Mobile Gaze Interaction: Gaze Gestures with Haptic Feedback

Akkil Deepak

University of Tampere
School of Information Sciences
Human Technology Interaction
M.Sc. thesis
Supervisor: Jari Kangas
December 2013

University of Tampere
School of Information Sciences
Computer Science / Interactive Technology
Forename Surname: Akkil Deepak
M.Sc. thesis, 55 pages, 5 index and 3 appendix pages
December 2013

There has been an increasing need for alternate interaction techniques to support mobile usage context. Gaze tracking technology is anticipated to soon appear in commercial mobile devices. There are two important considerations when designing mobile gaze interactions. Firstly, the interaction should be robust to accuracy problems. Secondly, user feedback should be instantaneous, meaningful and appropriate to ease the interaction. This thesis proposes gaze gesture input with haptic feedback as an interaction technique in the mobile context.

This work presents the results of an experiment that was conducted to understand the effectiveness of vibrotactile feedback in two stroke gaze gesture based mobile interaction and to find the best temporal point in terms of gesture progression to provide the feedback. Four feedback conditions were used, NO (no tactile feedback), OUT (tactile feedback at the end of first stroke), FULL (tactile feedback at the end of second stroke) and BOTH (tactile feedback at the end of first and second strokes). The results suggest that haptic feedback does help the interaction. The participants completed the tasks with fewer errors when haptic feedback was provided. The feedback conditions OUT and BOTH were found to be equally effective in terms of task completion time. The participants also subjectively rated these feedback conditions as being more comfortable and easier to use than FULL and NO feedback conditions.

Keywords: Mobile gaze interaction, Mobile HCI, Gaze gestures, Haptic feedback,

Acknowledgements

Many people have contributed to the progress and completion of this work. I would like to take this opportunity to acknowledge their contribution.

First and foremost I would like express my sincere gratitude to Dr. Jari Kangas, my supervisor, for his patient guidance from the start of this work till the end. Working on this thesis has been a significant academic challenge for me. I am convinced that without his constant support and encouragement, I could not have completed this thesis successfully.

I would like to thank my professor, Prof. Poika Isokoski, who has influenced this work in multiple ways. A major portion of this thesis was done as part of an internship in the Haptic And Gaze Interaction (HAGI) project group at TAUCHI. Firstly, for offering me the summer internship opportunity to work in this group. Secondly, for all his comments and feedback that has helped improve this work immensely. Lastly and most importantly, for introducing me to the basics of gaze tracking and experimental research in various courses as part of my master's degree experience.

I am immensely grateful to Mr. Jussi Rantala, researcher at TAUCHI, for all the insightful discussions and his meticulous comments that have helped enormously to improve this work.

I also acknowledge the contribution of Dr. Päivi Majaranta and Prof. Roope Raisamo for their expert suggestions and constructive feedback that has helped refine the research.

Last but not the least, I would like to acknowledge the support from the funding agency, The Academy of Finland, project number 260179, without which this work would not have been possible.

Table of Contents

List of figures

1. Introduction

During recent years, mobile technology has improved significantly. Mobile devices now cater to a variety of user needs and scenarios and are increasingly becoming an integral part of our life. Smartphones and tablet computers now come with a processing power comparable to desktop computers.

Currently touch screen interaction is the most common interaction technique in mobile devices. Even though fast and easy to use, the technique cannot be used efficiently in scenarios where one hand is occupied with other tasks or when the device is physically far from the user (for example, mobile device placed on a dock next to the user). Further, mobile devices are getting bigger and bigger. The dimensions of the Samsung Galaxy S4, a popular android mobile phone launched in 2013, are 136.6 mm x 69.8 mm [Samsung-S4] and the dimensions of tablet computers are usually even larger. Even though a large screen size is often desirable in mobile devices, it makes it hard to efficiently interact with these device with one hand. Figure 1 shows the difficulty of accessing all screen areas when interacting with a mobile device using touch with one hand.

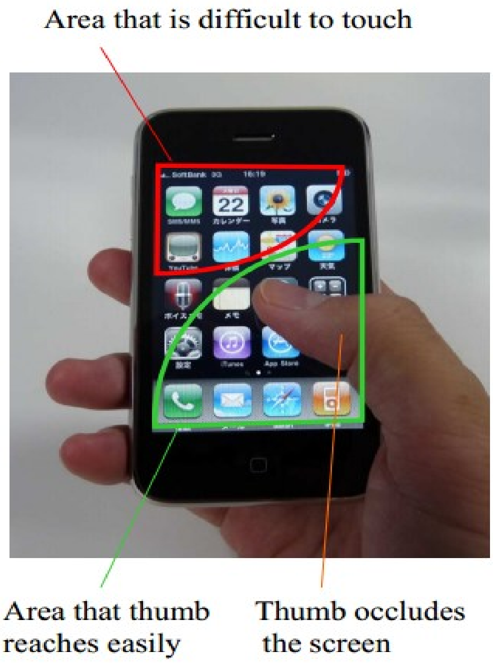

Figure 1 Difficulty of one hand touch screen interaction [Nagamatsu et al., 2010]

Mobile usage scenarios are very different from standard desktop computing. Oulasvirta *et al.* note that mobility often conflicts with mobile HCI [2005]. The Mobile context sometimes consumes some physical and cognitive resources which makes it difficult to use computing devices. In such scenarios, the user has to “make a place” for the device [Kristoffersen and Ljungberg, 1999]. For example, the user needs to stop the car to operate the mobile phone or the user needs to keep his coffee mug on the table and free both his hands to do a complex interaction with a tablet computer. The mobile usage paradigm introduces new interaction challenges and calls for more intuitive and natural interaction modalities [O'Grady et al., 2008].

Recently, there has been growing interest in alternate methods of interaction like voice commands, body gestures and eye-gaze interaction in mobile devices and these were also found effective in various usage scenarios [O'Grady et al., 2008].

Eye gaze based interaction has been available for more than 30 years [Majaranta and Räihä, 2002] but until recently, its use has been limited to severely disabled users. The interaction being natural and inherently fast, it has the potential to be used as an additional input channel in the mobile setting [Sibert and Jacob, 2000]. We anticipate that low cost miniature gaze tracking system would be available on mobile devices in the near future making this technology available to the mass user community.

There are many challenges in using gaze interaction. Two of these are extremely critical in the mobile context: Firstly, limited accuracy of gaze tracking due to frequent movement of the device and the user's head; Secondly, the need for appropriate, meaningful and instantaneous user feedback to make the interaction more intuitive.

The conventional gaze interaction method is using the dwell time. In this technique, the user focuses his gaze at a point for a predefined duration of time to invoke a predefined action, for example, the click of a button. Even though this technique is very intuitive and natural, it is highly susceptible to accuracy problems and may not be suitable for interactions with small screens [Drewes et al., 2007]. Gaze gestures, on the other hand, involve doing a predefined sequence of strokes using the eyes to invoke a command on the device [Drewes and Schmidt, 2007]. The gestures are often so designed to be distinct from natural eye movement. This helps distinguish between the conscious gestures and normal eye movement. Gaze gestures are a promising input method for mobile gaze interaction as they are more robust and tolerant to tracking inaccuracies [Bulling and Gellersen, 2010]. To overcome the accuracy problems, in this research work we use gaze gestures as the input modality.

The mobile device and the usage context present three potential feedback modalities i.e. visual, auditory and haptic. Both visual and audio feedback modalities have several shortcomings. Visual feedback may not always be appropriate for primarily two reasons. Firstly, in order to make optimum use of the screen area and not to overload the visual feedback channel. Secondly, while performing gaze interaction (e.g. off-screen gestures) users may not be looking at the screen of the device; this makes the visual feedback meaningless. Audio feedback, on the other hand, cannot be used in a noisy environment, situations where its usage is restricted due to social norms (e.g. in meeting rooms) or in situations where private feedback is required. Haptic feedback has the advantage that mobile device users are familiar with it and it provides for a private unobtrusive feedback channel. Haptic feedback is also known to be highly effective in situations of divided attention as it is processed at a low cognitive level [Hanson et al., 2009].

Gaze gesture input with haptic feedback is a novel combination and very little is known regarding the dynamics of the two modalities. The combination of the two modalities would be interesting especially in mobile usage scenario due to the inherent advantages of the two modalities individually.

This thesis focuses on gaze gesture interaction with haptic feedback on mobile devices. It highlights the suitability and challenges of using the combination of these input and feedback modalities. The main purpose of this work was to identify if haptic feedback helps the interaction and if so, what would be the best temporal point in terms of gesture progression to provide the feedback. As part of the work, an experiment was conducted to compare the usefulness of two stroke gaze gesture input with three different styles of haptic feedback against each other and also against the control condition of no haptic feedback on a mobile device. The thesis discusses the methodology and the findings of this experiment.

The reported study is based on a research work conducted in collaboration with other members of the HAGI project group, TAUCHI, University of Tampere. I was an intern in the group during the span of this work and was involved in the project from the ideation phase till the end of the study. My major contribution to the work includes the following:

- Taking active part in the conceptual phase of the project that eventually led to the design of the gestures, haptic feedback and the mobile application.
- Developing the mobile application that responded to the gaze events and provided the tactile feedback.
- Designing and developing the web socket interface between the mobile device and the gaze gesture recognizer that ran on the computer.
- Analyzing the experiment log files, formulating and calculating the ‘Gestures per Action’ metric.

This thesis work belongs to the field of Human Computer Interaction. The approach used in the thesis is primarily an extensive literature review and experimental research to identify the effectiveness of haptic feedback for two stroke gaze gesture input on mobile devices.

This thesis has seven chapters. Chapter 2 introduces the haptic and gaze interaction modalities in more detail. Chapter 3 summarizes the motivation for this research and describes the design consideration of the mobile application, gestures and feedback conditions used in the experiment. Chapter 4 describes the method of the experiment conducted, whose results are described in chapter 5. Chapter 6 presents a discussion of the results obtained in relation with the existing knowledge present in the literature. Chapter 7 summarizes the work and presents some of the future research opportunities in the field of mobile gaze interaction with haptic feedback.

2. Gaze and Haptic Interaction Modalities

This chapter introduces gaze and haptic interaction modalities. Gaze interaction in its fundamental form involves estimating where a person is looking and using this information in human computer transactions. With the advancements in computational power of devices and image processing technology, this interaction technique is gaining momentum as a powerful input modality in various visually-mediated applications [Duchowski, 2002]. Haptic interaction systems use the sense of touch in human beings as a channel to convey information to the users. In the following sections, we will take a closer look at the two interaction modalities.

2.1. Eye Gaze Interaction

2.1.1. The Human Eye

The eyes are an integral part of the human body. They serve the function of a sensory organ responsible for vision and also as an effective tool for social interactions and non-verbal communication. Our eyes provide a wealth of information to our communication partner or an onlooker regarding our point of attention and mental/emotional state. Eye contact is also known to be socially significant and an important component in effective face to face communication [Frischen et al., 2007]. Eyes hence function not just as an input system but also as an expressive communication channel.

The parts of the human eye that are visible from the outside include (figure 2):

- Cornea (the dome shaped outer layer over pupil and iris) [not shown in the figure]
- Sclera (the white colored region of the eye)
- Iris (the pigmented circular region)
- Pupil (the transparent circular region located in the center of the iris)

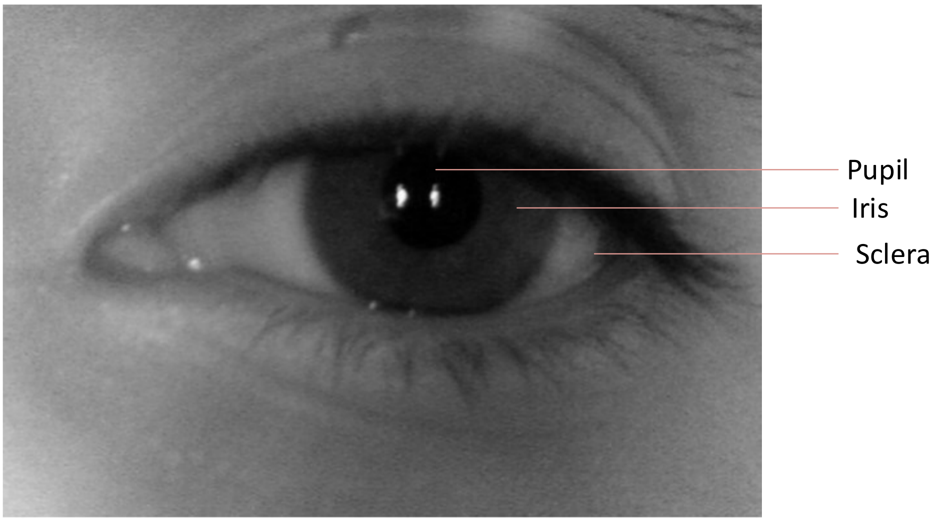

Figure 2 The human eye

Kobayashi & Kohshima studied the external morphology of the human eyes and found it to be very unique when compared to other primates [2001]. Human beings have a very widely exposed sclera region which is devoid of any pigmentation. Further, they note that human beings are the only species which has a very clear contrast difference between facial skin-sclera and sclera-iris. This contrast difference makes it easy to infer where a person is looking at. This is believed to be an evolutionary adaptation to enhance gaze signaling and communication using eyes.

There is a lot of literature explaining the anatomy of the human eye and the physiology of vision. For a comprehensive description of the anatomy of the human eye, see Oyster [1999]. This section only explains the basic anatomy of the eye that is required to explain the technology of gaze tracking.

2.1.2. Anatomy of Human Eye

The cornea in the human eye serves two basic functions. It protects the eye from dust and particulate matter and also accounts for a major part of the refractive power of the eye. This region acts as an outer lens and by virtue of its curvature and difference in refractive index with air, it refracts the incoming light rays through the pupil [Gross et al., 2008]. The iris is the muscular tissue that controls the amount of light entering the eye by controlling the diameter of the pupil. The light passing through the pupil falls on the lens which is the crystalline biconvex structure responsible for fine focusing of the light to the retina. The retina is the light sensitive region located in the inner surface of the eye. The central part of the retina is called "fovea", where visual acuity and color sensitivity are the highest.

Figure 3 shows the various parts of the eye. The imaginary line joining the fovea and the center of the cornea is called the visual axis or line of sight (LOS). The line connecting the center of the pupil, cornea and center of the eyeball is called the optical axis or line of gaze (LOG) [Drewes, 2010]. LOG and LOS intersect at the center of the cornea and the angle of intersection is specific to each individual as the location of the fovea can be anywhere between 4 to 8 degrees above the optical axis. LOS is believed to be the true direction of gaze [Hansen and Ji, 2010].

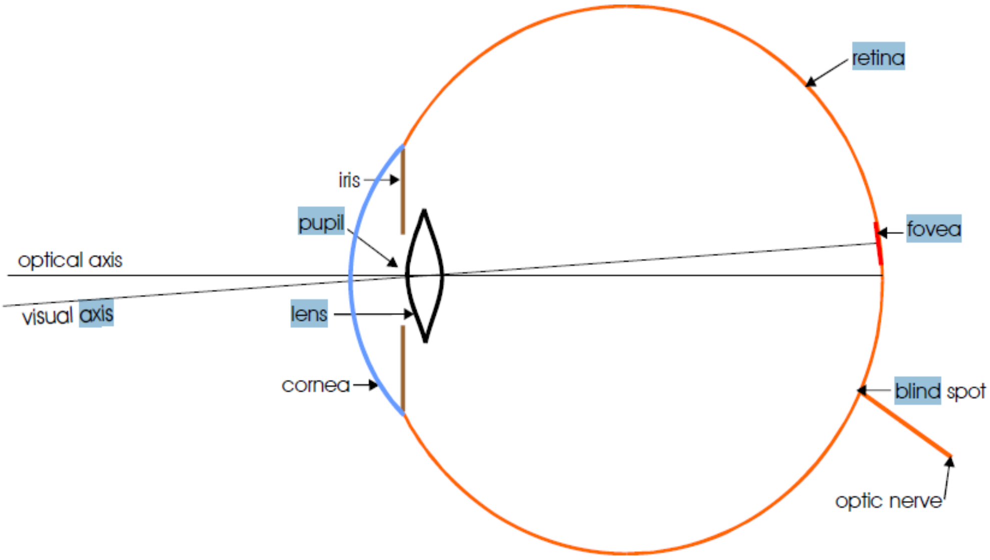

Figure 3 Parts of the eye [Drewes, 2010]

2.1.3. Eye Gaze Tracking Techniques

Gaze tracking techniques aim to either estimate the point of regard (POR) or the eye movement relative to head position. POR is defined as the point of intersection of the object being observed (e.g. on-screen objects) and the visual axis [Hansen and Ji, 2010]. There are mainly three different eye tracking techniques:

- Electrooculography (EOG) based technique relies on the fact that the front (cornea) and back (retina) of the eyes have a relatively steady standing potential difference of 0.4 - 1.0 mV, also known as corneo-retinal/corneo-fundal potential. Multiple electrodes are placed strategically at various points near the eye to record this potential. The potentials recorded at these locations change in relation with the eye movement. From the magnitude of potential variation in different electrodes, it is possible to ascertain the eye movement accurately. Such techniques measure eye movement relative to the head position and provide POR estimation only when combined with head tracking. EOG based gaze estimation is commonly used in clinical application and has a reported accuracy of two degrees [Morimoto and Mimica, 2005]. Figure 4 shows the EagleEyes, an EOG based gaze tracking system that uses five electrodes placed near the eyes of the user. One of the major disadvantage of such a system is that it requires electrodes in contact with the user [Gips and Olivieri, 1996]. Another disadvantage is that the corneo-retinal potential is not fixed but changes slowly with ambient lighting, fatigue etc. Hence, such a system may need to be frequently recalibrated [Brown et al., 2006; Malmivuo and Plonsey, 1995]. Bulling *et al.* have proposed the use of ambient light and physical activity sensors integrated into wearable EOG goggles to compensate for the EOG signal variations [2009].

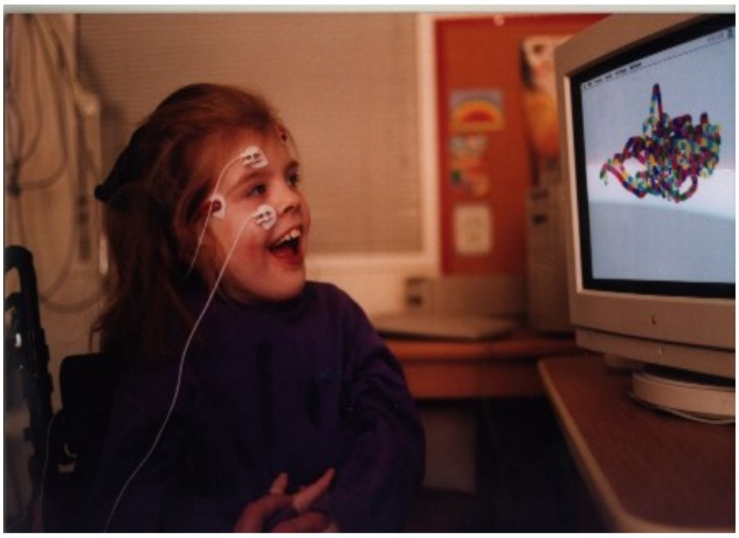

Figure 4 EOG based gaze tracking from the EagleEyes project [Gips and Olivieri, 1996]

- Scleral contact lens technique uses an optical or mechanical reference object attached to a contact lens worn on the eye to measure the eye movement. Such technique are very accurate, at the same time very invasive and uncomfortable for the user [Duchowski, 2007]. Such technique are seldom used in HCI and used only when a very accurate measurement is required for medical or psychological research.

- Video based technique is the most popular technique used for gaze tracking. Video-oculography (VOG) based technique uses a camera to ascertain the eye position [Duchowski, 2007]. They are enhanced by using infrared lighting to detect both the corneal reflection and the pupil to estimate the POR. Usually, infrared light source is placed on or off the optical axis of the video camera. This makes the pupil look bright (when IR light source is placed on axis) / dark (when IR light source is placed off axis) in contrast to the surrounding iris thereby enabling easy recognition of the pupil using image processing techniques [Drewes, 2010]. The light source is reflected at the four different layers of the eye. The reflections are called as Purkinje images. The first purkinje image appears at the outer surface of the cornea and is usually intense. It appears as a glint in the camera image. Due to the structure of the cornea, the glint remains static irrespective of the eye movement. By detecting the position of the glint and the pupil, the software deduces the gaze direction. [Drewes, 2010].

Video based gaze trackers usually use infrared lighting. Bright outdoor conditions may make pupil and glint detection using this technique difficult. Further, there are other video based techniques that use visible light, also called passive light approaches, to track the gaze point [Hansen and Ji, 2010]. Such methods could either work on the same corneal reflection technique or by extracting gaze information directly from the image using appearance based image processing techniques. These methods show promise as a viable outdoor gaze tracking solution and are being actively explored.

Depending on the physical set up, the video based gaze tracker could be either head worn or remote. Head worn tracker facilitates some level of mobility and may be the most suitable solution for mobile gaze tracking (figure 5). In remote gaze trackers, the camera and the infrared lighting are placed far away from the user (typically 50 to 80 cm) near a screen (figure 6). The gaze data quality in remote gaze trackers is known to degrade with relative head movement. Gaze tracking system integrated to mobile devices are also a viable solution in the mobile context if such challenges can be met.

Figure 5 Low cost head worn gaze tracker from the OpenEyes project [Li et al., 2006]

Figure 6 Tobii T60 remote gaze tracker [Tobii-T60]

2.1.4. Gaze Tracker Calibration and Accuracy

There is large anatomic variability of the eyes among individuals, for example radius of cornea, location of fovea etc. Both EOG and video based gaze tracking techniques require the gaze tracker to be fine-tuned for the subject to provide an accurate gaze estimation. This is done by a calibration process wherein the participant is shown multiple points on the screen and instructed to gaze at those points. With the collected gaze data, the algorithm fine tunes the system to the specific subject. The eye tracking accuracy largely depends on the calibration process. Generally, the larger the number of calibration points spanning the monitor, the better the accuracy of tracking. Some of the commercial gaze trackers, like Tobii T60, uses up to nine calibration points. However, from the perspective of the user, calibration with less number of on screen points is easier and preferred [Hansen and Pece, 2005].

Accuracy and precision are the two most widely used measures of quality of gaze data. Accuracy is defined as the closeness of the measured gaze point to the point that the tracked eye is looking at. Precision is defined as the ability of the tracker to re-produce the measurement. Precision of a gaze tracker largely depends on the system hardware and the algorithm used [Nyström et al., 2013]. Figure 7 shows a visualization of gaze data with different accuracy and precision characteristics.

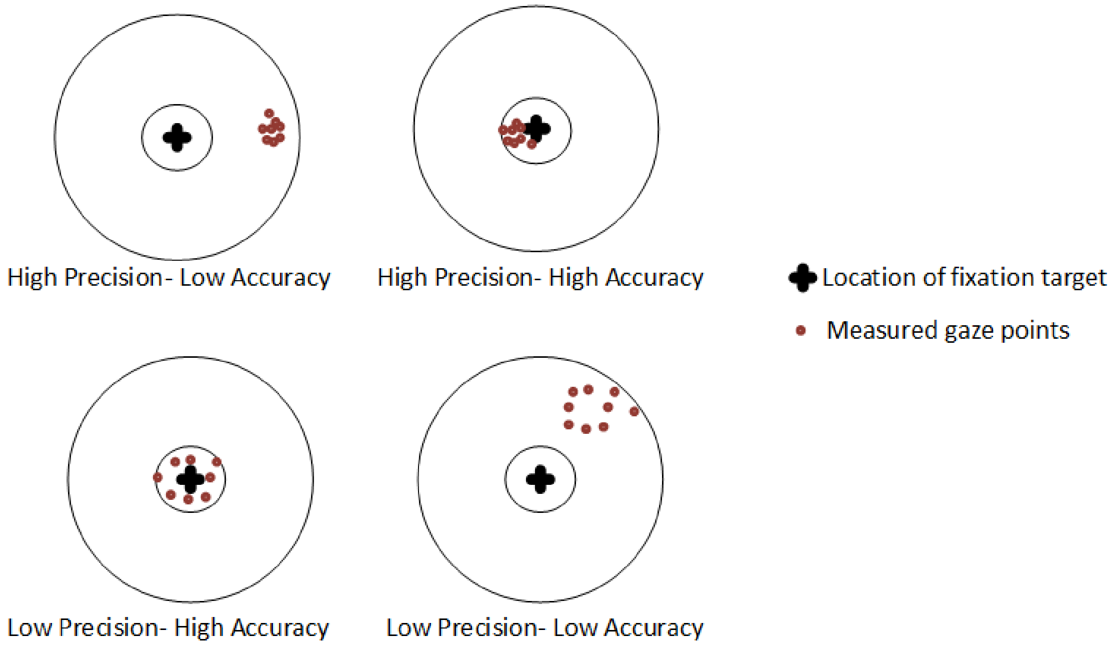

Figure 7 Accuracy and precision of gaze data

Modern commercial gaze trackers provide an ideal scenario accuracy of 0.5 degrees visual angle which is approximately 15 pixels on a 17 inch display placed at a distance of 70 cm and screen resolution of 1024 x 768 pixels [Majaranta, 2009]. There is a bottleneck on the maximum gaze tracking accuracy possible placed due to the size of the fovea and other characteristics of the human eye like drifting and micro-saccades. The foveal region is not perfectly circular and usually has an angular size of approximately 1 degree. In order to focus our gaze at any point, it is only required to have the image of the object somewhere on the fovea and not necessarily in the middle of it. This places a practical limitation on the maximum gaze tracking accuracy achievable. However, for a majority of HCI applications, this bottleneck does not have any significant implications.

2.1.5. Eye Gaze Interaction

Eye gaze is often associated with visual attention. Even then, it is sometimes possible for a person to dissociate his attention from foveal gaze direction and attend to an object of interest in the peripheral vision or look at something and mentally not attend to it at all. However, most eye tracking studies make a well accepted assumption that visual attention is linked to foveal gaze direction [Duchowski, 2007].

The studies of eye movement and gaze estimation would help psychological research and neural science to understand human visual perception and processing. The same technique is also used as a method to interact with computing systems. In this section, we will limit our focus on eye gaze as a human computer interaction modality.

Due to the physiology of vision, the eyes either remain stationary (fixation) to perceive an object, make a rapid movement (saccade) between fixations to perceive a scene or slowly move to follow a moving target (smooth pursuit). Thus voluntary eye movement mainly comprises of fixations, saccades and smooth pursuits.

In gaze based interaction with a computer, we generally use these three voluntary eye movements to perform predefined actions on a computer. Fixations and saccades are more commonly used in human computer interactions than smooth pursuits. In the following sections, we limit our focus to use of fixation and saccades based gaze interactions.

While using gaze as an input modality, it is important to distinguish between the natural eye movement and intentional commands [Majaranta, 2009]. This is known as the famous "Midas touch problem" in gaze based interaction. Two of the most common ways of using gaze input in HCI are dwell based interaction and gaze gesture based interaction.

2.1.5.1 *Dwell time based interaction*

Dwell based interaction uses prolonged gaze of a predefined duration ("staring") at a screen point to invoke a specific command e.g. the click of a button. If the dwell time is too long, it affects the performance as more time is required to invoke a command. On the other hand, a too short dwell time is likely to result in larger number of errors due to unintentional invocation of actions. Majaranta *et al.* notes that adjustable dwell time improves the performance considerably in eye typing applications. Novice users usually are comfortable with longer dwell time. For such users, the dwell time can be significantly reduced with some practice, thereby improving the eye typing performance [2009].

Another important aspect of dwell based interaction is the need for accurate and precise gaze tracking. A small offset in the gaze data can result in a triggering action on the adjacent screen element and a less precise gaze data can make detection of events like fixation and saccade difficult [Nyström et al., 2013]. Quality of gaze tracking is hence critical to successful dwell time based interaction.

2.1.5.2 *Gaze Gesture Based Interaction*

The concept of gesture is popular in human-human and human-computer interaction. Body gestures are known to play an important role in complementing speech in human-human communication. In HCI, there are already systems that use mouse gestures, pen gestures and body gestures to interact with a computer.

Gaze gestures consist of a sequence of saccadic eye movements, typically called strokes [Drewes and Schmidt, 2007]. Istance *et al.* define gaze gestures as:

"*A definable pattern of eye movements performed within a limited time period, which may or may not be constrained to a particular range or area, which can be identified in real-time, and used to signify a particular command or intent.*" [2010].

Gestures rely on relative eye movements and are known to be a robust alternative to dwell based interaction [Hyrskykari et al., 2012]. Gestures are less sensitive to gaze tracking inaccuracies. However, the true advantage of gestures is attained when strokes are of sufficient length. Within-screen gestures may not harness the full strength of this technique, especially when interacting with a small screen.

Isokoski proposed using off-screen targets for gaze based text entry [2000]. In order to enter text, the user needs to fixate briefly at physical targets placed around the screen area. The resulting eye movement is equivalent to gaze gestures with fixed end-of-stroke locations.

Drewes & Schmidt developed a generic gaze gesture recognizer inspired from mouse gesture plug-in for the Firefox web browser. It consists of eight strokes (figure 8) including the four diagonal strokes [2007].

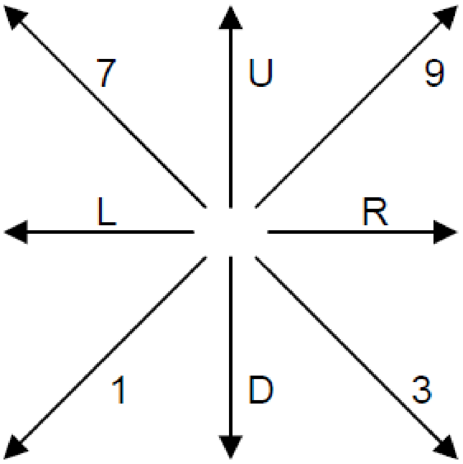

Figure 8 Strokes in gaze gesture recognizer

In a subsequent user study, participants were asked to perform three gaze gestures (figure 9) of varying difficulty with and without visual aids in the background to help perform gestures. The findings suggested that the gesture completion time is only dependent on the number of strokes in the gesture and independent of the complexity of the stroke or presence/absence of visual aids. Further, even though all participants could perform all the gestures with visual aids, only five out of nine participants could perform the most complex gesture with a blank background [Drewes and Schmidt, 2007].

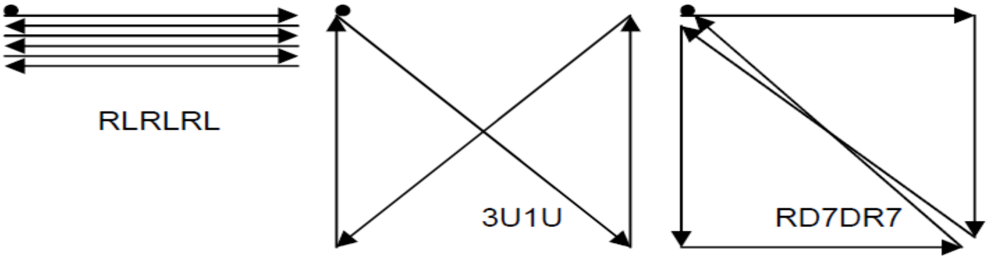

Figure 9 Gaze gestures used by Drewes and Schmidt [2007]

This suggests that without visual aids to assist the gesture, it is difficult to perform complex gaze gestures [Drewes and Schmidt, 2007; Isokoski, 2000]. This challenge is perhaps more prominent in the learning phase of the interaction as visual aids help users to direct their gaze to the predefined location and this movement could come naturally for an expert user. This should be taken into account while designing off-screen gesture based interactions either by providing visual cues to aid the fixation or by additional feedback through nonvisual channels.

2.2. Eye Gaze Interaction on Mobile Devices

Until recently, the use of gaze interaction has been primarily limited as an assistive technology to the disabled. However, several studies have shown that this technology could be beneficial to the larger user community in various scenarios [Drewes et al., 2007; Dybdal et al., 2012; Miluzzo et al., 2010; Nagamatsu et al., 2010]. Gaze interaction, when used as an additional input channel along with other modalities, could provide a richer interaction experience. Gaze information of the user could be as both an explicit input channel and an implicit input channel in HCI [O’Grady et al., 2008].

Explicit input is when user gives a command to the device to perform an action. In the case of gaze input, either by dwelling at a button or performing a gaze gesture. On the other hand, Implicit interaction is defined as “*an action performed by the user that is not primarily aimed to interact with a computerized system but which such a system understands as input*” [Schmidt, 2000]. For example, the system pauses the video when the user’s gaze wanders off screen or the system scrolls the webpage if it identifies that the user has reached reading the bottom of the page etc. Robust implicit interactions could result in “smart” devices that know what the user requires. Such methods may lead to a larger acceptance of the interaction technique by the user community. Gaze gestures fall in the broad category of explicit gaze interaction and in the subsequent sections, we focus only on this category.

Mobile usage scenarios are very different from standard desktop computing. The user could be stationary or mobile, use context could be indoor or outdoor, ambient environment and usage scenario could add a lot of noise to the system, the device processing power could be relatively low etc. Another important difference is in the usage characteristics. With the exception of gaming, mobile usage is often brief and concise, while a user tends to interact with a standard desktop computer for relatively long and uninterrupted period of time [O’Grady et al., 2008].

There are already some gaze tracking solutions available for mobile devices. Dickie *et al.* developed the eyeLook, a system that can detect user attention using gaze in a mobile device [2005]. The system detects whether the user is looking at the device using infrared illumination placed on and off camera axis and synchronized with the camera frames producing dark and bright pupil in adjacent camera frames. Such eye contact detecting system does not need considerable accuracy and hence does not need a calibration process. Even such application can be useful in mobile devices because interruptions of attention are very frequent in the mobile setting. For example : pausing a video when the user is not looking at the device or the device switching to sleep mode when no eye contact is detected for a predefined duration of time etc.

Miluzzo *et al.* developed the EyePhone, a mobile system that uses the front facing camera of the phone to detect and track the eye. The system was one of the first completely mobile phone based eye tracking prototypes developed. The system used template matching technique using the OpenCV libraries to track the eye and invoke a mobile application using wink [Miluzzo et al., 2010]. Eye phone could only detect POG of the resolution of nine regions in the mobile screen. In their study, the accuracy of the system was shown to degrade with ambient lighting, shake of the device induced due to user movement and even large variation in distance between eye and device. This indicates that further research is required to develop robust algorithms that can minimize the noise and efficiently track the eye in all conditions.

Stellmach *et al.* studied the use of gaze pointing along with touch and tilt sensors in mobile devices for visual exploration of large collection of images on a display screen [2012]. Even though the study did not use gaze as an input modality in mobile interaction, this was probably the first study to combine gaze input with sensors available in mobile devices. Their finding suggests that such gaze assisted interfaces allows for a more relaxed gaze interaction. The combination of tilt sensors in mobile device was found to be helpful in avoiding the Midas touch problem and removing the need for dwell based selection, which often slows down the interaction. The technique was also helpful in complex interactions like panning and zoom [Stellmach and Dachselt, 2012]. Most modern mobile devices are embedded with MEMS (micro electro-mechanical sensors) like accelerometers and gyroscopes, and could facilitate interactions where gaze input is smartly and seamlessly integrated with such sensors.

2.2.1. Challenges of Gaze Based Interaction on Mobile Device

There are several challenges of using gaze based interaction on mobile devices due to the context and style of use. Some of the major challenges are:

• ***Outdoor conditions and IR illumination***

Most of the research in gaze tracking is limited to stable indoor conditions with active IR illumination. These techniques do not work well in the outdoor conditions. Several alternatives based on visible light and eye appearance model has been proposed [Hansen and Ji, 2010]. However, the accuracy of such gaze trackers are still quite low. Further research would be required before stable gaze tracking is possible in outdoor conditions.

• ***Constant movement of head and device***

Movement of the head and the device are known to affect the tracking accuracy. In the mobile usage paradigm, we expect the users to be in motion and device itself not to be in a stable position. Further, due to the style of use, the relative distance between the mobile device and the user's eyes could vary considerably. By using the built in sensors in mobile phones and tablets, it is possible to differentiate between the movement of the device and movement of the head. Mobile gaze tracking is still in its nascent stage and substantial research and development is needed to overcome these problems [Dybdal et al., 2012; Hansen and Ji, 2010].

• ***Calibration Requirements***

We already discussed the need for the calibration process. In mobile device usage, it is common for the user to have frequent short and precise interactions with the device instead of a few long interactions. This would require the users to calibrate the device for every interaction, which is not practical. Gaze gesture based interaction is known to be tolerant to slight calibration shifts and could be used to solve this problem to an extent [Drewes et al., 2007].

- ***Screen size and Screen real estate***

Mobile devices often have small screen size compared to desktop computers and hence screen contents like links, thumbnails and icons are also smaller. The small screen size also means reduced screen real estate to provide visual feedback of user action. This poses a challenge to interaction designers. Solutions using eye gaze gestures have been proposed that could overcome these limitations and could provide easy interaction possibility to the user [Drewes et al., 2007]. Using non visual feedback is an alternative that should be explored further with such alternate interaction techniques.

2.2.2. Gaze Gesture Interaction on Mobile Phones

Many previous studies support the suitability of gaze gesture based interaction on mobile devices [Bulling and Gellersen, 2010; Drewes et al., 2007; Dybdal et al., 2012; Zhao et al., 2012].

Zhao *et al.* [2012] compared numerical text entry in mobile phones using gestures and dwell based gaze interaction. For an angular inaccuracy of 0.8 degrees of the tracker, they found that gestures were 60% more effective than dwell. The users could perform the numerical task entry faster and with lesser errors using gestures. Gaze gestures do not depend on absolute gaze point but on the pattern of eye movement and hence are less sensitive to tracker inaccuracies. In mobile device usage, movement of the device and user's head could result in poor quality gaze tracking. It would hence be desired that the interaction technique is tolerant to such problems.

Dydbal *et al.* [2012] compared gaze gesture and dwell based interaction on a mobile phone in a series of target selection tasks. Their results indicate that gaze gestures considerably outperformed dwell based interaction in terms of target completion time and error rate. Gesture based selection produced 21% fewer errors than dwell based selection technique. This could be due to the fact that dwell based interaction is sensitive to the target size. When the target is small, it is harder to select the target using dwell. While on the contrary, gaze gestures are independent of the target size. This can be crucial when interacting with small screen devices.

Drewes *et al.* [2007] studied gaze pattern in mobile phone interaction and found out that using a minimum stroke length of 80% of screen area and limiting the maximum duration of each stroke to 1 second could drastically reduce the chances of unintended invocation of gestures. Gaze gesture interaction on mobile devices also presents the possibility of using off-screen gestures. The accidental invocation of these can be reduced by using a time limit between the strokes. One of the major disadvantage of using dwell based interaction is that the user might invoke a command by accident when looking at the screen content. This is also the strength of gaze gestures [Drewes et al., 2007]. It is unlikely to invoke a command by mistake while using gaze gestures as they are designed to be quite different from the normal movement of the eyes.

Further, dwell based interaction requires the interaction object to be visually present on the mobile screen. This limits the number of objects that can be interacted with at a given point of time. This can be a major limitation in case of small screen devices. Gestures do not impose any such limitations due to screen size. The users could have many distinct gestures for different actions. For example, users could allocate predefined gestures as shortcuts to invoke certain applications. Such "non-visual" short cuts could also speed up the interaction. Gaze gestures do not need screen real estate and hence the screen area can be used for visual output [Drewes et al., 2007].

All these points support the suitability of gaze gestures in mobile devices. However, gestures are not without drawbacks. Some drawbacks mainly takes effect when large number of gestures are required to support the functionalities. The user needs to learn and remember all the gestures available for efficient interaction [Hyrskykari et al., 2012]. Additionally, more complex gestures would be required to support all the functionalities and such gestures may be difficult to perform for the user. However, these limitations are less pronounced when gestures are simple and the number of gestures is small.

Using gaze gestures can be cognitively more demanding than dwell based interaction [Dybdal et al., 2012]. The cognitive load could be reduced by providing appropriate user feedback. As the gestures become complex, users may need to be given feedback regarding gesture progression. This would allow users to stop the gesture once they know that a stroke is not recognized or is wrongly recognized. Visual feedback may not be suitable for this purpose as the eyes would be in motion. We will look at the qualities of a good user feedback for a gaze gesture based system in more detail in section 4.2.

2.3. Feedback in HCI

Donald Norman in his classic book " *The design of everyday things*" introduced the terms *gulf of execution* and *gulf of evaluation* in human system interaction [1988]. *Gulf of execution* is the degree of mismatch between the intention of the user and the actions supported by the system and is the measure of how well the system allows the person to do the intended action directly on the system. It indicates the difference between the mental model created in the user's mind and the actual system model that defines how user input is translated to real world action. *Gulf of evaluation* is the degree of effort required to interpret if a user input has created the intended real world actions. For effortless interaction with the system, the designers should bridge the gulf of evaluation and gulf of execution. A system that makes use of natural mapping between its controls and real world actions can reduce the gulf of execution and appropriate feedback to user actions is critical to bridging the gulf of evaluation [Norman, 1988].

The need for feedback is widely accepted even in human-human communication. Appropriate feedback helps satisfy communication expectation or "psychological closure" [Pérez-Quiñones and Silbert, 1996]. In normal conversation, each conversation partner provides cues of their state in order to maintain and repair the conversation flow. For example, the person listening could provide positive evidences such as nod of the head or utterances like *hmmm* or *ok* to convey that he or she has heard the speaker and understood what he said. If the person listening has not completely understood what was spoken, he could provide negative evidences like raising the eyebrows to show confusion or explicit utterances like "*what?*" to convey to the speaker that the conversation needs some repair.

Many of the feedback mechanisms in HCI are also modelled on this collaborative theory of human communication [Clark and Brennan, 1991]. Perez-Quinones and Sibert [1996] presented a collaborative model for feedback based on the linguistics theory of conversation for GUI. In their paper, they presented five feedback states (busy, processing, reporting, busy-no response and busy-delayed response) that must be communicated to meet the communication expectation of the user. Brennan and Hulteen [1995] also presented a feedback model for spoken language system in HCI which is also derived from human communication model.

2.3.1. Use of Feedback in HCI

In HCI, feedback is defined as:

"*Communication of the state of the system , either as a response to a user action, to inform the user about the conversation state of the system as a conversational partner, or as a result of some noteworthy event of which the user needs to be apprised*" [Renaud and Cooper, 2000].

The definition encompasses the fact that feedback need not always be in response to a user action. Feedback basically serves three functions in HCI [Renaud and Cooper, 2000]:

•***Response to user action***: An appropriate feedback to the user action conveys to the user that the system has accepted his input and is performing the corresponding action.

•***Modifying user behavior***: Often feedback could convey to the user that some fault has occurred. This enables the user to strategize their future actions. For example, if the system has wrongly accepted a user action as a command, the user knows the fault and can modify the next action so as to repair the interaction.

•***Promote Understanding***: Provide users with the understanding of the current state of the system, e.g. convey some system events to the user.

Continuous feedback is critical to simplify the interaction and for the user to have a sense of control over the interface. This is especially true during the learning process where the user familiarizes with the system. Gentner and Nielson notes that feedback in HCI should be flexible, continuous during the initial phases to instill confidence to the user and scaled down to special circumstances later on once the user is familiar with the system [1996].

Further, research has shown that appropriate feedback improves user performance. Majaranta *et al.* [2006] studied the effect of visual and audio feedback in dwell based eye typing application. The results suggests that feedback not only effects typing speed and accuracy of typing, but also the gaze pattern and subjective user preference. In eye typing applications, in the absence of clear feedback, the user need to point his/her gaze towards the text area to review the typed letter. In cases where feedback is adequate, the user is confident and can proceed with the task without the need for frequent review of the text entered. The study also stressed the need for context specific feedback. For example, when the dwell-time was longer (900 ms), a two level (focus and click) feedback combining both audio and visual feedback improved performance and was more liked by the users. However, for shorter dwell time (400 ms) a clear and crisp one level feedback worked best.

2.3.2. Feedback for Gaze Gesture Interaction in Mobile Context

Qualities of a good feedback are often task and context specific. Some feedback options that work well for a given situation may not work so well in others. For example, the ringtone based feedback to convey an incoming call in mobile phones may be the best option when the user is at his home and mobile device physically far away from him. The same feedback modality may not be very appropriate when the user is in a meeting room.

Gaze gesture interaction in a mobile context imposes some restrictions on the feedback options that can be provided. The following section lists the qualities of a good feedback in such a system.

•***Meaningful*:** Nielsen notes that gestural interfaces present a new challenge to interaction designers in terms of providing meaningful feedback [1993]. Confirmation feedback in these systems cannot be provided until the gesture is completely recognized, which means that feedback appears late to the user to help them complete the action [Nielsen, 1993]. It is important to provide feedback at meaningful positions as the gestures are being made. This type of progression feedback is even more helpful when the gestures are complex and user needs to know at each sequence if that part of the gesture is correctly understood by the system.

In gaze gestures, feedback could be provided at the end of each stroke. When compared to a simple confirmation feedback after gesture completion, these stroke completion feedbacks help user to detect and correct their errors sooner. The users no longer have to wait till the end to know if the gesture was correctly interpreted by the system

•***Instantaneous*:** In all communication, there is a response expectation and a strict time period within which the response is expected. Miller notes it is human nature to psychologically organize a tasks into multiple subtasks [1968]. For example, to call a contact from phone book, the subtasks could be to find the name in phone book and to dial the number. User has a temporary sense of task completion on finishing each subtask which is called as "psychological closure" [Miller, 1968; Pérez-Quiñones and Silbert, 1996] . In human computer transaction, we might tolerate an extended delay in response after a closure than during the process of attaining it. A delay in response can often be frustrating and also affect the task performance in HCI. This drop in performance is not linearly related to the response time but abrupt when the response time exceeds a threshold and can be because of the inability to connect the user action with the system response [Miller, 1968].

In Eye gesture interaction, because the interaction is inherently fast, it is important for the feedback to be instantaneous. More detailed research would be required to understand the acceptable response time to gaze gestures in different tasks.

•***Appropriate (Audio/Visual/Haptic...etc.)*:**

Appropriateness of a feedback modality depends on the individual and the usage context. In mobile usage, consideration should be given to the fact that users could be on the move and can be expected to be in any contexts. For example, from silent and stable meeting rooms to noisy environment or environment prone to frequent vibration etc. Feedback should be such that it is not excessive [too loud or strong etc.] but still easily perceivable in all environments [Linjama et al., 2005].

Performing different gestures using eyes often mean that visual feedback could be inappropriate as a feedback modality to convey gesture progression as the eyes are in constant motion. Audio feedback, even though helpful, may not be appropriate in all contexts. For example, noisy environments, silent meetings etc. which are common in the mobile usage. Further, sometimes it would be desired to have the feedback through a private channel. Haptic feedback provides a very unobtrusive feedback channel and could be used in all scenarios. Mobile devices are designed to be carried in hand when in use which provides a location for directly providing the haptic signal. Most of the mobile device users are also familiar with haptic feedback modality as it has long existed in mobile devices. However, one drawback of the modality is that given the state of mobile device haptic actuators, it limits the different types of haptic feedbacks that could be generated and recognized.

•Least cognitive load

Mobile user needs part of their visual, auditory and cognitive attention to safely navigate through the environment [Oulasvirta et al., 2005]. The feedback modality should be such that it does not further overload the user. Hanson *et al.* notes that in such scenarios, haptic modality works better than auditory and visual modalities [2009]. In a divided attention scenario, tactile stimulation is processed pre-attentively by the brain and is given more priority than visual or auditory channels by the nervous system [Hanson et al., 2009]. This also reduces the cognitive load associated with perceiving the feedback.

2.4. Haptic Feedback

The term “Haptics” is derived from the Greek word *haptikos* meaning to grasp or touch [Banter, 2010]. In its broadest definition, haptics refers to the study of touch sensing and also encompasses engineering different mechanical devices that provide touch stimuli.

For human beings, touch is a very personal medium of communication and is the only way for humans to directly manipulate real world objects. Touching an object provides a large amount of information, for e.g. dimension, weight, pressure, texture and warmth. The sense of touch in human beings is extremely complex and is in fact a combination of many closely related sensory mechanisms. All of these mechanisms fall into one of the two distinct category of senses: cutaneous senses and kinesthesis [Loomis and Lederman, 1986]. The cutaneous system receives information from the numerous mechanoreceptors and thermoreceptors present across the body surface to provide awareness about the skin stimulation. Kinesthetic system, on the other hand, uses the mechanoreceptors present in muscles, joints and tendons to provide awareness of limb position, limb movement and mechanical properties of objects they interact to.

To explain the physiology and psychology of touch in detail is beyond the scope of this document. For a more detailed discussion on physiology and psychology of touch, see Grunwald [2008]. Following section presents a brief overview of the sensitivity of human cutaneous system.

2.4.1. Temporal and Spatial Acuity of Human body

Like any other human sense mechanism, the human cutaneous system has its limitations. The cutaneous system has limited ability to resolve temporal and spatial details [Lederman and Klatzky, 2009].

Two of the most classical methods to evaluate the spatial acuity of the human body are the “two point touch threshold” and “point localization threshold”. The two point touch threshold is the smallest distance on the skin where two exact same stimuli can be rightly distinguished. The test is easy to administer and requires the participants to tell apart if the stimuli is applied to point-1 or point-2, two closely located points on the skin [Lederman and Klatzky, 2009]. The disadvantage of this method is that it relies on the subjective response of the participant. Point localization method involves applying a touch stimuli at a body location followed by another stimuli which may or may not be applied at the same location. The participants are required to tell apart if the stimuli was applied at the same point in both cases or different places [Lederman and Klatzky, 2009]. Both two point threshold and point localization threshold are highly correlated and good measures of spatial acuity of human cutaneous system. The point localization threshold is highly sensitive and the error in localization ranges from 1.5 mm in the fingertip to 12.5 mm in the back. The figure below shows the relative spatial acuity in terms of two point threshold and point localization threshold at various points in a female body. The spatial acuity for men follows the same pattern.

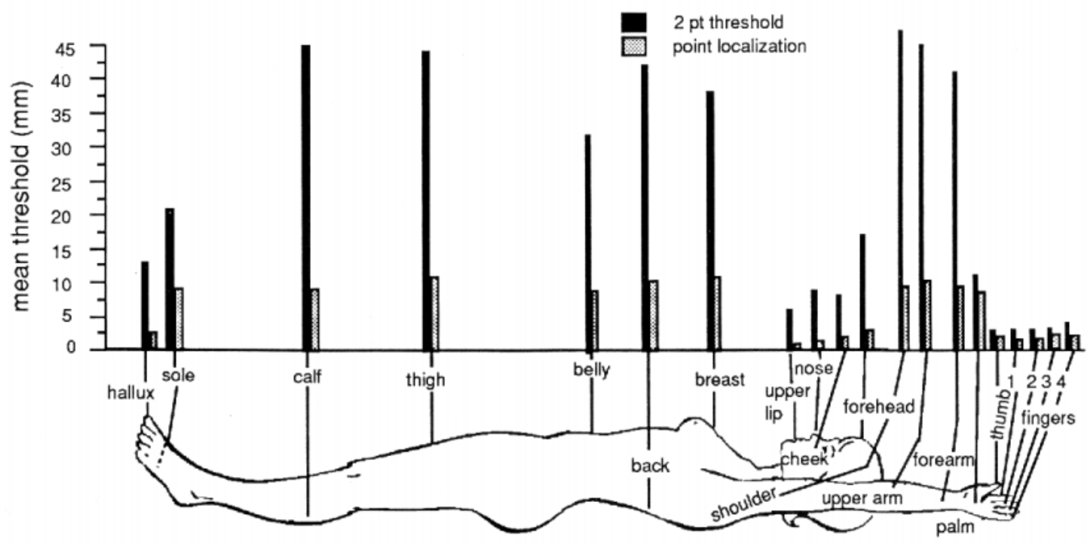

Figure 10 Spatial acuity in humans [Lederman, 1991]

The relative spatial acuity varies largely across the human body. The spatial acuity is high in the finger tips, face region and hands while relatively lower in back, shoulder and thigh region.

Studies on temporal sensitivity of the skin suggest that human beings can resolve two 1 msec tactile stimuli separated by as low as 5.5 msec. Overall, temporal sensitivity of cutaneous sense is better than vision but poorer than audio [Lederman, 1991].

2.4.2. Haptics in HCI

The computer keyboard, mouse and even stylus can be thought of as simple haptic devices. These devices, however, can only be used to perform actions on a computer and not as touch output devices that actively stimulate human touch senses. It is only recently that affordable haptic devices capable of providing more natural and believable touch stimuli have been made available.

In the beginning, teleoperation and telepresence were the two main domains in which haptic devices were extensively used. Teleoperation is "*the extension of a person's sensing and manipulation capability to a remote location*" [Stone, 2001] and telepresence is "*the ideal of sensing sufficient information about the teleoperator and task environment, and communicating this to the human operator in a sufficiently natural way, that the operator feels physically present at the remote site*" [Stone, 2001].

Currently, haptics has found usage in multitude of fields. For example, museum displays, virtual environments, various military applications and simulation studies, assistive technologies for visually impaired, automotive sector and commercial household devices.

Haptic devices in virtual reality systems provide users with a sense of touch of real world objects in virtual environments. When interacting with a real world object, different forces are exerted by the object to the skin, muscles and joints. That information is processed by the brain and leads to haptic perception. There are devices (e.g. SensAble Technologies PHANToM) available that mimics the various forces exerted by real world objects thereby resulting in believable haptic perception.

Haptics is also being increasingly used as an assistive technology for visually impaired users giving them a "sense of vision" through touch. Current GUIs rely on visual metaphors to make the interaction more intuitive. However, this makes such interfaces even more difficult to use for the visually impaired. Haptics can help the interaction in such cases. O'Modhrain and Gillespie [1997] presented Moose, a mouse like system capable of haptically enhancing the GUIs for use by both sighted and visually impaired user using haptic icons, controls and windows. Moose uses both cutaneous and kinesthetic touch sensation.

Some day to day consumer electronic devices also provide tactile sensation. Some due to the construction and operating mechanism (e.g. a drill, automatic shaver etc.) and others as an output modality to improve the interaction [Rovers and Essen, 2006]. For example, mobile devices that vibrate to convey an incoming call or game controllers that provide tactile feedback to increase the gaming experience or automotive controls that provide tactile sensation [Banter, 2010] .

In summary, haptics is steadily finding use in various devices and scenarios. The current state of the technology will only improve further with better haptic actuators capable of providing more natural and richer touch sensations.

2.4.3. Haptics in Mobile Devices

The focus of this section is limited to use of cutaneous touch sensing in mobile devices. It presents a brief overview of the vibrotactile actuation in mobile devices and some of the current literature about use of vibrotactile feedback in mobile HCI.

2.4.3.1 *Vibrotactile feedback in mobile devices*

Vibrotactile feedback has been present in mobile phones for a long time. We all are familiar with vibrating alerts to signify an incoming call or message. Usually the haptic actuator in mobile devices is a small DC motor with an eccentric weight attached to the shaft (figure 11). An electronic signal to the DC motor (figure 12) generates vibration that can be felt in the entire device. These motors, however, take a fraction of second to start up and stop. These actuators usually do not provide any control over the intensity of vibration. However, it is possible to create a few different distinguishable pulses using modulation of vibration ON and OFF pulse width in a standard mobile device (figure 13). Even then, the haptic capabilities of these devices are limited and not suitable for conveying complex messages.

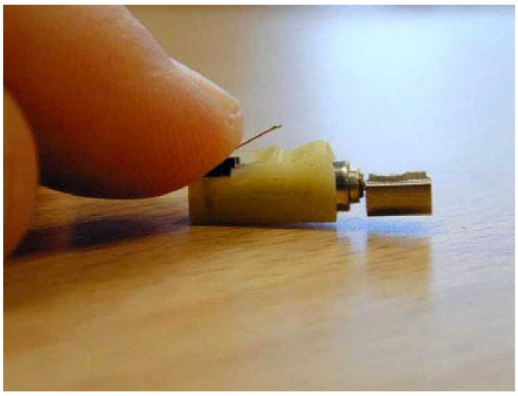

Figure 11 Mobile phone haptic actuator [Kaaresoja and Linjama, 2005]

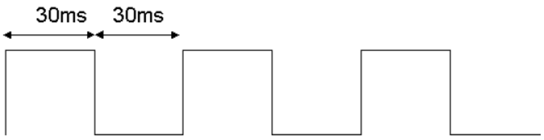

Figure 12 ON and OFF vibration pulse to phone [Brown and Kaaresoja, 2006]

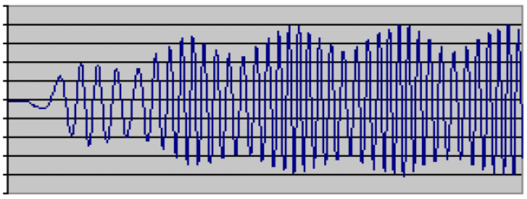

Figure 13 Output vibration of the phone [Brown and Kaaresoja, 2006]

Some of the exceptions are Samsung Anycall haptics mobile devices launched in South Korea in 2008 [Placencia et al., 2011] that has 22 different vibration patterns to provide a richer touch experience to the users. A few other devices offer similar capability but none have so far been successful in the mainstream consumer market.

Immersion is a company that has been working towards richer haptic experience in mobile devices. Immersion TouchSense® Haptic (Tactile) Feedback Technology and Integrator aims to provide crisp and realistic haptic feedback in mobile devices during various UI interactions including typing, scrolling , selecting and web browsing. Another module in the immersion toolkit called Reverb module automatically translates audio to haptic effects allowing users to feel their music and games [Immersion].

Immersion tactile presence technology enables haptic communication between two mobile devices allowing a mobile user to feel the touch of a remote person through the mobile device [Immersion]. This type of haptic communication, even though extensively studied, is new in commercial devices and has immense potential. This can be even more effective when coupled with other modalities like voice call or video call facility. This type of communication can help attain a feeling of co-presence, shared workspace and if creatively designed can result in an emotional experience.

To conclude, future of haptics in mobile communication devices like mobile phones and tablet computers seems to be bright with a variety of new innovations slowly emerging in the consumer market.

2.4.3.2 *Haptics in Mobile HCI*

Mobility often requires the system to provide feedback through an unobtrusive channel. Haptics has been serving this purpose on mobile phones for a long time already. However, there are not many studies on haptic perception in truly mobile context. Even though users do carry mobile phones and other similar haptic devices with them most of the time, the level of physical contact with these devices can vary significantly. We often hold the phone in the hands during use and otherwise keep it in the pocket, bag etc. Further, the environment of use can also add external noise and vibration. Linjama *et al.* [2003] studied subjective strength of tactile feedback in mobile devices when the device is in physical contact at different body locations. They proposed that vibration is felt by the movement of the mobile device and motion properties like velocity level is a suitable measure of human sensation to vibration. Their study also suggests that there is a relatively narrow range of stimuli strength that is optimal for use [Linjama et al., 2003]. A slightly higher intensity is often perceived as irritating and too strong while a slightly lower intensity is not perceived at all. This optimal stimuli intensity should be considered while designing HCI applications that incorporate haptic feedback.

Another design challenge arises from the fact that human senses are multimodal and therefore cues provided by the sense of touch should be consistent with the information provided by other input channels to result in robust perception [Ernst and Bülthoff, 2004]. For example, if a click of a button is designed to produce a haptic feedback, the haptic signal should be temporally and spatially synchronized with the visual cues associated with the click of a button. The haptic feedbacks should be consistent with the various laws of sensory integration thereby providing a natural multimodal interaction experience to the user [Linjama et al., 2005; Linjama and Kaaresoja, 2004]. For a more detailed discussion on sensory integration, see Ernst and Bülthoff [2004].

Haptic feedback has been studied in various mobile device interactions like touch typing. Current mobile devices do not have a physical keyboard and use touch screen for text entry. Physical keyboards facilitate different levels of feedback while typing, for example, feeling of the gap between keys indicate transition of finger, press and release of button indicates selection etc. In mobile devices, the user should constantly look at the onscreen keyboard area and the text entry window while entering text. This requires a lot of visual attention and results in a lot more text entry errors. Brewster *et al.* note that in most of the cases, these errors are not even noticed by the user due to the lack of appropriate feedback and also the cognitive load of the task itself [2007]. This is even more predominant when the user is mobile. Brewster *et al.* studied tactile displays in mobile devices in both static and mobile environment and found out that users made fewer errors when tactile feedback was available and more importantly users noticed and corrected more text entry errors with tactile feedback [2007]. Tactile feedback was found to be even more beneficial for error detection and correction in mobile situations [Brewster et al., 2007]. These findings suggest that tactile feedback improves performance and usability of on-screen keyboard interactions in touch screen devices. They provide a “sense of control” to the user as they know when a key is wrongly pressed or not pressed at all without looking at the text area.

Hoggan *et al.* compared text input using tactile soft keyboards, soft keyboards with multiple specialized actuators providing more localized tactile feedback and physical keyboards [2008]. Their findings support the previous work by Brewster *et al.* about the benefits of tactile feedback [2007]. They further found that the performance of tactile soft keyboards were comparable to real physical mobile keyboard and can be further improved using multiple specialized actuators providing localized feedback instead of a single actuator that vibrates the whole device [Hoggan et al., 2008].

Another interesting and novel use of haptics in mobile device is the Shoogle [Williamson et al., 2007]. The Shoogle allow users to naturally interact with devices using shakes and tilt. It provides information regarding the mobile device content using the audio-haptic channel facilitating a completely nonvisual interaction. For example, in a message box application, all the messages are rendered as “message balls” producing audio and haptic signals conveying bouncing and collision of these balls when the device is shaken or tilted [Williamson et al., 2007]. The size and weight of the balls can be used to convey the length and priority of the message producing a heavy feeling when a long or important message has been received.

In summary, haptics is a familiar feedback modality in mobile devices. Despite the limitation of the haptic actuator present in mobile devices, it has been shown to improve performance in various tasks like touch typing. The familiarity and unobtrusiveness of the feedback modality make it a very suitable candidate for use with other natural interaction techniques on mobile devices.

3. Gaze Gesture Interaction with Haptic Feedback in Mobile Devices

The combination of gaze interaction with haptic feedback has not been widely studied before. The only literature we are aware of is the work by Meers & Ward [2007]. They studied Haptic rendering of GUI elements for visually challenged users. Using the head position and orientation they estimated the virtual gaze position. They presented the screen object at that gaze point using haptic signals to provide the users with a 2D mental image of the screen [Meers and Ward, 2007]. This study uses "gaze" in an unconventional way and does not provide any information regarding the dynamics of two interaction modalities for an able bodied person.

One of the reasons that motivated this work was the fact that most mobile phone vendors are on the lookout for convenient and natural interaction techniques that can complement the existing touch screen interaction, for e.g. *Siri* - the voice based personal assistant in iPhone devices and touch free interaction in Samsung S4 smartphones including head gestures and air gestures [IPhone-Siri; Samsung-S4]. We predict that gaze tracking will soon find its way to commercial mobile devices. There is already news about several initiatives. For e.g. EyeTribe and Qualcomm Snapdragon SDK have announced their gaze tracking SDK for android mobile devices[EyeTribe; Snapdragon]. The main advantage of gaze in this context is that it facilitates hands free interaction.

As discussed in the previous sections, gaze gestures seems to be the most feasible gaze based interaction technique in mobile environment. Dybdal *et al.* noted that gaze gesture based interaction on mobile devices however results in a high cognitive load among users. They proposed that gaze gesture interfaces in mobile devices should provide adequate support and feedback to the users to reduce this mental load [2012].

The need for nonvisual feedback when using gaze interaction has been discussed [Dybdal et al., 2012]. The reason for not studying touch as a feedback modality in the standard desktop computing environment may have been because of the need for special hardware to provide the tactile feedback. However, Mobile devices have a built in tactile actuator and tactile feedback is one of the most natural, convenient and familiar feedback modalities in these devices. From the previous research, we know that tactile feedback can make the interaction more intuitive and improve the performance in various touch and simple gesture interaction in mobile devices [Hoggan et al., 2008]. This is the prime motivation of this work.

The subsequent sections describe the study that was conducted to evaluate the effectiveness of haptic feedback in two stroke gaze gesture based interaction in mobile devices.

3.1. Experiment Setup

The main purpose of this experiment was to answer the following research questions:

- Does haptic feedback help two stroke gaze gesture based interaction on a mobile device? If so, what would be the best temporal point for providing the feedback in terms of gesture progression?
- Does the users have any subjective preference towards any of the feedback conditions?
- Lastly, how does the user find the interaction? Would they use such an interaction technique if it was made available in a mobile device?

Next, we describe the details of the experiment.

3.2. Mobile Application Design

For this study, a mobile application was developed that can be operated by gaze gestures. The application resembles a typical phonebook with a list of names from which user can select a name and make a call to that person. Figure 14 shows the GUI of the application that was developed.

Figure 14 Mobile application GUI

The application first starts with a vertical list of contacts. The currently selected name is highlighted in the list. The first contact is selected by default when the application starts. Later, when the user scrolls the list using gaze gestures, the selection stays in the middle as in the figure, before moving further down to the bottom of the screen when the list approaches its end. The names in the list were alphabetically arranged and center aligned on screen. There were a total of 18 names in the phonebook application. This meant that not all names were visible at the same time on screen.

Once the user selects a name with gaze gesture, the application navigates to the contact preview page. From this page, the user could either go back to the contact list or proceed to call the previewed contact.

3.3. Gesture Design

Below we describe some of the considerations that were taken into account while designing the gaze gestures.

- ***Off-screen gestures***

When interacting with a mobile device, we usually hold the device at a distance of 20-40 cm from our eyes. The screen dimensions of the mobile device are rather small and usually mobile phone users do not hold their phone upright in front of their eyes. There is a small tilt and roll in the way we naturally hold these devices. Drewes *et al.* observed that in normal mobile phone usage, users tends to hold the phone with an approximate tilt angle of 20 degrees and roll angle of 10 degrees [2007]. This further reduces the effective screen area that is available. This means that eye movement associated with gazing the four corners of the screen display is very small.

If the length of the stroke is not sufficiently large, gaze gesture interaction is unlikely to be robust and suitable for mobile interaction. For this reason, the gestures used in our system were off-screen gestures. Each gesture starts from the center of the device and goes beyond the screen boundary and back to the device screen.

- ***Simple to perform***

It may not be possible to do all types of complex gestures with the eyes. Further, we anticipate that if the gestures are difficult to perform, the users would be discouraged to use the interaction. For the users to accept a new interaction technique, it is important that it is simple and intuitive.

For our study, we relied on simple vertical and horizontal eye saccades as the basic unit of the gestures. Each gesture was composed of two simple strokes starting from the center of the mobile screen in one of the four directions and back to the center of the device. Drewes & Schmidt in their study observed that all their participants could perform simple gestures like these even without any background visual cue or fixation points [2007].

- ***Avoiding Midas touch problem***

The gaze gestures used should be such that they don’t occur in the natural eye movement associated with interacting with the environment and navigating through the usage context. It is very common for a person using the mobile device to glance at a person or location beyond the screen of the device for a short period and then continue the interaction. In order to ensure that such situations do not result in accidental invocation of the gesture, we included a timeout of 500 msec between strokes. This means that once the first stroke is recognized by the gesture recognizer, it waits for up to 500 msec for the second stroke that completes the gesture. If the second stroke does not occur during this period, the first stroke is forgotten. The smaller the value of time-out, the lesser the chances of accidental invocation of the gesture. However, it becomes more demanding to the user to perform the gesture in this quick pace.

- ***Natural mapping to the performing action***

One of the drawbacks of gaze gestures and gesture based interaction in general is that the user should learn and memorize these gestures before interaction. A natural mapping between the gestures and the resultant action would help reduce the cognitive load associated with memorizing the gestures and can result in an intuitive interaction even for first time users. Natural mapping is the basis of response compatibility, a concept widely studied in cognitive psychology and even in different subfields of HCI like human factors [Norman 1988].

In our gaze gesture design, we relied on the principle of natural mapping. An UP gesture moved the focus upwards by one step and DOWN gesture moved the focus downwards by one step. A RIGHT gesture was associated with selection and LEFT gesture associated with cancellation of a selection. This provides a directional mapping for the user and was the basis of our gesture design.

Figure 15 shows the four different gestures. Gestures had a time-out value between strokes (shown as T in the figure)

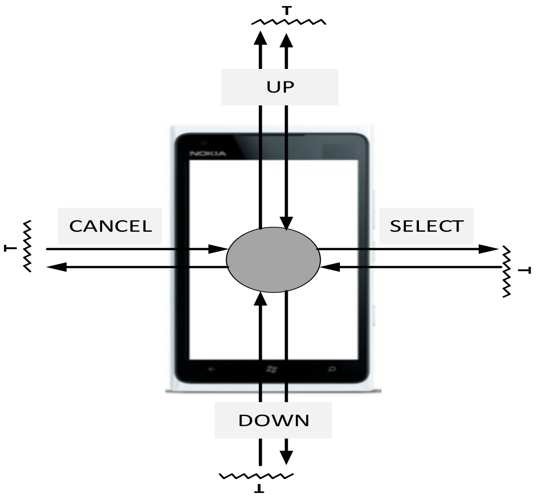

Figure 15 Gesture design

Not all gestures were available in all the pages of the application. The interactions with the application involved the following gestures (figure 16):

- *Contact list* page contains the list of name which the user can scroll using UP/DOWN gaze gestures. From this page, a SELECT gesture navigates the application to *Contact preview* that previews the highlighted contact name.
- From the *contact preview* page, the user can either do a SELECT gesture to call the previewed contact (*calling* page) or do a CANCEL gesture to navigate back to *contact list* page. When navigated back, the last contact previewed is the contact highlighted. UP/DOWN gestures are not available in this page.
- The only valid gesture in the *calling* page is the CANCEL gesture which takes the application back to *contact list*. *Calling* page also has an automatic return functionality which takes the application back to *contact list* page after 5 seconds.

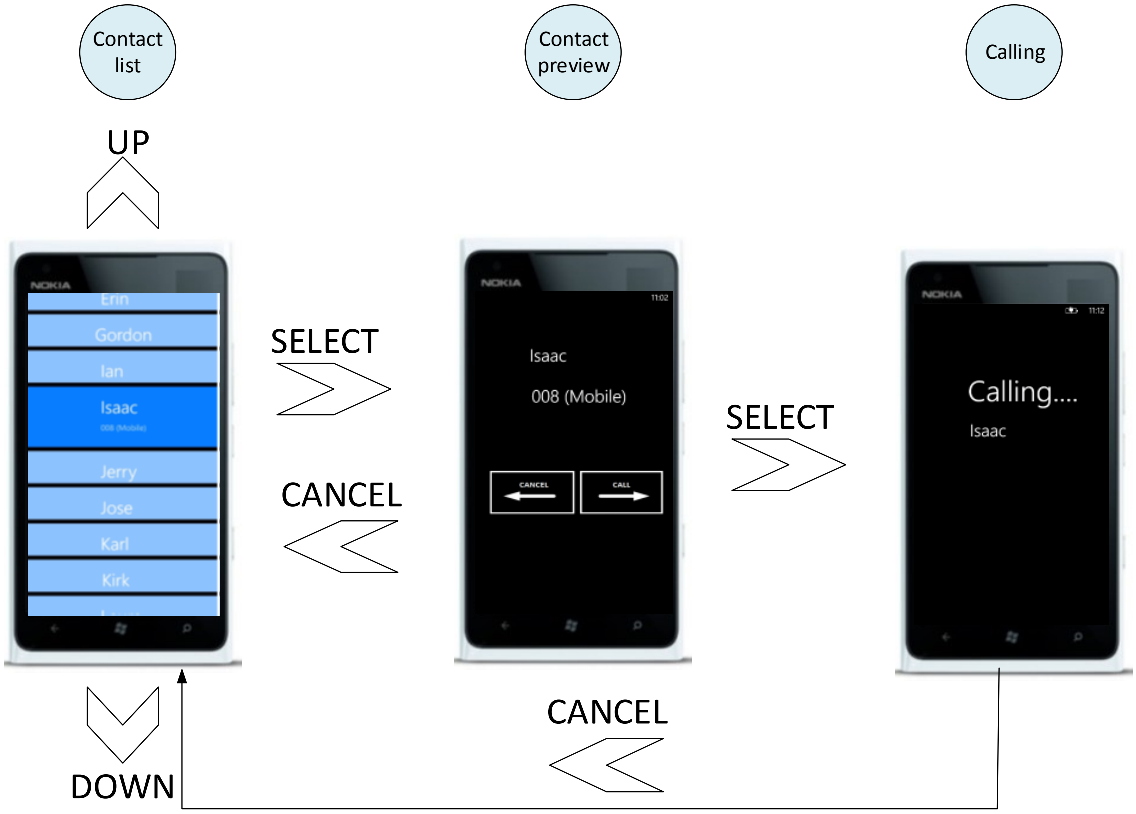

Figure 16 Gesture - Action mapping

3.4. Gesture modelling and recognition

Gestures were modelled as a sequence of spatio-temporal events. For each gesture, the recognition was performed using a finite state machine (FSM) model. The screen area of the gaze tracker was divided into different sectors and state transitions of the FSM were modelled based on gaze fixation duration and gaze saccades with regard to the sector boundaries.

The recognition worked in a similar way for all the gestures. However, for simplicity, in this section we explain only the SELECT gesture recognition. Figure 17 shows the screen sectors associated with the SELECT gesture recognition.

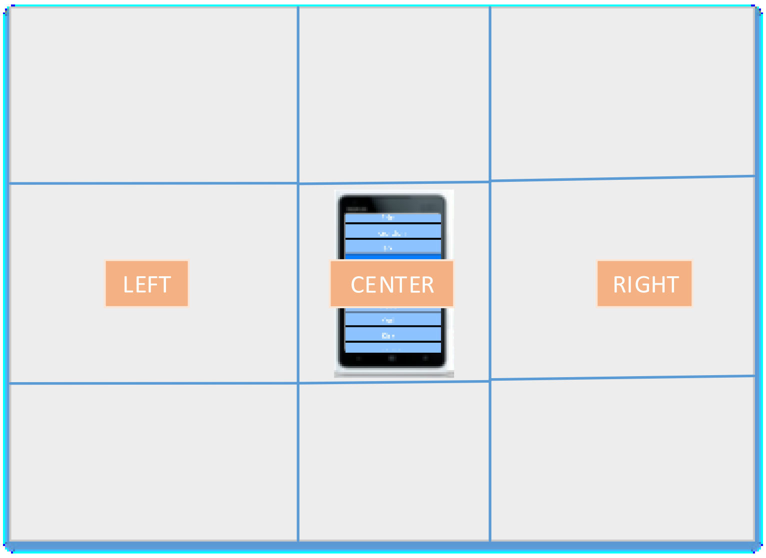

Figure 17 Screen sectors for gesture recognition

Figure 18 shows the corresponding FSM state transition diagram.

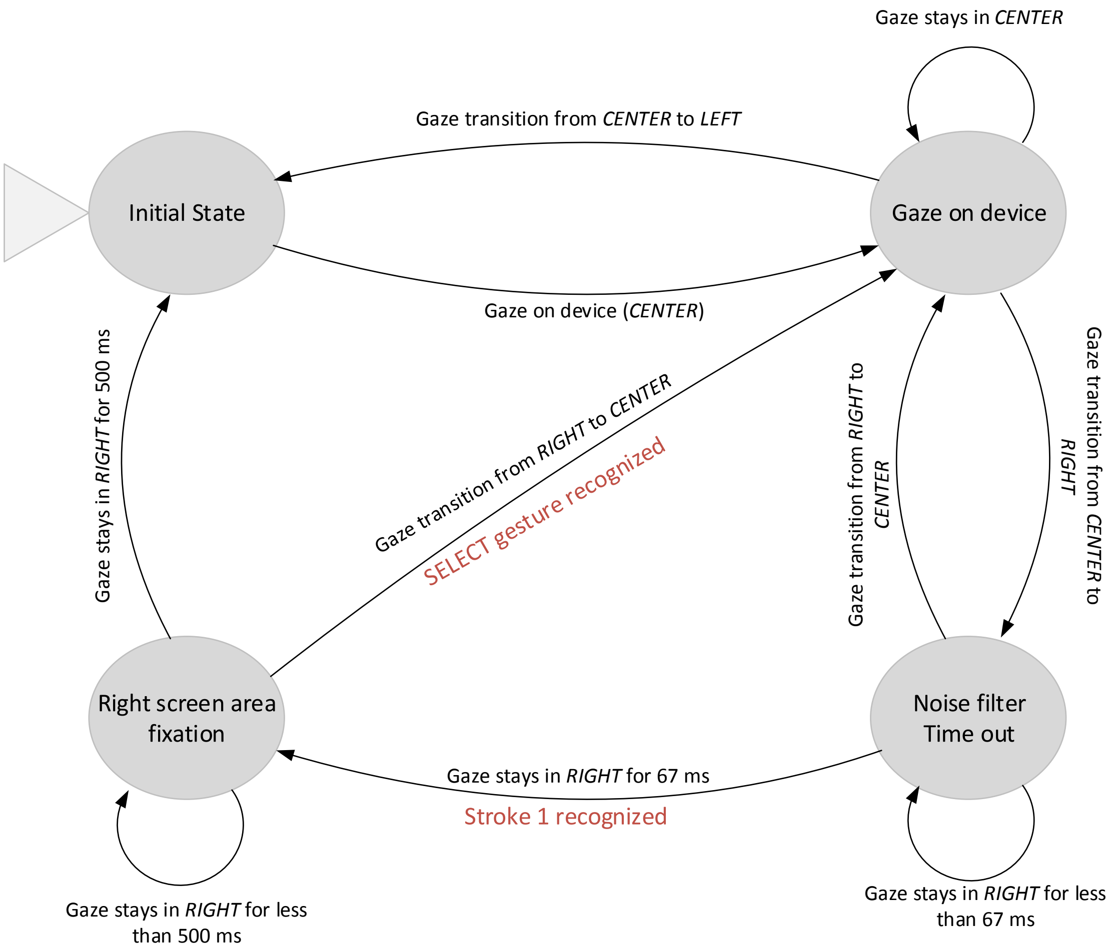

Figure 18 FSM state transition diagram

The following are the events associated with the gesture recognition:

- *Initial state* to *Gaze on device*: User gazes at the device (positioned in the *CENTER* area of the screen).
- The state machine can stay indefinitely in *Gaze on device* state as long as the user is fixating at the mobile device.
- *Gaze on device* to *noise filter time out* state: User makes a gaze saccade from *CENTER* to *RIGHT*.
- The state machine can stay in *noise filter time out* state for a predefined duration T1 (67 ms), if the gaze point continues to stay in *RIGHT* screen area.
- *Noise filter time out* state to *sector 2 fixation*: If T1 second is exceeded in *Noise filter time out* state, the FSM makes a transition to *sector 2 fixation* state. This transition completes the recognition of one stroke of the gesture.
- *Sector 2 fixation* to *gaze on device*: User makes a gaze saccade from *RIGHT* to *CENTER* area of the screen. This completes the SELECT gesture recognition.

Alternately, if the user makes a gaze saccade from *RIGHT* to *CENTER* screen area while in the *noise filter time-out* state, the FSM changes state to *gaze on device* and no gesture is recognized. The motivation for this design is to avoid accidental triggering of the gesture due to low gaze tracking data precision. In such cases, if the user fixates on the sector boundaries, there are chances that alternate gaze samples fall on different sectors causing unintended gestures.

Further, if the user continues to fixate at *RIGHT* screen area for more than 500 ms/30 samples, the FSM changes state from *Right screen area fixation* to *initial state* which resets the gesture recognition. This design was adopted as interruptions are very common in the mobile usage context. During the interaction, the user could momentarily shift his attention to an object of interest in the surrounding and later continue with the interaction. This design reduces the chance of triggering a gesture accidentally due to such attention shifts.

3.5. Feedback design

In order to evaluate the effectiveness of haptic feedback in the gaze gesture interaction and to find the most meaningful point for providing the feedback in terms of gesture progression, we designed 4 haptic feedback conditions. The vibrotactile feedback was provided using the built-in actuator of the mobile phone.

The haptic conditions are discussed below:

1. ***NO***: In this condition no haptic feedback is provided. However, upon a valid gesture completion, the system performs the corresponding action on screen [for example: scroll the focus by one name on completion of an UP gesture]. This provides a visual feedback for the user about the recognition of the gesture.

2. ***FULL***: The system provides a haptic feedback on gesture completion and also the visual feedback associated with performing the corresponding action on screen.

3. ***OUT***: The system provides a haptic feedback when the first stroke is successfully registered and a visual feedback when the full gesture is completed.

4. ***BOTH***: The system provides a haptic feedback on successful completion of the first stroke followed by another haptic pulse and visual feedback on gesture completion.

Figure 19 shows the four haptic conditions. For clarity, the figure only shows the haptic condition associated with the UP gesture. The feedback conditions are the same for all the four directional gestures.

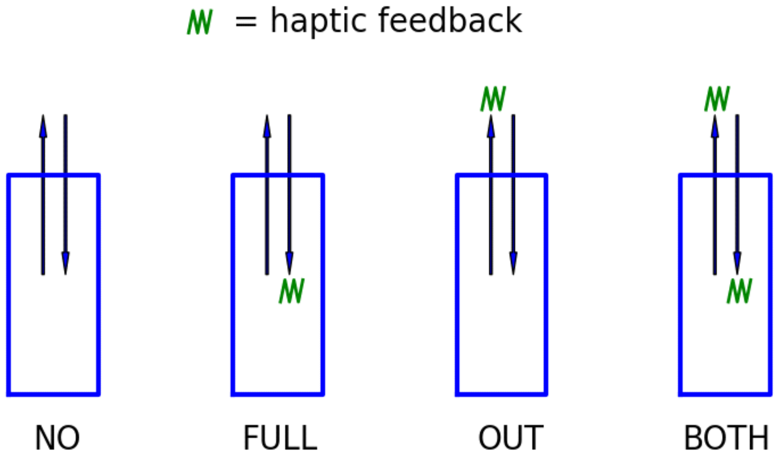

Figure 19 Haptic feedback conditions

3.6. System Design

We used a Tobii T60 remote binocular gaze tracker along with Nokia Lumia 900 mobile device to simulate a gaze tracking capable mobile phone. The screen of the gaze tracker was covered and the participants were asked to hold the mobile device at a particular location marked with a foam on the cover. The figure 20 shows the experimental set up.

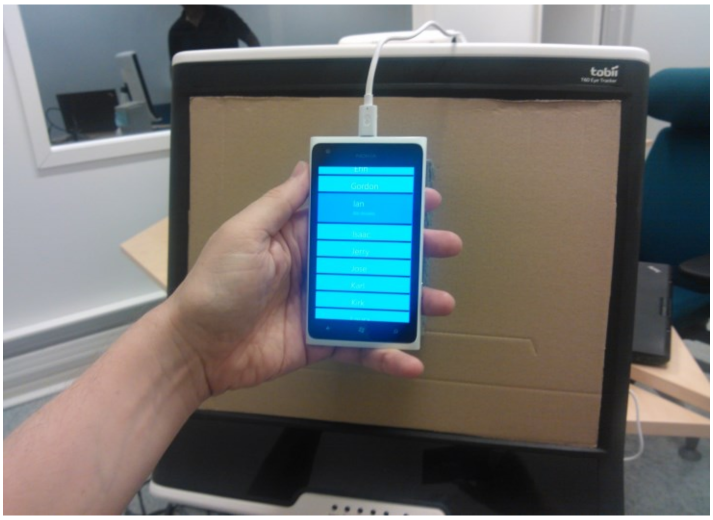

Figure 20 System set up

The Tobii T60 tracker which had a sampling frequency of 60Hz was connected to a laptop computer on which the gesture recognizer was running. The recognizer was a Microsoft windows form application written using .Net framework 4.0. The module retrieved the gaze coordinates and detected gaze gesture events like stroke completion and gesture completion. These events were transferred to the mobile device via a USB based socket connection.

All the application logic ran on the mobile device which responded to the gesture events by invoking the corresponding UI action and providing appropriate haptic feedback based on the condition.

4. Method

This chapter describes the participant demographics, the experimental method and metrics collected. The chapter also discusses the statistical test that was used to test the collected data for significance.

4.1. Participants

For the experiment, we recruited 12 able bodied participants from the university community. Table 1 shows the demographics of the participant.

Gender	Age group (years)	Familiarity with Gaze tracking	Vision	Sense of touch
Male	31-40	Yes	Normal	Normal
Male	<20	No	Normal	Normal
Male	20-30	Yes	Normal	Normal
Female	20-30	Yes	Normal	Normal
Male	20-30	Yes	Normal	Normal
Male	41-50	Yes	Normal	Normal
Male	31-40	Yes	Normal	Normal
Male	20-30	No	Normal	Normal
Male	31-40	Yes	Corrected	Normal
Male	<20	Yes	Normal	Normal
Male	31-40	Yes	Normal	Normal
Female	20-30	Yes	Corrected	Normal

Table 1 Participant Demographics

4.2. Method

The experimental task was designed to be similar to a real usage scenario. The task was to search for a name in the phonebook application and make a call to that person. This task was selected because it was familiar task for the mobile phone users and involved performing different types of actions like scrolling, selection and cancellation. For each haptic condition, the participants performed four such calls. After every successful call, the system waited for five seconds before automatically going back to the contact list with the same highlighted name as in the last successful call. During this five seconds, the participants were shown the next name to call on a paper.

The experiment followed a within-subject design. For a participant, one session consisted of four different test conditions. In order to eliminate the effect of the order of execution of the test conditions, the order of the test was counterbalanced. The table below shows the order of execution for conditions for each participant.

Participant	Test 1	Test 2	Test 3	Test 4
P1	FULL	NO	OUT	BOTH
P2	OUT	BOTH	NO	FULL
P3	BOTH	OUT	FULL	NO
P4	NO	FULL	BOTH	OUT
P5	NO	OUT	BOTH	FULL
P6	BOTH	FULL	OUT	NO
P7	FULL	BOTH	NO	OUT
P8	OUT	NO	FULL	BOTH
P9	OUT	BOTH	FULL	NO
P10	FULL	NO	BOTH	OUT
P11	NO	FULL	OUT	BOTH
P12	BOTH	OUT	NO	FULL

Table 2 Counter balancing scheme used for the experiment conditions

For a participant, the set of names to call were different in all four haptic conditions. The names were selected such that the minimum number of gestures required to complete the task was the same in all four conditions (27 gestures). However, all the participants were asked to call the same names for a given test slot.

Because this was a novel interaction technique, we anticipated considerable learning effect. We expected the performance and the perception of the user to change with the extent of time spent interacting with the system. This learning effect is predominant in the initial phases of the interaction. In order to avoid learning effect in the data collected, we repeated the session twice for each participant. The data for the first session was only used to evaluate the learning effect and all other comparisons regarding performance and user perception were based only on the data collected during the second session.

All the participants followed the same experimental procedure which is briefly explained below.

1. Filling the basic user background questionnaire (appendix A).

2. The participants were introduced to the experiment, the equipment, the gaze gestures and the haptic feedback. The moderator used visual representations for introducing gaze gestures and haptic conditions.
3. The participants were then calibrated to the gaze tracker using the Tobii built in 9 point calibration procedure.
4. All the four haptic test conditions were run twice one after the other.
5. Filling the post-experiment questionnaire (appendix C). In this, the participants compared the four different haptic conditions to answer questions like:
    - Which of the techniques was most comfortable?
    - Which of the techniques was easiest to use?
    - Which of the techniques was the best overall?

Each test condition consisted of the following steps:

1. A short practice session in which the users practiced the gestures and ensured the haptic feedback is felt.
2. Running actual test condition. During this, all test details including gestures identified and state of the mobile application were time stamped and logged separately for later analysis.
3. After the test condition, the users evaluated the test condition answering a brief questionnaire rating the comfort and ease of use of the interaction in 7 point Likert scale (appendix B).

4.3. Parameters investigated during pilot testing

The system was pilot tested multiple times to find the most suitable values for some important design parameters. The following decisions were reached based on the pilot tests.

- *Duration of tactile feedback*

  Duration of the tactile feedback is a key design parameter in our experiment. We expected that there could be some variability between people regarding the duration of the feedback that they can easily perceive. There is also a risk that a larger pulse duration could feel irritating to the users. We decided to use 20 ms long vibrotactile pulses for the feedback. This value was found to be easy to perceive and also comfortable.

- *Center alignment of contact name*
  Users need to focus their gaze on the names of the contact list to read the selected name. When the contact list is left or right aligned, the users are likely to focus their gaze on one side of the device screen. As a result of the pilot tests, we decided to center align the contact names as we expected that the system would be more robust to unintentional invocation of gestures if user's gaze is centered on the device while reading the names.

- *Gesture recognizer minimum and maximum fixation duration*
  The gesture recognizer was designed such that a valid gesture required a short fixation, after the first gaze saccade from the center of the device to outside of the device. The fixation duration had to be between a minimum and maximum value in order to reduce unintended gestures. The varying the maximum fixation duration results in a tradeoff between user convenience and chances of accidental gesture invocation. The values for the minimum and maximum fixation duration (67 msec and 500 msec respectively) were decided based on the pilot tests.

4.4. Metrics

The gaze data points during the session, gaze gestures performed by the user and mobile application events were all time stamped and logged in separate files. The participants also answered questionnaires providing their subjective evaluations of the feedback conditions. These files were analyzed to compute the following measures.

- ***Task completion time***

Task completion time is calculated as the time from the start of each test condition till the end of it (when the last name is successfully called). In every test condition, participants had to ideally do the same number of gestures. So, any noticeable difference in the task completion time between different conditions would signify that the different haptic feedback condition does influence the task performance (fewer errors or lesser time per gesture). The NO feedback task completion time could be used as a control condition to compare if the effect is positive or negative.

There could be large difference in the task completion between people. Hence the median of the task completion time would be a better measure than the mean and was used for the comparisons.

- ***Gestures Per Action (GPA)***

Keystrokes per character (KSPC) is a metric used in text entry research both as a characteristics of the interaction and also as a dependent measure. When used as a measure it signifies the errors committed and correction overhead of these errors during a text entry task. KSPC is defined as the ratio of keystrokes performed to produce the text and the minimum number of keystrokes required to produce the same [Soukoreff and MacKenzie, 2001]. KSPC has an ideal value of 1.

We devised a similar metric, Gestures per Action (GPA), to measure the errors committed and the effort invested in correcting these errors. We defined GPA as the ratio of the number of performed gestures to the minimum number of gestures required to complete the task. GPA has an ideal value of 1 when the task is completed with minimum number of gestures. The value of GPA increased if the user did wrong selections or overshoot the focus and needed further gestures in correcting these errors.

- ***Median value of subjective evaluations on comfort and ease of use***

Feedback conditions could have a positive or negative influence on the overall ease and comfort of the interaction. For any interaction to be accepted, it is important for it to be comfortable and easy to use for the users.

After each test condition, our participants were asked to rate the feedback condition in terms of comfort and ease of use in the Likert scale. The median value of these subjective evaluations was used to compare the feedback conditions and find if the participants particularly liked/disliked any condition in comparison to others.

4.5. Statistical Analysis

In order to test our results for statistical significance, we relied on non-parametric pairwise randomization tests. The task completion time varied largely between participants and an assumption of normality of distribution, which is required for the parametric approaches, was not practical. So we took the safer side of using a non-parametric approach.

In these tests, the null hypothesis (H_o) is that the difference scores of an observation is equally likely to be positive or negative. We draw a large number of samples (n=10,000) with replacement from the observed sample distribution and randomly assign a sign to the difference score [Howell, 2008]. From the resultant frequency distribution for the median, we find the probability of getting a median value as high as obtained in our observed data. A probability of $p < 0.05$ (two-tailed test) suggests that when H_o is true, it is highly unlikely to get a median value as that we observed in the data and hence we can proceed to reject the null hypothesis.

5. Results

This section details the results of the experiment conducted to evaluate the effectiveness of haptic feedback in two stroke gaze gesture interaction in a mobile device.

5.1. Data Considerations

To get data for 12 participants, we had to replace the data for 4 participants with new ones. Two participants could not complete the experiment due to eye tracking issues and data for two had to be left out due to problems in test execution.

For all the participants, the gaze data followed the same pattern. Figure 21 shows a visualization of all the gaze data from an experiment. The rectangle in the middle shows the location of the mobile devices. As expected, most of the gaze points were on the mobile device or in its vertical axis. There were also a few gaze points in the horizontal axis of the device. The task required the user to perform gestures following the vertical and horizontal axes. Some cluster of gaze points were also present in the corner of the device screen and could indicate short span of interruption in the interaction when the moderator presented the participant with the next name to call.

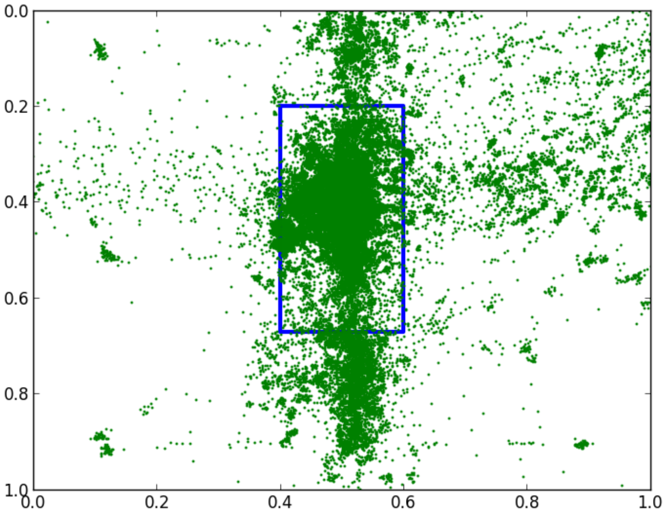

Figure 21 Visualisation of all gaze points from an experiment

5.2. Learning Effect

Figure 22 shows the boxplot for the completion time in seconds for the eight different sessions (T1 – T8). The median time taken to complete the task in slot 1 (T1) was considerably larger than the others, while for T2, it was slightly larger than the rest of the six sessions. The median task completion time for sessions (T5 – T8) is approximately the same

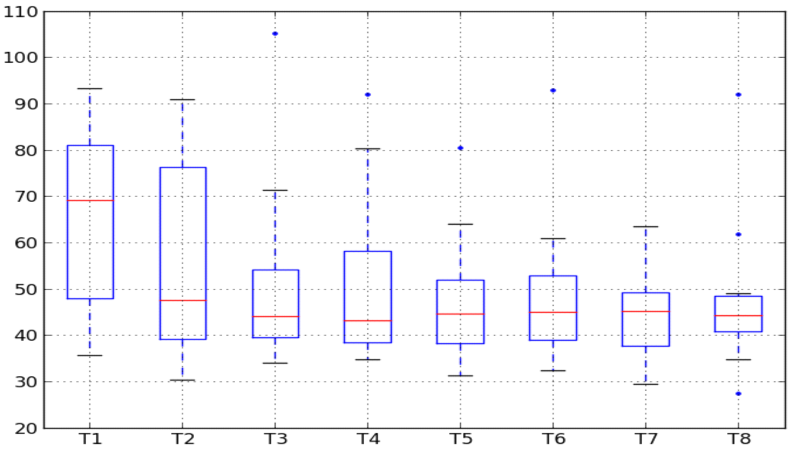

Figure 22 Task completion time per test slot

Figure 23 (best read in colour) shows the difference in task completion between the first session (T1 – T4) and second session (T5 – T8) for various feedback conditions. The reduction in task completion time was larger for conditions NO and FULL when compared to OUT and BOTH conditions.

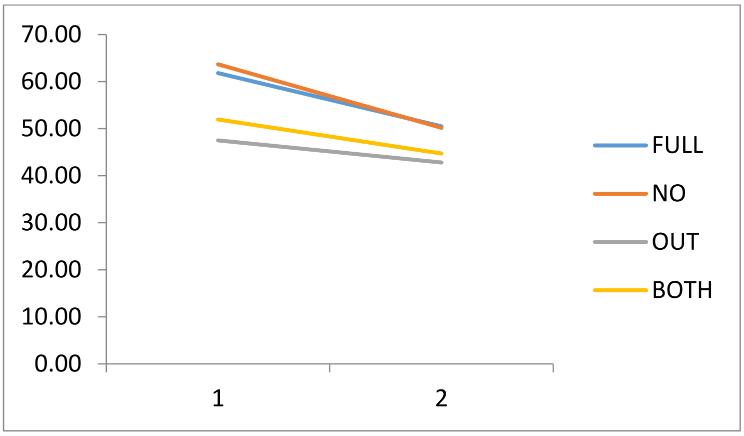

Figure 23 Difference in task completion time

In order to eliminate the learning effect in the results, only sessions T5- T8 are considered for further analysis and synthesis.

5.3. Task Completion Time

The task completion time for conditions NO and FULL was approximately the same. Similarly, task completion time for OUT and BOTH followed the same pattern. Figure 24 shows the boxplot of task completion time for different haptic conditions. The largest difference (17%) in median task completion time was between conditions OUT and FULL.

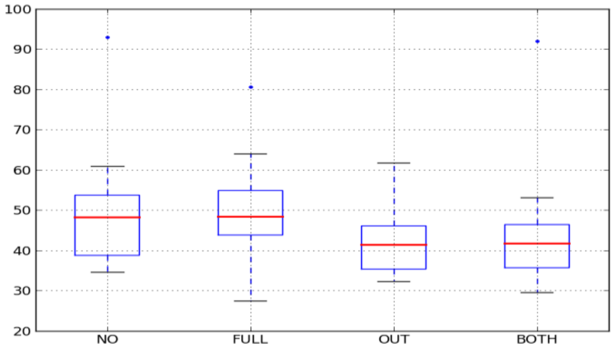

Figure 24 Task completion time for different conditions

A pair wise randomization test for the median showed no significant difference in task completion time between conditions NO and FULL (p=0.45), OUT and BOTH (p=0.39). However, the difference between conditions OUT and NO, and OUT and FULL was statistically significant, p<0.029 and p<0.019 respectively.

5.4. Gestures Per Action (GPA)

The participants performed more gestures to complete the task in the condition NO than in any other conditions. The biggest difference between median GPA was between conditions NO and FULL, about 17% more gestures were performed to complete the task in condition NO. The median GPA value for feedback conditions FULL, OUT and BOTH was relatively less when compared to condition NO. Figure 25 shows the boxplot of GPA for the different conditions.

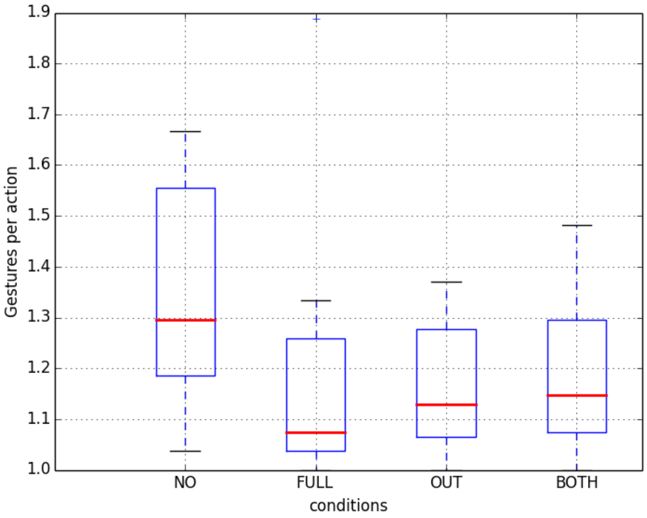

Figure 25 Gestures per action for different conditions

5.5. Subjective Evaluation

In the subjective evaluation for ease of use and comfort of user (appendix B), the NO haptic condition was rated relatively less comfortable and less easy to use compared to other haptic conditions. All other conditions were rated approximately the same in terms of ease of use. BOTH condition was rated slightly better than OUT and FULL in terms of user comfort. Figures 26 and 27 shows the boxplot of the ratings for each condition for the questions "*Was the technique easy to use?*" and "*Was the technique comfortable to use*?" in the condition evaluation questionnaire respectively.

The subjective evaluations on *comfort of use* for NO feedback condition was significantly different from feedback conditions FULL ($p<0.014$), OUT ($p<0.045$) and BOTH ($p<0.004$). Similarly, the subjective evaluations on *ease of use* for NO feedback condition was significantly different from feedback conditions FULL ($p<0.02$), OUT ($p<0.001$) and BOTH ($p<0.002$). This suggests that NO feedback condition was less comfortable and less easy to use when compared to others.

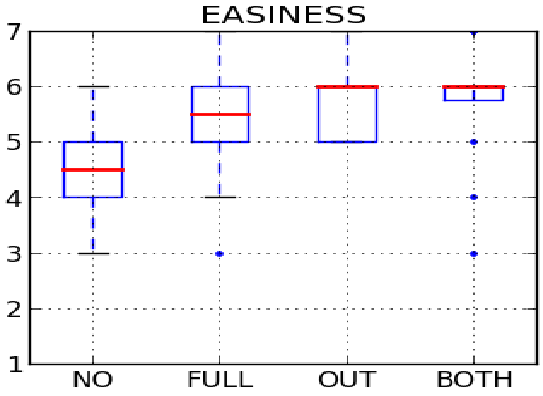

Figure 26 Subjective evaluation of ease of use

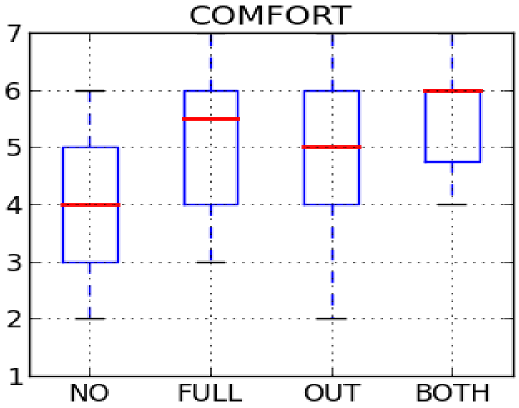

Figure 27 Subjective evaluation of user comfort

5.6. Other Results

In the post-experiment questionnaire (appendix C), 6 out of 12 participants felt that BOTH condition was overall the most comfortable and 8 out of 12 felt BOTH condition was overall the easiest to use. Notably, no participants felt that the NO condition was overall best in either of the criteria .

8 out of 12 participants felt that they would use gaze interaction on a mobile device if available. When asked about the usage scenario, most of the participants listed situations where hands were either not available (e.g. when user is wearing gloves in winter) or engaged with other task (e.g. carrying other materials in one hand).

Further, 10 out of 12 participants felt their eyes to be more tired after the interaction.

6. Discussion

Appropriate and meaningful feedback is known to improve human computer interaction [Brewster et al., 2007; Majaranta et al., 2006]. Not surprisingly, our results suggests the same. Haptic feedback helps in mobile gaze gesture interaction. Participants took longer time to complete the task in the NO feedback condition. The participants could also complete the tasks with fewer number of errors in all the haptic feedback conditions. The OUT and BOTH feedback conditions were found to be the best performing in terms of task completion time.

Norman notes that gestural interfaces should provide feedback at meaningful positions when gestures are being made for ease of interaction [Nielsen, 1993]. OUT and BOTH conditions provided a vibrotactile feedback when the gesture was half completed, that is when eyes moved from the screen to outside the border. Such gesture progression feedback could help in two ways.

- It conveys to the user that the system has recognized one part of the gesture and the user can do the other half for completing the gesture. This reduces the chances of users performing too long gestures or fixating for longer duration outside the screen area. Without this feedback, it could be difficult for users to know whether their fixation outside was long enough for the system to recognize the gesture and when to move the gaze back to the screen.

- It also indicates to the user when a normal gaze movement was wrongly identified as a stroke in the gesture. This provides users the opportunity to correct their gaze behavior before it results in a wrong gesture.

The participants performed tasks more efficiently when OUT feedback was provided when compared to FULL feedback condition. The participants also subjectively preferred the OUT and BOTH feedback in terms of ease of use and comfort of the interaction. The FULL feedback condition also resulted in low median GPA value. However, the FULL or completion feedback is provided only when the gesture is fully recognized and may appears too late for the user to help in the interaction. The median task completion time for the FULL feedback condition was relatively higher than OUT and BOTH. Part of this may be because the user fixated outside the screen for a larger duration of time than what is required.

Further, BOTH feedback was rated by majority of participants as the overall best feedback condition. This should be taken into account when designing gaze gesture based interactions on a mobile device especially when auditory feedback is not appropriate.

The average task completion time reduced considerably after the first time slot (T1) confirming the learning effect. The task completion time for slots (T5 - T8) was approximately the same, indicating that the effect of learning was not prominent in the second session. For all haptic conditions the task completion time reduced between the first session (T1 – T4) and second session (T5 – T8). The reduction was larger for NO and FULL feedback. This indicates that the effect of feedback on completion time is even more prominent during the learning phase.

8 out of 12 participants felt that they would use such an interaction if available on their mobile device. However, it should be noted that most of our participants were new to gaze tracking. The gaze gesture based interaction on the mobile device is a novel interaction technique. The interaction worked well in the stable indoor conditions of the laboratory. For the technology itself to be accepted and used by the user community, the technology should be consistent and the interaction intuitive. Our results suggests that, assuming that all the technical and interaction related challenges can be met, there may be potential for this technique in the consumer market.

Majority of our participants indicated that their eyes felt more tired after the experiment. Gaze gestures require users to perform unnatural eye movements which may lead to eye fatigue. We could expect more experienced users to perform such gaze gestures more easily than novice users. In such cases, their eyes may feel less tired. Nevertheless, such interaction technique may be best suited for short burst interactions when hands are occupied with other tasks. Such gaze based interaction with haptic feedback may also have promise when coupled with other interaction modalities in a mobile device.

Our study has its limitation. The study was done in stable indoor conditions, it could be beneficial to study the efficiency of gaze gesture input with haptic feedback in outdoor conditions with its inherent noise, vibrations and gaze tracking errors. Further, our study used simple two stroke off-screen gestures using haptic feedback modality. Further research is required to understand how feedback influences performance in interactions involving complex gaze gestures and different feedback modalities. Most of the participants used in the experiment were novice users and they used the system for as little as 5 to 10 minutes before the start of the experiment. Our results suggest that, with this small trial duration, we could eliminate most of the learning effect from the results. However, the need for feedback could be very different for an expert user with hours of practice. Such users after repeated use might understand the system so well that they could use the system confidently even without a gesture progression feedback. Further research would be required to know if that is really the case.

Further, in our study participants were instructed to perform a specific task. There were no outside interruptions that required the participants to shift their attention to any other objects in the surrounding while performing the tasks. This type of interruptions are common in the mobile usage context. When OUT feedback is used, there are chances that the system would falsely recognize such an attention shift as a stroke of the gesture and provide vibrotactile feedback. Such incorrect recognitions and resultant vibrotactile feedback may be considered as unwanted or frustrating by the users. One possible way to reduce such unintended gestures would be to limit the end of the stroke area for each gesture. This would result in less false recognition of strokes but on the other hand, would make performing a valid gesture more difficult. Our study did not consider such attention shifts and this will need to be explored further to understand how the trade-off affects the subjective evaluations of the users.

7. Conclusions and Future work

Both gaze based input and haptic feedback modality has its advantages which are discussed in the previous sections. Gaze interaction itself is gaining popularity as an interaction technique for the masses. Gaze interaction could be beneficial for the general users in mobile devices and other wearable devices as an additional input modality. In all these cases, user feedback is critical to make the interaction more effective and intuitive. This thesis work focused on the use of haptics as a feedback modality in gaze gesture based mobile interaction.

This chapter summarizes the thesis by presenting the findings in relation to the research questions that motivated this study. This chapter also provides insight into the future research possibilities in this domain.

The main purpose of this study was to answer the following research questions:

- Does haptic feedback help two stroke gaze gesture based interaction on a mobile device? If yes, what is the best point for providing the feedback in terms of gesture progression?

  Our results suggest that haptic feedback is beneficial for two stroke gaze gesture interaction on the mobile device. Feedback helped reduce errors and improved performance. OUT and BOTH feedback conditions were found to be equally effective in terms of task completion time. Both conditions gave a vibrotactile feedback when the system recognized one stroke or half the gesture, indicating gesture progression.

- Does the users have any subjective preference towards any of the feedback conditions?

  There were no statistically significant difference between the subjective evaluations of the three haptic feedback conditions. However, NO feedback condition was statistically different from all the other conditions. This suggests that our participants found the haptic feedback conditions to be significantly more comfortable and easy to use than the condition with no haptic feedback.

  BOTH feedback condition was rated by majority of our participants as the overall best in terms of comfort and ease of use.

- How does the users find the interaction? Would they use such an interaction technique if it was made available in a mobile device?

  Majority of our participants felt they would use such an interaction technique if available on a mobile device. It should be noted that there are many challenges before this interaction could be replicated in a truly mobile environment. However, our results suggest that, if the challenges are met, the technique could be usable for short burst interactions in scenarios where one hand is occupied with other tasks.

Currently the haptic capability of commercial mobile devices are limited to a fixed amplitude vibration of the whole device. It would be interesting to study how enhanced haptic feedback, for example multiple feedback patterns and localized tactile stimulation could benefit gaze interaction in these devices. Some of the usage scenarios could be differentiating contact category based on the vibration pattern while scrolling or differentiating read and unread emails scrolling through a mailbox etc. If gaze tracking system is available as a standalone interface to the mobile device (e.g. head worn trackers interfaced to mobile device), rich haptic feedback could help in eyes free interaction, where the user is not necessarily attending the phone display at all times.

Our study used a phonebook like application with 18 contact names. In real scenarios, the list could be longer and in those cases discrete gestures like those used in our study may not be suitable. Continuous scrolling could be used in those cases to reduce the number of gestures required to be performed. Continuous scrolling using gaze gestures could also be used in many other application scenarios, for example mailbox with large number of emails/messages or scrolling of webpages. The most suitable position for providing the haptic feedback with respect to gesture progression could be different when continuous scrolling is used. Further research is required to understand the effect of feedback in such scenarios.

Further, the true power of gaze interaction can only be harnessed when it is combined with other input modalities. It would also be interesting to study and evaluate a multimodal interaction technique on mobile device making use of touch and tilt in combination with gaze interaction with haptic feedback. Mobile gaming could be an exciting application area for such a system.

A more recent trend in HCI has been wearable computing. We anticipate that in future gaze tracking would be possible on head worn devices and eye gears which are not only simple and small but also fashionable. Such devices could also feature near eye displays. Google Glass is such a device which comes in the form factor of eye glass with a head mounted display [Google-Glass]. Google Glass currently does not support gaze tracking. Such devices would enable pervasive computing allowing simple gaze based interaction with real world objects. Feedback would be critical to make these interactions efficient and intuitive. Further, wearable haptic devices have also been emerging in various form factors like wrist watch and embedded in clothing etc. Haptic feedback would be one of the possible feedback modalities worth exploring further in these scenarios. Such wearable gaze trackers and haptic devices could enable effortless, hands-free and spontaneous interactions.

In summary, gaze interaction with haptic feedback has the potential to be used as a natural interaction technique in the mobile context. This study provides answers to a few questions, many other remain to be studied in this area. More research would be required in other mobile usage scenario and context to further refine the interaction design and harness the advantages of these modalities. To conclude, I am hopeful that this work and the results obtained would encourage more research in this area.

References

[Banter,2010] B. Banter, “Touch screens and touch surfaces are enriched by haptic force-feedback,” *Information Display*, 2010.

[Brennan and Hulteen,1995] S. E. Brennan and E. A. Hulteen, “Interaction and feedback in a spoken language system: A theoretical framework,” *Knowledge-Based Systems*, **8**. 143–151, 1995.

[Brewster et al.,2007] S. Brewster, F. Chohan, and L. Brown, “Tactile feedback for mobile interactions,” in *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI ’07*, 2007, 159.

[Brown and Kaaresoja,2006] L. Brown and T. Kaaresoja, “Feel who’s talking: Using tactons for mobile phone alerts,” *CHI’06 extended abstracts on Human factors in computing systems*, 2006, 604–609.

[Brown et al.,2006] M. Brown, M. Marmor, Vaegan, E. Zrenner, M. Brigell, and M. Bach, “ISCEV standard for clinical electro-oculography (EOG) 2006.,” *Documenta ophthalmologica. Advances in ophthalmology*, **113** (3), 2006, 205–12.

[Bulling and Gellersen,2010] A. Bulling and H. Gellersen, “Toward mobile eye-based human-computer interaction,” *IEEE Pervasive Computing*, **9**, 2010.

[Bulling et al.,2009] A. Bulling, D. Roggen, and G. Tröster, “Wearable EOG goggles: Seamless sensing and context-awareness in everyday environments,” *Journal of Ambient Intelligence and Smart Environments*, **1.2**, 2009, 157–171.

[Clark and Brennan,1991] H. H. Clark and S. E. Brennan, “Grounding in communication. Perspectives on socially shared cognition.,” in *Perspectives on socially shared cognition*, 1991, 127–149.

[Dickie et al.,2005] C. Dickie, R. Vertegaal, C. Sohn, and D. Cheng, “EyeLook: Using attention to facilitate mobile media consumption,” in *UIST ’05 Proceedings of the 18th annual ACM symposium on User interface software and technology*, 2005, 103–106.

[Drewes,2010] H. Drewes, “Eye gaze tracking for human computer interaction,” Ludwig Maximilian University of Munich, 2010.

[Drewes et al.,2007] H. Drewes, A. De Luca, and A. Schmidt, “Eye-gaze interaction for mobile phones,” in *Mobility ’07 Proceedings of the 4th international conference on mobile technology, applications, and systems and the 1st international symposium on Computer human interaction in mobile technology*, 2007, **07**, 364–371.

[Drewes and Schmidt,2007] H. Drewes and A. Schmidt, “Interacting with the computer using gaze gestures,” in *INTERACT’07 Proceedings of the 11th IFIP TC 13 international conference on Human-computer interaction*, 2007, 475–488.

[Duchowski,2002] A. T. Duchowski, “A breadth-first survey of eye-tracking applications.,” *Behavior research methods, instruments, & computers : a journal of the Psychonomic Society, Inc*, **34**, 2002, 455–470.

[Duchowski,2007] A. T. Duchowski, *Eye tracking methodology: theory and practice*. Springer, 2007, 373.

[Dybdal et al.,2012] M. L. Dybdal, J. S. Agustin, and J. P. Hansen, “Gaze input for mobile devices by dwell and gestures,” *Proceedings of the Symposium on Eye Tracking Research and Applications*, 2012, 225–228.

[Ernst and Bülthoff,2004] M. O. Ernst and H. H. Bülthoff, “Merging the senses into a robust percept.,” *Trends in cognitive sciences*, **8**, 2004, 162–169.

[EyeTribe] EyeTribe, “EyeTribe android based gaze tracking.” [Online]. Available: https://theeyetribe.com/. [Accessed: 03-Nov-2013].

[Frischen et al.,2007] A. Frischen, A. P. Bayliss, and S. P. Tipper, “Gaze cueing of attention: Visual attention, social cognition, and individual differences.,” *Psychological bulletin*, **133**, 2007, 694–724.

[Gentner and Nielsen,1996] D. Gentner and J. Nielsen, “The anti-mac interface,” *Communications of the ACM*, **39**, 1996, 70–82.

[Gips and Olivieri,1996] J. Gips and P. Olivieri, “EagleEyes: An eye control system for persons with disabilities,” *The Eleventh International Conference on Technology and Persons with Disabilities*, 1996.

[Google-Glass] Google-Glass, “Google glass official webpage.” [Online]. Available: http://www.google.com/glass/start/. [Accessed: 03-Nov-2013].

[Gross et al.,2008] H. Gross, F. Blechinger, and B. Achtner, *Handbook of optical systems, survey of optical instruments*, Volume 4. **4**. 2008.

[Grunwald,2008] M. Grunwald, *Human haptic perception*. Birkhaeuser Verlag AG, Boston Basel Berlin., 2008.

[Hansen and Ji,2010] D. W. Hansen and Q. Ji, “In the eye of the beholder: a survey of models for eyes and gaze.,” *IEEE transactions on pattern analysis and machine intelligence*, **32**, 2010, 478–500.

[Hansen and Pece,2005] D. W. Hansen and A. E. C. Pece, “Eye tracking in the wild,” *Computer Vision and Image Understanding*, **98** (1), 2005, 155–181.

[Hanson et al.,2009] J. V. M. Hanson, D. Whitaker, and J. Heron, “Preferential processing of tactile events under conditions of divided attention.,” *Neuroreport*, **20** (15), 2009, 1392–6.

[Hoggan et al.,2008] E. Hoggan, S. A. Brewster, and J. Johnston, “Investigating the effectiveness of tactile feedback for mobile touchscreens,” in *Proceeding of the twenty-*

sixth annual CHI conference on Human factors in computing systems - CHI ’08, 2008, 1573.

[Howell,2008] D. C. Howell, *Statistical methods for Psychology*, 7 th edition. Wadsworth Publishing, 2008, 661–670.

[Hyrskykari et al.,2012] A. Hyrskykari, H. Istance, and S. Vickers, “Gaze gestures or dwell-based interaction?,” *Proceedings of the Symposium on Eye Tracking Research and Applications*, (2010), 2012, 229–232.

[Immersion] Immersion, “Immersion - official webpage.” [Online]. Available: http://www.immersion.com/. [Accessed: 03-Nov-2013].

[IPhone-Siri] IPhone-Siri, “iPhone personal assistant SIRI official webpage.” [Online]. Available: http://www.apple.com/ios/siri/. [Accessed: 03-Nov-2013].

[Isokoski,2000] P. Isokoski, “Text input methods for eye trackers using off-screen targets,” *Proceedings of the symposium on Eye tracking research applications ETRA 00*, 2000, 15–21.

[Istance et al.,2010] H. Istance, A. Hyrskykari, L. Immonen, S. Mansikkamaa, and S. Vickers, “Designing gaze gestures for gaming: an investigation of performance,” *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, **1** (212), 2010, 323–330.

[Kaaresoja and Linjama,2005] T. Kaaresoja and J. Linjama, “Perception of short tactile pulses generated by a vibration motor in a mobile phone,” *First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. World Haptics Conference*, 2005.

[Kobayashi and Kohshima,2001] H. Kobayashi and S. Kohshima, “Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye.,” *Journal of human evolution*, **40**, 2001, 419–435.

[Kristoffersen and Ljungberg,1999] S. Kristoffersen and F. Ljungberg, 'Making Place' to make IT work : Empirical explorations of HCI for mobile CSCW,” in *Proceedings of the international ACM SIGGROUP conference on Supporting group work*, 1999, 276–285.

[Lederman,1991] S. J. Lederman, “Skin And touch,” in *Skin and touch*, 1991, 51–63.

[Lederman and Klatzky,2009] S. J. Lederman and R. L. Klatzky, “Haptic perception: a tutorial.,” *Attention, perception & psychophysics*, **71**, 2009, 1439–1459.

[Li et al.,2006] D. Li, J. Babcock, and D. J. Parkhurst, “openEyes: A low-cost head-mounted eye-tracking solution,” in *Proceedings of the 2006 symposium on Eye tracking research & applications*, 2006, 27–29.

[Linjama et al.,2005] J. Linjama, J. Hakkila, and S. Ronkainen, “Gesture interfaces for mobile devices-minimalist approach for haptic interaction,” *CHI Workshop: Hands on Haptics: Exploring Non-Visual Visualisation Using the Sense of Touch*, 2005.

[Linjama and Kaaresoja,2004] J. Linjama and T. Kaaresoja, “Novel, minimalist haptic gesture interaction for mobile devices,” *Proceedings of the third Nordic conference on Humancomputer interaction NordiCHI 04*, 2004, 457–458.

[Linjama et al.,2003] J. Linjama, M. Puhakka, and T. Kaaresoja, “User studies on tactile perception of vibrating alert,” *HCI International, Crete, Greece*, 2003, 1–5.

[Loomis and Lederman,1986] J. M. Loomis and S. J. Lederman, “Tactual perception,” in *Handbook of perception and human performances 2, 2*, 1986.

[Majaranta,2009] P. Majaranta, “Text entry by eye gaze,” University of Tampere, 2009.

[Majaranta et al.,2009] P. Majaranta, U.-K. Ahola, and O. Špakov, “Fast gaze typing with an adjustable dwell time,” in *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*, 2009, 357.

[Majaranta et al.,2006] P. Majaranta, I. S. MacKenzie, A. Aula, and K.-J. Räihä, “Effects of feedback and dwell time on eye typing speed and accuracy,” *Universal Access in the Information Society*, **5**. 199–208, 2006.

[Majaranta and Räihä,2002] P. Majaranta and K. Räihä, “Twenty years of eye typing: systems and design issues,” *Proceedings of the 2002 symposium on Eye tracking research & applications*, **1**, 2002, 15–22.

[Malmivuo and Plonsey,1995] J. Malmivuo and R. Plonsey, “Bioelectromagnetism: Principles and applications of bioelectric and biomagnetic fields,” in *Journal of medical Engineering Technology*, **20**, 1995, 438–441.

[Meers and Ward,2007] S. Meers and K. Ward, “Haptic gaze-tracking based perception of graphical user interfaces,” in *Information Visualization, 2007. IV’07. 11th International Conference*, 2007, (July), 713–719.

[Miller,1968] R. B. Miller, “Response time in man-computer conversational transactions,” *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*. 267–277, 1968.

[Miluzzo et al.,2010] E. Miluzzo, T. Wang, and A. T. Campbell, “EyePhone : Activating mobile phones with your eyes,” in *Proceedings of the second ACM SIGCOMM workshop on Networking, systems, and applications on mobile handhelds*, 2010, 15–20.

[Morimoto and Mimica,2005] C. H. Morimoto and M. R. M. Mimica, “Eye gaze tracking techniques for interactive applications,” *Computer Vision and Image Understanding*, **98**. 4–24, 2005.

[Nagamatsu et al.,2010] T. Nagamatsu, M. Yamamoto, and H. Sato, “MobiGaze: Development of a gaze interface for handheld mobile devices,” in *CHI’10 Extended Abstracts on Human Factors in Computing Systems*, 2010, 3349–3354.

[Nielsen,1993] J. Nielsen, “Noncommand user interfaces,” *Communications of the ACM*, 1993.

[Norman,1988] D. A. Norman, *The design of everyday things*. Basic Books, 1988.

[Nyström et al.,2013] M. Nyström, R. Andersson, K. Holmqvist, and J. van de Weijer, “The influence of calibration method and eye physiology on eyetracking data quality.,” *Behavior research methods*, **45** (1), 2013, 272–88.

[O’Grady et al.,2008] M. O’Grady, G. O’Hare, and S. Keegan, “Interaction modalities in mobile contexts,” in *Intelligent Interactive Systems in Knowledge-Based Environments*, **106**, 2008, 89–106.

[O’Modhrain and Gillespie,1997] M. O’Modhrain and B. Gillespie, “The moose: A haptic user interface for blind persons,” *Proc. Third WWW6 Conference*, 1997.

[Oulasvirta et al.,2005] A. Oulasvirta, S. Tamminen, V. Roto, and J. Kuorelahti, “Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI,” in *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI ’05*, 2005, 919–928.

[Oyster,1999] C. W. Oyster, *The human eye : Structure and functions*. Sinauer Associates, 1999, 766.

[Pérez-Quiñones and Silbert,1996] M. A. Pérez-Quiñones and J. L. Silbert, “A collaborative Model of Feedback in Human-Computer Interaction,” *Proceedings of the SIGCHI conference on Human factors in computing systems: common ground*, 1996, 316–323.

[Placencia et al.,2011] G. Placencia, M. Rahimi, and B. Khoshnevis, “Development of an economical haptic stimulus device,” in *HCI International 2011–Posters’ Extended Abstracts*, 2011, 257–261.

[Renaud and Cooper,2000] K. Renaud and R. Cooper, “Feedback in human-computer interaction-characteristics and recommendations,” *South African Computer Journal*, 2000, 105–114.

[Rovers and Essen,2006] A. Rovers and H. van Essen, “Using active haptic feedback in everyday products,” *Proc. EuroHaptics 2006*, 2006, 447–453.

[Samsung-S4] Samsung-S4, “Samsung S4 Android Phone official webpage.” [Online]. Available: http://www.samsung.com/global/microsite/galaxys4/. [Accessed: 03-Nov-2013].

[Schmidt,2000] A. Schmidt, “Implicit human computer interaction through context,” *Personal Technologies*, **4**. 191–199, 2000.

[Sibert and Jacob,2000] L. E. Sibert and R. J. K. Jacob, “Evaluation of eye gaze interaction,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2000, 281–288.

[Snapdragon] Snapdragon, “Qualcomm Snapdragon android based gaze tracking SDK.” [Online]. Available: https://developer.qualcomm.com/mobile-development/mobile-technologies/snapdragon-sdk-android.

[Soukoreff and MacKenzie,2001] R. Soukoreff and I. MacKenzie, “Measuring errors in text entry tasks: An application of the Levenshtein string distance statistic,” *CHI’01 extended abstracts on Human factors in computing systems*, 2001, 31–32.

[Stellmach and Dachselt,2012] S. Stellmach and R. Dachselt, “Look & touch: gaze-supported target acquisition,” *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2012, 2981–2990.

[Stone,2001] R. J. Stone, “Haptic feedback: A brief history from telepresence to virtual reality,” in *Haptic Human-Computer Interaction*, 2001, 1–16.

[Tobii-T60] Tobii-T60, “Tobii T60 official webpage.” [Online]. Available: http://www.tobii.com/en/eye-tracking-research/global/products/hardware/tobii-t60t120-eye-tracker/ .

[Williamson et al.,2007] J. Williamson, R. Murray-Smith, and S. Hughes, “Shoogle: Excitatory multimodal interaction on mobile devices,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2007, 1–7.

[Zhao et al.,2012] X. Zhao, H. Xie, and X. Zou, “Gaze-gesture Interaction,” in *International Conference on Computing and Convergence Technology (ICCCT)*, 2012, 1030–1033.

Appendix A: Background Questionnaire

Questionnaire:

Gender: Female Male

Age group: <20 years 20-30 years 31-40 years 41-50 years >50 years

Are you familiar with mobile devices? Yes No

Are you familiar with gaze tracking? Yes No

Are you familiar with haptic (touch) feedback? Yes No

Normal vision? Yes No

If not, what kind of problems?

____

Normal touch sense? Yes No

If not, what kind of problems?

____

Appendix B: Condition Evaluation Questionnaire

Questionnaire:

Was the technique comfortable to use?

1	2	3	4	5	6	7
Very uncomfortable			neutral			very comfortable

Was the technique easy to use?

1	2	3	4	5	6	7
Very difficult			neutral			very easy

Any general comments?

Appendix C: Post Experiment Questionnaire

Questionnaire:

Which of the techniques was most comfortable? 1 2 3 4
Which of the techniques was easiest to use? 1 2 3 4
Which of the techniques was the best, overall? 1 2 3 4

Did you find any of the techniques confusing? Yes No
If yes, why?

Did you feel your eyes tired? Yes No

Would you use this kind of interaction in your mobile? Yes No
If no, why?

If yes, why?

General comments?