Høgskolen i Østfold

Evaluating Conversational User Interfaces when Mobile

Master's Thesis in Applied Computer Science – Interaction Design
Faculty of Computer Science

Arman Balic

December 15th, 2022
Halden, Norway

Abstract

For the past 20 years, the evolution of technology has made computing move from our homes and desktops to join us everywhere we go in the palm of our hands. With the advancement of mobile technology and mobile phones becoming a bigger part of our everyday lives, we find ourselves utilizing smartphones in situations that impair us during smartphone interaction. With this advancement in mobile IT, it still introduces new challenges for mobile interaction with outside factors and multi-tasking situations interfering with our ability to perform tasks on the mobile device. However, with the advancement, the way we interact with such technology has also evolved. With alternative ways to engage with our mobile phones, we also discover that such alternatives has the potential to free us from some of the impairments we might face in situations where mobile interaction occurs simultaneously with other tasks. In this thesis, I explore the underlying issues pertaining to mobile and smartphone interaction in terms of situationally-induced impairments and disabilities, and how such situations can be improved with using a voice assistant when interacting with a smartphone. To achieve this, I conducted a controlled experiment with 15 participants to evaluate and compare regular touch-screen smartphone interaction with Siri-only smartphone interaction in terms of user performance on text message tasks and target selection tasks when walking a fixed path and standing. Results suggest that adopting Siri as an interaction modality when walking significantly improves the walking performance of the participants during both task types, particularly in terms of path accuracy and navigation. However, in terms of task performance, the Siri software is still lacking in fidelity to complete texting tasks efficiently when compared to touch-screen interaction leading to more effort and frustration with users. The results also suggest that Siri is a good option for simple tasks such asking for directions, but not reliable enough for more complex tasks like responding to messages.

Keywords: *Situationally-Induced Impairments and Disabilities, Mobile Interaction, Voice Interaction, Voice Assistant, Smartphone Interaction.*

Acknowledgments

For this thesis, I extend my appreciation to my supervisor, Georgios Marentakis. It has been a pleasure and a unique experience to be able to work with his support, guidance, and encouraging cooperation throughout this project. I would also like to extend my gratitude towards my good friend and classmate Fahad Faisal Said for additional guidance and rich discussions throughout the past year and providing me with tools for enabling my experiment for this project. Finally, I thank my family and close friends for their full support the past year in both good days and difficult days. Working on this project has given me an invaluable experience, and these people kept me motivated through tough and easy times throughout this project work.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

Born in the year of 1996, I have witnessed and experienced the evolution of technology. During my first experience with technology as a child, mobile phones were recognized as tools for communication. Now, it is practically a personalised computer that fits perfectly in your hand. Voice communication was replaced by text messages and buttons were replaced by interactive screens. In the later years, I gained an interest in how the advancement of technology has aided and improved the lives of humans. In my first year of the master´s program, I was introduced to the topic *"Interaction Design"*. In this course, I learned the importance of designing technology to make daily life easier for most people and, in some instances, improve human life in general through the invention of new methods for interaction. The field of Human Computer Interaction sparked my interest and motivation for the topic of this thesis.

During my years being a student at Østfold University College, I had to travel a long distance many times a week to attend my classes on campus. During my travels, I have often received important messages and e-mails that I have been unable to respond to when I travel as I cannot divert my attention from the road and to the mobile phone. Other than school and academic work, I am also a very active person. I like to do activities like running or hiking in my spare time. During these activities, I like to listen to music and, at times, I may again receive calls, e-mails and messages. Other times, I tend to receive important messages or e-mails when I walk outside in crowded areas like the city, when commuting or in a store. In the aforementioned situations, I am either unable to respond to the notifications right away, or I miss them completely as I might not hear them or I am unable to adequately respond to them. While I may not want to respond to a message or e-mail when I am running, I might however want to address some of the messages or e-mails when I am walking or driving from place to place. What if there was a way to adequately interact with my phone when the current use-context does not allow my to do so in the traditional sense by interacting with the touch-screen?

If we delve deeper into the driving example I presented in the previous paragraph, that situation impairs my driving and mobile computing abilities in two use ways; (1) it impairs me from using my phone properly, and (2) it impairs me from driving safely. In the walking situation example, the situation impairs me from interacting with the phone optimally, and focusing on the walking path so that I do not collide with objects or other people. In both of the examples, the context of walking and driving impairs me from utilizing my mobile phone optimally, and the context of mobile computing impairs me from driving and walking safely by diverting my attention away from the road. These impairments are more commonly known as *situationally-induced impairments and disabilities* [3]. Under these circumstances, I always asked myself if there was a way to interact with my phone, without having to divert my attention away from other tasks such as walking or driving? Since I am an owner of an iPhone, these situations have led to me adopting the use of Siri, a Conversational User Interface (CUI) developed by Apple. CUIs, also known as voice assistants or chatbots, are user interfaces that allows the user to issue commands to the system by speaking them rather than using the touch-screen or other physical forms of interaction with technologies like computers, mobile phones or even cars.

1.2 Purpose

For this project work, the main purpose is to evaluate whether the use of a conversational interface will lead to a reduction of situationally-induced impairments when the user is mobile, compared to using a traditional graphical interface in the same setting or situation. To accomplish this purpose, an experimental evaluation will be designed for the use of comparing the usability of a graphical interface and a conversational interface. As already mentioned in the previous section, a conversational interface uses speech and voice input to create novel interactions, while the graphical interface uses the physical touch for interaction. The resulting work from the experimental evaluation will explore how the task performance between the two modes of interaction compare to each other. Also, the resulting work will also explore the differences in path accuracy and cognitive load of the two different interaction modes and whether one is better than the other or if there is no difference between them.

In 2003, the term *Situationally-induced Impairments and Disabilities* was introduced by Andrew Sears and Mark Young [4] and is the second of four *"trends of increasing importance in Human-Computer Interaction"* [5]. As computing technologies are moving further away from the desktop and office space, and becoming more prevalent in mobile situations, the need for context-aware technology is increasing. Technology is being used in settings where outside factors like the environment or the surroundings interfere with how we interact with the mobile technology. Along with outside factors, also having your attention divided between mobile interaction and different tasks simultaneously can have a negative impact on how we interact with technology and complete the other task we do while interacting with the technology like the examples given in section 1.1.

Figure 1.1: Overview of environmental and context factors that affect mobile interaction [1]

Ever since its introduction, the impact and effects of situationally-induced impairments in the context of mobile interaction have been well documented [3, 6, 1, 7, 8], and efforts have been made to counter-act their impact [9, 10, 11, 12]. However, while some studies investigate speech and audio for reading comprehension [13], and some investigating the effects walking has on speech recognition [14], there have been very few studies conducted on the use of conversational user interfaces and voice interaction to counter-act the effects of situational impairments to the best of my knowledge. The aim of this project is therefore to investigate the factors that influence mobile technology interaction, and evaluate voice interaction as an alternative to reduce the effects of situational impairments. The work in this project will compare the task performance of the two interaction methods; graphical and conversational.

1.2.1 Research Question

Based on the introduction and the premises it presents, this study will take a quantitative evaluative approach with the following research question:

RQ 1: How can conversational user interfaces reduce the effects of situationally-induced impairments and disabilities when walking?

To answer this research question, further sub-questions were determined to reach a possible solution to the problem scope. The sub-questions are as follows:

- **RQ 1.1:** Will a conversational user interface improve walking performance during mobile smartphone interaction?
- **RQ 1.2:** Will conversational user interfaces improve smartphone task performance during mobile smartphone interaction?
- **RQ 1.3:** Will conversational interfaces reduce perceived workload when engaging in mobile smartphone interaction?

Furthermore, to answer this research question and the following sub-research questions, I present literature and research related to situationally-induced impairments, its definition and what causes them, related work and theory relevant to SIIDs such as the contextual and environmental factors and how to better understand and reduce the effects of situational impairments. In the subsequent Methodology chapter, I describe and present the methods utilized in this thesis to answer the research questions. After the methodology is presented, I then present how the evaluation is designed, set up and conducted. Later on, I present the results of this thesis in the Results chapter and discuss the finding further in the Discussion chapter. Finally, I conclude the findings of the study in the Conclusion chapter and propose future work and research in the Future Work chapter.

1.3 Research Area

The nature of the research in this thesis investigates situationally-induced impairments and disabilities in mobile computing settings. This project looks at the relationship between the two elements. Firstly, we have situationally-induced impairments, focusing on the aspect of mobile interaction and how it affects users during interaction. Secondly, we have mobile interaction and alternative input and output technologies available for reducing the effects of situationally-induced impairments. In addition, this study focuses on voice- and speech-interaction, which involves utilizing a user's voice for interacting with a mobile device.

1.4 Report Outline

This thesis has been structured into the following chapters:

Chapter 1 is the current chapter, where I introduce the background and personal motivation for this thesis. I present and explain the problem scope, which revolves around the subject of situationally-induced impairments, how it affects mobile IT use, and the incorporation of conversational user interfaces. After defining the purpose of this project, the research questions are then outlined.

Chapter 2 provides an extensive review on situationally-induced impairments and mobile computing, presenting a theoretical background for the project. The chapter will also provide insight on the alternative methods for combating the challenges created by situational impairments during mobile interaction tasks, which places my work in the context of the research questions stated in the introduction. The theory presented will help to establish evaluation methods and requirements to potential design concepts.

Chapter 3 pertains to the design of the experiment that will serve as an evaluation of using voice interfaces when walking. In this chapter, I go further into detail on how the experiment is structured, how the experiment will be set up, what measures will be made during the experiment, what tools, both hardware and software will be used, and how the experiment procedure will be carried out step-by-step.

Chapter 4 presents the results from the experimental evaluation. This findings are divided into several metrics used to answer the research questions outlined in Chapter 1, which include walking performance, task performance and mental workload. The results of these metrics are further analyzed using statistical analysis and the results from the analyses are presented in this chapter.

Chapter 5 pertains to the discussion of the findings from the previous chapter in relation to the research question and subresearch questions. The discussion will also be supported by the theoretical background and conceptual knowledge from the prior data collection techniques and literature review. In addition, I reflect on my experiences with the methodology used for the experiment and the design choices that were made throughout this project work. Furthermore, I also present the reflections of the participants of the experiment in terms of their own experience when conducting the experimental evaluation.

Chapter 6 presents the conclusion to the study, followed by a discussion on the limitations throughout the course of this study, including reflections on how the experiment could have been improved upon to achieve more detailed data. The thesis concludes with suggestions on directions forward for future work in the same scope and field of research as this study.

Chapter 2

Background

This chapter provides insight into the domains of situationally-induced impairments and disabilities (SIID), mobile accessibility, the theories behind the occurrence and origins of SIID's, and conversational user interfaces. Since the thesis concerns impairments when performing tasks on a smartphone in mobile situations, the chapter begins by presenting existing research and literature on accessibility in terms of mobile interaction followed by a review of background research on SIIDs and accessibility in the area of mobile interaction. The following sections and subsections pertain to the different factors that cause SIIDs, the theories of context and multiple resources as factors that determine the occurrence of SIIDs during mobile interaction, and give examples of how the different factors cause SIIDs during mobile interaction. In what context mobile interaction is taking place in, and which tasks are performed in combination with it, is relevant for the research area of this thesis. I will also present the importance of understanding of SIIDs and its relevancy in today´s world.

2.1 Theoretical Background

This section presents the theory behind situationally-induced impairments and disabilities, with focus on multiple resource theory. Furthermore, in subsequent sections, I provide a review over existing literature devoted into investigating the effects of various situational impairments that are relevant for this study.

As smartphones have become an indispensable part of our everyday lives, it is more common to utilize them in situations where contextual factors impairs the user during mobile interaction [6]. Smartphones and other mobile devices, for instance Personal Digital Assistants (PDAs), have made it possible for individuals to complete certain tasks (e.g., taking notes, answering e-mails, communication) while on the move [15]. While this has made life much simpler in terms of IT use and mobility, it still introduces new challenges by introducing outside factors that may interfere with user performance during interaction [15, 1]. For example, a person using a mobile device at the beach on a warm sunny day may struggle to read and interact the device´s screen due to the glare caused by bright sunlight while a person may walk the cold streets of Oslo during winter with gloves on, which can cause difficulties when interacting with the touch-screen of the smartphone. These challenges are known as *situationally-induced impairments and disabilities (SIIDs)*. This is one of many examples of when smartphone interaction occurs in combination with other outside factors, most often in combination with other tasks.

While many have described mobile interaction SIIDs as a consequence of contextual factors [16, 4, 17, 3], in terms of walking, it is more productive to consider mobile interaction as a multi-tasking problem due to the fragmentation in attention when navigating a path and engaging in smartphone interaction succesfully [18, 2, 19]. Due to contextual restrictions, user's needs for smartphone usage may either be delayed or completely forgotten, while other times, the users may even be creating restrictive situational contexts for themselves by engaging in multi-tasking where their attention is divided between an active task and a task to address their informational needs [19].

Furthermore, as walking keeps exposing the user to different contextual and environmental factors due to the user moving through different environments (e.g. entering and exiting buildings, being in environments with a lot of noise, different lighting levels, etc.), it is also usually accompanied by other SIIDs that may further impair a user's ability to perform mobile interaction.

2.1.1 Multiple Resource Theory

When conducting research within the realm of SIIDs and mobile interaction, it is important to note that the challenge for small-device users is not due to severe permanent disability. Rather, it is the result of the environment that the users are working in and the current context of the situation [4]. A small-device user impaired by surroundings, environment and other contextual factors, perform the same amount of errors during task performance as a user with the aforementioned permanent impairments [4]. While broad definitions of context have been offered, rather limited definitions are used in practice and the definitions that are used usually focus on the underlying technologies as opposed to how the difficulties users may encounter when interacting with the technologies [20].

While most research continues to delve primarily into one of the three dimensions presented by Schmidt and colleagues [16] which consist of environment, applications and human factors, the concept of SIID can only be fully addressed and understood if the effects of all three dimensions are considered [20]. Jumisko-Pyykko and Vainio [17] present the five categories for contextual factors in terms of mobile interaction and SIIDs:

- *Physical* - this includes apparent features of a situation or physical circumstances such as location, gradient and altitude, physical objects, orientation, weather and lighting.
- *Temporal* - describes factors of past, present and future situations which includes time of day, week, month or season. This context also covers the time for task

completion and, in the context of mobile interaction, can be classified as either hurried, normal or waiting when interacting in the temporal dimension.

- *Social* - represents other people, their characteristics, their roles and interpersonal interaction during mobile interaction.
- *Task* - represents the multitasking and possible interruptions to mobile interaction related to execution of a task.
- *Technical* - this context focuses on the devices, available infrastructure, facts and system assumptions, sensors and network services. It can be related to hardware, software or other products.

For the purpose of this study, I will focus on the task context category in the context of mobile interaction and SIIDs. The task context can be summarized under the factors multi-tasking, task domain and temporal interruptions, with multi-tasking being considered one of the leading causes of SIIDs [21]. In experimental studies, one of the most common techniques for investigating the effects of multi-tasking is to have the participants walk while completing a set of given tasks. The various studies use different walking conditions, while participants are asked to either maintain walking speed, follow a specific route or avoid obstacles along the way. With this in mind, mobile interaction can be thought of as a dual-task, or multi-tasking, context. The next subsection will delve further into the theory on multi-tasking and temporal interruption, called multiple resource theory.

As this project focuses on whether the use of conversational user interfaces can alleviate the effects that are present when smartphone interaction is performed whilst the user is walking, it is appropriate to consider this situation a dual-task context. As previously mentioned, multi-tasking belongs within the task category of context theory and is one of the leading causes of situational impairments in users who conduct small screen interaction in mobile situations. Furthermore, multiple resource theory can be viewed as an extension to context theory as multi-tasking and temporal interruption is a form of context [17]. To better understand how multi-tasking affects the user in terms of task performance and mental workload, Wickens [2] proposed the *theory of multiple resources* to explain the mechanism and process of multi-tasking. As previously mentioned, context theory explains that context comes from the relationship between the environment, the individual user and its current situation, and the application that is being used by the individual user. However, sometimes users create contextual restrictions on themselves by engaging in multiple tasks simultaneously (i.e. multi-tasking). A common example of this kind of dual-task situation is a driver interacting with their smartphone while driving a car. Paying attention to the road and paying attention to the contents of the smartphone can prove very challenging when engaging in both tasks at the same time. Will the driver be succesful in engaging in conversation? Would the phone conversation impair the driving safety? Would the driver be more successful and experience less demand with another interface? In such a practical sense, multiple resource theory stems from the individual person's ability to operate and perform in multi-task environment with a high workload. In a theoretical sense, the importance of the theory lies within the ability to predict levels of dual-task interference between two simultaneously performed tasks, where task performance is consistent with neurophysiological mechanisms and task interference is accounted for by variations that cannot be explained by simpler theories (e.g. "Bottleneck" or "Filter" theories) [2].

In both practical and theoretical implications, it is important to distinguish between the terms "multiple" and "resources". Within this theory, "resources" are referred to as something that is limited and can be allocated between tasks, while "multiple" refers to the processes that are either separate, parallel or relatively independent [2]. To further explain this, Wickens introduced a model for multiple resources consisting of four dimensions; (1) *processing stage*; (2) *processing codes*; (3) *modalities*; and (4) *visual channels* [2, 22]. The processing stage indicates that cognitive and perceptual tasks use different resources than selection and execution of action. Processing codes indicates that spatial activities use different resources than linguistic and verbal activities. The modalities dimension indicate that the resources used for visual perception are different from the ones used for auditory perception. Finally, the visual channels dimension indicates that focal vision and ambient vision perceive different subjects in the visual spectrum. For instance, focal vision involves object recognition and acuity perception which involves reading text and recognizing symbols, while ambient vision involves orientation and movement which further includes task such as walking upright in targeted directions and keeping within lanes on a highway [2]. Figure 2.1 illustrates the multiple resource model as proposed by Wickens.

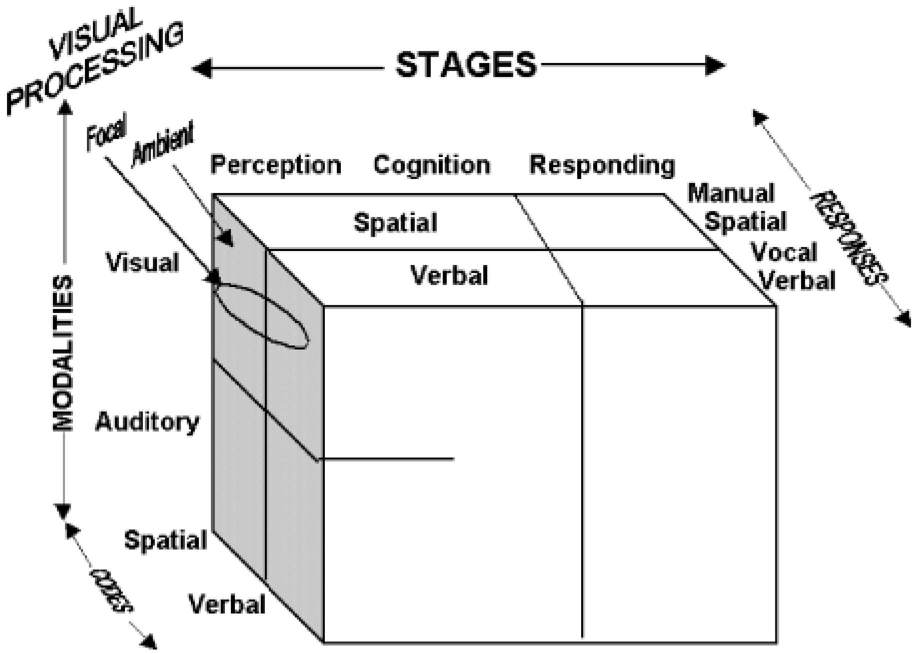


Figure 2.1: Three dimensional representation of the structure of multiple resources by Wickens [2]. The dimension of visual processing is nested within visual resources

While keeping this model for multiple resources in mind, Wickens further explains when two tasks where both demand a level of a given dimension, they will interfere with one another more than two tasks that demand different levels of a given dimension [23]. Studies that have been conducted on this topic found advantages in time-sharing in cross-modals as opposed to intra-modals. In other words, it was observed that attention can be better divided between one visual channel and one auditory channel (eye and ear) than between two visual channels (eyes only) and two auditory channels (ears only) [23]. Additionally, Wickens also explains that it is more due to peripheral factors than perceptive factors that make intra-modal time-sharing less advantageous than cross-modal time-sharing. In other words, if two tasks that are within the visual channel dimension, they will either demand that you divide your attention between them if they are too far apart, or mask the contents of each other if they are too close together. Parkes and Coleman further investigated the time-sharing effectiveness between cross-modal and intra-model time sharing by observing driving and operation of vehicles and found that discrete route guidances were better presented through cross-modal means (auditory and visual channels) than intra-modal means (visual only) [24, 2].

While the theory supports the notion that cross-modal time-sharing is more effective in time-sharing than intra-modal time-sharing, if visual scanning is carefully controlled, cross-modal displays do not often result in better time-sharing between the modalities [2]. However, in a real world setting, visual scanning is enough of a factor to impose dual-task interference. Such interference can be reduced by adopting a cross-modal display of time-sharing where some information channels are moved from a visual channel to a auditory channel [25]. With mobile interaction being a dual-task context of interaction, we can assume that such interaction can be improved in terms of dual-task interference by adopting a cross-modal form of interaction. Since smartphone interaction when walking requires divided attention between the small-screen and path orientation (two visual channels; focal vision and ambient vision), and such interference is improved by moving some of the information from visual channels to auditory channels, there is a possibility to reduce such dual-task interference by adopting a voice based interaction approach to mobile interaction [2, 22, 23, 25, 24]. In the following sections, I explore the theoretical background behind multi-tasking and dual-task interference as well as existing literature on SIIDs with mobile interaction. Additionally, I also explore how a conversational user interface (voice asssistant) can be used to reduce the effects of dual-task interference.

2.2 Situationally-Induced Impairments and Disabilities

This section of the background chapter covers the related work and previous studies conducted on SIIDs and follows the order presented by Wobbrock [26] for investigating SIIDs and that includes: understanding, sensing, modelling and adapting. While effects of ambient light, ambient noise, encumbrance and temperature is not investigated in this project work, it is still worth mentioning as walking is usually accompanied by one or more of these environmental factors during mobile interaction [27]. Each of these topics will presented in the same order and focus on literature pertaining specifically to SIIDs in regards to mobile interaction.

2.2.1 Understanding

This subsection pertains to understanding the effects of SIIDs, which involves existing literature that investigates in what way various SIIDs affect mobile interaction.

The Effects of Walking on Mobile Interaction

The effects of mobile interaction with smartphones have been widely studied. One of the many ways that mobility affects how users interact with their mobile phones, smartphones and PDA's is through task performance and perceived mental workload. In a study where participants were asked to perform mobile phone tasks while walking on a treadmill, Kane et al. [12] investigated the effects of walking on mobile interaction performance when using soft buttons of varying sizes. The participants were given the task of selecting music by scrolling through a playlist to find a specific song and play the song. Their findings showed a significant effect of button size and an interaction of button size and movement on the performance of the participants. The authors also found that button size affects the error rates and completion time of the tasks, and that movement determines the effect size of the buttons.

Additionally, Lin et al. [15] investigated how walking affects the performance on tapping tasks during mobile interaction. The authors had the participants walking in a given path through obstacles while performing the tapping tasks and found that walking in an environment with obstacles negatively impacted target selection accuracy, reduced the walking speed and showed an increase in perceived workload with the participants.

Schildbach and Rukzio [28] further studied the negative effects walking presented on selection and reading tasks when the users are engaging in mobile interaction. The focus was aimed towards the negative effects such as weaker task performance as well as higher mental workload. Their study demonstrated that walking increased the participants´ error rate by 24% and target selection time increased by 31%. Additionally, walking further decreased the participants´ reading speed by 19% and also increased cognitive load by 16%. Similarly to Schildbach and Rukzio, Mustonen et al. demonstrated how walking deteriorates visual performance when reading and when accomplishing visual search tasks [29]. Additionally, Bergstrom-Lehtovita et al. studied the relationship between walking speed and mobile interaction in terms of target acquisition performance [30]. The authors found that participants walking between 40% to 80% of their preferred walking speed produced a desirable target acquisition performance when mobile.

When investigating the effects of walking in isolation from other contextual factors, divided attention and distraction effects are fairly prevalent when engaging in mobile interaction. Similarly to using hand-held devices when driving, walking while interacting with a hand-held device may cause similar distractions experienced by drivers when engaging in multi-tasking which includes mobile interaction while the user is mobile [31]. Harvey and Pointon further researched the effects of walking on mobile interaction in terms of distractions in attention which negatively impacts task performance [32].

The Effects of Fragmented Attention on Mobile Interaction

Beyond studying the external factors for the effects of SIIDs, also "internal" factors (i.e., meaning the factors coming from the user's internal states [33]) were studied of SIIDs during mobile interaction. As previously mentioned by Wickens [2, 22, 23], users often expose themselves to SIIDs by engaging in more than two tasks at the same time where the tasks affect each other's performances (i.e. multi-tasking).

In one study conducted by Oulasvirta and colleagues, the effects of fragmented attention caused by moving through urban environments were investigated on mobile interaction via having the participants complete mobile browsing tasks while moving through urban environments [18, 34]. When the participants performed the mobile browsing tasks in this study, the outcome resulted in the participants focusing their attention on the tasks for 6-16 seconds and had intermittent breaks between the tasks for 4-8 seconds. In conclusion, the results of the study show that competition on attention resource is real and also constrains mobile interaction which also gives further support the theory presented by Wickens in terms of multiple resources.

Furthermore, some authors also suggest that fragmented attention due to multi-tasking effects can also be considered a visual problem in certain situations. A. Mack presented the term "Inattentional Blindness" which argues the notion that during multi-tasking situations, the task that is not focused on, the user is essentially blind to it's contents and environment. A good example includes smartphone interaction while driving, where a driver is blind to the road ahead of him when he is focusing on interacting with smartphone while driving. With this in mind, multi-tasking can be viewed as a visual problem where voice and gesture interaction modalities might aid in improving task performance with mobile interaction. Existing literature on this is further expored in Section 2.3.2.

The Effects of Encumbrance on Mobile Interaction

A common situation users will find themselves in when interacting with a smartphone is when also being encumbered (e.g., carrying shopping bags, travel bags etc.). Since this is a common occurence, Ng and colleagues have conducted mulitple investigations when researching the effects of encumbrance on mobile interaction. In one of the studies conducted by Ng et al., they investigated the effect encumbrance has on mobile interaction. They found that encumbrance negatively impacts accuracy during target selection tasks where participants performed more errors when carrying a small and medium bag combined with a thin and thick box. Furthermore, mobility was an added effect to the study and was included by having the participants walk a defined path while also being encumbered. The researchers found the walking speed decreased by 41% compared to completing the target selection tasks when encumbered in a standing position [35]. Ng et al. further expanded their research by investigating the effects of encumbrance on various interaction postures, including two-handed with an index finger, two-handed with both thumbs, and one-handed with only one thumb [36]. The authors found that accuracy in target acquisition tasks dropped by 48.1% when interacting with one-hand using the index finger, while the error rate of one-handed thumb interaction increased by 40% when interacting with a mobile phone while encumbered.

When investigating the effects mobility and encumbrance has on standard, commonly performed, gestures on touchscreens (such as tapping, dragging, spreading, pinching and rotating), Ng et al. found that encumbrance and walking had a negative effect on each gesture in terms of accuracy [37]. For this study, the authors usede Fitt's law to measure and quantify the effects on the gestures. They found that the performance of each gesture were negatively impacted by the encumbrance and mobility, except for the rotational gestures where the participants performed well. Furthermore, the participants were slower when completing tasks when performing the tasks by tapping, dragging and rotating. Also, the authors further demonstrate that it is important to account for physically demanding contexts when designing for interaction methods for mobile devices.

The Effect of Ambient Light on Mobile Interaction

A contextual factor that has also been researched in terms of the effects from SIIDs is ambient lighting. In a study conducted by Lee et al., the authors demonstrate the effects lighting has on the performance of visual searching tasks during reading [38]. In this study, the authors further demonstrated that searching speed increased as the surrounding lighting increased. Furthermore, this implies that dark ambient light resulted in slower searching speed in participants, while brighter ambient light increased the searching speed of the participants during the visual search tasks [38]. Additionally, the authors also touched on the size of characters on the search tasks. They showed that bigger characters lead to a higher search speed when the participants were completing the visual search tasks. As a result of this finding, the authors recommend a minimum size of 3.3 millimeters for optimal search performance on mobile devices [38].

In another study conducted on the effects different levels of illumination on mobile interaction in terms of measuring the performance on character detection tasks, Liu et al. discovered that the participants performed the character detection tasks worse under bright levels of illumination compared to low levels of illumination due to the glare of the device screen [39]. The authors demonstrated the importance of accounting for ambient lighting on mobile interaction, particularly for cases where mobile interaction may interfere with other critical tasks, for example in a situation where a medical operation requires a mobile device to interact with other technology during a operation [39].

The effects of lighting during mobile interaction has not been thoroughly researched in the context of situationally-induced impairments. However, there is existing literature that touches on the impact of lightning during mobile interaction where the environment that is being researched is not centered around lighting. A study conducted by Barnard et al. [40], where participants performed tasks on small devices while either walking on a treadmill or walking a specific path, showed that lighting levels impact the users during interaction. The results of the study indicated that response time and the number of times the participants used the scroll bars were significantly higher in low light scenarios than in high light scenarios.

The Effects of Ambient Temperature on Mobile Interaction

Some research has also been conducted to focus on environmental factors that may contribute to SIIDs on mobile interaction. In this aspect, cold environments have been investigated for effects on mobile interaction. Cold ambience is already known for causing impairments to fine-motor skills during mobile device interaction. This impairment is caused by fine-motor dexterity impairment that occurs when the finger used for interacting with the screen drops in temperature [41, 42]. While the number of studies conducted on the effects of temperature on mobile interaction is limited, it is shown that SIIDs caused by cold temperatures affects user performance during tapping tasks from low temperatures on fingers [41] where the participants performed tasks such as finding application icons on the smartphone to quantify vigilance in a search task and tapping circles to quantify the fine-motor performance of the participants. The results showed that the cold environments increased the task completion time for the participants as well as showing that the participants were significantly slower in memorising an icon than in warm environments [42]. Sarsenbayeva et al. [41] further investigated the relationship between finger and battery temperature in cold environments. The results demonstrated that smartphone battery temperature is highly correlated to a users´ finger temperature. This can be further used to infer potential cold-induced SIIDs during mobile interaction in cold environments, while interacting either one- or two-handed [41].

Goncalves et al. further expanded on the previous study by Sarsenbayeva et al. [42] and investigated the effect cold ambience has on mobile interaction in terms of performance in target selection tasks [43]. The study shows that performance is negatively affected during smartphone interaction in cold environments and cold ambience as the throughput drops which then causes an increase in error rate. Based on the findings in this study, the researchers suggest using ambient temperatures as a performance predictor in Fitts' Law [44].

2.2.2 Sensing

This subsection focuses on literature pertaining to mobile device sensors that have been widely used in the communities of HCI and UbiComp for different purposes. This subsection also provides an overview over existing literature based on sensing SIIDs. For instance, some of the work focuses on using device sensors for detecting context (e.g., [16, 45, 46], while some of the work focuses on using smartphone sensors for detecting and recognizing activity (e.g., [47, 48, 49]. Similar to the studies mentioned above, smartphone sensors have also been used for detection of SIIDs.

One example of research focused on sensing SIIDs, Goel et al. used a built-in accelerometer on a smartphone to detect walking when a user is interacting with the smartphone [9]. Based on the sensing observed in this study, the authors developed an adapted keyboard to accommodate for the walking effects on mobile interaction. More specifically, the authors used the displacement and acceleration of the mobile device for the classification model of the text entry. In the study, the authors demonstrate that successful detection of walking can lead to the development of adapting interfaces that can reduce the effects of SIIDs that occur when mobile interaction is conducted.

Furthermore, the smartphone's built-in accelerometer has also been used for the detection of hand tremors. Denault et al. used the smartphone accelerometer to collect and process the data from the accelerometer to compare with the laboratory accelerometer data [50]. In their study, they discovered that tremors with amplitudes lower than 1mm were not detected by the sensors on the smartphone. However, hand tremors with amplitudes higher than 1mm were detected and the correlation between the smartphone data and the laboratory data were relatively high (e.g. r ¿ 0.88). The above examples make use of accelerometers to detect both SIIDs caused by walking and permanent impairments (in this example, hand tremors). This means that instruments used to detect SIIDs can also be used to detect permanent impairments and disabilities.

Some work also went in the direction of identifying the grip and holding posture of the users on mobile devices. Goel et al. presented "GripSense" - a system which uses a gyroscope, vibration motor, user touch size and swipe shape to detect hand-posture of users [10]. The authors demonstrated that their system had extremely good accuracy when detecting if the device was on the table or in the hand of the user. The reported accuracy of "GripSense" between table position and hand posture was 99.7%. The authors also reported a slight drop in accuracy when users were switching between hand postures, where the reported accuracy of "GripSense" was 84.3%. The authors further suggested that their system could be used for detecting SIIDs when the users are encumbered and are only able to use their smartphone with one hand [10]. Similarly to Goel and colleagues, Gupta et al. also focused on detecting the grip strength of users. The authors present a novel haptick feedback system called "SqueezeBlock" that can be used for various use-case scenarios, for example changing the ringer volume by squeezing the device [51]. When designing the system, the authors suggested using virtual spring technology to detect grip strength when holding the device with one hand to enable eyes-free interaction. This concept can be used to accommodate for situations where the user is encumbered by limiting the interaction to a squeezing gesture [51].

The system suggested by Gupta et al. has also been implemented into Google smartphones, to detect grip force and, depending on the force of the grip, launch Google Assistant and allowing for quick interaction with the voice assistant [52]. This feature is called "Active Edge", and can be useful in situations where the users cannot adequately interact with their smartphones with both hands or require quick activation of the Google Assistant due to either encumbrance or divided attention between tasks. After evaluating the "Active Edge" feature, the authors report that the system was found to be comfortable, easy and reliable for tasks such as silencing alarms, take photos and issue voice commands to the device [52].

Mobile device sensors have also been suggested to detect surrounding context, which means context that derives from outside environments. Yi et al. utilised a tri-axial accelerometer for detecting and determining the mobile state of the user as well as detect ambient light [53]. By using a single tri-axial accelerometer attached to a mobile device to sense contextual information, they were able to successfully achieve this. Similarly, Mass and Madaus suggest using a smartphone pressure sensor to detect pressure from environmental factors [54]. Mass and Madaus claim that the smartphone network could provide reliable pressure data as it would not be influenced by other external factors [54]. Additionally, Overeem et al. demonstrated that smartphone detected temperature data (via battery temperatures) can also be used to predict average air temperature [55].

Reis et al. also presented two context-aware prototypes for adjusting volume automatically based on the detected levels of ambient noise [56]. The interfaces used in this study use the smartphone's microphone for detecting the levels of the surrounding noise of the environment. Additionally, the interfaces also accounted for the settings and modifications set by the users based on their preferences. The authors demonstrate a positive response from the users towards the interfaces, which shows that accounting for SIIDs from ambient noise is important when designing technology for interaction [56].

2.2.3 Modelling

In terms of SIIDs, modelling pertains to changing either the user behaviour or the surrounding environment to further expand the understanding of SIIDs. Additionally, modelling can also include the creation for machine learning models for predicting SIIDs and their effects on mobile interaction or human behaviour when affected by SIIDs. For this aspect of understanding SIIDs, several researchers have designed models for creating adaptive user interfaces to acommodate the effects of SIIDs.

In one study by Flatla and Gutwin, individual models were presented for the purpose of colour differentiation caused by either permanent impairments (such as colour blindness) or situational visual impairments (caused by bright light) [57]. Flatla and Gutwin demontrated that their model effectively detected the users' individualistic colour differentiation abilities and also improved the colour adaptability of the screen on the mobile device [57].

Mott and Wobbrock also presented a touch model for improving the touch accuracy of users with motor impairments [11] called the "Cluster Touch". The general model of the "Cluster Touch" personalises the model based on the touch data it receives from multiple users based on their touch behaviour when interacting with the mobile device. The study shows that "Cluster Touch" improved accuracy with both motor-impaired users and users impaired by SIIDs. With motor-impaired users, the accuracy improved by 14.65%, while accuracy improved by 6.81% with users experiencing SIIDs when compared to a native touch baseline. Additionally, in an offline analysis the accuracy was reported to be improved by 8.21% for permanently impaired users and 4.84% for users impaired by SIIDs. Mott and Wobbrock also add that this work can be extended to support users experiencing SIIDs when walking [11].

In addition to the works mentioned above, several works have focused on using language models to adjust key pressing probabilities. For example, Buschek and Alt designed a graphical user interface framework for mobile devices called "ProbUI" which defines touch behaviours, evaluate them in terms of probability, and finally infers touch intentions based on the previous steps [58]. The authors suggest that this framework can be used to improve gesture and touch accuracy of users where SIIDs affect their touch accuracy, for example, when walking.

In another study conducted by Buschek et al., the authors note the importance of considering individual characteristics and human behaviour during touch interaction [59], since users generally perform tasks differently when interacting with their mobile devices. The authors demonstrate that the touches performed with the thumb are more individualistic than the ones performed with the index finger. As a result, the authors suggest that mobile devices should be able to differentiate between different device-holding postures and also adapt accordingly depending on the hand posture during interaction [59].

In addition to personalising hand gestures and hand posture, one example of successfully personalising text entry was with the system "Text Text Revolution" designed by Rudchenko et al. [60]. In this user study, the authors had participants train the system target resizing on touch points collected by the participants. The participants perform 10 rounds of tasks in the TTR system, and then simulated personalised target resizing models in the next 10 rounds on the TTR. The results from the tests showed that the accuracy improved on text entry as the error rate dropped by 21.4% [60].

2.2.4 Adapting

The aspect of adapting in terms of SIIDs pertains to the research directed at creating adaptive interfaces to accommodate for the effects from various SIIDs. This subsection provides an overview over several adaptive interfaces from existing literature that are used to reduce the effects of SIIDs such as walking, encumbrance, and situational visual impairments that are present due to environmental factors.

Interface Adaptations for Walking

Since walking is the SIID that has been the most researched, it is only natural that most suggestions of adaptive interfaces have been directed towards mitigating the effects of mobility. In one example, Kane et al. designed a *walking user interface* for music players that alters target sizes based on the motion of the user to mitigate for the effects of walking on song selection [12]. The adaptive interface makes the targets smaller when the user is standing still, and increases the size of the targets when the user is walking to improve readability and touch accuracy [12]. The evaluation of the adaptive interface showed that task completion time improved among the participants, however the authors suggest that individual characteristics and task difficulty need to be accounted for when designing an adaptive interface to mitigate effects from walking [12].

Other proposals for overcoming SIIDs caused by walking includes using techniques for screen stabilisation to stabilise the contents of the mobile device's screen. For this purpose, Rahmati et al. [61] developed an interface they named "NoShake", which is a system utilized for detecting if the device was shaking by using the smartphone's accelerometer and then compensate for the shaking motion by shifting the contents of the screen in the opposite direction. The results show that the user experience improved after using the adaptive interface when shaking the mobile device. Furthermore, the authors suggest that the system could also improve the user experience with users who suffer from diseases that cause hand tremors, such as Parkinson's Disease [61].

Yamabe and Takahashi also proposed a user interface adaption approach where the interface would change the size of the screen elements based on movement of the display [62]. The authors further suggest to account for individual characteristics when designing adaptive interfaces to accommodate for walking SIIDs [62]. Brewster expanded on the suggestions by Yamabe and Takahashi by suggesting that small buttons should be accompanied with sound feedback to enhance the usability of the adaptive interface [63]. This can be useful in situations where touch interaction has a higher rate of inaccuracy due to walking [63].

Other adaptive interfaces were created to focus on improving text entry methods as this is one of the most common tasks users perform on their mobile phones. One adaptive interface for improving text entry performance was developed by Goel et al. and is called the "WalkType" [9]. This adaptive text entry system was developed for smartphones to accommodate for SIIDs that occur when walking and interacting by performing text entry tasks on the smartphone. The system utilizes the built-in accelerometer on the smartphone to detect if the user is walking. The system proved effective in reducing the effects of SIIDs on text entry tasks when walking by reducing the error rates by 45.2% and improving typing speed by 12.9% [9]. Similarly to Goel et al., Himberg and colleagues developed an adaptive keyboard which adjusts the key positions with the spatial distribution of the keystrokes of the user [64]. The results of the evaluative study conducted by Himberg et al. demonstrates that the changes in the keyboard are consistent with the user of the keyboard. Based on this, the authors suggest taking personalisation into account when designing adaptive keyboards [64]. Additionally, Go and Endo developed a customizable and adaptive keyboard interfaces for touchscreen mobile devices. They developed a customizable and adaptable touchscreen keyboard named "CATKey" - a software which adapts each key's centroid to the centroid of the recorded keystroke points [65]. Despite of the "CATKey" not showing any improved efficiency in text entry evaluation, the participants still expressed a preference for the "CATKey" adaptive keyboard compared to the traditional "QWERTY" keyboard [65].

Interface Adaptations for Encumbrance

As previously mentioned, one of the main challenges of encumbrance during mobile interaction is the forced one-handed interaction with mobile device due to the other hand being occupied with holding various objects at the same time. For this, several researchers have made suggestions for improving one-handed interaction with a mobile device.

In one study, Buschek et al. presented and evaluated dynamic adaptions of mobile touch interfaces to accommodate for inconveniences that arise with interacting with a large screen using only one hand [66]. In their work, they present three techniques to locate interface elements on the screen; "Roll", "Bend", and "Move". The authors demonstrate the effectiveness of their proposed techniques due to the participants experiencing increased usability comfort, reduction in fatigue, and easier grip during interaction [66]. According Ng et al., these techniques can be used to overcome challenges that arise when interacting with a mobile device when encumbered [37].

Other adaptive interfaces designed for enabling one-handed interaction were "AppLens" and "LaunchTile". Both of the interfaces used the zoom function to improve one-handed interaction, however the zoom is used differently in the two interfaces [67]. "LaunchTile" uses a pure zoom to be used by performing gestures with the thumb when interacting with the mobile device. "AppLense", on the other hand, uses a tabular fisheye. The authors demonstrate that the participants preferred using the "AppLense" more than the "LaunchTile", as well as showing improved performance with the "AppLense" in terms of task completion time [67].

There were also adaptive interfaces and interaction techniques developed for improving performance in certain task categories when the user is encumbered. One such technique is "ThumbSpace", presented by Karlson and Bederson, which is an interaction method involving one-handed interaction with the thumb [68]. "ThumbSpace" was positively received by the participants of the evaluation study and improved their performance when performing target acquisition tasks on a smartphone with small target sizes. The authors further suggest that this interaction technique can be beneficial when users are encumbered, as the interaction technique allows for one-handed interaction and effectively complete interaction [68].

Other interaction techniques have also been developed for text entry tasks, such as "Twiddler" - a text entry technique presented by Lyons et al. [69]. Similar to the "AppLens", "LaunchTile", and "ThumbSpace", this system can be used in situations where the user is encumbered and enables the user to perform text entry tasks with one-handed interaction of the mobile phone [69]. The authors' system proved effective and intuitive as demonstrated in their study [69]. Another adaptive system developed for accommodating for encumbrance when performing text entry tasks is "Unigesture". "Unigesture" is a tilt-to-write system that enables one-handed text entry by using accelerometer data to detect and determine the tilting position of the mobile device, and use the input characters that is mapped to the specific tilt [70]. Similar to previous studies mentioned above, individual characteristics had a great effect on the performance on the text entry tasks. Again, the authors suggest that individual characterstics should be accounted for when designing adaptive technology for text entry tasks [70].

Besides encumbrance, interaction difficulties may also arise when the user is interacting with a small screen [71, 72], causing something known as the "Fat finger problem" [73]. To solve the issue with "Fat finger problem", some researchers suggest interacting with the back side of the mobile device via touch-interaction [74]. In a user study conducted by Baudisch and Chu, they showed that using the back of the device for mobile interaction leads to higher accuracy and lower error rates when completing target selection tasks [74]. Additionally, using the back of the device of a mobile device can also allow for more efficient and rapid text entry as it involves all ten fingers during interaction [75]. This solution [75] can also be used to accommodate for SIIDs that are caused by interacting with a small screen mobile device [63, 72].

As some research has shown that holding posture can either improve or detoriate the performance on text entry tasks, some additional work has been conducted that account for holding posture in text entry tasks during mobile interaction [76]. "ContextType" - a system that uses the users' hand posture to improve the interactivity in terms of text entry on mobile touch screen devices, such as smartphones. This system provides four modes of interaction: two thumbs, left thumb, right thumb, and the index finger of either hand [76]. The system switches between touch-models based on the inference of the hand posture when the user is holding the mobile device, and also uses the leverage of previous adaptive interfaces introduced by Goel et al., in this case "GripSense" [10], which is based on the detection of holding posture of the device [76].

Interface Adaptions for Situational Visual Impairments

In certain situations it would important for users to be able to perform eyes-free interaction. In other words, when the user is not able to interact with the mobile device due to mobility, divided attention, or situational visual impairments. To overcome these situations, several researchers have attempted to create adaptive techniques and interfaces to accommodate for these challenges. One such method was introduced by Jain and Balakrishan, who developed a text entry interaction method that uses bezel gestures to allow for eyes-free interaction with a mobile device [77]. The authors demonstrate that beze-gesture based text entry was similar to traditional text entry methods in terms of speed, accuracy, and how easy it was to learn its use. Their study also showcased its adaptability in terms of novice users learning and mastering the method after one hour of training [77]. This study serves as an example of how eyes-free interaction can be enabled via gesture interaction.

Similarly to the bezel-gesture method mentioned above [77], Chen et al. presented another method for text entry called "Swipeboard". This is also an eyes-free text entry method that uses swiping-gestures for interaction instead of bezel-gestures for text input [78]. The technique consists of two swiping gestures: (1) specifying the region of the character location, and (2) selecting the desired character. The study showed that the participants performed the text entry tasks 15% faster than with traditional baseline text entry methods, after going through extensive training with the "Swipeboard" [78].

2.3 Conversational User Interface

This section explores conversational user interfaces (CUI), or voice assistants, and their role in mitigating the effects of SIIDs. While the amount of research that has been conducted on voice assistants as a potential tool to reduce the effects of SIIDs are relatively sparse, there are some studies conducted on its effects on individuals with permanent physical impairments. This section begins by describing what CUIs are and what they are generally used for, followed by how they can be used to make smartphone interaction more accessible for users with disabilities. This section will provide further background as to why I chose to voice assistants as a interaction method for mitigating SIIDs as opposed to some other forms of interaction (e.g. interaction via gestures).

Even though the current generation of IPAs, such as Alexa, Cortana, Google Voice and Siri, have only been around for less than a decade and needed time to mature in terms of capabilities, they have seen a rapid rise in popularity. In April 2018, 41.4% of surveyed US adults reported using the IPA on their smartphone, while 19.7% were found to use a smart speaker IPA []. This popularity is mirrored in terms of increasing research attention for IPAs and other conversational agents. Usage of IPAs was studied by, among others, Garcia et al. [8], who conducted a questionnaire about IPA usage in Argentina, Brazil, Chile, Germany, Spain, the UK and the US. They found that IPA usage in other countries lags behind the US, but that around 50-60% of IPA users use them several times a week. Bogers et al. [] performed a similar study in Denmark, and found that Siri was by far the most popular IPA due to support for the Danish language, but that only 20% of respondents used Siri more than once a month.

2.3.1 Definition

Conversational User Interfaces (CUI's), also known as *Voice Assistants* (VA's), are the realization of many people's science fiction dream of being able to interact with a computer or system by talking to them [79]. Examples of popular voice assistants such as these include; Apple's Siri, Amazon's Alexa, Microsoft's Cortana and Google's Assistant. The way these software agents work is that they all wait for a wake-up word, which is unique to each voice assistant (i.e. "Hey, Siri" for Apple's Siri, or "Hey Google" for Google Assistant) and records the user's voice, sends it to a specialized server which then interprets the command that was issued by the user. Depending on the command that is given by the user, the server will supply the VA which then reads back to the user the appropriate information, plays the requested media, or completes the tasks asked by the user with various connected services and devices [79]. Ever since the introduction of Apple's Siri in 2010, the number of services that support voice commands has been growing rapidly with manufacturers of Interntet-of-Things (IoT) services and products also including voice control built-in to their products (i.e. cars with voice control, smart homes).

All of the mentioned VA's all work in a similar fashion, while they each also have unique features to them. Since these CUI's are always connected to the internet, they can respond to a larger number of commands and sentences than previous voice-activated technologies [79]. As the result of advancements in natural language processing, voice assistants are able to create meaningful responses quickly as opposed to earlier voice-activated devices that relied on a smaller set of phrases and commands that were "built-in" to the systems. The recent improvements to natural language processing can be credit to the following four developments presented by Hirschberg and Manning [80]: (1) a large increase in computing power, (2) the increased availability of large amounts of linguistic data, (3) advanced developments in successful machine learning methods (ML), and (4) an improved and richer understanding of human language structure and its deployment in social contexts.

Due to personal computers becoming cheaper and more powerful over the years, and more text online has been created for analysis purposes, scientists have been able to use text to train algorithms in voice assistants to better respond to request in a more meaningful and naturally linguistic way. Due to this, CUIs like Siri and Google Assistant are able to parse requests phrased in various ways and accurately interpret the wishes of the user [79]. For example, asking Siri to set a reminder for a specific event, a user can either say "Set a reminder for 9 A.M. tomorrow" or "Can you remind at 9 tomorrow morning?" and Siri will interpret both phrases as a reminder that will notify the user the next day at 9 A.M. Additionally, users can also ask the same question phrased differently; by asking for example "What's the weather like tomorrow"" and "How many degrees will there be tomorrow?" both trigger the same response from Siri, reporting the weather forecast for the next day. Natural language processing like this helps users avoid frustration when interacting with a CUI, which is a big improvement over earlier voice recognition systems that required specific words and phrases in order to achieve desired outcomes [79].

Voice Assistant Abilities

While all of the mentioned CUIs offer unique features, they also share multiple similarities and are able to perform basic tasks such as the following [79]:

- send and text messages and e-mails, make phone calls.
- provide informational queries (weather forecasts, news etc.).
- set timers, alarms, reminders, calendar entries.
- make lists, notes and solve basic math equations.
- control various media playback (e.g. Spotify, Netflix etc.)
- control IoT devices such as lights, alarms and clocks.
- tell jokes and stories

Additionally, voice assistants also offer other features that expand on their abilities by interacting with other programs. For example, Amazon's Alexa can play Jeopardy with a user, order a drink from a local Starbucks, and also order an Uber. Google Assistant offers a way for users to create personalized features by using services like Tasker and IFTTT (If This Then That), which allows users to create features that automate social media posts, turn devices on and off, and many other possible features. Apple's Siri has recently made it possible for users to hear and respond to messages when the smartphone is in a locked state and connected to wireless earphones. This enables a form of hands-free and eyes-free interaction with a smartphone in terms of text entry abilities.

2.3.2 Voice Assistants for Accessibility

With the advancements that voice assistants and CUIs have made over the past years, they have the potential to radically change how we interact with computers, smartphones and other IoT-devices [79]. For many users, due to various impairments and disabilities, reading and typing poses a challenge for accessing information. For certain users who are cognitively impaired, research has shown that voice assistants can aid patients of dementia by providing a constantly present voice that can answer the same questions multiple times without submitting to frustration and loss of patience [81]. For other individuals, reading from computer screens and smaller smartphone screens can prove difficult due to visual impairments. Building a voice that can read the contents of a screen into currently available consumer technologies would be more cost effective than building a device solely for the purpose of aiding disabilities. With the rapid improvement in the vocal qualities of voice assistants, to the point where they sound less and less robotic for each iteration, can result in not being off-putting for the user to more easily adapt to using a voice assistant for reading texts and writing messages [79, 81].

In addition to the text dictation and screen reading abilities of voice assistants, they also have the potential to apply further accessibility to disabled individuals through voice input and output beyond the scope of text entry and text comprehension. This includes controlling the lighting or door locks of a user's home who suffers from mobility impairments, and asking for weather forecasts and time for blind users [82]. For users with other kinds of visual impairments, voice assistants provide an effective solution for text entry purposes and browsing purposes. In terms of text entry, voice assistants have also been proven useful to aid users with motor impairments in tasks that involve text input on desktop computers and smartphones. They have also been useful in aiding the same group of impaired users with controlling wheel-chairs as well as aid in "free-hand" drawing.

To further expand on how voice assistants can assist users with motor impairments, Corbett and Weber addressed the user experience challenges that arise when using a mobile voice user interface (M-VUI) [83]. They argue that for users with motor impairments, a M-VUI would allow for a hands-free, voice only alternative to mobile interaction which would provide a high level of accessibility and independence in terms of mobile interaction.

Other examples of adaptive interfaces for the visually-impaired, includes "Touch n Talk" - a interface which assist visually-impaired individuals to complete computer tasks [84]. The study showed that the system was well accepted among visually-impaired users and also improved their task performance in accessing menus and tasks involving edition as opposed to terminals with key-based talking.

2.3.3 Voice Assistants for Addressing Situational Impairments

While there has been conducted research on how voice assistants can help in making interaction with desktop computers and smartphones more accessible for users with permanent physical impairments and disabilities, it's effects on SIIDs on users interacting with a phone in different contexts and situations remains largely unexplored. As already mentioned in Section **??**, contexts and use-situations such as walking, cold temperatures, noise, encumbrance and divided attention are the most common sources in which temporal situations impair a users ability to interact with a smartphone. While voice assistants have not been researched in these use-situations directly, there are other use-situations that have been studied in terms of interacting with a voice assistant.

Some research has been conducted on to what extent a voice assistant can reduce the effects of divided attention of user when interacting with a smartphone while driving. As voice assistants and IPAs are being more and more pushed by car manufacturers as they provide the ability to perform digital tasks while keeping the attention focused on the road, which results in safer interaction with smartphones when driving. However, studies have found that engaging in hands-free interaction still causes distraction similarly to hand-held touch interaction while driving despite the potential of reducing bio-mechanical distractions when engaging in hands-free interaction [85].

This project attempts to investigate how voice assitants and conversational user interfaces can be used to reduce the effects of dual-task interference during mobile interaction. By revising previous work pertaining to the topic of SIIDs in terms of mobile interaction, we can address the research gap that can be supplemented wit the contributions from this research project.

2.4 Summary

In this chapter, I have provided an overview of situationally-induced impairments and disabilities and the theory behind its concept which includes multiple resource theory which extends from context theory, followed by a overview over existing literature that investigated its effects on mobile interaction and literature attempting to create adaptions to these effects. I have also shed light on conversational user interfaces and voice assistants, as well as how these have been used to aid in permanent impairments in users which provides insight in how they can be used in adapting to SIIDs. Based on the theoretical background and the existing literature on the topic, I see that there has not been extensive studies conducted on how voice assistants can be used to reduce the effects of SIIDs in mobile interaction and I also see the need to explore possibilities for alternative modalities to mobile interaction as smartphone usage in mobile conditions is becoming more and more prevalent as time goes on.

Chapter 3

Methodology

In this chapter, I present the research approach that I have applied for this project work. Based on the core methodology described in this project, I outline the methods utilized for answering the research questions presented in Section 1.2.1. For this project, I use an experimental design method for evaluating the the conversational interface as it allows me to compare the cause and correlation between situational impairments and mobile interaction during different states of mobility by comparing quantifiable data. As explained in Chapter 1, different forms of interaction are becoming more popular

The challenges of finding methods and adaptions to mitigate effects present by dual-task interference on mobile interaction such as mobility create grounds to investigate whether alternative modalities for mobile interaction can be used to reduce multi-tasking effects. To aid in my research, I design tasks that resemble real world representation of tasks performed by smartphone users in everyday life. Furthermore, I use the study by Barnard and colleagues [40] as inspiration for designing the experimental setup and procedure for this study.

3.1 Experiment Design

This section of the methodology presents how the evaluation is designed and how it will be performed. The section begins by describing the evaluation method chosen for this project, as well as the dependent and independent measures, and the hypotheses of the experiment. Following, I introduce the procedure of the evaluation, the participants involved in the evaluation, the stimuli, the setup of the experiment and what tasks were performed and how they were performed during the experiment. The experiment for this project work will consist of two experimental tasks: a text entry task and a target acquisition task. These two tasks will be referred to as texting task trials and selection task trials throughout the study. For the both tasks, the same group of participants complete the text entry and comprehension tasks and target selection task scenarios, where the participants use both the CUI and touch interaction to complete the tasks while either standing or walking. Each participant in these tasks complete at least 8 trials per task where one trial requires the participant to complete one of the tasks listed in section 3.5. The participants also participate in both mobility conditions, meaning if a participant is given the touch interaction condition then they will perform 1 trial standing and three trials walking and the same for voice interaction. The texting tasks and selection tasks are randomized between the 16 trials, meaning some might have the first trial be a text entry and comprehension task, while other participants might have the first trial be a selection scenario.

3.1.1 Independent Variables

For this experimental evaluation, three independent variables are defined for this experiment. The first independent variable is the type of interaction (conversational and visual), the second independent variable is the type of mobility (standing and walking) and the third independent variable is task type (texting and selection). The standing condition will allow for a base-line assessment of each participant as well as allow for mobility comparisons, and the walking condition will allow for comparisons of how SIIDs affect mobile interaction for use in motion. Each task will be performed under the following conditions: conversational-walking, conversational-standing, visual-walking, and visual-standing. The first independent variable allows us to compare interaction with a CUI (in this case a voice assistant) while walking to interacting with a smartphone in the traditional sense with touch-screen interaction while walking to assess whether or not the alternative interaction modality will reduce the effects of SIIDs. The second independent variable allows us to determine how the act of walking itself influences how users interact with their smartphones either by using a voice assistant or by engaging in touch interaction. We do this by having the same tasks being performed both when standing still in one place, and by having the participants walk a predefined path while performing the tasks with both interaction modalities (Siri and touch-screen). This experiment is designed as a 2x2x2 within-subjects factorial design with the following factors and levels: Task (texting, selection), Modality (touch, voice), and Mobility (standing, walking).

The tasks for this experiment are designed based on the tasks that are most completed with by smartphone users on a daily basis [1] and that can also be performed on a smartphone by using voice assistant. The details of how the tasks are carried out in this experiment is described in Section 3.5.

3.1.2 Dependent Variables

The goal of this study is to compare participants's task and walking performance between traditional touch-screen interaction and using a conversational user interface (in this case Siri). For this experiment, I utilize both qualitative and quantitative measures. To determine how to properly measure usability for this evaluation, I follow the ISO

[1]Leading Mobile Phone Activities in 2018 https://www.statista.com/statistics/187128/leading-us-smartphone-activities/

9241-11[2] for examining the quality of how well tasks are performed and fulfilled by the participants. The components for the framework consists of: *effectiveness*, *efficiency*, *satisfaction*, and *learnability*. For the sake of the experimental approach to this study, I avoid learnability by preventing order effects.

For assessing the effectiveness between touch interaction and voice interaction, I measure task errors and task completion time for both task types of the experiment. In the text entry tasks, the error rates will be recorded as the number of spelling errors the participants perform when replying to the four text messages, whether the participants reply with more than one text message before receiving the next text message, and whether the text message addresses the question from the received text. For the selection trials, skipping one or more steps before completing a task scenario, performing one or more additional steps before completing a task scenario and not performing a step correctly are recorded as task errors. Furthermore, task completion time is measured as the total time from when a participant begins their given task to when they complete it. For example, the time a participant takes to respond to all four messages in a texting trial is, from when the participant receives the first message to when they send the last response, is recorded as the task completion time for that trial.

Additionally, total distance of the travelled path is calculated by the number of paths the participant completes (full or partial) before completing the tasks. Path accuracy represents the number of times the participant stepped on or outside the lines of the path, the amount of times a participant lost track of the path, went in the wrong direction or stopped walking the path completely. The average walking speed is determined by dividing the total distance walked by the total time to complete all the tasks during the trial. This is measured by task completion time and number of laps made during a trial.

To assess perceived cognitive workload and user satisfaction, each participant will be asked to complete the standard NASA Task Load Index (NASA-TLX). The NASA-TLX

[2]ISO 9241 https://en.wikipedia.org/wiki/ISO$_9$241$ISO_9241 - 11$

is a questionnaire used to measure subjective workload ratings. Previous literature show that it is a reliable and valid measure of the workload imposed by a task. The NASA-TLX consists of 6 scales of measurement: mental demand, physical demand, temporal demand, performance, effort, and frustration. Each one of the scales have 21 gradations. For each scale of the questionnaire, individuals rate the demand the task imposed during the trials. The participants for this study will fill out a NASA-TLX questionnaire for each of the task that were completed during the trials.

3.2 Participants

When searching for participants for the experimental procedure, close friends and work colleagues were invited to participate in the experiment. Out of all the invited individuals, 15 participants agreed to take part in the experiment. The participants were not controlled for any demographic factors (i.e. gender, native language, etc.). All participants were controlled for any permanent impairments that may impact the results of the experiment, such as speech-impairments (i.e. chronic stuttering) or motor impairments (i.e. reduced finger and hand mobility). For this reason, 15 participants agreed to participate, however one of the participants had to withdraw due to having one of the permanent impairments that could possibly impact the results. Time to complete the experiment ranged from 45 to 65 minutes, depending on the participants needing a 15 minute break in the middle of the experiment. All of the 15 data sets that were generated from the experiment were used for the study. Throughout the experiment, no technical difficulties were experienced with the experimental hardware, as all the hardware was controlled for issues prior to starting the experiment for each participant.

The age of the 15 participants ranged in age between 18 to 50 years, with 10 out of 15 participants ranging in age between 18 to 27 in age, with 5 out of the 15 participants ranged in 37 to 50 years in age. All of the participants were native Norwegian speaking/reading which motivated the decision of setting Siri to speaking Norwegian as to keep the experiment in a language the participants were most comfortable with. Furthermore, 9 of the participants were male and 6 of the participants were female. Additionally, 13 out of the 15 participants were iPhone owners and 2 of the participants were Android smartphone owners.

3.3 Apparatus and Materials

In this section of the design chapter, I will present all the hardware and software that was used during the experiment phases. I will also present how the experiment was set up as well as showcase the environment and the design of the walking path that is used throughout this experiment.

3.3.1 Walking Path

This experiment evaluates the performance of the participants in two mobility conditions: standing and walking. The walking is conducted by drawing and marking a specific path on the floor where the participants will walk during the interaction tasks. Before starting the evaluation, the participants will be allowed to conduct 2-3 walking trials to familiarize themselves with the path they will walk during the evaluation. The path was located at the campus of Østfold University College, and the path was layed out in a laboratory room containing obstacles that was placed around the path. The reason for utilizing an established path for walking instead of conducting the walking on a treadmill is to achieve more realistic measures in terms of performance, accuracy and workload, as a treadmill is not sensitive enough to contextual factors [40]. Also, similarly to the study by Barnard and colleagues [40], the direction the participants walked on the path (clockwise and counter-clockwise) alternated between trails, but stayed constant within the trial (meaning that one the participant started walking in one direction, they could not switch direction midway through the trial). Figure 3.2 showcases an illustration of the layout of the fixed path the participants will walk. While having the walking condition be away from the treadmill and on a specific path increases the sensitivity to contextual factors, the researchers still want to keep the walking condition in a controlled setting. This is to keep the focus on the walking aspect, since if we have the participants walk freely inside and outside of the campus, we then have to account for additional contextual factors alongside walking for example; ambient noise, ambient light, cold environments etc.

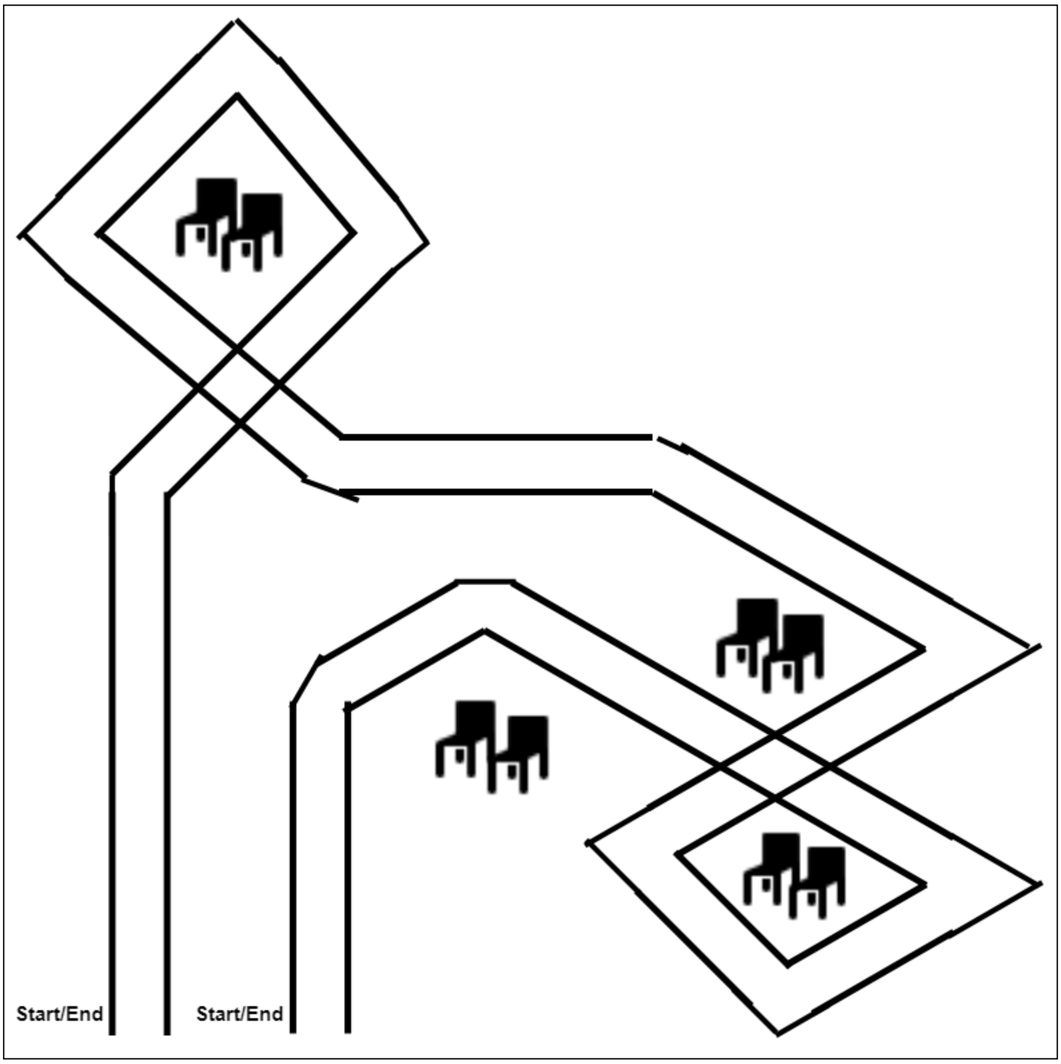


Figure 3.1: Schematic illustration of the walking path for the experiment

As shown in Figure 3.2, the walking path (illustrated by the two lines forming a walking lane) consists of obstacles in the form of chairs and have paths that make basic loops and sharp turns. The total length of the path is 30 meters. There will also be equipment against the walls that the participants have to take caution in not walking into them or have excessive contact with the equipment. If the experimenter fears the participant will crash into the equipment, then the experimenter will notify the participant to stay on the path. The walking took place in the Virtual Reality Lab at Østfold University College. A 30cm wide path was taped on the floor which navigated the participants around in the classroom, using furniture such as tables and chairs as obstacles. Further images of the actual path and lab environment are shown below in Figure 3.3 and 3.4.

Figure 3.2: Image showing the path from the starting point perspective

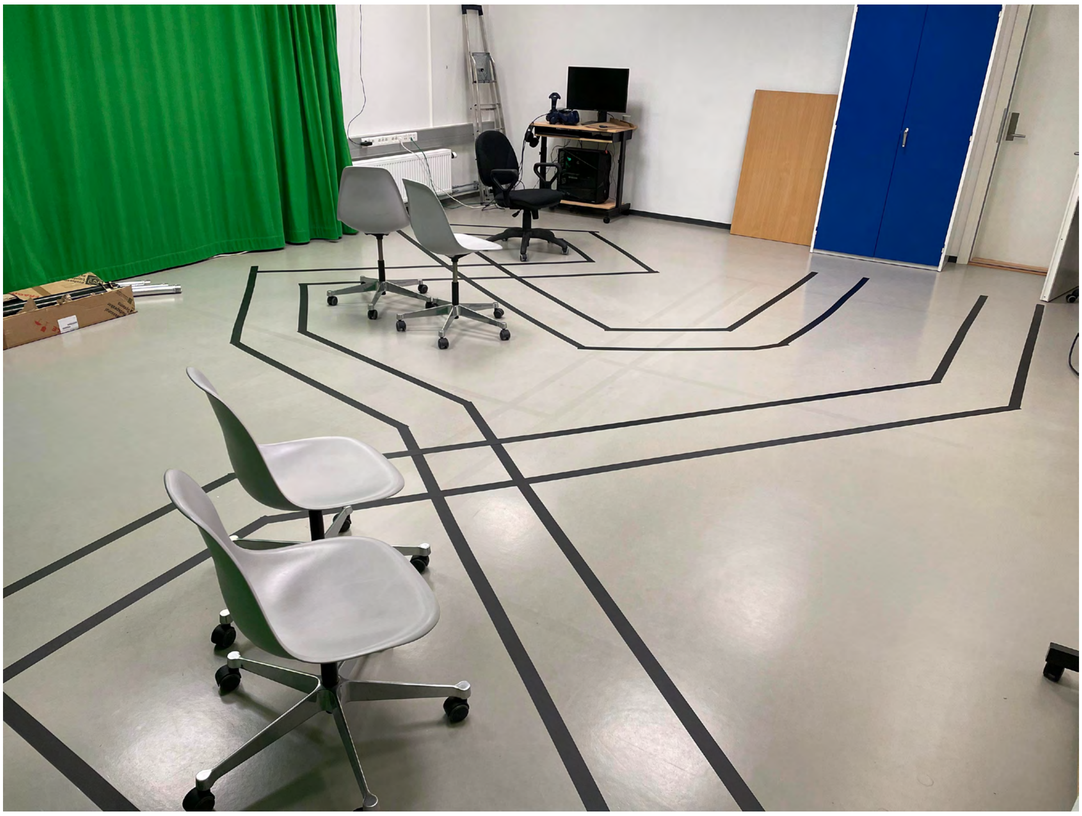


Figure 3.3: Image showing the path from the perspective of the experimenter

3.3.2 Hardware

Hardware that was used in the set-up of the experiment was a GoPro Hero 8 camera that was placed at the top corner of the Virtual Lab to record the whole experimental process of the participants. The purpose of the GoPro Hero 8 camera was to record the performance of the participants in terms of walking accuracy, walking speed, and task completion time. This also allowed for a more thorough review of the experiment and log the performance data of the participants from each of the performed trials.

Figure 3.4: Image of the GoPro Hero 8 used in this experiment

Along with the GoPro Hero 8 camera, to be able to fully evaluate the usability of Siri, a pair of second generation AirPods were also used during the experiments, mainly during the Siri trials. Airpods are wireless in-ear earphones developed by Apple that provide hands-free functionality with iPhones by using Siri. With the AirPods connected to the iPhone the participant will use, Siri will then have added functionality as in announcing incoming messages to the participant which allows for hands-free and eyes-free communication.

Figure 3.5: Image of the AirPods used during the Siri trials of the experiment

3.3.3 Software

Of the software that will be used, Siri will be the software that represents the conversational user interface in this study. Access to Siri is provided using an Apple iPhone 12 Pro with a iOS version of 15.2 and with the language set to English as the English voice of Siri is the most advanced in terms of Natural Language Processing. Siri will be reset after each trial of the experiment. Voice activation of Siri is enabled by enabling the "Hey Siri" command, which will be issued by the participants in Norwegian or English. The participants will hold the iPhone in their dominant hand during the trial, and are allowed to interact with the phone with both of their hands during the visual conditions of the experiment. During the trials, the screen on the iPhone 12 Pro will be recorded, utilizing the built-in screen recorder on the iPhone 12 Pro.

For analysis of the video footage captured by the GoPro Hero 8 camera, a data analysis software named MaxQDA was used to analyze each individual trial that was recorded of the participants. With this software, I was able to perform a more thorough analysis of the recorded trials the participants performed throughout the experiment and conduct a more thorough assessment of their performance on all four conditions of the experiment. The analysis process on MaxQDA included observing whether the participants followed the path as instructed, the amount of times the performed walking errors (stepping on or over the tape outlining the walking path) and if they walked into, or in any way came into contact, any obstacles that were laid out around the path.

3.4 Procedure

This section presents the procedure of the evaluation from start to finish. The experiment began for each participant with a brief introduction to the experiment and an interview for assessing the participants's prior experience with using voice assistants, the experiment's process step-by-step, and finally an introduction to the NASA-TLX questionnaire. The procedure in this experiment started with a pre-experimental questionnaire that each participant answered. Later on, the participant completed some training trials before starting the actual trials where the trials were recorded and used for further analysis. Finally, at the end of the condition trials, the participants filled out a NASA-TLX questionnaire for each of the modality, mobility and task conditions of the experiment.

3.4.1 Pre-Experimental Questionnaire

Before starting the experiment, the participants will first be presented what the experiment is about and its purpose. After the experiment presentation, the experimenter will conduct a short interview with the participants before describing and explaining the tasks the participants will perform. The purpose of this interview is to assess the prior experience of CUIs and mobile interaction of the participants. They will be asked if they use their phone while on the move, how often they use their phone when they are on the move, if they have used a CUI, if they generally use CUIs from time to time and how often they use a CUI. After this is completed, the experimenter will move on to describe the tasks and conditions to the participant.

After the introduction, interview and descriptions have been completed with the participants, the participants will then sign the necessary consent forms before starting the training session. After the consent forms have been filled out, we then move on to a training session.

3.4.2 Training

The training sessions are designed to instruct the participant on how to complete the different tasks of the experiment and to clarify any questions about the tasks. First, the researcher introduces and explains to the participants the different tasks that they will be performing on during the experiment. Later on, the researcher will walk the path with the participant once while explaining further how to complete the tasks during the actual experiment. This will count as the first trial run. Next, the participant will walk the path on their own the second time, and the participant will be allowed to ask questions to the researcher during this trial. This same training process will be repeated with the CUI trials. Finally, at the beginning of each of the eight conditions (voice-standing-texting, voice-standing-selection, voice-walking-texting, voice-walking-selection, touch-standing-texting, touch-standing-selection, touch-walking-texting and touch-walking-selection), the participants complete one additional training trial in a respective mode, for example, for the conversational-walking condition, a participant completes a practice trial with the CUI while walking around the path.

3.4.3 Experiment

For establishing the base-line data for the natural walking speed, both the walking speed along the path and the number of steps off the path are collected at the beginning of the experiment (the first trial after training has concluded) and at the end of the experiment (after the final condition has been concluded). The participants are instructed to walk once around the path first in a clockwise direction, and then in a counter-clockwise direction at a comfortable walking pace while trying to stay within the path. The time to complete each lap is recorded, as well as the number of steps on or over the lines of the defined path.

At the beginning of each trial condition, the experimenter will configure the software and hardware as needed. The experimenter will also remind the participant they will complete six trials in a row. For the walking conditions, the participants are instructed to continue walking until they have completed all the trials, and are also instructed not to stop walking in between the trials. The participants will be allowed to either speed-up or slow down the walking as desired, but should refrain from stopping until all of the trials are completed. As the participants are walking the path, they will be equipped with a GoPro camera, which will be strapped to their head and act as a monitor for the performance of the participant, and the footage will be used for further investigation. Figure 3.6 and Figure 3.7 shows participants performing the experiment. Additionally, the order of the conditions is structured as a Latin Square of Order 8, where the conditions include: Touch-Standing-Texting, Siri-Standing-Texting, Touch-Walking-Texting, Siri-Walking-Texting, Touch-Standing-Selection, Siri-Standing-Selection, Touch-Walking-Selection, Siri-Walking-Selection.

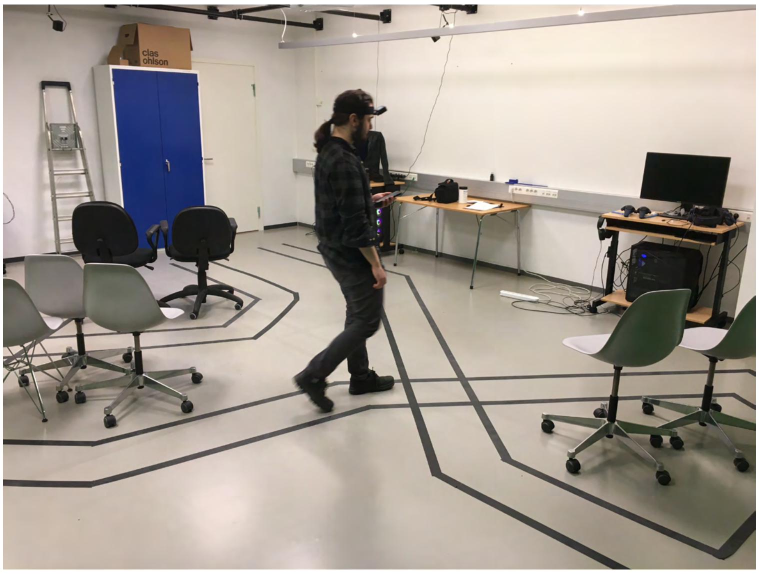


Figure 3.6: Participant performing touch interaction trial

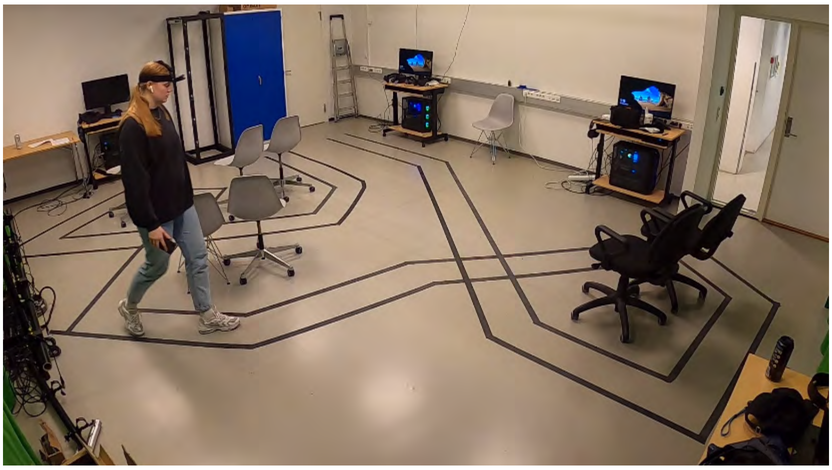


Figure 3.7: participant performing voice interaction trial

3.4.4 NASA-TLX Questionnaire

At the end of each condition, the experimenter will direct the participant to the NASA-TLX survey, where they will complete survey while keeping in mind the recent trials that were done during the condition that was completed prior to the survey. After the TLX is conducted, the procedure was repeated with the remaining interaction and mobility conditions. Finally, at the end of the experiment, the experimenter will ask the participant if they want to share any additional comments they may have about their experience during the experience and with any of the interaction models and the tasks that were performed.

3.5 Tasks

When designing the tasks the participants will perform during the evaluation experiment, it is important to design tasks that are viable for both manual touch-screen interaction and voice interaction using a voice assistant. This means the tasks need to be possible to perform in both modes of interaction (conversational and visual). In this context, both the Siri and the touch-screen conditions need to have the same tasks in this experiment. Additionally, the tasks need to be relevant in the sense that they represent common tasks that are completed by users in everyday settings. With this in mind, I design the task based on statistics provided by Statista [86] to get a better understanding of how to design tasks that best resemble such tasks. Some of the most common and fundamental mobile interaction tasks, that are mostly completed on a smartphone, can be put into the following categories of task types:

- *Text Entry and Text Comprehension* where participants may enter a form of text on the smartphone or read a from of text on the smartphone. This may typically be text messages or e-mails. An example of this may be a text snippet that is displayed on top of the screen of the smartphone, and the participants will type in the text following the case and punctuation.

- *Target Acquisition* where the tasks consist of selecting an icon or application, and opening the application on the mobile phone. An example of how this task will be evaluated is by having the participants open the "Maps" application on their smartphone and request for specific directions.

- *Visual Search* where the tasks consist of a participant selecting an application icon on the smartphone. This is done for memorisation purposes by having an application icon being selected at random and shown to the participant, and having the participant later search among 24 icons randomly distributed on the smartphone screen and finding the selected icon.

Based on these descriptions of the types of tasks that will be utilized during the evaluation in this project, the tasks that were chosen for the experiment phases focus mainly on text entry and target acquisition as the visual search counteracts the point of CUI's being a hands-free option for interaction (based on the description provided above). I will go into further details about how these tasks will be conducted in the following subsections.

3.5.1 First Task: SMS Reply

The first task will be a form of text entry and text comprehension. These tasks will consist of participants receiving and replying to SMS (Short Message Service) messages on the smartphone, where the participant receives a text message on the "Messages" app where each message contains a question that the participants address in their reply. During this task, the participants receive four text messages in one trial and participants have to respond to all four messages. The messages consist of questions that are selected from a list of 15 pre-written questions and each question involve reading a single line of text (which constitutes approximately 5 to 7 characters each message) and the replies can contain anything from a single word to simple sentences. The only requirement is that the reply addresses the question they are asked. For example, one message may read "What was the last film you watched?", and an accepted response can either be "The Gladiator" or "The last film I watched was The Gladiator". The responses of the participants do not have to be a copy of the example already given, but the responses should be a understandable variant of the given response example. Table 3.1 shows the various text messages sent to the participants during the texting trials. The messages are all written and ready to be sent before the participant is finished responding to the previous message to keep the interval of 15 seconds between each sent message consistent throughout the experiment.

Text ID	**Prewritten Message**
1	"At what time will you be home?"
2	"What are you doing right now?"
3	"Where are you?"
4	"What was the last city that you visited?"
5	"What city were you born in?"
6	"What is your favourite book?"
7	"What was the last film you watched?"
8	"Where are you going on vacation this summer?"
9	"Do you have any pets? If so, what animal is your pet?"
10	"What do you enjoy doing in your free time?"
11	"I am going to the grocery store, what can I get you?"
12	"Can you name two kinds of fruit?"
13	"What is your favourite thing to eat?"
14	"What type of music do you enjoy listening to the most?"
15	"What year were you born?"

Table 3.1: List of the pre-written sentences used during the texting task trials

For this task, no specific instructions are given for grammatical structure or abbreviations that are allowed in the response. However, the participants are told to provide responses to the questions as naturally as possible as they otherwise would in a text conversation they normally would have with a friend or a family member. The participants are also required to maintain a consistent response style throughout the experimental trials and each response must address the question fully. In other words, the participants are not allowed to send multiple text messages to respond address the same text message, unless a premature message was sent inadvertently. Before each trial, the participants are also reminded to stay within the walking path the best they can and walk at their preferred walking speed.

The purpose of this task is to test and compare the writing and reading ability of the participants between interaction and mobility conditions and compare the task performance of the participants between the mobility and modality conditions. For the writing, we want to compare the performance between writing by interacting with the touch-screen and writing using voice-based text entry via a CUI, and how the context of walking affects the performance during both interaction models. For comprehension, we want to observe how well the participants comprehend text when reading it and when it is spoken to them. This is determined by the amount of times participants ask Siri to repeat the message or when they ask the experimenter to resend the message. The text messages containing the questions are sent by one of the experimenters with a time frame of 15 seconds after the message that has been sent has received a reply.

When the participant is conducting the task in the conditions that include touch interaction, they will perform the tasks in a traditional sense while either standing or walking. In general, they will perform the touch interaction condition by simply interacting with the touch-screen of the smartphone. The standing-touch and walking-visual conditions will have the following task flow during the text message task:

1. Experimenter sends message
2. Participant receives message
3. Participant reads question: *"What was the last film you watched?"*
4. Participant writes the reply using the touch-screen keyboard: *"It was The Gladiator"*
5. Participant presses send

6. Experimenter receives reply and waits 30 seconds before sending the next message

During the conversational conditions of this task, the participants are required to respond to the messages by using Apple's voice assistant Siri. Performing the texting task with Siri requires the participants to first listening to Siri read out the incoming message from the experimenter, and respond to the message by asking Siri to respond to the message and dictate the response with their speech. During these conditions, the participants are allowed to say "Repeat" to Siri if they experience difficulties understanding the TTS (Text-to-Speech) rendering or if they experience any laps in attention in the trial run. The task flow of the conversational conditions are similar to the visual ones, with difference being the interaction method for sending the message. During this task, the interaction should be hand-free and eyes-free, where the participant will only listen to the message and speak the response.

Prior to the Siri trials, the participants will conduct two training trials to familiarize themselves with the Siri functionality. The training trials are further detailed in subsection 3.4.2.

3.5.2 Second Task: Selecting Scenarios

The second task category will cover the target acquisition task. The target acquisition category will consist of 5 different tasks that will be combined into scenarios that the participants will perform. When one scenario is completed, then this will count as one trial run. The four tasks the scenarios consist of include: (1) the participants opening the "Maps" app on the smartphone and ask for the nearest bus stop, (2) the participants opening the music application and selecting a specific song to play, (3) selecting the "Weather" app and checking how the weather will look like throughout the current week, and (4) setting a reminder. Here, the purpose is to test and compare the performance of target selection between both modes of interaction. During the touch interaction trials, the participants will complete these scenarios in a traditional sense by using a finger to interact with the smartphone screen, while during the voice interaction trials, the participants will the voice assistant Siri to complete the same scenarios. Table 3.2 showcases the tasks the scenarios will consist of, and will be used in combination to create a more realistic use-case of them in a mobile setting.

ID	Task	Description
1	Directions	Using the "Maps" or "Google Maps" to find travel time and the closest bus stops.
2	Weather	Using the "Weather" app to find the weather conditions for the following day.
3	Music	Using the "Spotify" app to control music by starting and stopping a song, controlling the volume and playing another song.
4	Reminder	Using the "Reminders" app to set reminders with accurate dates and times, and a name describing what the participant will be reminded on.
5	Message	Using the "Messages" app to send a message to the experimenter. The participant will not receive a response during selecting trials.

Table 3.2: List of target acquisition tasks used in this evaluation

When performing these tasks with the CUI, the participants will perform them by issuing commands to the CUI in similar fashion as the texting task trials. The participant will issue them by activating the CUI by calling it by its wake word first ("Hey Siri") and then give the task command to the CUI. The screen will be recorded during these tasks for further reviewing and analysis of performed task errors. The tasks chosen for the target acquisition part of the experiment were based on the common tasks that voice assistants can perform and that make sense with what tasks are common to perform on the smartphone [79]. When the participants perform a test trial, they will perform one of the three scenarios showcased in the following subsections.

Scenario 1: Route Planning

In the first scenario, the participants are required to plan their way from the campus at Østfold University College to their home. First, the participants have to find out how long it will take to get home (via the Maps app). After they have assessed how long it will take to get home, they need to know what time it is and set a reminder to leave in 30 minutes to be able to get home in time. When it is time to leave, the participant will ask for bus routes for travelling home, choose the fastest option, and start the GPS for getting home via bus.

Scenario 2: Music Selection

In the second scenario, the participants are required to select between different songs and playlists while during a walk. When starting the walk, the participants will select a specific song to play. After listening for 15-20 seconds, they realize they enjoy the song and proceed to increase the sound. They want to hear what else the artist has to offer, so they select "Next" on their music player. After listening for another 15-20 seconds, they decide that the song was not enjoyable enough and proceed to play a song of their liking.

Scenario 3: Planning an Outdoor Hiking Trip

In the third scenario, the participants will plan a hiking trip outside with a friend. This scenario will also involve some text entry along with the other tasks to realize a full realistic use-case scenario. The scenario will start by the participant checking the weather for the next day. They find out that the weather will be cold and decide to set a reminder to dress well tomorrow before leaving for the trip. After the reminder is set, they also need to send a message to their friend to remember to dress well for the next day.

3.6 Stimuli

This section represents the stimuli the experiment will induce on the participants during the trials. As the purpose of this study is to investigate whether a voice-based interface can relieve user's of dual-task interference with walking and smartphone interaction, I attempt to create conditions that induce multi-effects in isolation from other outside factors that might be present with mobile interaction. Here, I will go into detail on how the stimulus was achieved in this experiment, and what the their purpose is for this project work.

3.6.1 Multi-tasking Stimulus

While having a walking path for stimulating the effects of walking on mobile interaction, we also want to have some ways to measure what the effects of walking is when interacting with a mobile phone during the walking. For that purpose, we have designed tasks that the participants will complete when walking on the predefined path. These tasks combined with the walking aspects of the experiment will serve as the multi-tasking stimulus the participants will experience when completing the experimental trials. This combination of tasks and walking allows us to further measure, quantify and analyze the effects walking has on mobile computing by measuring user performance on the different tasks. Furthermore, the visual and conversational interaction types will allow for further analysis of whether a different form of interaction, in this case conversational, will lead to better task performance when walking or whether it will be equal or worse than performing mobile computing in the traditional way. The tasks that will assess the performance of the participants during the various walking and interaction conditions are detailed in section 3.5.

3.7 Hypotheses

For this experiment, I perform a statistical analysis over the data that is collected from the experiment. For testing the research questions presented in Section 1.2.1, hypotheses are formulated based on these research questions and also based on the findings from the existing literature in Chapter 2.

3.7.1 Walking Performance

In the theory by Wickens over multiple resources, it is theorized that two tasks competing for the same cognitive resources (two visual channels) will interfere with each other more than two task that recruit separate cognitive resources (one visual channel and one auditory channel) [2, 22]. Moreover, previous literature also showcases how walking and mobile interaction negatively impacts walking in terms of walking performance due to fragmented attention [18, 34]. Based on the findings from the literature review, as well as the theory by Wickens over multiple resources, I present the following hypotheses for walking conditions:

- *Hx:* There will be significantly less performed path errors when using voice interaction compared to touch interaction on both tasks.

In addition to path accuracy, I also compare the walking speed for both interaction modalities where I test the following hypothesis:

- *Hx:* Participants will walk faster when completing smartphone tasks with voice interaction compared to touch interaction on both tasks.

3.7.2 Task Performance

With reports indicating that voice assistants have the ability to aid permanently visually impaired users in reading and texting tasks [81] and text entry with desktop computers in motor impaired users [83], it can be assumed that a voice assistant may improve task performance in texting trials in terms of accuracy and errors in multi-tasking situations as divided attention can be considered a visual impairment [87]. Based on the same theory by Wickens in terms om multiple resources and performance [2] and the findings in previous literature, I present one hypothesis for standing conditions (Hs) and one for walking conditions (Hw):

- *Hs:* There will be fewer performed task errors when using voice interaction compared to touch interaction on texting tasks.
- *Hw:* There will be fewer performed task errors with voice interaction compared to touch-screen interaction on texting tasks.

Additionally, with other reports also suggest that voice-based interfaces may aid visually impaired users in target selection and browsing tasks [88], it can also be assumed that CUIs may improve task performance on selection tasks in terms of errors in multi-tasking situations since divided attention can be viewed as a visual impairment [87]. It has also been shown that CUIs may aid blind users in asking for weather forecasts and time of day [82]. Furthermore, it is also expected that the participants will perform less errors with touch interaction when standing due to the lack of multi-tasking stimuli. Based on the same theory by Wickens, and the findings in previous literature pertaining to how voice assistants help permanently impaired users in performing smartphone tasks such as target selection and browsing [88], I present the following hypotheses for selection tasks and task errors:

- *Hs:* There will be fewer performed task errors when using touch interaction compared to voice interaction on selection tasks.

- *Hw:* There will be fewer performed task errors with voice interaction compared to touch interaction on selection tasks.

In terms of task completion time, it is expected that the participants will perform tasks with touch interaction faster than voice interaction when standing due to the lack from multi-tasking stimuli as there is no secondary task such as walking. However, it is expected that the task completion time will be shorter with voice interaction on the walking trial conditions due to the expected reduction in mental demand since voice intefaces have the ability to reduce the effects of multi-tasking [22]. With this in mind, I present the following hypotheses:

- *Hs:* Task completion time will be longer with voice interaction compared to touch interaction for texting tasks.

- *Hw:* Task completion time will be shorter with voice interaction compared to touch interaction for texting tasks.

Additionally, I also present two hypotheses for task completion time pertaining to selection tasks based on the same premises as with texting tasks:

- *Hs:* Task completion time will be shorter with touch interaction compared to voice interaction for selection tasks.

- *Hw:* Task completion time will be longer with touch interaction compared to voice interaction for selection tasks.

3.7.3 Perceived Workload

In terms of perceived mental workload with mobile interaction, multiple resource theory argues that mental workload will be reduced when separate cognitive resources are recruited by two different tasks that are performed at the same time as opposed to two tasks competing for the same cognitive resource when performed at the same time [22]. Since SIIDs have the same effect on mobile interaction as permanent impairments, and voice assistants have been shown to aid patients of dementia without the patients being submitted to frustrations and loss of patience [81], we can assume that a voice assistant might reduce frustrations with both texting and selection tasks compared to touch interaction. With this in mind, it is reasonable to assume that an alternate modality for smartphone interaction during mobile interaction may reduce frustrations as a result of multi-tasking stimuli. Based on previous literature by Wolters [81] and colleagues and the theory by Wickens, I present the following hypotheses:

- *Hs:* Perceived frustration will be significantly lower with touch interaction compared to voice interaction for both tasks.

- *Hw:* Perceived frustration will be significantly lower with voice interaction compared to touch interaction for both tasks.

In terms of perceived physical demand, it is reasonable to assume that standing conditions paired with traditional touch interaction on smartphones results in less physical demand due to no divided attention that is present with walking as well as interacting with the smartphone in a traditional fashion. It is also reasonable to assume that physical demand will be reduced when walking during voice interaction for both tasks due to recruited resources being shared between visual and auditory channels [2, 22]. With this in mind, I present the following hypotheses for physical demand:

- *Hs:* Perceived physical demand will be lower with touch interaction compared to voice interaction for both tasks.
- *Hw:* Perceived physical demand will be significantly lower with voice interaction compared to touch interaction for both tasks.

In terms of mental demand, the theory of multiple resources pertaining to mental workload by Wickens argues once again that divided attention during multi-tasking situations is reduced when two tasks are not competing for the same cognitive resources which can be achieved when the tasks are shared between visual resources and auditory resources [22]. With this in mind, I present the following hypotheses for mental demand:

- *Hs:* Perceived mental demand will be lower with touch interaction compared to voice interaction for both tasks.
- *Hw:* Perceived frustration will be lower with voice interaction compared to touch interaction for both tasks.

Additionally, based on the theories by Wickens in terms of performance and mental workload, I also present the following hypotheses for perceived effort:

- *Hs:* Perceived effort will be lower with touch interaction compared to voice interaction for both tasks.
- *Hw:* Perceived effort will be lower with voice interaction compared to touch interaction for both tasks.

In terms of how the participants perceive their own performance, the theory of multiple resources on mental workload and task performance [2, 22] suggests that the participants will experience better performance with a CUI on walking conditions. However, due to the lack of a multi-tasking stimuli in the standing conditions, it is suggested that participants will experience better performance with a modality that they are used to. With this in mind, I present the following hypotheses:

- *Hs:* Perceived performance will be better with touch interaction compared to voice interaction for both tasks.
- *Hw:* Perceived performance will be lower with voice interaction compared to touch interaction for both tasks.

In terms of temporal demand, where the participants report how rushed or stressed they felt when performing the tasks during the experiment, I expect the participants to feel less temporal demand when using a CUI during the walking condition trials. However, due to familiarity, I also expect the temporal demand to be lower with touch interaction during the standing condition trials. Therefore, I present the following hypotheses:

- *Hs:* Perceived temporal demand will be lower with touch interaction compared to voice interaction for both tasks.
- *Hw:* Perceived temporal demand will be lower with voice interaction compared to touch interaction for both tasks.

3.8 Ethical Considerations

For any research that is conducted, it is important to be knowledgeable of the ethical implications when conducting scientific experiments with human subjects. Since the work on my project is being done on behalf of Østfold University College, I will adhere to the guidelines for research ethics [3]. The research guidelines for research ethics within Østfold University College state that the main researcher, which in this context represents myself,

[3] Research Ethics Guidelines - Østfold University College: https://www.hiof.no/for-ansatte/english/work-support/research-support/research-ethics/index.html

must bear the responsibility for ensurance that my research work is within general ethical principles. Additionally, the research work must have a beneficial value to society and avoid any controversial implications, while still contributing to relevant research fields. This research project has no negative or harmful implications for society.

Involving human subjects concerns the process of inviting potential users for testing and inquiring data for the research work. Before starting the experiment with subjects, I obtained informed consent on paper along with a signature form the individuals I collected data from, providing them with details about the project and the purpose for using the data collected from them before and after the data collection process.

During the data collection process, I wanted to measure data that would help me assess the performance and mental workload between the two modes of mobile interaction. Along with this, I also asked myself what data would help me answer my research questions and give me answers to my hypotheses. Before starting any data collection process, The Norwegian Centre for Research Data [4] requires researchers to send in an application for collection of personal data and review the purposes for the collection of personal data. As my research involves collecting data by using questionnaires for collecting age and gender for the purpose of assessing the demographic for the test subjects, as well as video recording the participants during their experiment trials for further performance analysis, I have to register the personal data that will be collected from my test subjects as I relied on these data for my research. Based on the guidelines from Østfold University College, the research performed in this experiment are within ethical principles.

[4] Norwegian Center for Reseach Data https://www.nsd.no/

3.9 Summary

This chapter has presented the methodology and design of the experiment for the project work. Furthermore, the chapter also describes the setup of the experiment in terms of apparatus and materials which involve GoPro cameras, an iPhone 12 smartphone, and AirPods Pro for conducting the experiment. Additionally, the thesis will utilize the experiment to test hypotheses based on the theory by Wickens [2, 22] in terms of task performance and walking performance as well as investigate mental workload of the participants. As I take a quantitative approach to collect data from the methods presented in this chapter, the analysis will be conducted in a similar manner by utilizing statistical analysis. Subsequent setions describe the design of the experiment in terms of tasks the participants will perform in both touch and voice interaction, including the procedure of the experiment. 16 accepted invitations to take part in the experiment, however only 15 performed the experiment due to one participants having an impairment which could impact the data. The chapter concludes with an evaluation of the ethical considerations of the faculty which were followed in order to be able to perform the experiment and project.

Chapter 4

Results

This chapter presents the results from the experimental evaluation described in Chapter 3. I present the results and findings of the experiment based on the walking and task performance of the participants as well as the mental workload of the participants during the experiment. The chapter begins by presenting the statistical analyses of the walking and task performances of the participants in the texting and selection tasks, in addition to the significant differences analyzed between the tasks and conditions of the experiment. Finally, I present the statistical analysis of the mental workload of the participants throughout the different tasks and conditions of the experiment that are determined by the NASA-TLX scores given by the participants after each condition has been completed.

4.1 Pre-Experimental Questionnaire

This section covers the findings pertaining to the questionnaires given to the participants prior to conducting the experimental evaluation. As previously mentioned in Section 3.4.1, the pre-experimental questionnaire acts as an preliminary assessment of the participants in terms of their previous experience with using voice user interfaces such as Apple's

Siri or Google Assistant. The questionnaire asked the participants about their gender and age, as well as their current smartphone of use, and finally if they are using a voice interface or have used one and their experience with the voice interface.

The results from the pre-experimental questionnaire showed that all of the participants were iPhone users, where all of the participants use Apple's Siri as their main voice user interface. All of the participants reported to have been using a smartphone for longer than 5 years when questioned beyond the questionnaire of their smartphone use experience. In terms of the participants' prior experience to using conversational user interfaces and voice assistants, 16.7% of the participants reported to using Apple's Siri actively, 16,7% reported to using Google Assistant actively, and 66,7% reported to not using any voice assistant or conversational user interface actively. Furthermore, 25% of the participants also reported to have never used a voice assistant before, 16,7% reported to only have used a voice assistant once or twice since owning a smartphone or device with a voice assistant, while 41,7% report to have used voice assistants on occasion and 16,7% report to have used voice assistants many times.

Out of those that have answered to have used a voice assistant to interact with devices, 66,6% report to rarely use their voice assistant. Out of the participants that actively use a voice assistant prior to participating in the evaluation, 33,4% report that they use their voice assistant on occasion when either driving or at home. The same participants also reported to using voice assistants primarily for sending and responding to text messages, controlling music players by stopping or starting music and shifting through songs, and conducting web searches via the voice assistant. Other tasks that were less commonly performed by the participants included setting reminders, asking for directions and locations from a GPS, and checking the weather.

Of the 16 participants that answered the pre-experimental questionnaire, 50% reported to have faced difficulties when using a voice assistant for mobile interaction. Of the difficulties experienced by the participants, they reported speech recognition errors and general misunderstandings to be the most experienced difficulties with voice interaction. 50% of the participants reported misunderstandings as their most experienced difficulty, where they explained that the voice assistant would at times issue the wrong command or not respond. However, only 16,7% reported to have experienced speech recognition issues in regards to text entry tasks which include sending text messages and writing notes. Finally, all of the participants were asked to rate their experience with using a voice assistant from 0 to 7, where the average result from the 16 participants resulted in an average score of 3.9375 and a median score of 4.5.

4.2 Effects of Text Messages

For the texting trials, I also wanted to analyze whether the types of text messages would impact the results of the participants in terms of task and walking performance. For this reason, I examined the effects of the various text messages on task errors, path errors, and the message response time. After the data had been collected by analyzing the performance of the participants from the GoPro recordings and the performed task errors and text response times from the screen-recordings on MaxQDA, a series of repeated measures ANOVAs were conducted to compare the scores of task errors, path errors and response time between the touch interaction and voice interaction.

A two-way repeated measures ANOVA *(Text ID (15) x Modality (2))* was performed on text response time. The results of the ANOVA showed that the main effect of text type ($p = 0.367$) and the interaction with modality ($p = 0.694$) were not significant . Additionally, the ANOVA also showed a significant main effect of interaction modality on text response time ($F(1, 14) = 63.197$, $p < 0.001$), because text response time was significantly longer with voice interaction compared to touch interaction (Mean Difference (MD) = 17.262, Standard Deviation (SD) = 2.171, $p < 0.001$). Figure 4.1 illustrates the graph showing the mean scores of response time between the various text messages and the two interaction modalities.

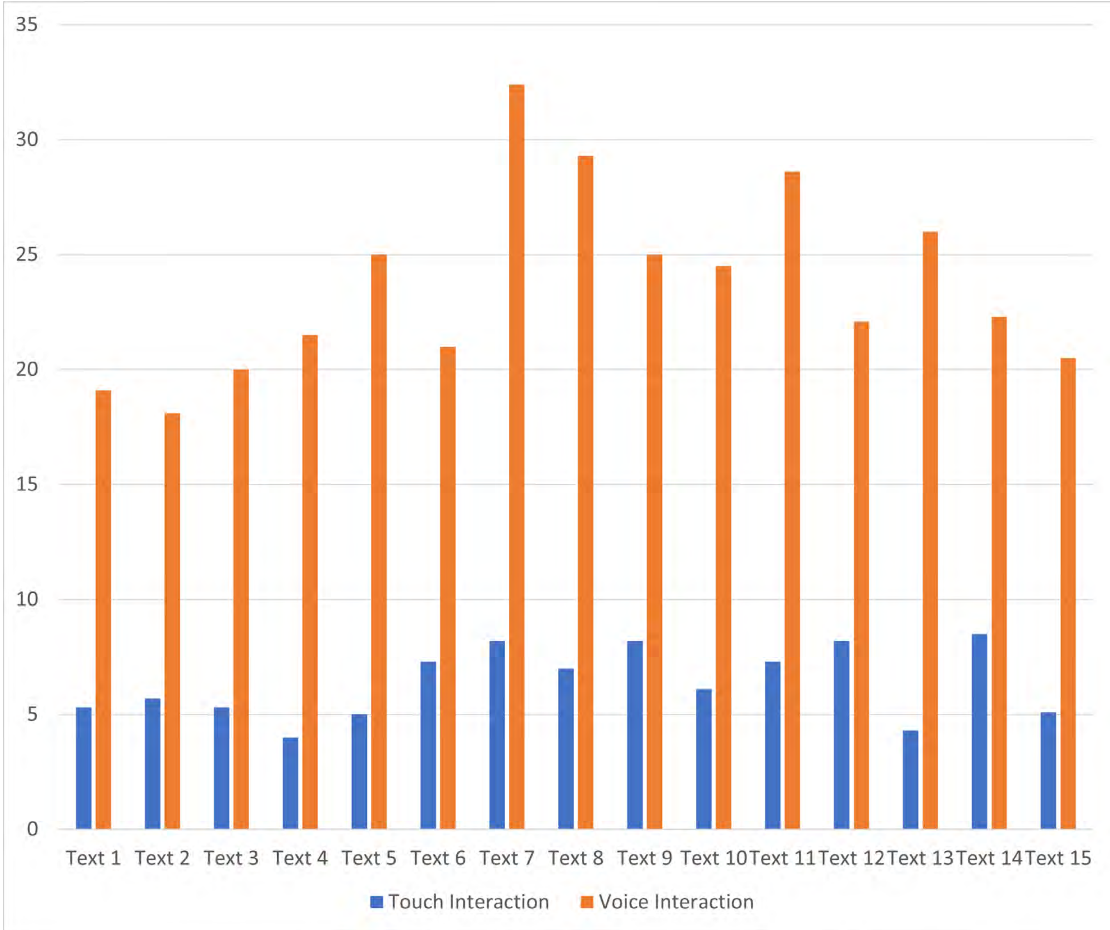


Figure 4.1: Plot representation of the mean response times of the text messages

Additionally, a two-way repeated measures ANOVA *(Text ID (15) x Modality (2))* was performed on texting task errors. The results from the ANOVA showed that the main effect of text type ($p = 0.615$) and the interaction with modality ($p = 0.457$) were not significant. Additionally, the ANOVA also showed that the main effect of interaction modality was significant ($F(1, 14) = 10.344$, $p = 0.006$), because participants performed more errors with voice interaction compared to touch interaction on the various text messages ($MD = 0.480$, $SD = 0.149$, $p = 0.006$). Figure 4.2 illustrates the graph showcasing the mean scores of task errors between the various text messages and the two interaction modalities.

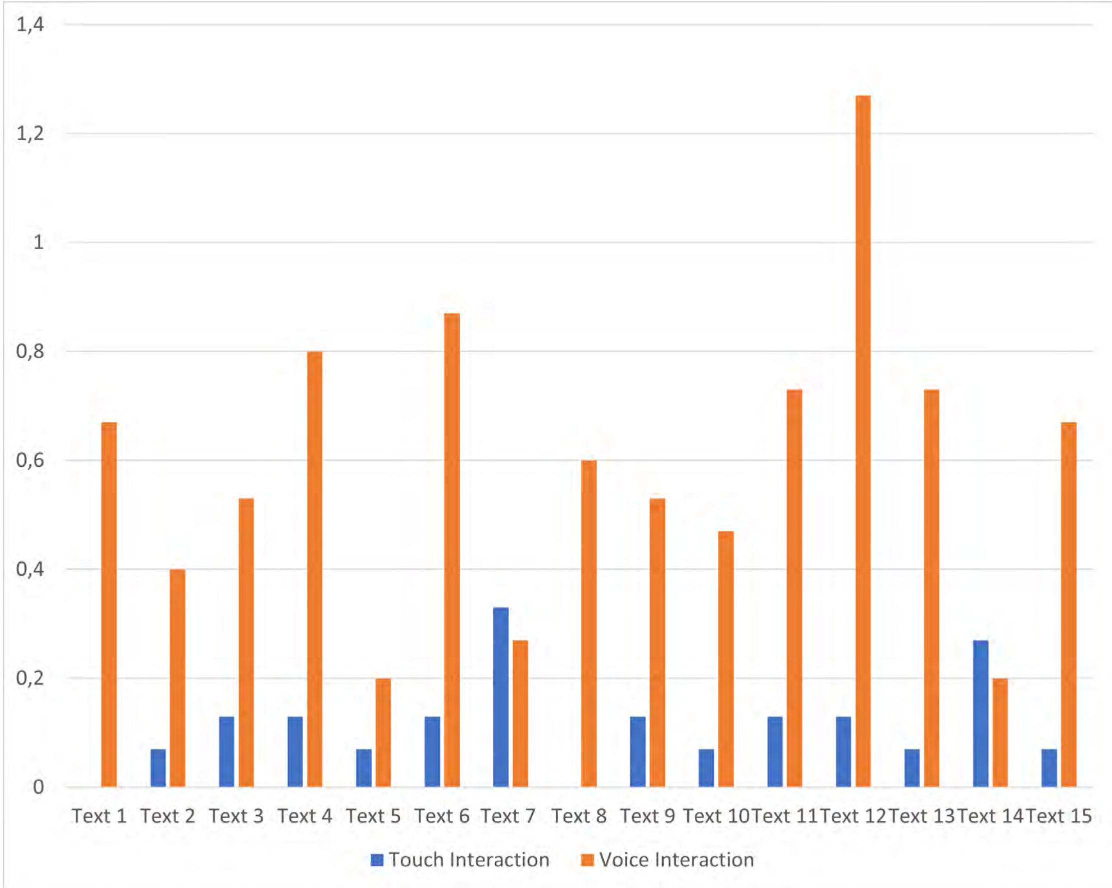


Figure 4.2: Plot representation of task error means present with each text message.

Finally, a two-way repeated measures ANOVA *(Text ID (15) x Modality (2))* was performed on path errors of the texting trials. The results of the ANOVA show that the main effect of text type ($p = 0.417$) and the interaction with modality ($p = 0.605$) were not significant a main effect of interaction modality was significant ($F(1, 14) = 8.307$, $p = 0.012$) for path errors, where participants performed more path errors when engaging in touch interaction compared to voice interaction when replying to text messages ($MD = 0.387$, $SD = 0.134$, $p = 0.012$). Figure 4.3 showcases the graph illustrating the difference in mean scores of path errors between the various text messages and interaction modalities.

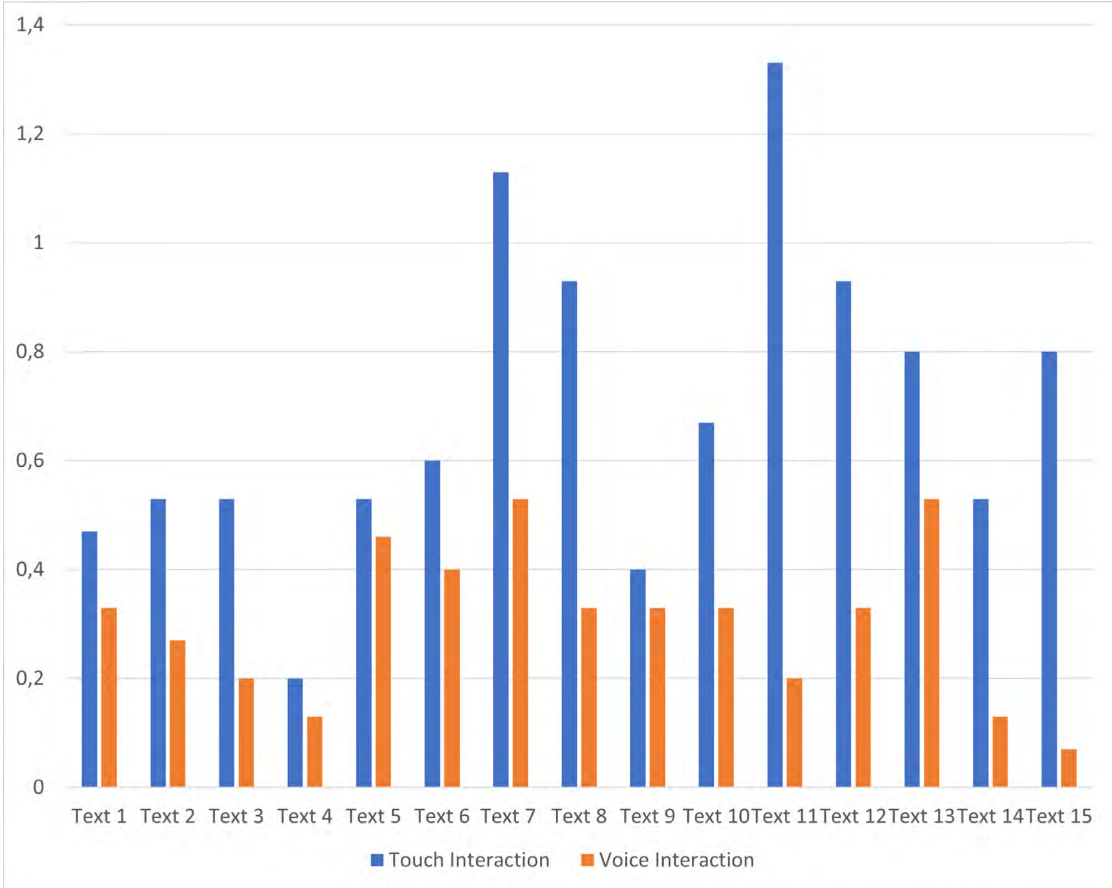


Figure 4.3: Plot representation of path error means present with each text message.

4.3 Walking Performance

This section covers the results of the walking performances of the participants when completing the experiment. When analyzing the walking performance of the participants, I wanted to compare the amount of walking errors performed and the walking speed between natural walking, walking with manual interaction, and walking with voice interaction of the participants for both texting and selection tasks of the experiment.

4.3.1 Path Errors

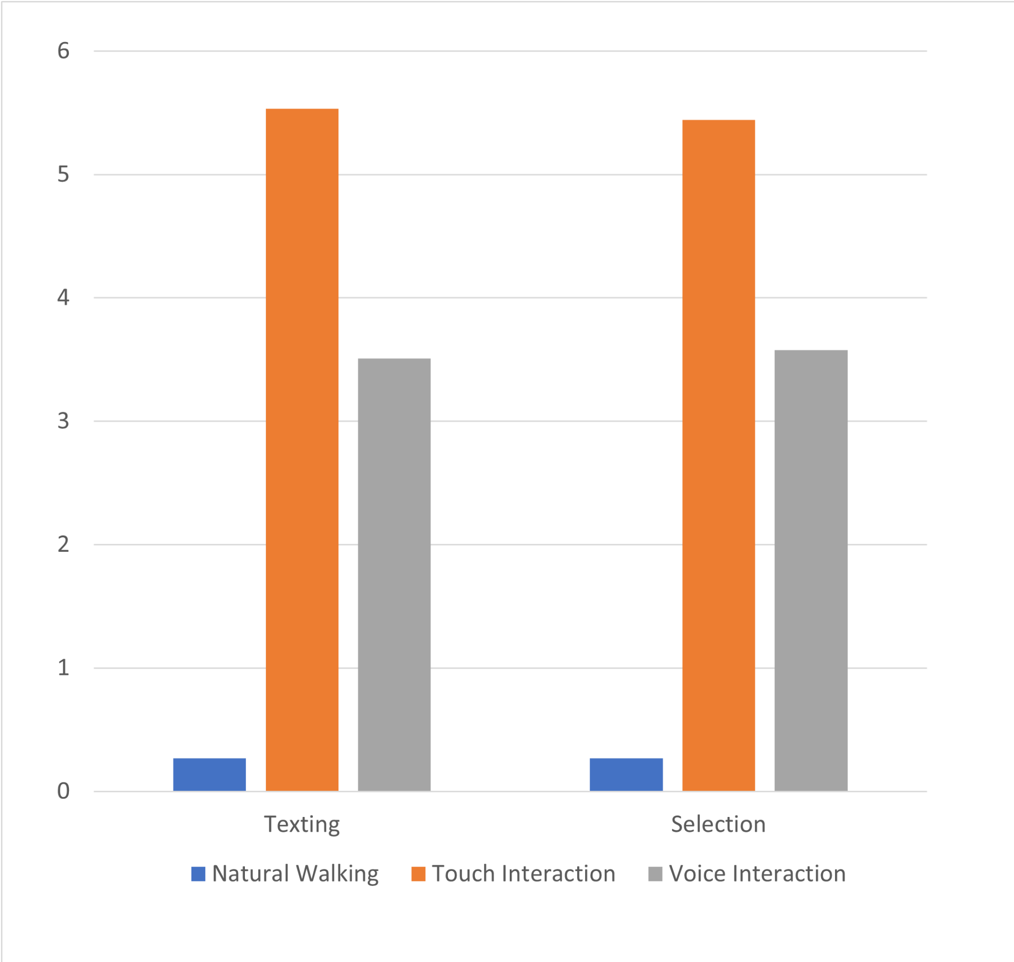


Figure 4.4: Mean score values of path errors from texting and selection task trials

A two-way repeated measures ANOVA *(Task (2) x Modality (3))* was performed on the number of path errors in the walking trials. Figure 4.4 showcases the mean scores of path errors between the different interaction modalities and task types.

The results show no significant interaction between the interaction modality and the task type in terms of path errors performed by the participants ($p = 0.465$). Additionally, the ANOVA results also show that the main effect of task type was not significant in terms of path errors ($p = 0.654$). However, the results showed that the main effect of interaction modality was significant ($F(2, 28) = 16.226$, $p < 0{,}001$). The results of pairwise comparisons show that participants generally performed less path errors when walking naturally compared to interacting with the smartphone with touch interaction (MD = 5.221, SD = 1.209, p = 0.002). The pairwise comparisons also show that participants also performed more path errors when engaging in voice interaction with the smartphone compared to walking naturally (MD = 3.276, SD = 0.867, p = 0.006). However, participants also performed significantly less path errors when using voice interaction compared to using touch interaction (MD = 1.945, SD = 0.601, p = 0.018), which confirms my hypothesis presented in Section 3.7 which states that there will be less path errors performed with voice interaction compared to touch interaction during mobile interaction. Table 4.1 showcases the mean values and standard deviations between both interaction modalities and natural walking, as well as both task types.

	Natural Walking	Touch Interaction	Voice Interaction
Texting	0.27 (1.03)	5.53 (5.12)	3.50 (3.81)
Selection	0.27 (1.03)	5.44 (5.17)	3.57 (3.71)

Table 4.1: Means and standard deviations in path errors from texting and selection trials

4.3.2 Walking Speed

In terms of walking speed, this was measured by counting the amount of full laps and partial laps the participants completed, where the amount of laps was used to measure the distance that was travelled. After measuring the distance, the distance was then divided by the completion time the participant for each trial. This gives us the walking speed measured in meters per second (m/s) for each walking trial for each participant. The mean walking speed for each walking trial was compared between interaction modalities and task types during the statistical analysis. Figure 4.5 illustrates the graph showcasing the mean score values of walking speed for the texting and selection trials.

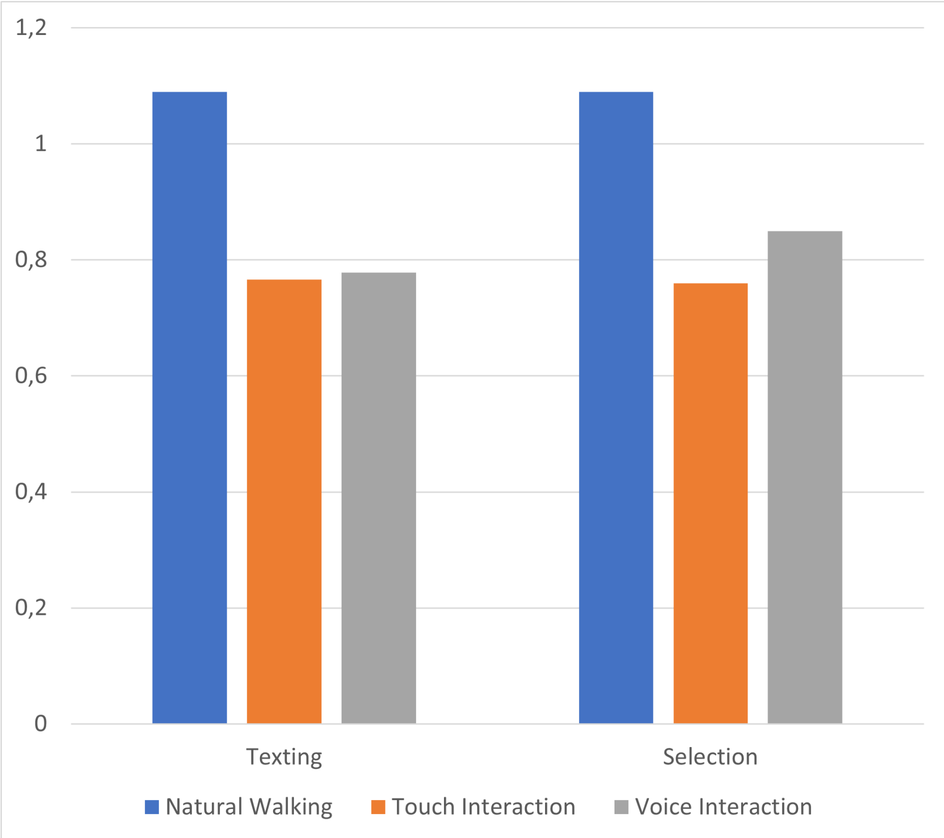


Figure 4.5: Mean score values of walking speed from texting and selection task trials

A two-way repeated measures ANOVA *(Task (2) x Modality (3))* was performed on walking speed of the walking conditions, where touch interaction and walking conditions were used as a baseline. The results show a main effect of interaction modality on walking speed ($F(2, 28) = 55$, $p = 0.019$), because participants generally walked faster without engaging in smartphone interaction compared to both touch interaction (MD = 0.326, SD = 0.042, $p < 0.001$) and voice interaction (MD = 0.276, SD = 0.035, $p < 0.001$). The pairwise comparisons also showed that the difference in walking speed between touch interaction and voice interaction was not significant (MD = 0.051, SD = 0.019, $p = 0.057$). Furthermore, the results also show that a main effect of task type was not significant ($p = 0.056$). Table 4.2 shows the mean values and standard deviations of the walking speed between the interaction modalities and natural walking, as well as between the two task types.

The results also show that the interaction between interaction modality and task type was significant in terms of walking speed ($F(2, 28) = 4.778$, $p = 0.016$). The participants walked faster with voice interaction compared to touch interaction when completing selection tasks (MD = 0.089, SD = 0.024, $p = 0.006$). However, the difference in walking speed between touch- and voice interaction was not significant for texting tasks (MD = 0.012, SD = 0.028, $p = 1.00$). Additionally, pairwise comparisons also show that participants walked slower when interacting with a smartphone for texting tasks with both touch interaction (MD = 0.323, SD = 0.046, $p < 0.001$) and voice interaction (MD = 0.311, SD = 0.041, $p < 0.001$) compared to natural walking without engaging in smartphone interaction. Pairwise comparisons show that the same is also true for touch interaction (MD = 0.329, SD = 0.040, $p < 0.001$) and voice interaction (MD = 0.240, SD = 0.34, $p < 0.001$) for selection tasks compared to natural walking without smartphone interaction.

	Natural Walking	Touch Interaction	Voice Interaction
Texting	1.09 (0.09)	0.77 (0.18)	0.77 (0.15)
Selection	1.09 (0.09)	0.76 (0.17)	0.85 (0.13)

Table 4.2: Means and standard deviations of walking speed in both task trials

4.4 Task Performance

This section presents the results of the task performance of the participants when conducting the experiment. For assessing the task performance of the participants, both the number of errors and the task completion times were measured for both the texting task trials and the selecting task trials.

4.4.1 Task Errors

When analyzing the task performance of the participants in terms of errors performed, the video recordings were used to observe the amount of errors performed on MaxQDA where both the screen-recordings and the recordings from the camera that recorded the participants's walking were used to observe and log the amount of errors that were performed during the experiment.

Before comparing the effects of task, modality and mobility on task errors, I investigated the effects of interaction modality for the texting trials in the walking conditions, and whether there were significant differences between the three trials with walking and texting task conditions. This is due to the randomization of text messages sent to the participants between texting trials. A two-way repeated measures ANOVA *(Trial (3) x Modality (2))* was performed on task errors from the trials with texting task and walking conditions. The results show that the main effect of modality was significant in terms of task errors during texting task trials with walking conditions ($F(1, 14) = 8.940$, $p = 0.010$). A pairwise comparison with Bonferroni correction shows that the participants performed significantly more task errors with voice interaction compared to touch interaction with texting tasks on walking trial conditions ($MD = 2.244$, $SD = 0.751$, $p = 0.010$). Additionally, the results also show that the main effect of trial ($p = 0.540$), and the interaction with interaction modality ($p = 0.576$), were not significant in terms of task errors.

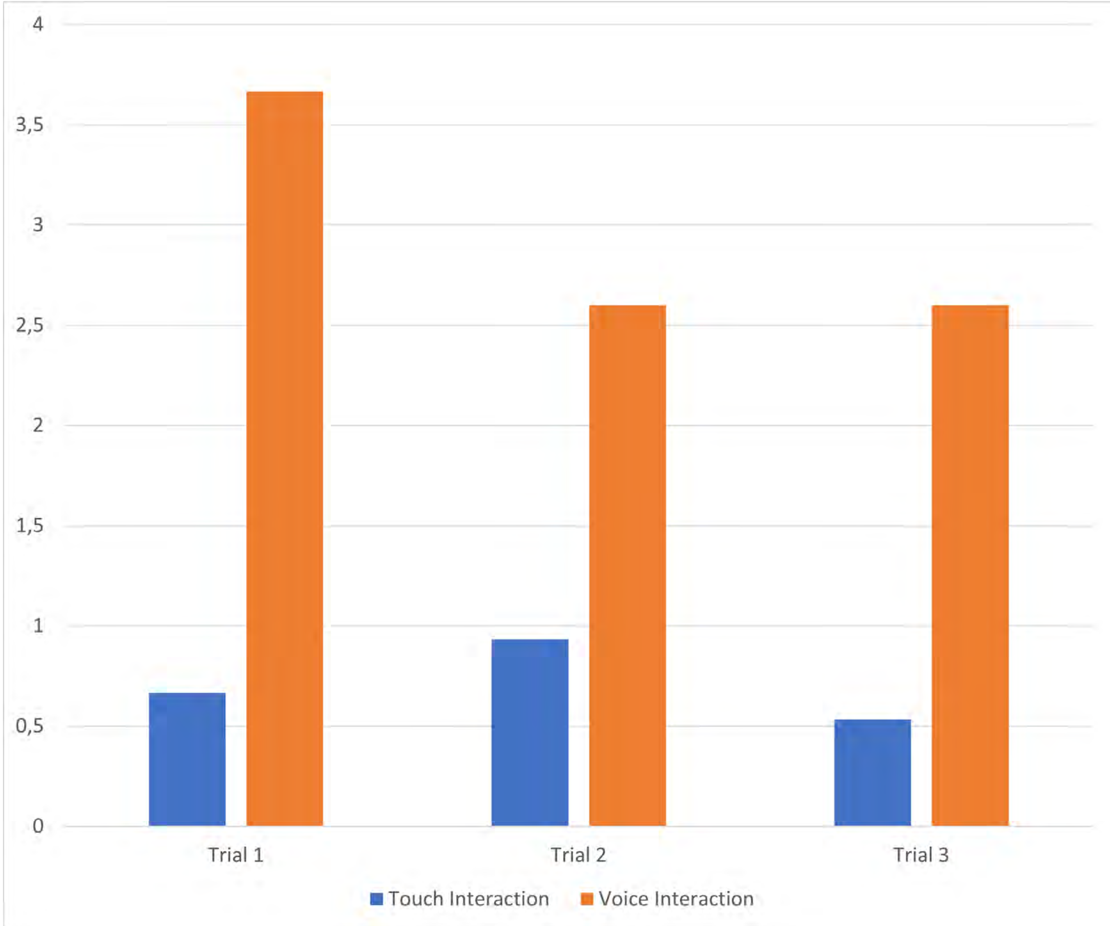


Figure 4.6: Mean task errors between texting trials

	Touch Interaction	Voice Interaction
Trial 1	0.67 (0.82)	3.67 (5.01)
Trial 2	0.93 (1.49)	2.60 (3.43)
Trial 3	0.53 (0.64)	2.60 (3.43)

Table 4.3: Means and standard deviations of task errors in texting trials

Since there was no significant trial effect with the texting tasks, I investigated the effects of interaction modalities on both tasks in terms of task errors. When investigating the effects on task errors, another two-way repeated measures ANOVA

(Task (2) x Modality (2)) was performed on the mean task errors for both tasks in the walking conditions. The results from the ANOVA show a significant main effect of interaction modality ($F(1, 14) = 10.466$, $p = 0.006$), where pairwise comparisons with Bonferroni correction show that participants performed less task errors in general with voice interaction compared to touch interaction for both texting and selection tasks (MD = 1.945, SD = 0.601, $p = 0.006$). Figure 4.7 shows the average task errors performed on the walking trials, while Table 4.7 shows the mean and standard deviations for the performed errors on both tasks and interaction modalities.

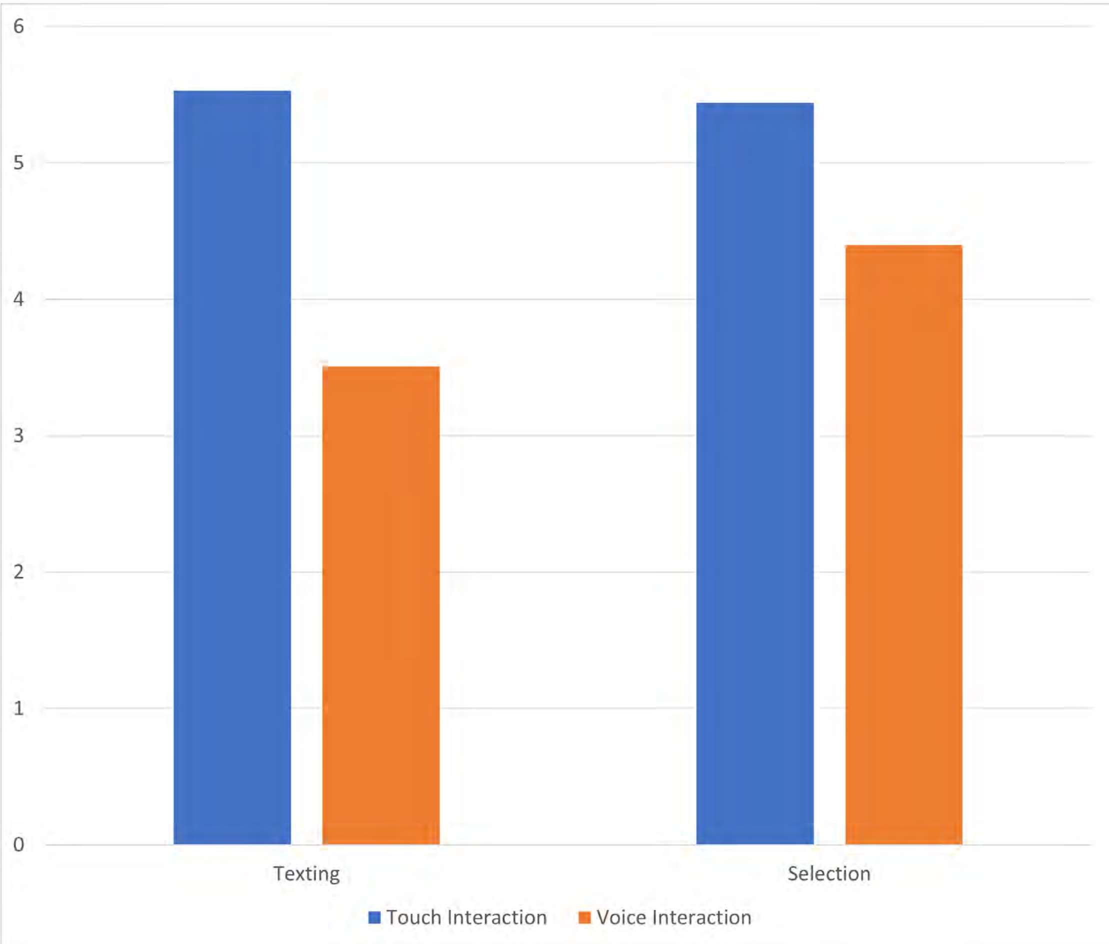


Figure 4.7: Mean score values of average task errors for walking trials

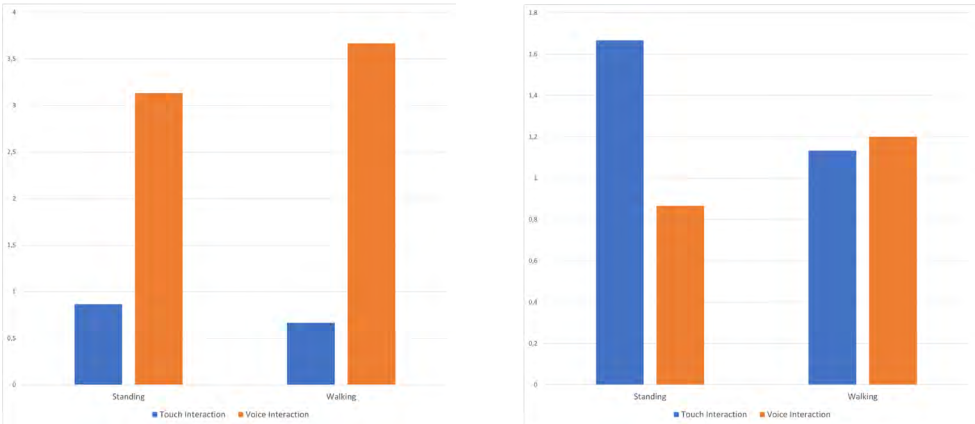
	Touch Interaction	Voice Interaction
Texting	5.53 (5.12)	3.51 (3.81)
Selecting	5.44 (5.18)	3.58 (3.71)

Table 4.4: Means and standard deviations of task errors in walking trials

Finally, the effects of mobility in terms of task errors were analyzed by comparing one of the three trials of the walking conditions with the standing condition trials in an effort to make the data analysis balanced. For the comparison, I chose the last trial run of the walking conditions when comparing the effects. This is due to there only being one standing trial conducted for each task type and interaction modality, and three walking trials conducted for each task type and interaction modality.

A three-way repeated measures ANOVA *(Task (2) x Modality (2) x Mobility (2))* was performed on task errors. The results show that there was a significant main effect of task type ($F(1, 14) = 6.345$, $p = 0.025$) because participants performed more errors on texting tasks compared to selection tasks (MD = 0.867, SD = 0.344, $p = 0.025$). There was also a significant main effect of interaction modality ($F(1, 14) = 5.008$, $p = 0.042$), because participants performed more errors with voice interaction compared to touch interaction (MD = 1.133, SD = 0.506, $p = 0.042$). However, the main effect of mobility was not significant in terms of task errors ($p = 0.944$).

Additionally, there was a significant interaction between modality and task type in terms of task errors performed by the participants ($F(1, 14) = 8.025$, $p = 0.013$), because participants performed more task errors with voice interaction on texting tasks compared to touch interaction (MD = 2.633, SD = 0.952, $p = 0.015$), while the difference in mean amount of errors between interaction modalities for selection tasks was not significant (MD = 0.367, SD = 0.410, $p = 0.386$). Additionally, participants performed significantly more errors in texting tasks compared to selection tasks with voice interaction (MD = 2.367, SD = 0.802, $p = 0.011$), while the difference in task errors between touch- and voice interaction for selection tasks were not significant (MD = 0.633, SD = 0.392, p = 0.128). Furthermore, the interaction between modality, mobility and task (p = 0.935), as well as the interaction between mobility and modality (p = 0.311) were not significant.



((a)) Texting ((b)) Selection

Figure 4.8: Graphs of the average task errors for texting and selection trials

Table 4.5 show the mean scores and standard deviations for each task type, interaction modality and mobility type.

	Standing-Touch	Walking-Touch	Standing-Voice	Walking-Voice
Texting	0.87 (1.64)	0.67 (0.82)	3.13 (3.64)	3.67 (5.01)
Selection	1.67 (1.45)	1.13 (1.13)	0.87 (1.13)	1.20 (1.61)

Table 4.5: Means and standard deviations of performed errors in selection trials

4.4.2 Task Completion Time

Along with measuring the amount of task errors performed by the participants, the time it took for the participants to complete the tasks was also measured. Similarly to the analysis of the task errors, the walking trials of both tasks have been compared first due to having similar amount of trials (3), where the average completion time has been compared. Later on, one of the walking trials were compared with the standing trials for both tasks. The task completion time was measured in seconds, and is also reported in seconds throughout the study.

First, the trial effect was investigated with regards to task completion time on texting trials with walking conditions. A two-way repeated measures ANOVA *(Trial (3) x Modality (2))* was performed on task completion time for texting and walking condition trials. The results showed that a main effect of interaction modality was significant ($F(1, 14) = 47.565$, $p < 0.001$) in terms of task completion time for the texting and walking trial conditions. Pairwise comparisons showed that participants performed texting tasks significantly faster with touch interaction compared to voice interaction when walking (MD = 75.867, SD = 11.00, $p < 0.001$). Additionally, the main effect of trial ($p = 0.266$), as well as the interaction with modality ($p = 0.052$), were not significant.

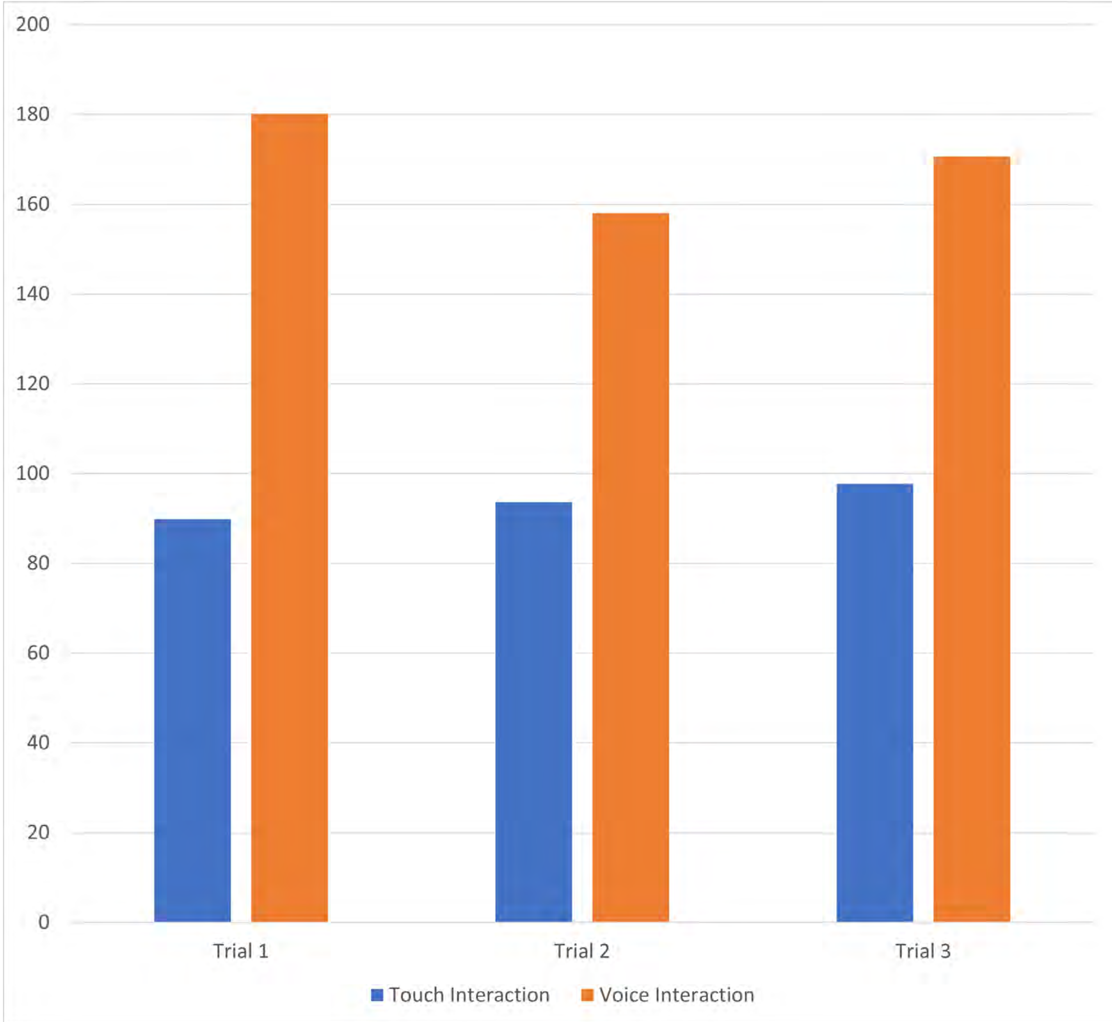


Figure 4.9: Mean completion time between texting trials

	Touch Interaction	Voice Interaction
Trial 1	89.87s (21.94)	180.20s (40.23)
Trial 2	93.60s (13.36)	158.00s (40.31)
Trial 3	97.73s (25.80)	170.60s (42.85)

Table 4.6: Means and standard deviations of completion time in texting trials

A two-way repeated measures ANOVA *(Task (2) x Modality (2))* was performed on the mean task completion time of the walking trials to examine the effect of interaction modality and task type. The results show that there was a significant main effect of task type ($F(1, 14) = 144.047$, $p < 0.001$) and a significant main effect of modality ($F(1, 14) = 22.661$, $p < 0.001$) on task completion time when walking. Pairwise comparisons show that participants generally performed selection tasks faster than texting tasks during mobile interaction ($MD = 65.133$, $SD = 5.427$, $p < 0.001$). Additionally, the pairwise comparisons also show that participants generally performed both texting and selection tasks faster with touch interaction than with voice interaction ($MD = 33.533$, $SD = 7.044$, $p < 0.001$).

Also, the interaction between task type and interaction modality was significant ($F(1, 14) = 24.374$, $p < 0.001$), because the participants performed texting tasks faster with touch interaction compared to voice interaction ($MD = 65.267$, $SD = 11.183$, $p < 0.001$), while the difference in task completion time between touch- and voice interaction for selection tasks were not significant ($MD = 1.800$, $SD = 7.537$, $p = 0.815$).

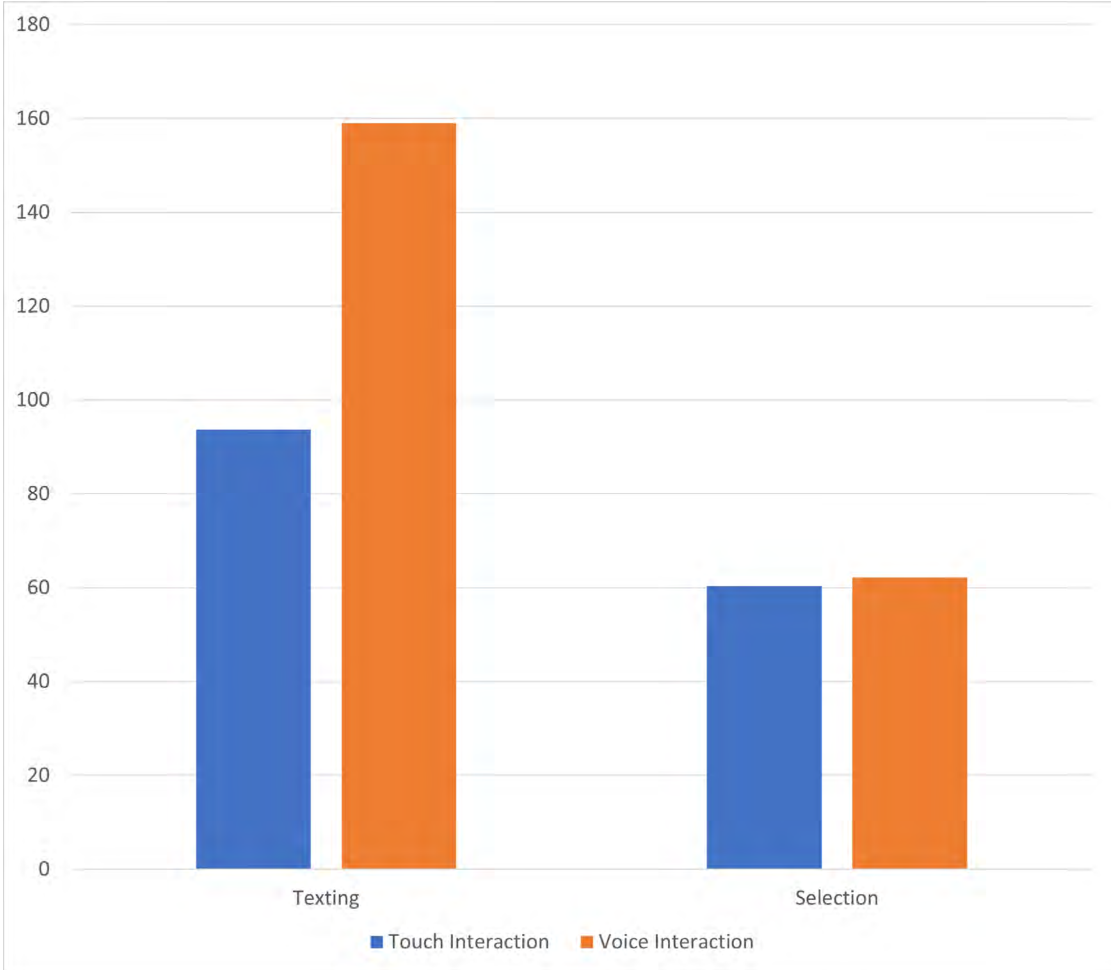


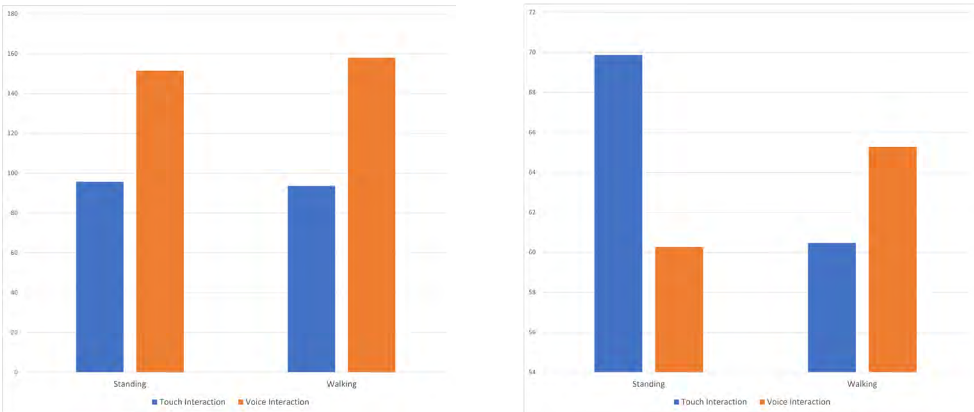
Figure 4.10: Mean task completion time for the walking trials

	Touch Interaction	Voice Interaction
Texting	93.73s (18.13)	159.00s (37.20)
Selecting	60.33s (15.40)	62.13s (23.38)

Table 4.7: Means and standard deviations of completion time in walking trials

For comparing the effects of mobility, interaction modality and task type, a three-way repeated measures ANOVA *(Task (2) x Modality (2) x Mobility (2))* was performed on task completion time. The results show that the main effect of interaction modality was significant ($F(1, 14) = 33.325$, $p < 0.001$) in terms of task completion time, where a pairwise comparison shows that participants in general performed tasks faster with touch interaction compared to voice interaction (MD = 28.833, SD = 4.995, $p < 0.001$). The results also showed that a main effect of task type was also significant in terms of task completion time ($F(1, 14) = 176.432$, $p < 0.001$), where a pairwise comparison shows that task completion time significantly decreased from texting tasks to selection tasks (MD = 60.733, SD = 4.572, $p < 0.001$). However, the results also show that the main effect of mobility was not significant ($p = 1.00$).

The results also show a significant interaction between task type and interaction modality on task completion time ($F(1, 14) = 47.680$, $p < 0.001$), where pairwise comparisons show that participants performed texting tasks faster with touch interaction compared to voice interaction (MD = 60.067, SD = 7.911, $p < 0.001$). Additionally, the differences in task completion time for selection tasks between touch interaction and voice interaction were not significant (MD = 2.400, SD = 5.313, $p = 0.658$). The pairwise comparisons also show that participants performed selection tasks faster than texting tasks with both touch interaction (MD = 29.50, SD = 3.186, $p < 0.001$) and voice interaction (MD = 91.967, SD = 8.520, $p < 0.001$). The results also show that the interaction between task type, interaction modality and mobility type was not significant ($p = 0.783$). Additionally, the interaction effect between mobility type and interaction modality was not significant ($p = 0.161$). The interaction between task and mobility was also shown to be not significant ($p = 0.686$).



((a)) Texting ((b)) Selection

Figure 4.11: Graphs of the average completion time for texting and selection trials

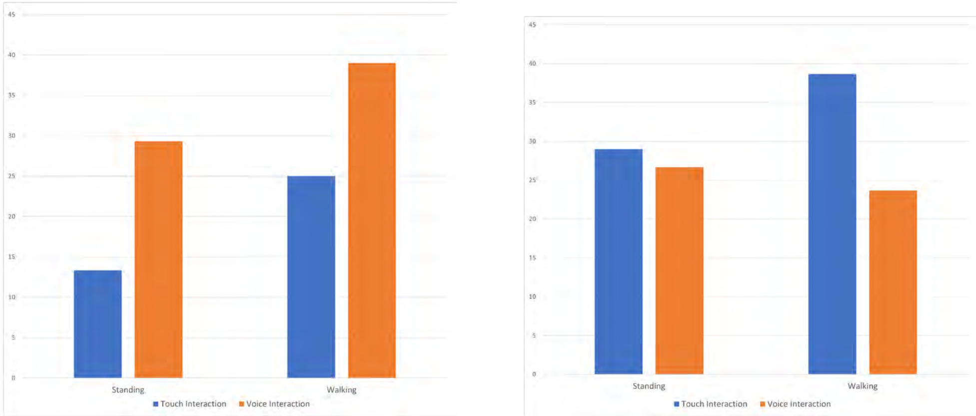
	Standing-Touch	Walking-Touch	Standing-Voice	Walking-Voice
Texting	95.73s (19.70)	93.60s (13.40)	151.50s (33.87)	158.00s (40.31)
Selection	69.87s (26.11)	60.47s (19.95)	60.27s (24.49)	65.27 (27.83)

Table 4.8: Means and standard deviations of completion time in all trials

4.5 NASA-TLX Scores

This section covers the results of the NASA-TLX questionnaires that were answered by the participants after every condition in both texting and selecting task trials. For the statistical analysis of the NASA-TLX scores, in each condition the scores of the 6 scales were calculated for each participant by using the given NASA-TLX formula: counting the number of lines the participants marked, subtracting by 1 and multiplying by 5. These scores were then compared to each other in terms of mobility and mode of interaction.

4.5.1 Mental Demand



((a)) Texting ((b)) Selection

Figure 4.12: Graphs illustrating mean mental demand between both tasks

When analyzing the mental demand of the participants from the experimental trials, a three-way repeated measures ANOVA *(Task (2) x Modality (2) x Mobility (2))* was performed on mental demand scores. Figure 4.12 illustrates the plot showcasing the mean scores of mental demand between task types, modality and mobility.

The results show that the main effect of task type was not significant for mental demand ($p = 0.367$). The main effect of interaction modality was also not significant in terms of mental demand ($p = 0.540$). There was a main effect of mobility in terms of mental demand ($F(1, 14) = 4.901$, $p = 0.044$). Unsurprisingly, the participants reported to experiencing a greater mental demand with walking as opposed to standing (MD = 7.00, SD = 3.162, $p = 0.044$).

The results also show that the interaction between mobility, modality and task was not significant ($p = 0.423$). Additionally, the results also showed that the interaction between mobility and task type was not significant ($p = 0.259$). Also, the interaction between mobility and modality was not significant ($p = 0.080$). The results also show that there was a significant interaction between modality and task type ($F(1, 14) =$

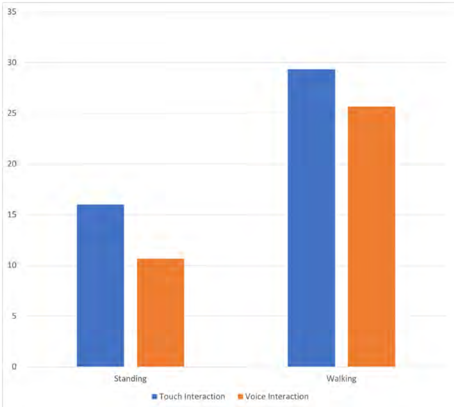
23.357, $p < 0.001$) in terms of mental demand, because the participants reported to experiencing a higher mental demand with voice interaction compared to touch interaction with texting tasks (MD = 15.00, SD = 6.045, p = 0.026). But, the participants did not report a significant difference in mental demand between touch- and voice interaction for selection tasks (MD = 8.667, SD = 5.117, p = 0.112). Additionally, the participants found the selection tasks to be more mentally demanding than texting tasks with touch interaction (MD = 14.667, SD = 3.739, p = 0.002), and they also found the texting tasks to be more mentally demanding than selection tasks with voice interaction (MD = 9.00, SD = 4.059, p = 0.044).

	Standing-Touch	Walking-Touch	Standing-Voice	Walking-Voice
Texting	13.33 (10.97)	25.00 (16.58)	29.33 (24.34)	39.00 (29.53)
Selection	29.00 (20.63)	38.67 (21.67)	26.67 (25.82)	23.67 (16.31)

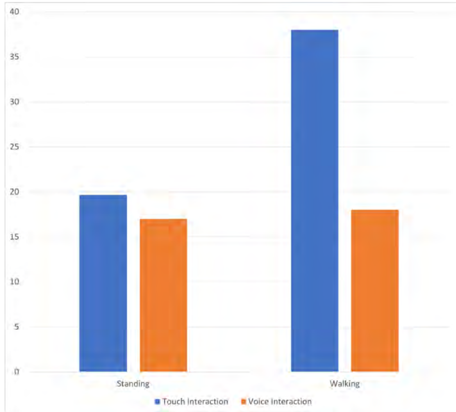
Table 4.9: Means and standard deviations of mental demand in all trials

4.5.2 Physical Demand

When analyzing the scores for perceived physical demand, a three-way repeated measures ANOVA *(Task (2) x Modality (2) x Mobility (2))* was performed on physical demand scores. Figure 4.13 illustrates the mean scores of physical demand between task types, mobility, and modality.



((a)) Texting



((b)) Selection

Figure 4.13: Graphs illustrating mean physical demand between both tasks

The results show a significant main effect of mobility in terms of physical demand ($F(1, 14) = 8.379$, $p = 0.012$), because the participants reported to experiencing greater physical demand when walking compared to standing (MD = 11.917, SD = 4.117, $p = 0.012$). Also, the main effect of task type ($p = 0.360$) and the main effect of interaction modality ($p = 0.114$) were not significant in terms of perceived physical demand.

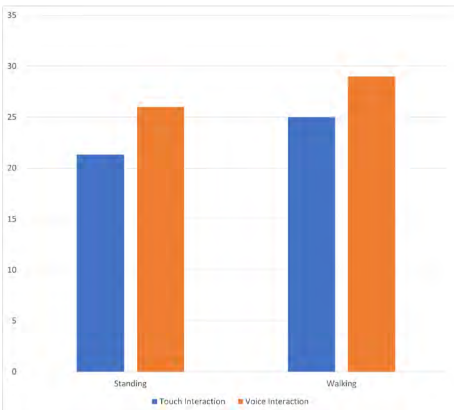
The results also show that the interaction between task type, interaction modality and mobility type was significant ($F(1, 14) = 4.765$, $p = 0.047$). Unsurprisingly, the participants generally reported to experiencing greater physical demand with walking trial conditions as opposed to standing trial conditions texting tasks with both touch- and voice interaction and for selection tasks with touch interaction. However, rather surprisingly, the reported difference in physical demand between standing and walking conditions for selection tasks with voice interaction was not significant (MD = 1.00, SD = 5.350, $p = 0.854$). Pairwise comparisons also show that participants experienced greater physical demand with touch interaction on selection tasks compared to voice interaction on selection tasks when walking (MD = 20.00, SD = 7.868, $p = 0.023$).

	Standing-Touch	Walking-Touch	Standing-Voice	Walking-Voice
Texting	16.00 (14.41)	29.33 (24.34)	10.67 (7.04)	25.67 (22.59)
Selection	19.67 (17.57)	38.00 (27.89)	17.00 (18.30)	18.00 (14.24)

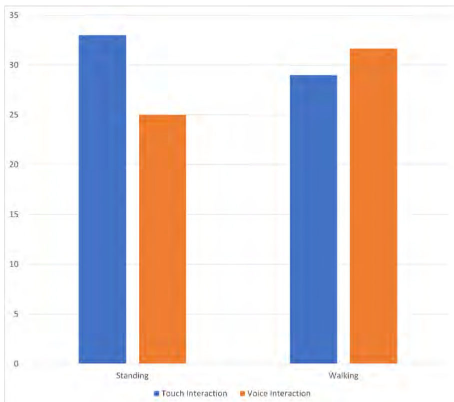
Table 4.10: Means and standard deviations of physical demand in all trials

4.5.3 Temporal Demand

When analyzing the scores for the perceived temporal demand, a three-way repeated measures ANOVA *(Task (2) x Modality (2) x Mobility (2))* was performed on temporal demand scores. Figure 4.14 illustrates the mean scores of temporal demand between task types, mobility, and modality.



((a)) Texting



((b)) Selection

Figure 4.14: Graphs illustrating mean temporal demand between both tasks

The results show that the main effect of task type was not significant ($p = 0.236$). Additionally, the main effect of mobility was also not significant ($p = 0.329$). Finally, the main effect of interaction modality was also not significant ($p = 0.787$).

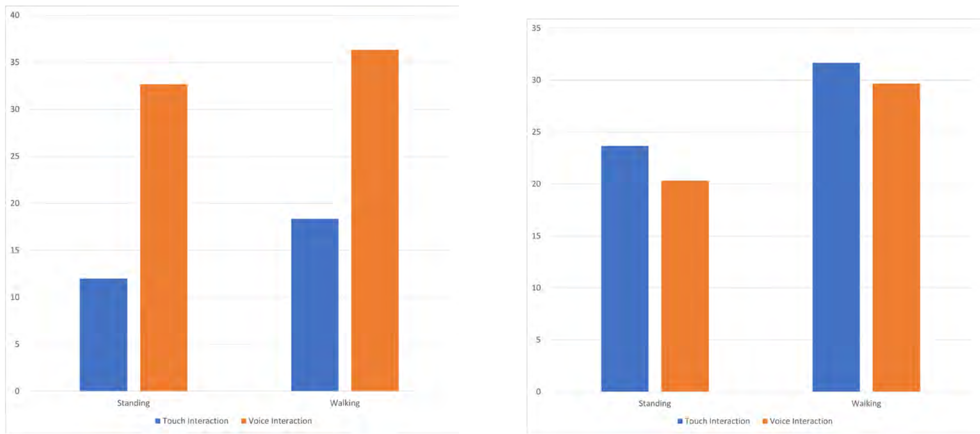
The results also show that the interaction between task type, interaction modality and mobility type was significant ($F(1, 14) = 6.939$, $p = 0.020$). However, pairwise comparisons with Bonferroni correction show no significant difference in the interaction between task type, modality, and mobility. There was also no significant interaction between mobility and modality conditions ($p = 0.179$). Additionally, there was no interaction effect between task type and interaction modality ($p = 0.279$). Finally, there was no significant interaction effect between task type and mobility type ($p = 0.597$).

	Standing-Touch	Walking-Touch	Standing-Voice	Walking-Voice
Texting	21.33 (19.59)	25.00 (20.53)	26.00 (19.10)	29.67 (21.75)
Selection	33.00 (20.42)	29.00 (20.20)	25.00 (19.82)	31.67 (24.47)

Table 4.11: Means and standard deviations of temporal demand in all trials

4.5.4 Perceived Performance

When analyzing the perceived performance scores of the participants, a three-way repeated measures ANOVA *(Task (2) x Modality (2) x Mobility (2))* was performed on the performance scores. Usually, the scales in the NASA-TLX are labelled as 0 being the most negative score and 20 being the most positive scores when filling out the form. However, the performance scale is flipped where 0 is the most positive score and 20 is the most negative score. Due to the TLX being designed this way, it is important to keep in mind that higher numbers equals worse performance when analysing the performance scores for the NASA-TLX. Figure 4.15 illustrates the mean scores of perceived performance between task types, modality and mobility.



((a)) Texting ((b)) Selection

Figure 4.15: Graphs illustrating mean of perceived performance for both tasks

The results show that the main effect of mobility was significant ($F(1, 14) = 5.676$, $p = 0.032$), because participants felt they performed tasks better when standing compared to walking (MD = 6.833, SD = 2.868). Also, the results show that the main effect of task type was not ($p = 0.642$), and the main effect of modality ($p = 0.061$), were not significant.

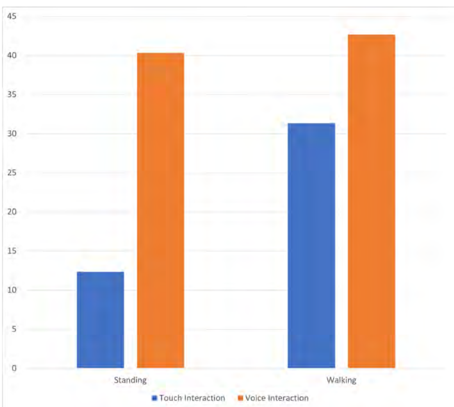
The results also show a significant interaction between task type and interaction modality ($F(1, 14) = 21.846$, $p < 0.001$) for perceived performance. The participants reported to experiencing better performance with touch interaction compared to voice interaction for the texting tasks (MD = 19.333, SD = 5.172, $p = 0.002$), because the difference in perceived performance between touch- and voice interaction for selection tasks were not significant (MD = 2.667, SD = 4.222, $p = 0.538$). Pairwise comparisons also showed that the participants performed texting tasks better than selection tasks with touch interaction (MD = 12.50, SD = 3.888, $p = 0.006$), while they also experienced better performance with selection tasks compared to texting tasks with voice interaction (MD = 9.50, SD = 3.988, $p = 0.032$). Also, the interaction between task type, mobility type and interaction modality ($p = 0.562$), as well as the interaction between mobility and modality (p = 0.864) and the interaction between task and mobility (p = 0.393), were not significant.

	Standing-Touch	Walking-Touch	Standing-Voice	Walking-Voice
Texting	12.00 (17.81)	18.33 (13.32)	32.67 (23.44)	36.33 (25.03)
Selection	23.67 (20.04)	31.67 (18.09)	20.33 (17.06)	29.67 (23.64)

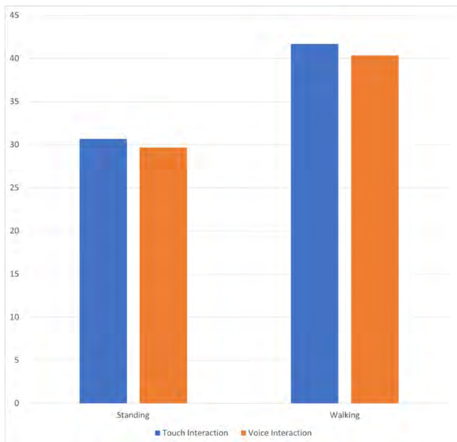
Table 4.12: Means and standard deviations of perceived performance in all trials

4.5.5 Perceived Effort

When analyzing the perceived effort scores of the participants, a three-way repeated measures ANOVA *(Task (2) x Modality (2) x Mobility (2))* was performed on the scores of perceived effort. Figure 4.16 shows the mean scores of perceived effort between task types, modality and mobility.



((a)) Texting



((b)) Selection

Figure 4.16: Graphs illustrating mean of perceived effort for both tasks

The results show that the main effect of mobility was significant (F(1, 14) = 7.259, p = 0.017) because participants felt they had to put more effort into the walking trials as opposed to the standing trials (MD = 10.750, SD = 3.990). Also, the results show that the main effect of task type ($p = 0.083$), and the main effect of interaction modality ($p = 0.088$), were not significant.

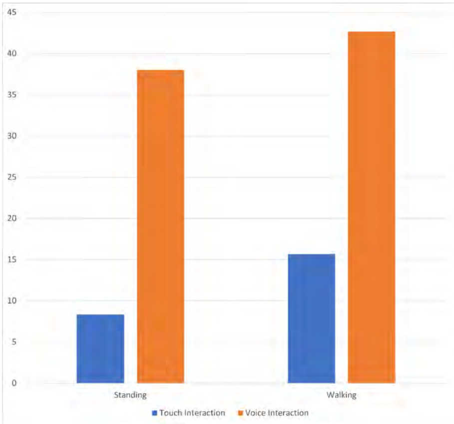
The results also show that the interaction between task and modality was significant ($F(1, 14) = 7.489$, $p = 0.016$), because participants felt they needed to put more effort into completing selection tasks than texting tasks when engaging in touch interaction (MD = 14.333, SD = 4.064, $p = 0.003$), while they felt that texting and selection tasks required similar levels of effort when engaging in voice interaction (MD = 6.500, SD = 4.608, $p = 0.180$). Additionally, pairwise comparisons also show that participants felt they needed more effort in completing texting tasks with voice interaction compared to touch interaction (MD = 19.667, SD = 7.237, $p = 0.017$), while the perceived effort when completing selection tasks were similar between touch- and voice interaction (MD = 1.167, SD = 5.255, $p = 0.827$). Furthermore, the results also show that the interaction between mobility, modality and task type ($p = 0.219$), as well as the interaction between mobility and modality ($p = 0.207$) and the interaction between task and mobility ($p = 0.980$), were not significant.

	Standing-Touch	Walking-Touch	Standing-Voice	Walking-Voice
Texting	12.33 (9.61)	31.33 (22.08)	40.33 (27.61)	42.67 (26.52)
Selection	30.67 (21.20)	41.67 (23.27)	29.67 (24.89)	40.33 (23.41)

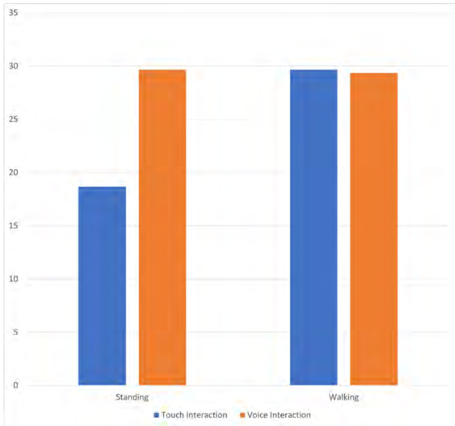
Table 4.13: Means and standard deviations of perceived effort in all trials

4.5.6 Perceived Frustration

When analyzing the perceived frustration scores of the participants, a three-way repeated measures ANOVA *(Task (2) x Modality (2) x Mobility (2))* was performed on scores of perceived frustration. Figure 4.17 illustrates the mean scores of perceived frustration between task types, modality and mobility.



((a)) Texting



((b)) Selection

Figure 4.17: Graphs illustrating mean of perceived frustration for both tasks

The results show that a main effect of mobility was significant (F(1, 14) = 6.098, p = 0.027), because participants experienced greater levels of frustration when walking compared to standing (MD = 5.667, SD = 2.295). Additionally, the results also showed that the main effect of modality was also significant in terms of perceived frustration (F(1, 14) = 6.580, p = 0.022), because participants experienced greater levels of frustration with voice interaction compared to touch interaction (MD = 16.833, SD = 6.562). However, the results also show that the main effect task type was not significant (p = 0.837).

The results also show that the interaction between task type and modality was significant (F(1, 14) = 9.503, p = 0.008), because participants experienced greater levels of frustration when using voice interaction for texting tasks compared to using touch interaction (MD = 28.333, SD = 7.648, p = 0.002), while the difference in frustration levels between touch- and voice interaction for selection tasks was not significant (MD = 5.333, SD = 7.448, p = 0.486). Additionally, pairwise comparisons also show that participants experienced increased levels of frustration when performing selection tasks with touch interaction compared to texting tasks (MD = 12.167, SD = 3.406, p =

0.003), while the difference in frustration levels between texting and selection tasks for voice interaction was not significant (MD = 10.833, SD = 6.049, p = 0.095). Also, the interaction between task type, interaction modality and mobility type (p = 0.467), as well as the interaction between mobility and modality (p = 0.195) and the interaction between task and mobility (p = 0.895), were not significant.

	Standing-Touch	Walking-Touch	Standing-Voice	Walking-Voice
Texting	8.33 (4.88)	15.67 (15.80)	38.00 (32.45)	42.67 (35.65)
Selection	18.67 (16.63)	29.67 (20.48)	29.67 (31.59)	29.33 (29.63)

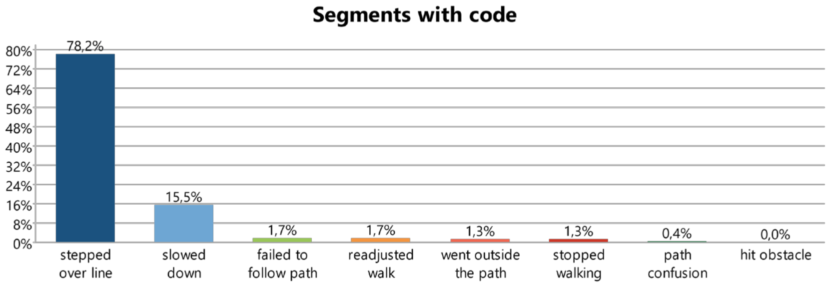
Table 4.14: Means and standard deviations of perceived frustration in all trials

4.6 Qualitative Results

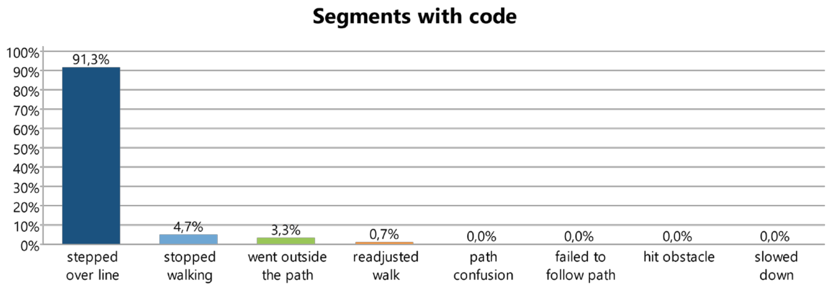
To gain further insight into the task performance, walking performance and results of the experiment, additional qualitative observation was conducted using the video footage of the participants on the MaxQDA analysis software. When observing and counting various task errors and path errors performed by the participant, the errors were categorized and labelled as an effort to observe what kind of errors were more prevalent and their frequency between the interaction modalities, mobility types and task types.

4.6.1 Path Errors

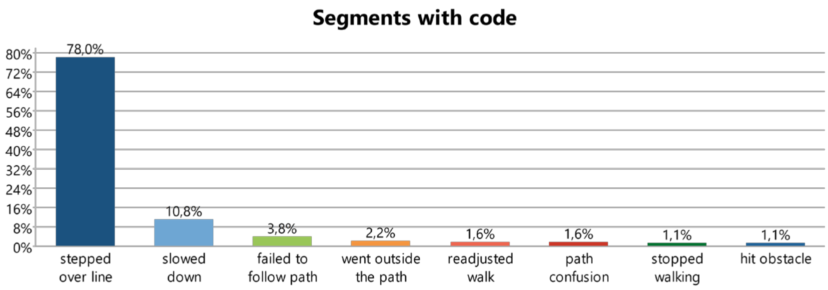
When observing the type of path errors performed during the walking trials, the following errors were observed: stepping on or over the line, slowing down the walking speed, readjusting their walking path, walking outside the outlined path, stopping to walk completely, path confusion (failing to orientate themselves when walking), and failing to follow the path the way the participants were instructed to. Figure 4.18 shows the percentage coverage of the various path errors performed by the participants during the texting trials with touch interaction.



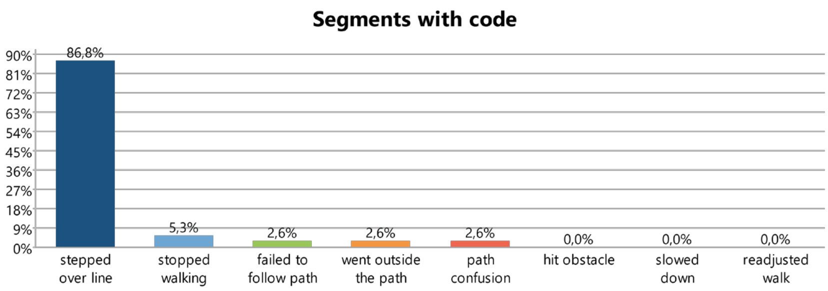
((a)) Texting and Touch



((b)) Texting and Voice



((c)) Selection and Touch



((d)) Selection and Voice

Figure 4.18: Percentage representation of path errors from both tasks and modalities when walking

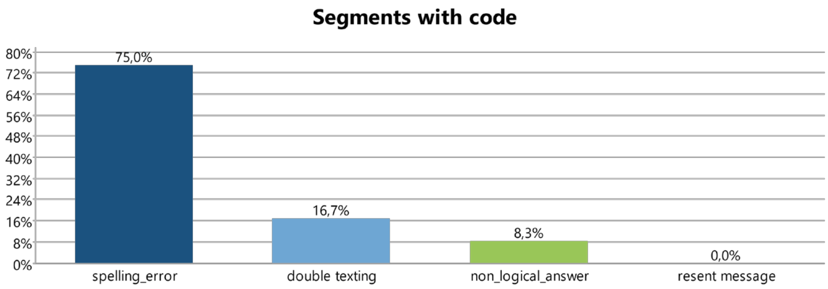
Of all the various errors performed during the walking texting trials with touch interaction, stepping on or over the line of the outlined path was the most commonly performed path error. Slowing down the walking speed was also one of the more prevalent path errors performed by the participants, however the amount of path errors that involved failing to follow the path, walking outside the outlined path, readjusting the walking path, stopping to walk and path confusion remained relatively uncommon.

Similarly to the texting trials with touch interaction, the path errors that were performed during the voice interaction and texting trials consisted mostly of stepping on or over the line. Other errors that were performed during the texting trials with voice interaction were stopping their walk, walking outside of the outlined path, and participants readjusting their walking path during the trials.

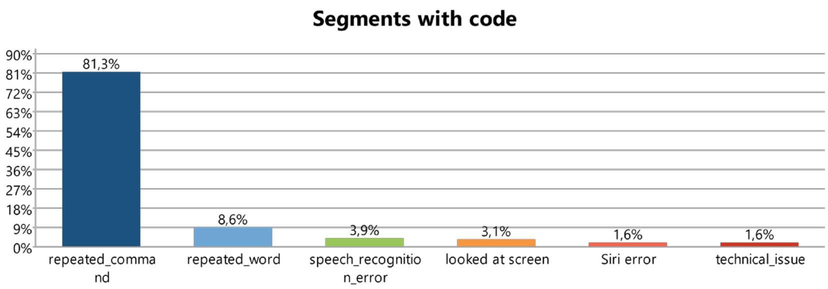
For the selection trials with touch interaction, the most common path error performed by the participants was stepping on or over the line. Other less common path errors that were performed included slowing down the walk, failing to follow the path, walking outside of the path, showing path confusion, stopping to walk and hitting an obstacle. With voice interaction, the most common path error was also stepping on or over the lines outlining the walking path. Other path errors that were considerably less common for these trial conditions included stopping to walk, failing to follow the path, walking outside of the path and showing path confusion.

4.6.2 Task Errors

When observing the type of task errors performed during both the walking trials and the standing trials, the following texting errors were observed: spelling errors on text replies, not addressing the question they were asked (i.e. giving an illogical answer), double texting/replying to the same message multiple times, and sending the same message multiple times. For the texting trials with voice interaction, a different set of errors were observed: Repeated commands, repeating words or task for Siri, speech recognition errors, participants needing to look at the screen, Siri not responding or stopping mid-process, and Siri time-outs. Figure 4.19 showcases the various error types and their coverage for the performed errors in both touch interaction trials and voice interaction trials for the texting task.



((a)) Texting-Touch

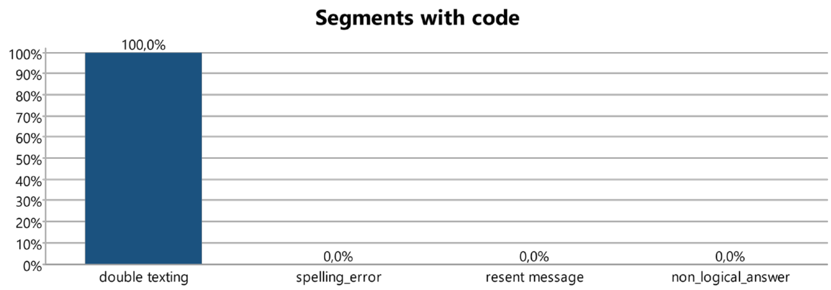


((b)) Texting-Voice

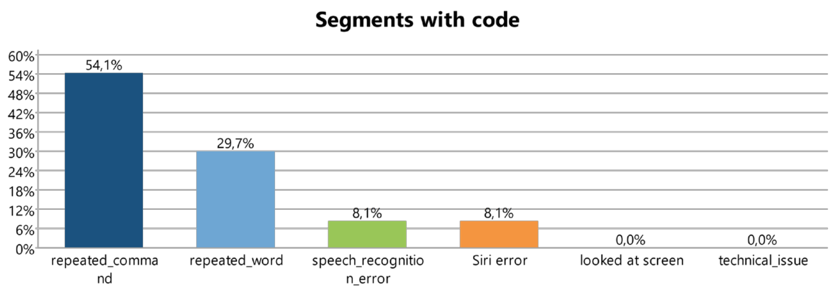
Figure 4.19: Percentage of task errors for texting tasks with walking conditions

For the texting trials with touch interaction, spelling mistakes were the most common errors performed by the participants. Other errors that were also performed were double texting (where one text message was addressed with more than one reply) and not addressing the question when replying to the received text message, however these errors were not as common. For the texting trials using voice interaction, the most common error was repeating the same command multiple times before it was performed by Siri, as this indicates flaws with software which should be adhered to. Other errors that were not as common was repeating a specific word to Siri, speech recognition issues, Siri experiencing time-outs or not responding, and the participants looking at the iPhone screen during the trials.

When observing the errors performed by the participants during the standing trials of the texting task using touch interaction, only double texting was observed as performed errors. In this case, there were only two instances where double texting occurred. The following errors were observed when participants performed texting tasks with voice interaction while walking: repeated commands, repeated words, speech recognition errors and Siri time-outs. Figure 4.20 showcases the percentage representation of the various task errors from the texting trials involving standing and using touch- and voice interaction.



((a)) Touch-Standing



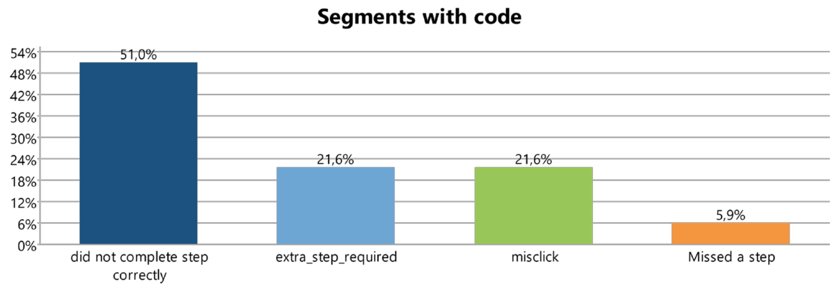
((b)) Voice-Standing

Figure 4.20: Percentage of errors on texting tasks with standing conditions

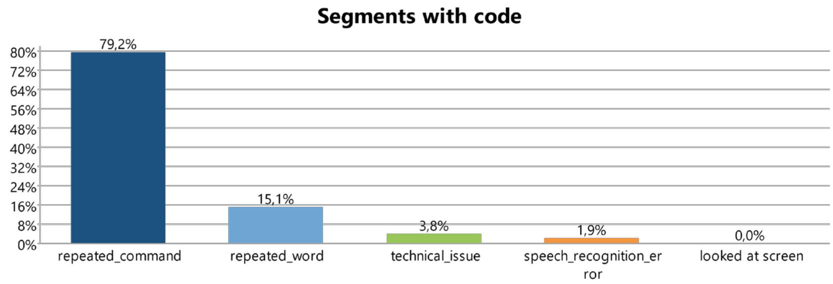
Similarly to the walking trials, repeating commands and repeating words multiple times to Siri were the most common errors that occurred during the standing and texting trials. The errors also consisted of speech recognition errors and Siri experiencing time-outs and processing issues, but these errors were considerably less common.

When observing the task performance for the selection trials, the following task errors were observed: not completing the steps of the scenario as instructed, performing extra steps to complete the scenario task, mis-clicking or opening the wrong application, skipping a step to complete the scenario. Out of the errors performed during the selection tasks with walking and touch interaction conditions, not completing the task scenario as instructed was the most common task among the participants. Other errors that were moderately common in these conditions were the participants having to perform extra steps to complete a task scenario or mis-clicking on targets. Missing certain steps of the task scenarios were also performed, but this error was considerably less common than the others.

For the voice interaction and walking condition, the most commonly performed errors were repeating commands to Siri. Other less commonly performed errors were repeating words to Siri, Siri experiencing time-outs and processing issues, and speech recognition issues. Figure 4.21 showcases the percentage representation of the various task errors performed during the selection tasks with walking and voice interaction conditions.



((a)) Touch-Walking

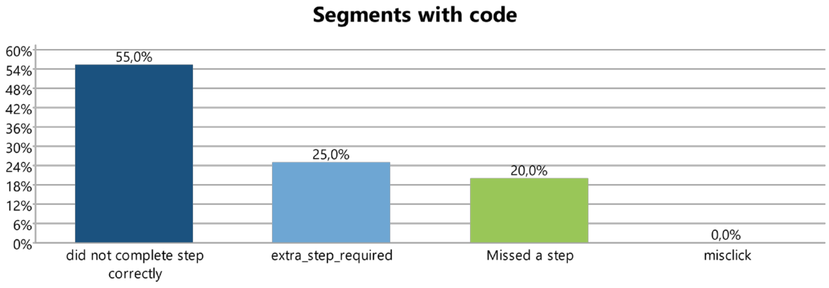


((b)) Voice-Walking

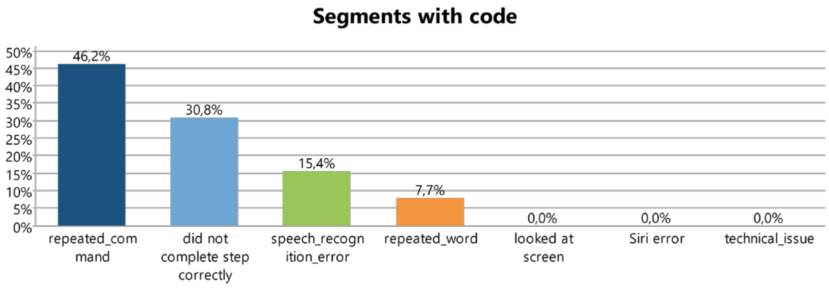
Figure 4.21: Percentage of errors for selection tasks with walking conditions

For the standing trials of the selection tasks with touch interaction, the most common error was participants not completing certain steps of a selection scenario correctly. Other errors that were moderately common as well were participants requiring extra steps before completing a task scenario, or skipping a step before completing a task scenario.

For standing selection task trials with voice interaction, the most common errors were participants required to issue repeated commands to Siri and also not completing steps correctly in task scenarios. Other less common errors include speech recognition errors and participants having to repeat words to Siri when issuing commands. Figure 4.22 showcases the percentage of various errors performed on selection tasks with standing and voice interaction conditions.



((a)) Touch-Standing



((b)) Voice-Standing

Figure 4.22: Percentage of errors with selection tasks with standing conditions

Chapter 5

Discussion

In this chapter, I discuss the research questions I presented in Chapter 1 in light of the analyzed findings from the results of the experimental evaluation in Chapter 4. I also use the findings of the literature review conducted in Chapter 2 to build on the discussion on the results from the experiment. To reiterate, the research questions for this thesis are as follows:

RQ 1: How can conversational user interfaces reduce the effects of situationally-induced impairments and disabilities when walking?

This research question is further divided into three sub-questions:

RQ 1.1: *Will a conversational user interaface improve walking performance during mobile smartphone interaction* I answer this subresearch-question in Section 5.1 using the results from the experiment and the analysis of the walking performance results in terms of performed path errors and walking speed.

RQ 1.2: *Will a conversational user interface improve task performance during mobile smartphone interaction?* I answer this sub-research question in Section 5.2 by presenting and discussing results of the statistical analyses performed on the scores from the experimental evaluation in terms of task errors and task completion times.

RQ 1.3: *Will a conversational user interfaces reduce perceived workload during mobile smartphone interaction?* I answer this sub-research question in Section 5.3 by presenting and discussing results of the statistical analyses performed on the scores from the NASA-TLX in combination with the task- and walking performance scores to gain additional context to NASA-TLX scores given by the participants.

In the following sections, I summarize the findings of the analyses performed on the results from the experimental evaluation and discuss the implications of these findings on the research questions raised by this study. For each of the following sections, this begins by briefly describing the variables that were measured and what effects I wanted to analyze, presenting the results from the analyses that are relevant to the research question that is addressed and how the analysis results address the research question. Later, I again use the results from the analyses performed on the results from the experimental evaluation and the theoretical background to address the second sub-research question. Additionally, I also shed light on the methodology used in this study and reflect on the design choices I made throughout this project.

5.1 Voice Interaction and Walking Performance

This section addresses the first sub-question by reviewing the implications of the results from the experimental evaluation pertaining to walking performance in terms of path accuracy and walking speed when using voice interaction compared to touch interaction during mobile interaction. To summarize, I measured the number of path errors performed by the participants during the walking trials to determine path accuracy during mobile interaction, while walking speed was measured by counting the number of laps (full and partial) and divided that number by the task completion time of the participants.

When looking at the results from the experiment of this study, specifically in terms of walking performance, the participants moved through the outlined path more accurately when interacting with a CUI compared to traditional touch-screen interaction. This is evidenced by the participants performing significantly less walking path errors when using Siri as an interaction modality when completing both texting and selection tasks while walking, which supports my hypothesis *Hx* which states using a voice-based interaction modality will reduce the amount of performed path errors compared touch interaction during mobile interaction. This could be due to reducing the effects of multi-tasking and divided attention by having the eyes and attention solely focused on navigating the outlined path during the experiment, instead of constantly switching the visual focus back and forth between the small-screen of the smartphone and navigation of the outlined path while also avoiding obstacles on the path. This notion is also supported by Oulasvirta and colleagues [18], where the participants kept shifting the attention between the mobile browsing and surrounding environment when moving through urban environments. Anecdotally, the participants seemed more comfortable navigating the path when having the eyes-free from having to focus on two things at once, some participants even commenting on "feeling dizzy" when performing smartphone tasks when mobile with touch interaction due to the constant shifting of focus between two tasks. This gives further support to the multiple resource theory which argues that multi-tasking is better achieved by not having two different tasks competing for the same cognitive resources simultaneously and instead have the two tasks recruit separate cognitive resources when performed at the same time [2, 22].

In addition, the results from investigating the path errors between natural walking, touch interaction, and voice interaction suggests that effects from walking SIIDs are still present with mobile interaction with a CUI. This is evidenced by the significantly higher number of performed path errors with voice interaction trials and the lower number of performed path errors with natural walking where the participants do not complete smartphone tasks as shown in Section 4.3. Reasons for this may include the large experience gap between traditional touch-screen interaction and voice-based interaction via the Siri software, as the pre-experimental questionnaire from this study revealed that the participants had very limited experience with voice-based interfaces such as Siri, Google Assistant and Alexa. Other reasons might also be the usability issues present with Siri which adds to the frustration levels of the participants when performing smartphone tasks when walking. This is also shown in the work by Larsen and colleagues [89], where participants that experienced frequent Siri time-outs were less accurate in both task performance and driving performance. The authors even argued that the added frustration and effort that occur when Siri experiences frequent time-outs adds to the SIID effects of mobile interaction [89].

When looking at the types of path errors performed by the participants, they showcased more path confusion when using touch interaction which includes failing to follow the path, walking outside of the path and hitting obstacles on the path. There were less errors pertaining to failure to follow the path when the participants were using voice interaction, hitting obstacles along the path or re-orientating themselves when walking. These qualitative results gives support to the concept of ”Inattentional Blindness” presented by Mack [87], where an individual under dual-task situations is blind to the environment and context of the task this is not in focus when performing two tasks simultaneously. Based on the concept presented by Mack, I argue that due to the eyes having to wander between the small-screen and the outlined path during the walking trials with touch interaction, this could be the cause for hitting obstacles and troubles following the path as the eyes were focusing on the contents of the small-screen, the participants became blind to the environment around them such as chairs and the boundaries of the path.

In terms of walking speed between natural walking, mobile interaction with touch interaction, and mobile interaction with voice interaction, the results show that participants walked faster when walking naturally as opposed to engaging in either touch- or voice interaction. Furthermore, the results also showed that the participants walked slightly faster with voice interaction compared to touch interaction, but the difference was not significant. This does not support my hypothesis *Hx* which states that participants would walk faster with voice interaction compared to touch interaction on both tasks. Based on these results, I had to accept the null hypothesis due to there not being a significant difference in walking speed between touch- and voice interaction.

The slight increase in walking speed with voice interaction is probably due to the reduced fragmentation of attention when engaging in mobile interaction with a CUI due to removing the attentional need on the small-screen which touch interaction requires [2, 22]. However, the insignificant difference in walking speed between touch interaction and voice interaction might be due to novelty of using a CUI to perform smartphone tasks which possibly increases the mental demand of the participants due to unfamiliarity with the Siri software. Other factors may include the added frustrations that come with the time-outs and unresponsiveness of the Siri software.

5.2 Voice Interaction and Task Performance

This section addresses the second subresearch question with a statistical analysis of the task performance results from the participants who underwent the experimental evaluation. To summarize, I measured the task errors by counting the amount of mistakes, categorized and described in Section 3.1.2 and Section 4.6, performed by the participants by reviewing and analyzing the video recordings from the GoPro camera and smartphone screen recordings. Task completion time was timed from when the participant starts the task to when the task is completed.

The results from the experiment of this study suggest that CUIs might not be the best option for reducing SIID effects in terms of task performance on mobile interaction. As indicated by the statistical analysis, how effective participants were in performing smartphone tasks when mobile heavily depends on the type of task that is performed and whether they are using touch interaction or Siri interaction. This is evidenced by the significantly higher number of task errors with voice interaction compared to touch interaction on texting tasks, and the insignificant difference in task errors between touch- and voice interaction on selection tasks. This goes against my hypotheses on task errors on texting tasks (Hs and Hw) that say there would be fewer task errors with voice interaction on texting tasks when both standing and walking. The results also oppose my hyptheses on task errors on selection tasks (Hs and Hw) that state there would be fewer task errors with voice interaction on selection tasks when walking but fewer with touch interaction when standing. For the selection tasks, the null hypotheses for Hs and Hw could not be rejected.

The higher number of errors with voice interaction in terms of texting tasks are mostly due to the time-outs and unresponsiveness of Siri when conducting texting tasks, which implies that the task performance is more due to the unreliability of the Siri software and not the ability of the user. This is further evidenced by the error descriptions in Section 4.6 where the errors performed during touch interaction were due to individual mistakes (sending more than one message to address the received message, spelling mistakes etc.), while the errors performed during voice interaction were due to Siri errors (No response, time-outs, speech recognition errors, etc.). With the voice interaction errors being performed by the software itself, it is likely this increases the frustration and effort of the participants to complete tasks using CUIs. This notion is also supported in the findings by Larsen and colleagues [89].

In terms of task completion times, the results once again show that the best option for modality during mobile interaction is determined by what task the user engages in. This evidenced by the significantly increased task completion time with voice interaction compared to touch interaction on texting tasks, and the insignificant difference in task completion time between touch- and voice interaction on selection tasks. The increased completion time with voice interaction on texting tasks is in large part due to the additional steps required by the Siri software to reply to one incoming text message, as well as Siri forcing the user to follow the software's pace in text comprehension (i.e. Siri reads out the message in it's own pace, whereas user's can read the message in their preferred pace with touch interaction). Other factors that also add to the task completion time are the frequent time-outs and the at times non-responsiveness of the Siri software when the participants were issuing commands to the software. This resulted in participants spending more time replying to one text message than the participants would like to, leading to added frustrations and effort from the participants. As mentioned in Section 4.6, some participants had to resort to looking at the small-screen mid trial to open and lock the screen of the smartphone to restart the Siri software, defeating the purpose of eyes-free interaction.

During the selection tasks with voice interaction, Siri did not display as many time-outs or unresponsiveness as in the texting trials, which is probably due to performing such tasks with Siri has been possible since it's introduction while complete hands-free and eyes-free text messaging is a relatively new implementation. However, the majority of the errors during the Siri interaction trials still consisted of time-outs and unresponsiveness of Siri which led to delays in completing selection tasks. This led to frustration in some participants, but the frequency of these errors were very minimal and did not significantly impact the satisfaction of the participants. Surprisingly, some participants commented that they considered adopting Siri as a primary mode of interaction for completing selection tasks after completing the experiment. They reported it was easier to say one short sentence to complete such a task rather than having to point and click, and sometimes input text, numerous times.

One factor that also needs to be considered is the experience gap between touch-screen interaction and Siri interaction. As most smartphone users have had many years of experience in performing text entry tasks and target selection tasks in mobile situations, it makes sense that there would be less errors and frustrations with performing the tasks in a familiar manner. From the results of the pre-experimental questionnaires that were issued to all participants, all of the participants reported to having minimal to no experience in using voice assistants such as Siri, Google Assistant or Alexa.

Based on the results from the experiment in terms of task completion time and task errors, as well as the comments from the participants, Siri remains too unreliable in terms of fidelity to be considered a great option for improving task effectiveness and efficiency during mobile interaction. While splitting the cognitive resources between visual channels and auditory channels might improve certain aspects of the dual-task interference between walking and smartphone interaction, the low fidelity of Siri keeps it from being an optimal solution for reducing effects from walking SIIDs. Further improvement and optimization needs to be conducted on Siri's responsiveness, as well as increase its time-out window before it can be considered a alternative to traditional touch interaction in mobile situations.

5.3 Voice Interaction and Mental Workload

This section addresses the answers from the participants on the NASA-TLX questionnaires, where each participant filled out a NASA-TLX questionnaire consisting of seven scales ranging from 0-20 after each experimental condition on both texting task trials and selecting task trials. These questionnaires were given after every condition as the participants would still have fresh memory from their feelings of mental workload after having completed a condition. The scores given by the participants on the NASA-TLX scales were used to perform statistical analysis to compare the effects mobility type, task type and interaction mode have on mental workload and level of distraction. The combination of task and walking performance metrics will be used to give further context to the perceived workload reported by the participants. For the NASA-TLX scores, the six scales were analyzed individually to gain more insight into how the participants felt when completing the experiment.

5.3.1 Mental Demand

In terms of mental demand scores, the results indicate that the mental demand is largely dependent on what task a user is performing, what modality the user is using, and whether the user is standing or walking. This is evidenced by the varied results of the perceived mental demand given by the participants, where mental demand was generally higher with voice interaction on texting tasks, and also higher with touch interaction on selection tasks. These results give support to my hypothesis *Hs* pertaining to texting tasks, and to my hypothesis *Hw* pertaining to selection tasks.

5.3.2 Physical Demand

When looking at the physical demand scores, the results indicate that participants, rather unsurprisingly, experienced greater physical demand from walking trials than standing trials for both modalities on texting trials, and for touch interaction on selection trials. However, the results also indicate that participants feel similar levels of physical demand between standing and walking trials on selection tasks when using voice interaction. This is evidenced by the greater physical demand scores on walking trials compared to standing trials, with the exception on selection tasks with voice interaction where standing and walking had similar scores. The *Hs* hypothesis is not supported with these results due to the voice interaction trials having lower physical scores than touch interaction. But it does support the *Hw* hypothesis due to the higher physical demand scores with touch interaction.

The results of physical demand between standing and walking with voice interaction on selection tasks were rather surprising, and is most probably due to the simplicity in performing selection tasks with Siri compared to performing them with touch-screen interaction. Moreover, the simplicity of the task combined with the small-screen not demanding any visual attention from the participant during the Siri and walking trials of the selection tasks could be a factor in why the difference in physical demand between standing and walking is not significant in terms of selection tasks with Siri interaction.

5.3.3 Perceived Frustration

When looking at the perceived frustration levels of the participants, the results indicate that the participants generally felt a higher level of perceived frustration with voice interaction compared to touch interaction. This is evidenced by the higher levels of frustration with voice interaction compared to touch interaction in both mobility conditions for the texting tasks, and for the standing conditions on the selection tasks. However, the walking conditions on selection tasks had similar frustration levels between interaction modalities. These results oppose my hypothesis (Hw) which states that perceived frustration would be lower with voice interaction on both tasks, however the results also support my hypothesis (Hs) that perceived frustration would be lower with touch interaction on both tasks when standing. One of the main causes of this is most probably due to the technical errors experienced by participants when using Siri as opposed to human errors they performed when using the software. With the time-outs from Siri, and otherwise Siri's failure to respond when called upon by the participant, when the participants wanted to respond only increased frustrations with the participants as they felt their ability to perform the task was determined more by the software and less their own perceived ability. Moreover, the longer time for audio is not a surprising finding due to the participants needing to listen to the incoming messages at the pace of software while the speech of the software being more linear than natural human speech.

This is further supported in the findings by Vadas and colleagues where the audio tasks also required more time to complete tasks than visual tasks [13]. While the experiment has shown that voice assistant have the ability to reduce dual-task interference by improving walking performance, its user experience involving increase in frustration and effort with users when replying to text messages makes voice assistant not the optimal choice for accommodating SIIDs during mobile interaction.

5.3.4 Perceived Performance

In terms of perceived performance, the participants felt they performed the texting tasks better with touch interaction compared to voice interaction. This is evidenced by the higher performance scores of voice interaction on texting tasks, which indicates that participants perceived to have performed worse with voice interaction in both mobility conditions when texting. In terms of selection tasks, the participants felt they performed the task better when standing as opposed to walking with both interaction modalities with little difference in performance scores between the interaction modalities. These results support the standing *Hs* hypothesis in terms of texting tasks, but not in terms of selection tasks. Additionally, the results do not support the walking *Hw* hypothesis on either task.

The main factor to consider for this result is the different categories of task error types that were present between the two interaction modalities. Due to the frequent time-outs by Siri and the software at times being unresponsive during the texting tasks, it is unsurprising the participants felt their performance on these tasks were worse than when performing them in a manner that they have more experience in. Furthermore, with the large decrease in Siri time-outs and unresponsiveness in the selection tasks, as well as the more simple nature of performing such tasks on Siri as opposed to touch-screen interaction, it is unsurprising that the participants found their performance to improve with selection tasks via Siri interaction.

5.3.5 Perceived Effort

When looking at the scores for perceived effort, the participants felt they had to put more effort into the walking trials compared to the standing trials. This is evidenced by the higher effort scores on the walking conditions given by the participants. Additionally, the participants also felt they had to put more effort into the voice interaction compared to touch interaction as evidenced by the higher effort scores on voice interaction on both tasks and mobility conditions with the exception on standing-selection condition.

The added effort in using Siri for texting tasks is also largely due to Siri requiring extra steps for responding to one single incoming message, and Siri dictating the pace of the task flow for such tasks. The results also suggested that the participants experienced an increased level of frustration and needed effort with walking trials as opposed to the standing trials, giving further support to the dual-task interference that occurs during smartphone interaction and walking in terms of mental workload [22].

5.3.6 Temporal Demand

In terms of temporal demand, the participants felt that the temporal demand was similar between all mobility, modality and task conditions. This is evidenced by the very similar scores in temporal demand given by the participants for all conditions. In this case, the null hypotheses for both the standing Hs and the walking Hw conditions cannot be rejected.

In terms of temporal demand which assesses the feeling of pressure during the experiment, the differences between interaction modality, task types and mobility types were insignificant probably due to the participants not feeling much pressure when completing the tasks during the experiment. This is in large part due to the participants not being incentivised or pressured to perform the tasks to the best of their abilities (e.g.

prizes for best performance), or being rushed by being induced to stressful conditions (e.g. time-limits).

The results from the TLX scores given by the participants suggest that participants felt a higher level of perceived frustration with using a voice assistant, specifically for the texting task trials. This is evidenced by the higher mental workload score The added effort in using Siri for texting tasks is also largely due to Siri requiring extra steps for responding to one single incoming message, and Siri dictating the pace of the task flow for such tasks. The results also suggested that the participants experienced an increased level of frustration and needed effort with walking trials as opposed to the standing trials, giving further support to the dual-task interference that occurs during smartphone interaction and walking in terms of mental workload [22].

When looking at the selection tasks however, the participants experienced less frustration with using Siri to complete selection tasks as opposed to using touch-screen interaction. This is in large part due to Siri requiring nothing more than a single sentence by the participants to perform a selection task, which otherwise would require multiple pointing and clicking actions from the participant to complete the same task. For example, setting a reminder on Thursday at 12:00 a.m. would only require the participant to say "Hey Siri, can you remind on thursday at 12 am?", while with touch-screen interaction the participants would have to locate the correct app, click on it, click on the correct day and time on the calendar, set the time, and confirm. Due to this difference in simplicity and steps required, it is not surprising that the participants found the selection tasks less frustrating and requiring less effort than the texting trials when performing them with Siri. Surprisingly, the participants felt they had to put more effort into completing selection tasks with touch interaction as opposed to texting tasks with touch interaction. This is probably a result of not being too familiar with some of the applications used in the selection tasks (e.g. Reminders and Maps) as most participants spent some time looking for the correct apps on the smartphone when performing selection tasks with touch interaction.

Chapter 6

Conclusion

In this thesis, I have explored the concept of situationally-induced impairments and disabilities and its relationship with smartphone interaction and mobile computing with a theoretical background. Furthermore, I have examined and observed how SIID's affect users of smartphones while walking as well as how to reduce these effects by adapting a different mode of interaction. For this project, I chose voice interaction as a alternative method for interaction and this decision was based on the theoretical background presented in Chapter 2. Throughout this thesis, I applied the methodology of experimental research as the core methodology for answering my research questions, and conducted a true experiment to collect data. The activities that followed included creating a true experiment where I had the participants walk a fixed path with obstacles in a lab environment while also completing general smartphone tasks. The experiment consisted of 16 trial runs where 8 trials were dedicated to touch interaction and 8 trials were dedicated to voice interaction. The tasks that were performed included; (1) responding to four text messages individually to simulate a text conversation; and (2) completing use-scenarios that involved selection tasks such as checking the weather, setting reminders etc. After the experiments were concluded, the data was then used to analyze the participants performance and mental workload.

I hope to contribute by both presenting my results of the experiment, as well as by providing the documentation of my design process in regards of using an experimental approach as a methodology for evaluating interaction methods. Furthermore, the contributions from this thesis will also include findings from the post - and pre-experimental questionnaires and an extensive literature review on SIIDs in mobile computing the theories behind SIIDs. Finally, the following sections in this chapter will provide suggestions for future work, as well as discuss the limitations of the study.

6.1 Limitations

This section discusses the limitations of the conducted study. These limitations may have impacted the design of the experiment and the results to some degree.

6.1.1 Participants

In the experiment designed for this project, there were 16 participants in total. All 16 participants took part in all of the parts of the conducted experiment, which includes the pre-experimental questionnaire, the 16 usability testing trial runs, and the post-experimental questionnaire and brief interview. This project would benefit from acquiring more participants for conducting statistical analyses of the collected data, as well as look at a variation of perspectives from users by looking at different age groups. Furthermore, running the experiment with different age groups could also broaden the understanding of how the various age groups interact with the smartphone. The age groups varied between the 16 participants in this experiment, which may have impacted the results and should be regarded as a possible limitation.

6.1.2 Data Collection

This subsection expands further on the subsection of participants. Along with recruiting more participants for a more accurate statistical analysis, the project would have benefited from using other forms of data collection. Using an accelerometer during the walking-conditions of the experiment, would give further insight into the participants' performance during the trial runs of the experiment.

6.2 Future Work

This section outlines suggestions for future work for research in a similar scope within mobile computing. The suggestions presented in this section builds upon the contributions provided by this research. The suggestions are rooted in existing literature and the available data that are available from the conducted experiments in this project work.

6.2.1 Expanding The Scope Of Research

Results from the experiment evaluation suggest that the research scope should be expanded to various demographics as the different age demographics reported different feelings towards the adaption of CUIs for more accessible interaction. It would be interesting to explore how younger users's experience vary from the experience of older users when using a voice assistant when engaging in mobile interaction.

Exploring other environments of mobile interaction

In this study, I limited the experiment to a laboratory setting where I explored the effects of walking on mobile interaction only. This study suggests future work to explore mobile interaction in more realistic environments where mobility is taking place. For instance, conducting experiments where the participants are instructed to walk in an outside environment where ambient noise, ambient lighting and low temperatures are present factors in the experiment. This will give a more realistic assessment of performance metrics due to most of mobile smartphone interaction takes place outside of the home where a user might be exposed to different contextual factors during the walk from location to location.

Exploring other modes of interaction

In this study, I wanted to focus on voice interaction through conversational user interfaces as a adaptive method for overcoming impairments from contextual and environmental factors, in this case walking. The results from this study suggests that other adaptive methods for interaction should be explored and compared with the results of this study for future work. From previous literature, it would be interesting to explore how interaction with gestures can be used to accommodate for the effects of situational impairments when engaging in mobile smartphone interaction.

6.2.2 Collecting More Data

With any conducted research, it will always be beneficial to collect more data, however, in this context it would be appropriate to gather more accurate data from a larger sample of participants when assessing performance. Additional tools for measuring walking speed and divided attention could have been used for a more accurate assessment of these metrics. Tools such as an accelerometer for walking speed measurements and an eye-tracking software to measure and observe how often the participants switch their focus points during interaction, where their attention shifts to and the duration between focus points.

It would have also have been beneficial for this study to adapt a more mixed methods approach to give further context to the performance data from the participants. This could be achieved by including an extensive post-experimental interview where the participants give a more detailed response on how they viewed their experience with using a voice assistant to perform tasks they otherwise do on a daily basis.

Bibliography

[1] Z. Sarsenbayeva, N. van Berkel, C. Luo, V. Kostakos, and J. Goncalves, "Challenges of situational impairments during interaction with mobile devices," in *Proceedings of the 29th Australian Conference on Computer-Human Interaction*. Brisbane Queensland Australia: ACM, Nov. 2017, pp. 477–481. [Online]. Available: https://dl.acm.org/doi/10.1145/3152771.3156161

[2] C. D. Wickens, "Multiple resources and performance prediction," *Theoretical issues in ergonomics science*, vol. 3, no. 2, pp. 159–177, 2002.

[3] J. O. Wobbrock, "Situationally-Induced Impairments and Disabilities," in *Web Accessibility: A Foundation for Research*, Y. Yesilada and S. Harper, Eds. London: Springer London, 2019, pp. 59–92. [Online]. Available: https://doi.org/10.1007/978-1-4471-7440-0

[4] A. Sears and M. Young, *Physical Disabilities and Computing Technologies: An Analysis of Impairments.* USA: L. Erlbaum Associates Inc., 2002, p. 482–503.

[5] J. O. Wobbrock, "The future of mobile device research in hci," 2006.

[6] Z. Sarsenbayeva, V. Kostakos, and J. Goncalves, "Situationally-Induced Impairments and Disabilities Research," p. 5, 2019.

[7] G. W. Tigwell, Z. Sarsenbayeva, B. M. Gorman, D. R. Flatla, J. Goncalves, Y. Yesilada, and J. O. Wobbrock, "Addressing the Challenges of Situationally-Induced Impairments and Disabilities in Mobile Interaction," in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. Glasgow Scotland Uk: ACM, May 2019, pp. 1–8. [Online]. Available: https://dl.acm.org/doi/10.1145/3290607.3299029

[8] L. Barnard, J. S. Yi, J. A. Jacko, and A. Sears, "Capturing the effects of context on human performance in mobile computing systems," *Personal and Ubiquitous Computing*, vol. 11, no. 2, pp. 81–96, Jan. 2007. [Online]. Available: http://link.springer.com/10.1007/s00779-006-0063-x

[9] M. Goel, L. Findlater, and J. Wobbrock, “WalkType: using accelerometer data to accomodate situational impairments in mobile touch screen text entry,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Austin Texas USA: ACM, May 2012, pp. 2687–2696. [Online]. Available: https://dl.acm.org/doi/10.1145/2207676.2208662

[10] M. Goel, J. Wobbrock, and S. Patel, “GripSense: using built-in sensors to detect hand posture and pressure on commodity mobile phones,” in *Proceedings of the 25th annual ACM symposium on User interface software and technology - UIST ’12*. Cambridge, Massachusetts, USA: ACM Press, 2012, p. 545. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2380116.2380184

[11] M. E. Mott and J. O. Wobbrock, “Cluster Touch: Improving Touch Accuracy on Smartphones for People with Motor and Situational Impairments,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Glasgow Scotland Uk: ACM, May 2019, pp. 1–14. [Online]. Available: https://dl.acm.org/doi/10.1145/3290605.3300257

[12] S. K. Kane, J. O. Wobbrock, and I. E. Smith, “Getting off the treadmill: evaluating walking user interfaces for mobile devices in public spaces,” in *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services - MobileHCI ’08*. Amsterdam, The Netherlands: ACM Press, 2008, p. 109. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1409240.1409253

[13] K. Vadas, N. Patel, K. Lyons, T. Starner, and J. Jacko, “Reading on-the-go: A comparison of audio and hand-held displays,” in *Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services*, ser. MobileHCI ’06. New York, NY, USA: Association for Computing Machinery, 2006, p. 219–226. [Online]. Available: https://doi.org/10.1145/1152215.1152262

[14] K. J. Price, M. Lin, J. Feng, R. Goldman, A. Sears, and J. A. Jacko, “Motion does matter: an examination of speech-based text entry on the move,” *Universal Access in the Information Society*, vol. 4, no. 3, pp. 246–257, Mar. 2006. [Online]. Available: https://doi.org/10.1007/s10209-005-0006-8

[15] M. Lin, R. Goldman, K. J. Price, A. Sears, and J. Jacko, “How do people tap when walking? An empirical investigation of nomadic data entry,” *International Journal of Human-Computer Studies*, vol. 65, no. 9, pp. 759–769, Sep. 2007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1071581907000572

[16] A. Schmidt, K. A. Aidoo, A. Takaluoma, U. Tuomela, K. Van Laerhoven, and W. Van de Velde, “Advanced interaction in context,” in *Handheld and Ubiquitous Computing*, H.-W. Gellersen, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 89–101.

[17] S. Jumisko-Pyykkö and T. Vainio, "Framing the context of use for mobile hci," *International journal of mobile human computer interaction (IJMHCI)*, vol. 2, no. 4, pp. 1–28, 2010.

[18] A. Oulasvirta, "The fragmentation of attention in mobile interaction, and what to do with it," *interactions*, vol. 12, no. 6, pp. 16–18, 2005.

[19] T. Sohn, K. A. Li, W. G. Griswold, and J. D. Hollan, "A diary study of mobile information needs," in *Proceedings of the sigchi conference on human factors in computing systems*, 2008, pp. 433–442.

[20] J. Jacko and C. Stephanidis, *Human-Computer Interaction: Theory and Practice*. CRC Press, Sep. 2003, google-Books-ID: XP8Y0S0q0ccC.

[21] E. Akpinar, Y. Yeşilada, and S. Temizer, "The Effect of Context on Small Screen and Wearable Device Users' Performance - A Systematic Review," *ACM Computing Surveys*, vol. 53, no. 3, pp. 1–44, Jul. 2020. [Online]. Available: https://dl.acm.org/doi/10.1145/3386370

[22] C. D. Wickens, "Multiple resources and mental workload," *Human factors*, vol. 50, no. 3, pp. 449–455, 2008.

[23] C. D. Wickens, D. L. Sandry, and M. Vidulich, "Compatibility and resource competition between modalities of input, central processing, and output," *Human factors*, vol. 25, no. 2, pp. 227–248, 1983.

[24] A. Parkes, N. Coleman *et al.*, "Route guidance systems: A comparison of methods of presenting directional information to the driver," 1990.

[25] F. J. Seagull, C. D. Wickens, and R. G. Loeb, "When is less more? attention and workload in auditory, visual, and redundant patient-monitoring conditions," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 45, no. 18. SAGE Publications Sage CA: Los Angeles, CA, 2001, pp. 1395–1399.

[26] J. O. Wobbrock, S. K. Kane, K. Z. Gajos, S. Harada, and J. Froehlich, "Ability-based design: Concept, principles and examples," *ACM Transactions on Accessible Computing (TACCESS)*, vol. 3, no. 3, pp. 1–27, 2011.

[27] S. Kristoffersen and F. Ljungberg, "Mobile use of it."

[28] B. Schildbach and E. Rukzio, "Investigating selection and reading performance on a mobile phone while walking," in *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services*, ser. MobileHCI '10. New York, NY, USA: Association for Computing Machinery, 2010, p. 93–102. [Online]. Available: https://doi.org/10.1145/1851600.1851619

[29] T. Mustonen, M. Olkkonen, and J. Hakkinen, "Examining mobile phone text legibility while walking," in *CHI'04 extended abstracts on Human factors in computing systems*, 2004, pp. 1243–1246.

[30] J. Bergstrom-Lehtovirta, A. Oulasvirta, and S. Brewster, “The effects of walking speed on target acquisition on a touchscreen interface,” in *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, 2011, pp. 143–146.

[31] J. Mwakalonge, S. Siuhi, and J. White, “Distracted walking: Examining the extent to pedestrian safety problems,” *Journal of traffic and transportation engineering (English edition)*, vol. 2, no. 5, pp. 327–337, 2015.

[32] M. Harvey and M. Pointon, “Perceptions of the effect of fragmented attention on mobile web search tasks,” in *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, 2017, pp. 293–296.

[33] J. O. Wobbrock, K. Z. Gajos, S. K. Kane, and G. C. Vanderheiden, “Ability-based design,” *Communications of the ACM*, vol. 61, no. 6, pp. 62–71, 2018.

[34] 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*, 2005, pp. 919–928.

[35] A. Ng, S. A. Brewster, and J. Williamson, “The impact of encumbrance on mobile interactions,” in *IFIP Conference on Human-Computer Interaction*. Springer, 2013, pp. 92–109.

[36] A. Ng, S. A. Brewster, and J. H. Williamson, “Investigating the effects of encumbrance on one-and two-handed interactions with mobile devices,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2014, pp. 1981–1990.

[37] A. Ng, J. Williamson, and S. Brewster, “The Effects of Encumbrance and Mobility on Touch-Based Gesture Interactions for Mobile Phones,” in *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, ser. MobileHCI ’15. New York, NY, USA: Association for Computing Machinery, Aug. 2015, pp. 536–546. [Online]. Available: https://doi.org/10.1145/2785830.2785853

[38] D.-S. Lee, K.-K. Shieh, S.-C. Jeng, and I.-H. Shen, “Effect of character size and lighting on legibility of electronic papers,” *Displays*, vol. 29, no. 1, pp. 10–17, 2008.

[39] P. Liu, F. Zafar, and A. Badano, “The effect of ambient illumination on handheld display image quality,” *Journal of digital imaging*, vol. 27, no. 1, pp. 12–18, 2014.

[40] L. Barnard, J. S. Yi, J. A. Jacko, and A. Sears, “An empirical comparison of use-in-motion evaluation scenarios for mobile computing devices,” *International Journal of Human-Computer Studies*, vol. 62, no. 4, pp. 487–520, Apr. 2005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1071581905000145

[41] Z. Sarsenbayeva, N. van Berkel, A. Visuri, S. Rissanen, H. Rintamaki, V. Kostakos, and J. Goncalves, "Sensing Cold-Induced Situational Impairments in Mobile Interaction Using Battery Temperature," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, pp. 98:1–98:9, Sep. 2017. [Online]. Available: https://doi.org/10.1145/3130963

[42] Z. Sarsenbayeva, J. Goncalves, J. García, S. Klakegg, S. Rissanen, H. Rintamäki, J. Hannu, and V. Kostakos, "Situational impairments to mobile interaction in cold environments," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp '16. New York, NY, USA: Association for Computing Machinery, Sep. 2016, pp. 85–96. [Online]. Available: https://doi.org/10.1145/2971648.2971734

[43] J. Goncalves, Z. Sarsenbayeva, N. van Berkel, C. Luo, S. Hosio, S. Risanen, H. Rintamäki, and V. Kostakos, "Tapping task performance on smartphones in cold temperature," *Interacting with Computers*, vol. 29, no. 3, pp. 355–367, 2017.

[44] P. M. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement." *Journal of experimental psychology*, vol. 47, no. 6, p. 381, 1954.

[45] A. K. Dey, "Understanding and using context," *Personal and ubiquitous computing*, vol. 5, no. 1, pp. 4–7, 2001.

[46] G. D. Abowd, A. K. Dey, P. J. Brown, N. Davies, M. Smith, and P. Steggles, "Towards a better understanding of context and context-awareness," in *International symposium on handheld and ubiquitous computing*. Springer, 1999, pp. 304–307.

[47] A. Anjum and M. U. Ilyas, "Activity recognition using smartphone sensors," in *2013 ieee 10th consumer communications and networking conference (ccnc)*. IEEE, 2013, pp. 914–919.

[48] J. Wannenburg and R. Malekian, "Physical activity recognition from smartphone accelerometer data for user context awareness sensing," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 12, pp. 3142–3149, 2016.

[49] X. Su, H. Tong, and P. Ji, "Activity recognition with smartphone sensors," *Tsinghua science and technology*, vol. 19, no. 3, pp. 235–249, 2014.

[50] J.-F. Daneault, B. Carignan, C. É. Codère, A. F. Sadikot, and C. Duval, "Using a smart phone as a standalone platform for detection and monitoring of pathological tremors," *Frontiers in human neuroscience*, vol. 6, p. 357, 2013.

[51] S. Gupta, T. Campbell, J. R. Hightower, and S. N. Patel, "Squeezeblock: using virtual springs in mobile devices for eyes-free interaction," in *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, 2010, pp. 101–104.

[52] P. Quinn, S. C. Lee, M. Barnhart, and S. Zhai, "Active edge: Designing squeeze gestures for the google pixel 2," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–13.

[53] J. S. Yi, Y. S. Choi, J. A. Jacko, and A. Sears, "Context awareness via a single device-attached accelerometer during mobile computing," in *Proceedings of the 7th international conference on Human computer interaction with mobile devices & services*, 2005, pp. 303–306.

[54] C. F. Mass and L. E. Madaus, "Surface pressure observations from smartphones: A potential revolution for high-resolution weather prediction?" *Bulletin of the American Meteorological Society*, vol. 95, no. 9, pp. 1343–1349, 2014.

[55] A. Overeem, J. R. Robinson, H. Leijnse, G.-J. Steeneveld, B. P. Horn, and R. Uijlenhoet, "Crowdsourcing urban air temperatures from smartphone battery temperatures," *Geophysical Research Letters*, vol. 40, no. 15, pp. 4081–4085, 2013.

[56] T. Reis, L. Carriço, and C. Duarte, "Mobile interaction: automatically adapting audio output to users and contexts on communication and media control scenarios," in *International Conference on Universal Access in Human-Computer Interaction*. Springer, 2009, pp. 384–393.

[57] D. R. Flatla and C. Gutwin, "Individual models of color differentiation to improve interpretability of information visualization," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2010, pp. 2563–2572.

[58] D. Buschek and F. Alt, "Probui: Generalising touch target representations to enable declarative gesture definition for probabilistic guis," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, pp. 4640–4653.

[59] D. Buschek, A. De Luca, and F. Alt, "Evaluating the influence of targets and hand postures on touch-based behavioural biometrics," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016, pp. 1349–1361.

[60] D. Rudchenko, T. Paek, and E. Badger, "Text text revolution: A game that improves text entry on mobile touchscreen keyboards," in *International Conference on Pervasive Computing*. Springer, 2011, pp. 206–213.

[61] A. Rahmati, C. Shepard, and L. Zhong, "Noshake: Content stabilization for shaking screens of mobile devices," in *2009 IEEE International Conference on Pervasive Computing and Communications*. IEEE, 2009, pp. 1–6.

[62] T. Yamabe and K. Takahashi, "Experiments in mobile user interface adaptation for walking users," in *The 2007 International Conference on Intelligent Pervasive Computing (IPC 2007)*. IEEE, 2007, pp. 280–284.

[63] S. Brewster, "Overcoming the lack of screen space on mobile computers," *Personal and Ubiquitous computing*, vol. 6, no. 3, pp. 188–205, 2002.

[64] J. Himberg, J. Häkkilä, P. Kangas, and J. Mäntyjärvi, "On-line personalization of a touch screen based keyboard," in *Proceedings of the 8th international conference on Intelligent user interfaces*, 2003, pp. 77–84.

[65] K. Go and Y. Endo, "Catkey: customizable and adaptable touchscreen keyboard with bubble cursor-like visual feedback," in *IFIP Conference on Human-Computer Interaction*. Springer, 2007, pp. 493–496.

[66] D. Buschek, M. Hackenschmied, and F. Alt, "Dynamic ui adaptations for one-handed use of large mobile touchscreen devices," in *IFIP Conference on Human-Computer Interaction*. Springer, 2017, pp. 184–201.

[67] A. K. Karlson, B. B. Bederson, and J. SanGiovanni, "Applens and launchtile: two designs for one-handed thumb use on small devices," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2005, pp. 201–210.

[68] A. K. Karlson and B. B. Bederson, "Thumbspace: generalized one-handed input for touchscreen-based mobile devices," in *IFIP Conference on Human-Computer Interaction*. Springer, 2007, pp. 324–338.

[69] K. Lyons, T. Starner, D. Plaisted, J. Fusia, A. Lyons, A. Drew, and E. Looney, "Twiddler typing: One-handed chording text entry for mobile phones," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2004, pp. 671–678.

[70] V. Sazawal, R. Want, and G. Borriello, "The unigesture approach one-handed text entry for small devices," in *International Conference on Mobile Human-Computer Interaction*. Springer, 2002, pp. 256–270.

[71] S. Harper, Y. Yesilada, and T. Chen, "Mobile device impairment … similar problems, similar solutions?" *Behaviour & Information Technology*, vol. 30, no. 5, pp. 673–690, Sep. 2011. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/01449291003801943

[72] Y. Yesilada, S. Harper, T. Chen, and S. Trewin, "Small-device users situationally impaired by input," *Computers in Human Behavior*, vol. 26, no. 3, pp. 427–435, May 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563209001861

[73] K. A. Siek, Y. Rogers, and K. H. Connelly, "Fat finger worries: how older and younger users physically interact with pdas," in *IFIP Conference on Human-Computer Interaction*. Springer, 2005, pp. 267–280.

[74] P. Baudisch and G. Chu, "Back-of-device interaction allows creating very small touch devices," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2009, pp. 1923–1932.

[75] D. Buschek, O. Schoenleben, and A. Oulasvirta, "Improving accuracy in back-of-device multitouch typing: a clustering-based approach to keyboard updating," in *Proceedings of the 19th international conference on Intelligent User Interfaces*, 2014, pp. 57–66.

[76] M. Goel, A. Jansen, T. Mandel, S. N. Patel, and J. O. Wobbrock, "Contexttype: using hand posture information to improve mobile touch screen text entry," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2013, pp. 2795–2798.

[77] M. Jain and R. Balakrishnan, "User learning and performance with bezel menus," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Austin Texas USA: ACM, May 2012, pp. 2221–2230. [Online]. Available: https://dl.acm.org/doi/10.1145/2207676.2208376

[78] X. Chen, T. Grossman, and G. Fitzmaurice, "Swipeboard: a text entry technique for ultra-small interfaces that supports novice to expert transitions," in *Proceedings of the 27th annual ACM symposium on User interface software and technology*, 2014, pp. 615–620.

[79] M. B. Hoy, "Alexa, siri, cortana, and more: an introduction to voice assistants," *Medical reference services quarterly*, vol. 37, no. 1, pp. 81–88, 2018.

[80] J. Hirschberg and C. D. Manning, "Advances in natural language processing," *Science*, vol. 349, no. 6245, pp. 261–266, 2015.

[81] M. K. Wolters, F. Kelly, and J. Kilgour, "Designing a spoken dialogue interface to an intelligent cognitive assistant for people with dementia," *Health informatics journal*, vol. 22, no. 4, pp. 854–866, 2016.

[82] A. Pradhan, K. Mehta, and L. Findlater, ""accessibility came by accident": Use of voice-controlled intelligent personal assistants by people with disabilities," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 1–13. [Online]. Available: https://doi.org/10.1145/3173574.3174033

[83] E. Corbett and A. Weber, "What can I say? addressing user experience challenges of a mobile voice user interface for accessibility," in *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*, ser. MobileHCI '16. New York, NY, USA: Association for Computing Machinery, Sep. 2016, pp. 72–82. [Online]. Available: https://doi.org/10.1145/2935334.2935386

[84] D. R. Hill and C. Grieb, "Substitution for a restricted visual channel in multimodal computer-human dialogue," *IEEE transactions on systems, man, and cybernetics*, vol. 18, no. 2, pp. 285–304, 1988.

[85] Y. Ishigami and R. M. Klein, "Is a hands-free phone safer than a handheld phone?" *Journal of safety research*, vol. 40, no. 2, pp. 157–164, 2009.

[86] "Smart speakers: use case frequency U.S. 2020." [Online]. Available: https://www.statista.com/statistics/994696/united-states-smart-speaker-use-case-frequency/

[87] A. Mack, "Inattentional blindness: Looking without seeing," *Current directions in psychological science*, vol. 12, no. 5, pp. 180–184, 2003.

[88] V. Ashok, Y. Borodin, S. Stoyanchev, Y. Puzis, and I. Ramakrishnan, "Wizard-of-oz evaluation of speech-driven web browsing interface for people with vision impairments," in *Proceedings of the 11th Web for All Conference*, 2014, pp. 1–9.

[89] H. H. Larsen, A. N. Scheel, T. Bogers, and B. Larsen, "Hands-free but not eyes-free: A usability evaluation of siri while driving," in *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, 2020, pp. 63–72.

Pre-Experimental Questionnaire

This is a questionnaire where we gather general information about the participant, and the prior experience with using a smartphone when on the move, and prior experience with using voice assistants/conversational user interfaces.

What is your gender/sex? *

- ◯ Male
- ◯ Female
- ◯ Other

How old are you? *

[ ]

What type of smartphone are you currenty using? *

- ◯ iPhone
- ◯ Android
- ◯ Other

Are you actively using a voice assistant? If so, which one? *

- ◯ Apple's Siri
- ◯ Google Assistant
- ◯ Amazon Alexa
- ◯ Microsoft Cortana
- ◯ I am not using a voice assistant

If you are not actively using a voice assistant, have you ever used one before? *

Consider other technologies that have a built-in voice assistant like cars, smart houses or other IoT technologies.

- ◯ Never
- ◯ Once or twice
- ◯ A couple of times
- ◯ Many times

If you are actively using a voice assistant, how often do you use it? *

Just like the previous question, also consider other technologies that have a built-in voice assistant like cars, smart houses or other IoT technologies.

- ○ Never
- ○ Rarely
- ○ Occasionally
- ○ Sometimes
- ○ Often
- ○ Always

If you are actively using, or have used, a voice assistant, which of the following tasks do you complete/have completed? *

- ☐ Sending text messages/responding to text messages
- ☐ Setting reminders
- ☐ Asking for gps locations/directions
- ☐ Playing/stopping/generally controlling a music player
- ☐ Checking the weather
- ☐ web searching/asking questions to the voice assistant

Have you ever faced difficulties in using voice assistants? *

- ○ Yes
- ○ No

If you answered yes in the previous question, which difficulties have you faced?

- ☐ Speech recognition
- ☐ Misunderstandings
- ☐ Embarrassment
- ☐ Other

Do you use a voice assistant for other purposes than the suggestions mentioned in the previous question? If so, describe them shortly.

How would you rate your experience with the voice assistant you use, or have used?

0 1 2 3 4 5 6 7

Value

Figure 8.6

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

Are you interested in taking part in the research project

”Evaluating Conversational Interfaces when Mobile”?

This is an inquiry about participation in a research project where the main purpose is to evaluate the use of a conversational user interface, such as Siri. In this letter we will give you information about the purpose of the project and what your participation will involve.

Purpose of the project
This project is a master’s thesis within Applied Computer Science, which is specialized within interaction design and human-computer interaction. The goal of this master’s thesis is to evaluate the extent to which conversational interfaces. Conversational user interfaces, which use speech input and output to create novel interactions, are user interfaces which are found in several devices nowadays and they are still rather novel forms of interaction. The project will be an evaluation study where users perform similar tasks using a typical GUI or using a conversational interface and measure the emerging differences.

Who is responsible for the research project?
Østfold University College (Høgskolen I Østfold) department of Information Technology is the institution responsible for the project.

Why are you being asked to participate?
The participants asked to partake in this evaluation study have been selected based on the person’s technological experience, their daily use of a smartphone for common tasks, and if they have any prior experience to using a conversational user interface like Siri, Google Assistant or Alexa. The participants inquired for the project are students, particularly students within the field of Information Technology and Computer Science, and everyday people who use their smartphone for accomplishing everyday tasks. Since we want to investigate alternative methods for interaction with a mobile phone while walking or travelling, we want participants where alternative methods for mobile interaction may be viable.

What does participation involve for you?
If you choose to take part in this project evaluation, which will be a controlled experiment, you as the participant will be filling out an initial questionnaire about your prior experience with using digital voice assistants, interacting with a mobile phone both traditionally and with a voice assistant, and finally fill out a post-experimental questionnaire about your interaction experience. While conducting the experiment, you will be asked to give your phone number so that you receive information about the tasks in each trial and your interactions will be video- and sound recorded from a third- and firstperson perspective.

These recordings will be used after the experiment has concluded for further reviewing, data collection and statistical analysis. After the videos have been reviewed, the data collected including the videos and sound recordings and your phone number will be permanently deleted. Performance data ontained through analysing the video and audio recordings (time to complete trials, error rate, questionnaire responses) obtained in the analysis may be retained until publication in a numeric form. It will not be possible to retrace the test person from this data. Each participant will be given an ID and the link between the personal data and the ID will be destroyed at the end of the project.

Participation is voluntary

Participation in the project is voluntary. If you chose to participate, you can withdraw your consent at any time without giving a reason. All information about you will then be made anonymous. There will be no negative consequences for you if you chose not to participate or later decide to withdraw.

If you are a student from the university mentioned in this information letter (Østfold University College), or a student form any other university/college, your decision to terminate your participation will not affect your relationship with your university/college, teacher/professor, or your relationship with the university in general.

Your personal privacy – how we will store and use your personal data

We will only use your personal data for the purpose(s) specified in this information letter. We will process your personal data confidentially and in accordance with data protection legislation (the General Data Protection Regulation and Personal Data Act).

- For this project, only the student and the supervisor will have access to the collected data of the experiment. The data will not be shared by any outside parties that are not directly involved with the project work, as well as details that may contribute to identifying a participant will also not be shared outside of the experiment.
- The data collected from this project will be stored in a private filesystem within the university's encrypted server within an encrypted file with limited access where only the student conducting the project and the supervisor will have access to the data.

The participants will not be recognizable in publications, and only the published information about the participants will be their gender and age in a collective context, and not individually. Names and contact information will not be published in any publication papers after the experiment is concluded.

What will happen to your personal data at the end of the research project?

The project is scheduled to end by 15th of December, 2022. After the project is concluded, the personal data stored for each participant will be deleted and anonymised. Further use of the personal data will not be allowed in follow-up studies, only non-personal data presented in a publication will be available for future reference.

Your rights

So long as you can be identified in the collected data, you have the right to:

- access the personal data that is being processed about you
- request that your personal data is deleted
- request that incorrect personal data about you is corrected/rectified
- receive a copy of your personal data (data portability), and
- send a complaint to the Data Protection Officer or The Norwegian Data Protection Authority regarding the processing of your personal data

What gives us the right to process your personal data?
We will process your personal data based on your consent.

Based on an agreement with Østfold University College, NSD – The Norwegian Centre for Research Data AS has assessed that the processing of personal data in this project is in accordance with data protection legislation.

Where can I find out more?
If you have questions about the project, or want to exercise your rights, contact:

- Østfold University College via Georgios Marentakis by e-mail: (georgios.marentakis@hiof.no) or by telephone: +47 696 08 398
- Our Data Protection Officer: Martin Gautestad Jakobsen by e-mail: (martin.g.jakobsen@hiof.no)
- NSD – The Norwegian Centre for Research Data AS, by email: (personverntjenester@sikt.no) or by telephone: +47 53 21 15 00.

Yours sincerely,

Project Leader (Researcher/supervisor)	Student
Georgios Marentakis	Arman Balic

Consent form

I have received and understood information about the project *"Evaluating Conversational Interfaces when Mobile"* and have been given the opportunity to ask questions. I give consent:

- ☐ to participate in *an evaluation experiment*
- ☐ to participate in *an online questionnaire and a paper-based questionnaire*
- ☐ *for my participation to be video recorded for further analysis purposes.*
- ☐ *for my personal data to be stored after the end of the project for follow-up studies*

I give consent for my personal data to be processed until the end date of the project, approx. December 2022

-- (Signed by participant, date)