullivan, S., Dickey, S. and Leider, C. (2011). A Low-Cost, Low-Latency Multi-Touch Table with Haptic Feedback for Musical Applications. *Proc. of Int. Conf. on New Interfaces for Musical Expression*, Citeseer.

Nashel, A. and Razzaque, S. (2003). Tactile Virtual Buttons for Mobile Devices. *CHI2003*. Ft. Lauderdale, Florida, USA, ACM Press**:** 854-855.

Nelson, M. and Thom, B. (2004). Interactive MIDI: Real-time Performance Evaluation. *New Interfaces for Musical Expression*, Hamamatsu, Japan.

Ng, A., Annett, M., Dietz, P., Gupta, A. and Bischof, W. F. (2014). In the blink of an eye: investigating latency perception during stylus interaction. *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, ACM.

Ng, A., Lepinski, J., Widgor, D., Sanders, S. and Dietz, P. (2012). Designing for low-latency direct-touch input. *UIST'12*. St Andrews, UK, ACM**:** 453-464

Nichols, S. J. V. (2007). "New Interfaces at the Touch of a Fingertip." *Computer* 40(8): 12-15.

OptoFidelity (2014). OptoFidelity WatchDog Automatic User Experience Tester.

Paek, T., Chang, K., Almog, I., Badger, E. and Sengupta, T. (2010). A practical examination of multimodal feedback and guidance signals for mobile touchscreen keyboards. *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, ACM.

Pakkanen, T., Raisamo, R., Raisamo, J., Salminen, K. and Surakka, V. (2010). Comparison of three designs for haptic button edges on touchscreens. *Haptics Symposium, 2010 IEEE*, IEEE.

Park, G., Choi, S., Hwang, K., Kim, S., Sa, J. and Joung, M. (2011). Tactile effect design and evaluation for virtual buttons on a mobile device touchscreen. *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, ACM.

Pavlovych, A. and Stuerzlinger, W. (2009). The tradeoff between spatial jitter and latency in pointing tasks. *Proceedings of the 1st ACM SIGCHI symposium on Engineering interactive computing systems*. Pittsburgh, PA, USA, ACM**:** 187-196.

Pesqueux, L. and Rouaud, M. (2005). Vibration level of mobile phones' silent alerts Master's Thesis, Aalborg University.

Poupyrev, I. and Maruyama, S. (2003). Tactile interfaces for small touch screens. *UIST 2003*, ACM**:** 217-220.

Power, S. E. (2011). Temporal illusions—philosophical considerations. *Multidisciplinary Aspects of Time and Time Perception*, Springer**:** 11-35.

Rao, S. (2012). "High-definition haptics: Feel the difference!" *Analog Applications (Texas Instruments)* 2012(3Q): 29-32.

Rohde, M. and Ernst, M. O. (2013). "To lead and to lag–forward and backward recalibration of perceived visuo-motor simultaneity." *Frontiers in psychology* 3.

Rose, C. and Smith, M. (2002). "The Trivariate Normal." *Mathematical Statistics with Mathematica, New York: Springer-Verlag*: 226-228.

Rose, C. and Smith, M. D. (1996). "The multivariate normal distribution." *Mathematica Journal* 6(1): 32-37.

Saffer, D. (2009). Designing Gestural Interfaces. Sebastopol, Canada, O'Reilly.

Schomaker, L., Nijtmans, J., Camurri, A., Lavagetto, F., Morasso, P., Benoît, C., Guiard-Marigny, T., Goff, B. L., Robert-Ribes, J., Adjoudani, A., Defée, I., Münch, S., Hartung, K. and Blauert, J. (1995). A Taxonomy of Multimodal Interaction in the Human Information Processing System, A Report of the ESPRIT PROJECT 8579.

Sears, A. (1991). "Improving touchscreen keyboards: design issues and a comparison with other devices." *Interacting with computers* 3(3): 253-269.

Sears, A., Plaisant, C. and Shneiderman, B. (1991). "A new era for touchscreen applications: High precision, dragging icons, and refined feedback." *Advances in Human-Computer Interaction* 3.

Seow, S. C. (2008). Designing and engineering time: the psychology of time perception in software, Addison-Wesley Professional.

Shea, J. M. (2015). "Joint Density of Trivariate Gaussian Random Variables." from the Wolfram Demonstrations Project http://demonstrations.wolfram.com/JointDensityOfTrivariateGaussianRandomVariables/. Published under http://creativecommons.org/licenses/by-nc-sa/3.0.

Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. *Visual Languages, 1996. Proceedings., IEEE Symposium on*, IEEE.

So, R. H. and Chung, G. K. (2005). Sensory motor responses in virtual environments: Studying the effects of image latencies for target-directed hand movement. *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, IEEE.

Speckert, G., McGoldrick, P., Nowell, D. and Wade, J. (1979). Human interface to a major fusion experiment, a graphics solution, California Univ., Berkeley (USA); Lawrence Livermore National Lab., CA (USA).

Spence, C. (2009). "Explaining the Colavita visual dominance effect." *Progress in Brain Research* 176: 245-258.

Spence, C., Shore, D. I. and Klein, R. M. (2001). "Multisensory prior entry." *Journal of Experimental Psychology Gen.* 130: 799-832.

Steed, A. (2008). A simple method for estimating the latency of interactive, real-time graphics simulations. *Proceedings of the 2008 ACM symposium on Virtual reality software and technology*, ACM.

Stone, J. V., Hunkin, N. M., Porrill, J., Wood, R., Keeler, V., Beanland, M., Port, M. and Porter, N. R. (2001). "When is now? Perception of Simultaneity." *Proceedings of the Royal Society of London: Series B*(268): 31-38.

Stuart, A. and Ord, J. K. (1994). "Kendall's Advanced Theory of Statistics, Volume I: Distribution Theory." *Arnold, London*.

Sugano, Y., Keetels, M. and Vroomen, J. (2010). "Adaptation to motor-visual and motor-auditory temporal lags transfer across modalities." *Experimental brain research* 201(3): 393-399.

Teather, R. J., Pavlovych, A., Stuerzlinger, W. and MacKenzie, I. S. (2009). Effects of tracking technology, latency, and spatial jitter on object movement. *3D User Interfaces, 2009. 3DUI 2009. IEEE Symposium on*, IEEE.

Tikka, V. and Laitinen, P. (2006). Designing Haptic Feedback for Touch Display: Experimental Study of Perceived Intensity and Integration of Haptic and Audio. *Workshop on Haptic and Audio Interaction Design*, Glasgow, Scotland, UK.

Tsai, W.-C. and Lee, C.-F. (2009). A study on the icon feedback types of small touch screen for the elderly. *Universal Access in Human-Computer Interaction. Intelligent and Ubiquitous Interaction Environments*, Springer**:** 422-431.

Tukey, J. W. (1977). Exploratory Data Analysis, Addison-Wesley.

Valk, M. A. (1985). An experiment to study touchscreen "button" design. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, SAGE Publications.

Walker, P. M. B., Ed. (1995). Dictionary of Science and Technology. New York, Larousse.

van Wassenhove, V., Grant, K. W. and Poeppel, D. (2007). "Temporal window of integration in auditory-visual speech perception." *Neuropsychologia* 45(3): 598-607.

van Vugt, F. T. and Tillmann, B. (2014). "Thresholds of auditory-motor coupling measured with a simple task in musicians and non-musicians: was the sound simultaneous to the key press?" *PloS one* 9(2): e87176.

Wankhade, L. and Dabade, B. (2010). Quality uncertainty and perception: information asymmetry and management of quality uncertainty and quality perception, Springer Science & Business Media.

Weisstein, E. W. (2015). "Trivariate Normal Distribution." MathWorld--A Wolfram Web Resource. http://mathworld.wolfram.com/TrivariateNormalDistribution.html.

Windows (2015). Guidelines for buttons. *Windows Dev Center, https://msdn.microsoft.com/en-us/library/windows/apps/xaml/hh465470.aspx*.

Winter, R., Harrar, V., Gozdzik, M. and Harris, L. R. (2008). "The relative timing of active and passive touch." *Brain Research* 1242(Special Issue: Multisensory Integration): 54-58.

Vitense, H. S., Jacko, J. A. and Emery, V. K. (2003). "Multimodal feedback: an assessment of performance and mental workload." *Ergonomics* 46(1-3): 68-87.

Vogels, I. M. L. C. (2004). "Detection of temporal delays in visual-haptic interfaces." *Human Factors* 46(1): 118-134.

Wright, J. and Brandt, E. (2001). System-level MIDI performance testing. *Proc. 2001 Intl. Computer Music Conf.(ICMC-01).*

Wright, M., Cassidy, R. J. and Zbyszynski, M. F. (2004). Audio and Gesture Latency Measurements on Linux and OSX. *International Computer Music Conference* Miami, FL, USA**:** 423-429.

Wu, S. and Yi, D. (2013). Touch Panel Industry & Cost Analysis Report - H2 2013, IHS Electronics & Media**:** 273.

Yates, A. J. (1963). "Delayed auditory feedback." *Psychological Bulletin* 60(3): 213.

Zampini, M., Guest, S., Shore, D. I. and Spence, C. (2005). "Audio–visual simultaneity judgments." *Perception & Psychophysics* 67(3): 531-544.