AALTO UNIVERSITY

School of Science,
Department of Biomedical Engineering and Computational Science

Joanna Bergström-Lehtovirta

Multimodal Flexibility in a Mobile Text Input Task

Master's Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Technology.

Espoo, March 18, 2011

Supervisor: Professor Mikko Sams
Instructor: Antti Oulasvirta, Ph.D.

Aalto University
School of Science

AALTO UNIVERSITY
SCHOOL OF SCIENCE

Abstract of the Master's Thesis

Author: Joanna Bergström-Lehtovirta	
Name of the thesis: Multimodal Flexibility in a Mobile Text Input Task **Date:** March 18, 2011	**Number of pages:** 49
Department: Department of Biomedical Engineering and Computational Science **Professorship:** S-114 Cognitive Technology	
Supervisor: Professor Mikko Sams **Instructor:** Antti Oulasvirta, Ph.D.	

The mobile usability of an interface depends on the amount of information a user is able to retrieve or transmit while on the move. Furthermore, the information transmission capacity and successful transmissions depend on how flexibly usable the interface is across varying real world contexts. Major focus in research of multimodal flexibility has been on facilitation of modalities to the interface. Most evaluative studies have measured effects that the interactions cause to each other. However, assessing these effects under a limited number of conditions does not generalize to other possible conditions in the real world. Moreover, studies have often compared single-task conditions to dual-tasking, measuring the trade-off between the tasks, not the actual effects the interactions cause.

To contribute to the paradigm of measuring multimodal flexibility, this thesis isolates the effect of modality utilization in the interaction with the interface; instead of using a secondary task, modalities are withdrawn from the interaction. The multimodal flexibility method [1] was applied in this study to assess the utilization of three sensory modalities (vision, audition and tactition) in a text input task with three mobile interfaces; a 12-digit keypad, a physical Qwerty-keyboard and a touch screen virtual Qwerty-keyboard. The goal of the study was to compare multimodal flexibility of these interfaces, assess the values of utilized sensory modalities to the interaction, and examine the cooperation of modalities in a text input task.

The results imply that the alphabetical 12-digit keypad is the multimodally most flexible of the three compared interfaces. Although the 12-digit keypad is relatively inefficient to type when all modalities are free to be allocated to the interaction, it is the most flexible in performing under constraints that the real world might set on sensory modalities. In addition, all the interfaces are shown to be highly dependent on vision. The performance of both Qwerty-keyboards dropped by approximately 80% as a result of withdrawing the vision from the interaction, and the performance of ITU-12 suffered approximately 50%. Examining cooperation of the modalities in the text input task, vision was shown to work in synergy with tactition, but audition did not provide any extra value for the interaction.

Keywords: Multimodal flexibility, mobile interaction, multimodal interface, text input, human-computer interaction, sensory modality

AALTO-YLIOPISTO
PERUSTIETEIDEN KORKEAKOULU

Diplomityön tiivistelmä

Tekijä: Joanna Bergström-Lehtovirta

Työn nimi: Multimodaalinen joustavuus mobiilissa tekstinsyöttötehtävässä

Päivämäärä: 18.3.2011

Sivumäärä: 49

Laitos: Lääketieteellisen tekniikan ja Laskennallisen tieteen laitos

Professuuri: S-114 Kognitiivinen Teknologia

Valvoja: Professori Mikko Sams

Ohjaaja: FT Antti Oulasvirta

Mobiili käytettävyys riippuu informaation määrästä jonka käyttäjä pystyy tavoittamaan ja välittämään käyttöliittymän avulla liikkeellä ollessaan. Informaation siirtokapasiteetti ja onnistunut siirto taas riippuvat siitä, kuinka joustavasti käyttöliittymää voi käyttää erilaisissa mobiileissa käyttökonteksteissa. Multimodaalisen joustavuuden tutkimus on keskittynyt lähinnä modaliteettien hyödyntämistapoihin ja niiden integrointiin käyttöliittymiin. Useimmat evaluoivat tutkimukset multimodaalisen joustavuuden alueella mittaavat vuorovaikutusten vaikutuksia toisiinsa. Kuitenkin ongelmana on, että ensinnäkään käyttöliittymän suorituksen arviointi tietyssä kontekstissa ei yleisty muihin mahdollisiin konteksteihin, ja toiseksi, suorituksen vertaaminen tilanteeseen jossa kahta tehtävää suoritetaan samanaikaisesti, paljastaa ennemminkin tehtävien välillä vallitsevan tasapainoilun, kuin itse vuorovaikutusten vaikutukset.

Vastatakseen näihin ongelmiin multimodaalisen joustavuuden mittaamisessa, tämä diplomityö eristää modaliteettien hyödyntämisen vaikutuksen vuorovaikutuksessa mobiilin käyttöliittymän kanssa. Samanaikaisten, toissijaisten tehtävien sijaan modaliteettien hyödyntämisen mahdollisuus suljetaan kokonaan vuorovaikutuksesta. Multimodaalisen joustavuuden arvioinnin metodia [1] käytettiin tutkimuksessa osoittamaan kolmen aistikanavan (näön, kuulon ja tunnon) käyttöasteita mobiilissa tekstinsyöttötehtävässä kolmella laitteella; ITU-12 näppäimistöllä, sekä fyysisellä ja kosketusnäytöllisellä Qwerty -näppäimistöllä. Työn tavoitteena oli määrittää näiden käyttöliittymien multimodaalinen joustavuus ja yksittäisten aistikanavien arvo vuorovaikutukselle, sekä tutkia aistien yhteistoimintaa tekstinsyöttötehtävässä.

Tutkimuksen tulokset osoittavat, että huolimatta ITU-12 näppäimistön hitaudesta kirjoittaa häiriöttömässä tilassa, sillä on ylivertainen mukautumiskyky toimia erilaisten häiriöiden vaikuttaessa, kuten oikeissa mobiileissa konteksteissa. Kaikki käyttöliittymät todettiin hyvin riippuvaisiksi näöstä. Qwerty –näppäimistöjen suoriutuminen heikkeni yli 80% kun näkö suljettiin vuorovaikutukselta. ITU-12 oli vähiten riippuvainen näöstä, suorituksen heiketessä noin 50%. Aistikanavien toiminnan tarkastelu tekstinsyöttötehtävässä vihjaa, että näkö ja tunto toimivat yhdessä lisäten suorituskykyä jopa enemmän kuin käytettynä erikseen. Auraalinen palaute sen sijaan ei näyttänyt tuovan lisäarvoa vuorovaikutukseen lainkaan.

Avainsanat: Multimodaalinen joustavuus, mobiili vuorovaikutus, multimodaalinen käyttöliittymä, moniaistinen vuorovaikutus, tekstinsyöttö, ihmisen ja koneen vuorovaikutus, aistikanava

Contents

1 Introduction

Wireless landline phones first allowed the users to walk around home while communicating, in a fairly safe environment. Soon afterwards mobile phones made it possible to communicate from a variety of new environments, including outdoors. In mobile environments attentional demands and safety risks are much greater than sitting indoors. Interacting with a device in mobile context increases these demands, as the user is constantly sharing the interaction with a trade-off between the environment and the interface. Mobile interaction can be negatively affected when crossing the road, carrying a shopping bag, or running to a meeting, as these events require the use of sensory modalities. Users often have to make an effort to maintain the level of interaction in an environment where attention shifts distract the utilization of sensory modalities to the interaction. Flexible interface contributes to the challenges that the mobility sets on the use, supporting the interaction even while the use is encumbered by the context.

The mobile usability of an interface depends on the amount of information a user is able to retrieve or transmit while on the move. Moreover, the information transmission capacity depends on how flexibly usable the interface is across varying real world contexts. One meta-review [2] suggested, that most empirical mobile usability studies are focused on measuring efficiency and effectiveness, efficiency interpreted as a degree of quick and effective task performance the system enables, and effectiveness as the accuracy and completeness that the user is able to perform with the system in a specific context of use.

Research on multimodal interfaces is traditionally dedicated to improving the information transmission firstly by examining efficient modalities for information transmission, and secondly by examining the cooperation of modalities to maximize information throughput. In other words, research focus is on facilitating efficient modalities and modality combinations to the interfaces for different information types.

The multimodal flexibility is commonly interpreted as the ability of a system to adapt in varying environments maximizing the amount of contexts it can be used in. Again, the major focus has been on facilitation of modalities, but more to maximize the amount of interactions than the actual information throughput. Research on multimodal flexibility of mobile interfaces is focused on improving interactions by (1) allowing user to select interaction modalities, (2) interface's ability to dynamically adapt to user's context or (3) utilizing interaction modalities that are assumed to be free to be allocated to the interface in any context.

These approaches focus on the interface and it's abilities to adapt to as many contexts as possible, assuming that the context of use sets constraints on the interaction with the interface. For example, a bright light can hamper the ability of a user to retrieve visual information from screen, or a traffic noise can mask the aural notification of an incoming phone call. On the other hand, the interaction with the interface can limit a user's abilities to perform tasks in the context. For example, satisfying text entry speed with a touchscreen interface might be achieved only by slowing down walking speed. So is it the environment that is limiting the interaction with the mobile, or the mobile that is limiting the interaction with the environment?

Most studies evaluating multimodal flexibility are focused on measuring effects that two interactions (e.g. typing and walking) cause to each other. In other words, measuring the effect that the interaction with the context has on the interaction with the device, or vice versa. However, assessing these effects under a limited number of conditions does not generalize to other possible conditions (which are infinite in the real world). Moreover, the tests compare single-task conditions to dual-tasking, measuring the trade-off between the tasks, not the actual effects the interactions cause. The problem is, that if the interface's performance is measured while conducting another task (e.g. walking, attending to context, hearing noise), dual-task interference exists. Not only the resources are withdrawn from one interaction, but also allocated to another. As a result, the measured effect of utilization of some users resource includes the effect caused by cognitive load from dual-tasking. Then how to measure the utilization of modalities without setting a condition where modalities are allocated away from the interface to interaction with another task? The only way is to isolate the effect of modality utilization in the interaction with the interface; instead of using a secondary task, the modalities have to be withdrawn from the interaction.

The approach on multimodal flexibility applied [1] in this thesis focuses on how flexible the *interaction* is to adapt to the contexts of use. The difference to previous research is to evaluate interaction that the interface enables, not the interface itself. The purpose is to study, how free sensory modalities (vision, audition, tactition) are from one interaction to be allocated to another. This approach leads to more a generic measure of multimodal flexibility, as it does not depend on the context, but only on the interface's utilization of modalities.

As the question of multimodal flexibility contributes more to research on mobile than stationary devices and contexts, this thesis examines flexibility of mobile interfaces. Text input was chosen to be experiment task, as it is a typical task in mobile interaction. Text is typed into text messages, e-mails, web browsers and calendars among other mobile applications. Furthermore, text input performance also represents target selection speed and accuracy when both, the errors and speed of key presses are considered. Virtually all mobile text input interfaces are either 12-digit keypads or Qwerty keyboards. The 12-digit keypad was used already in landline phones, and Qwerty keyboard in typewriters and in tabletop computers (Figure 1). The alphabetical 12-digit ITU keypad is the oldest mobile text input interface. In the 21st century, a full Qwerty-keyboard first penetrated corporate and heavy user mobile phone markets. Now the global mobile text input interface markets are shared by ITU-12 keypads, physical Qwerty keyboards and the touchscreen virtual keyboards.

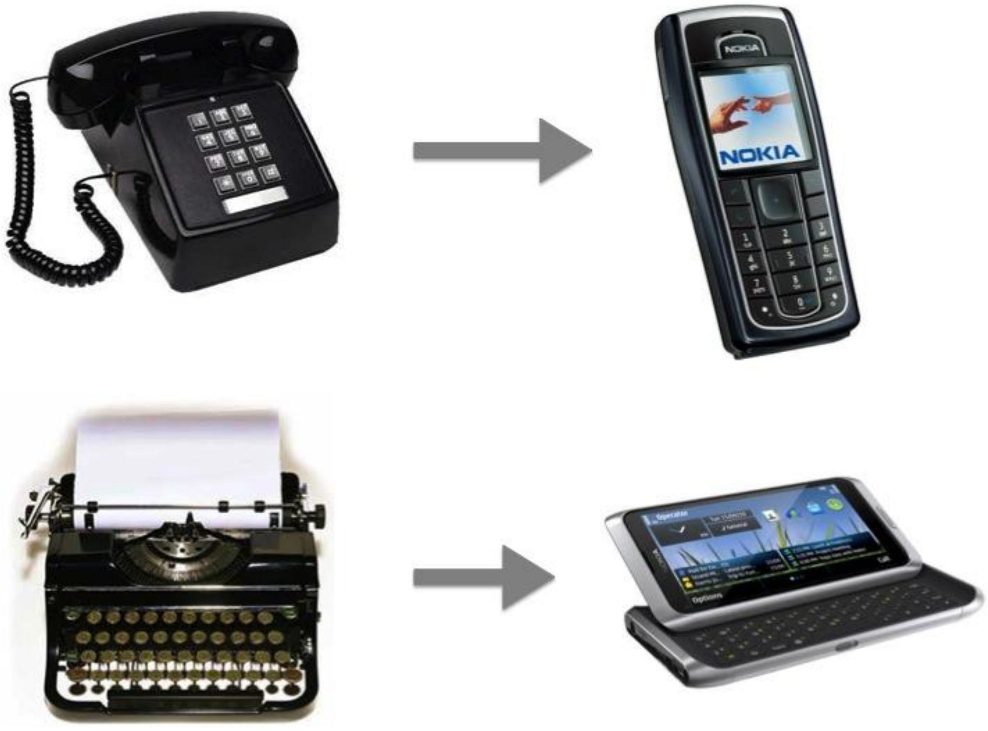

Figure 1. Keyboards from landline phone to 12-digit alphabetical keypad and from typewriter to mobile Qwerty-keyboard.

To assess mobile interfaces' utilization of modalities, the multimodal flexibility of these three common mobile text input interfaces is compared. The primary research question is:

> **How flexible are the three mobile text input interfaces in multimodal interaction?**

As the multimodal interaction approach focuses on efficiency of modalities, the values of single modalities are also examined. The secondary research question is:

> **How much is each modality utilized with each interface in a text input task?**

Finally, effectiveness is considered by studying the cooperation of modalities. The third research question is:

> **How the modalities are cooperating in a mobile text input task?**

The background on multimodal flexibility is given in the following, second chapter, indicating the different approaches on the subject by first introducing multimodal interaction and then presenting previous research on multimodal flexibility. The third chapter describes the methodology applied in this thesis and indicates the novelty of approach and importance of the study. The fourth chapter presents the experiment and shows the results. The last, fifth chapter, concludes the thesis discussing the validity and importance of both, the method and the study, as well as the possible focuses of future work.

2 Related Work

The mobile user is constantly trade-offing sensory attendance between the interface and the environment. Oulasvirta et al. [3] implied that the duration of continuous visual attendance to mobile interface decreases from approximately 14 seconds in a laboratory context to as low as 4 seconds in a real world mobile context. Real world contexts increase the need for dual-tasking and decrease user ability to devote attention to the interactions. Even memorizing word lists while walking is shown to become more difficult with age because balance and gait are in greater need of attentional resources [4], not to mention mobile interaction encaging more modalities to be attended.

2.1 Multisensory Perception

Sensory modalities appear to operate together, but it is not known if the perception results from linked but separate unimodal sensations, or from a single, supramodal sensation [5]. Cognitive-load theory assumes that information is processed within a limited working-memory. Thus, the theory suggests, design of information presentation should focus on reducing the load on working-memory (e.g., [6, 7, 8]). The proposed techniques to reduce the load include dual-mode presentation [7], where information is transmitted utilizing two different modalities. Wickens' Multiple Resource Theory (MRT) also supports the idea that multimodal information transmission could be more effective than a single modality one [9]. Although Wickens has criticized that cross-modal audiovisual time-sharing, for instance, does not necessarily overcome intra-modal visual-visual or auditory-auditory one [10].

In multimodal information transmission, sensory modalities can have different relations [11]. Modalities can either (1) work identically, transmitting the same information, (2) work synergistically, by sending partially different information and thus adding information to each other or (3) interfere and induce each other altering received information [12].

Dual-mode perception increases performance in some cases. For example, audio-visual feedback is shown to result in more effective learning, than when employing only visual material [13]. Mayer et al. studied audio-visual dual-mode instructions in technical material [14, 15]. They implied, that audio-visual instructions overcome single-modality audio and visual instructions, but only when information is presented simultaneously. Blake et al. [16] found synergy between tactile and visual modalities; somatosensory information was shown to be able to disambiguate information when visual cues were conflicting, but only when stimulated simultaneously. Although simultaneous multimodal presentation is in some cases shown to be better than sequential, the cross-modal cues can, however, interfere and induce each other or result in an illusion, where stimuli are not perceived correctly. Between vision and audition, a well-known illusion is the McGurk effect, where sensed sound and image are mismatching, causing a synthesized perception. For example, if a seen face pronounces a phoneme and a heard voice pronounces another phoneme, the perceived phoneme can be either of them, or a totally new, synthesized one [17]. Another illusion effect between vision and audition is the ventriloquist effect. This effect relates to the spatial location of sound source, which is perceived in synthesis with visual perception. For example, speech from a video is perceived coming from the people seen on the screen, when in reality it comes from the speakers [18]. There is also a potential interference between tactile and visual modalities in information processing [19]. Vision is shown to alter the perception of tactile modality, for example by affecting the perceived location of a finger pointing [20], or in an effect called “rubber hand illusion” [21]. In addition, the processing of tactile cues is dependent on visual processing in orientation [22]. It is also implied, that performance on discriminating tactile and visual targets decreases, if the stimulus is invalidly cued to the other modality [23].

2.2 Multimodal Interaction

We encounter multimodal interaction in our everyday lives. Even normal face-to-face communication is multimodal, employing speech and gestures [24, 25] and sensory modalities of vision and audition accordingly. As in conversation between humans, there is input and feedback in interaction between the human and the interface.

According to Perakakis et al. [26], the synergy of modalities in a multimodal interface can result in better performance than a constituent unimodal ones: “A synergistic multimodal interface is more than the sum of its parts”. Although the authors are focused on graphical user interface (GUI) and speech as interaction modalities, their thoughts are applicable considering other modalities as well. The authors define, that a multimodal system can become more efficient in task performance by (1) input modality choice (either by user or by adaptive system), (2) improving the presentation of output and (3) correcting errors of one modality by perception had with another. A major focus of multimodal interaction research is related to these three areas. The presentations of output and feedback modalities, as well as the integration of input modalities are developed to maximize the information throughput, i.e. the efficiency of the information transmission. The modality integration, on the other hand, also relates to the effectiveness of the information transmission, i.e. the amount of correctly transmitted information.

2.2.1 Multimodal Feedback

Previous research has shown that both auditory (e.g. [27]) and tactile feedback (e.g. [28]) can improve performance in a visual task. Prewett et al. [29] conducted meta-analysis comparing visual single-modality feedback to multimodal visual-tactile feedback. Results indicated that visual-tactile feedback enhanced task effectiveness more than visual feedback. Visual-tactile feedback was suggested to be particularly effective at reducing reaction time and increasing performance. However, it was not shown to substantially reduce the number of errors in task performance.

Jacko et al. [30] compared computer task performance with three sensory modalities (audition, haptic, visual) and combinations thereof utilized as a feedback. The task was to move an object (file icon) on a computer screen to the target location (folder). The feedback was indicating the object to be positioned correctly over the target location as follows; the auditory indicator being a sound mark, visual a highlighting color, and haptic a mouse vibration. The effects of feedback was compared in performance between visually healthy older adults and adults suffering from age-related ocular disease. The results show, that multimodal feedback aided performance with both, healthy and visually impaired users, compared to unimodal conditions. Visual unimodal performance was worse than auditory-haptic dual-mode performance with both groups. Auditory feedback was shown to be synergistic in performance of dual-modal and multimodal conditions with both, haptic and visual feedback. However, significant differences were not observed between any unimodal conditions. In addition, tactile feedback did not significantly improve the performance when added to visual feedback compared to the visual-only condition. Finally, the authors noted that the addition of non-visual (auditory, haptic, or both) feedback to visual feedback resulted in improved performance for both groups.

2.2.2 Modality Integration

Nigay et al. [31] defined a design space for multimodal systems for the design of modality integration to the interface. The idea is to consider information transmission according to the types of information and the integration types of modalities. For example, synergy of modalities can only occur in simultaneous (parallel) utilization of the modalities, as sequential modalities providing additive information on each other are alternating instead of synergistic [31]. Moreover, synergy depends on the data type; independent data might interfere being concurrent with the other data and competing on the transmission capacity (Figure 2). This approach on multimodal information transmission implies that effectiveness and efficiency of multimodal interaction varies depending on the modality integration pattern and the type of transmitted data, contributing to the focus of most multimodal interaction research on facilitation of modalities to improve transmission speed and avoid transmission errors.

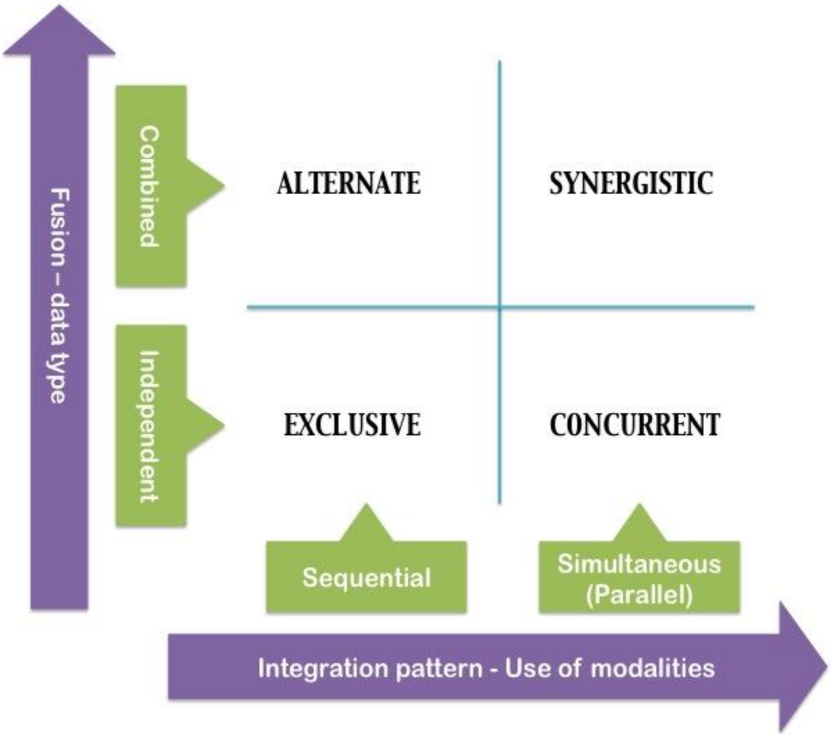

Figure 2. Multimodal interaction design space, applied from [31].

Oviatt et al. [32], on the other hand, have studied optional integration patterns. The authors have examined user preferences on the utilization of modalities. The study suggests, that the user's predominant modality integration pattern (whether its sequential or simultaneous) differs individually, and moreover, is delivered quite consistently (93.5% of the time). This result adds even more challenge to the integration of modalities to multimodal systems, as it seems that user preference has an influence on modality cooperation in addition to the information type delivered with the modality.

2.3 Multimodal Flexibility

A major focus of multimodal interface development is on operating the modalities of interfaces to maximize the information throughput. Multimodal flexibility is commonly interpreted as the ability of a system to adapt in varying environments maximizing the amount of contexts it can be used in. Again, the major focus is on facilitation of modalities, but more to maximize the amount of interactions than the actual information throughput. Research on multimodal flexibility of mobile interfaces is focused on improving interactions by (1) allowing user to select interaction modalities (e.g., [33, 34]), (2) interface's ability to dynamically adapt to user's context (e.g., [35]) or (3) utilizing interaction modalities that are assumed to be free to be allocated to the interface in any context (e.g. [36, 37]). Additionally, an interface's dynamic adaptation to the context can happen in two ways. The first is, that the system itself modifies the output within the modality (e.g. brightening the display when used under sunlight, illuminating the keyboard when used in the dark, or increasing the icon size when used in a bumpy context). And second one is, that the system selects the modality according to contextual attributes (e.g. audio in bumpy context, vision in stabile context).

2.3.1 User's Choice of Modality

According to Oviatt [33] users know naturally the most efficient modality in their contexts of use. However, in some situations, being able to choose might benefit the user less than the choosing costs effort. Oviatt has suggested, that users prefer to interact multimodally rather than unimodally noting, that 95% to 100% of users preferred to interact multimodally when they were free to use either speech or pen input in a spatial domain [38]. In addition, in [32] the authors imply, that multimodal interface users spontaneously respond to dynamic changes in their own cognitive load by shifting from unimodal to multimodal input communication as the load increases.

Hoggan et al. [34] studied users preference of tactile and audio cross-modal feedbacks with vision, and a vision-only feedback condition in situ. The preferences were studied by (a) measuring the utilization of different feedbacks in different places and (b) in different levels of context's attributes, and (c) asked the reasons for preferences. Results indicated users to choose an added tactile feedback 82% of the time, and added audio 18% of the time over the contexts. Vision-only feedback was never chosen. There were four locations, three in which tactile feedback was most preferred (home, work, and restaurant), and one in which audio and tactile feedbacks were preferred equally (commuting). Tactile feedback was preferred in contexts where vibration level was lower, and audio in contexts with higher vibration levels. For contexts with high and low noise levels, tactile feedback was preferred over audio, whereas in medium noise levels audio was more preferable. The reasons preferring audio included bumpy (vibrating) contexts, and tactile preference the social acceptability issues. In addition, most participants stated to find the feedback modalities equally good, and to prefer to use those simultaneously.

2.3.2 Interface's Adaptation to Context

In addition to user preferences, the interface can adapt to the context dynamically. Research has been conducted to improve the existing interaction modalities in the interfaces to gain better performance, but also the systems' automatic adaptation to the context has been considered. Text input task performance has been improved in existing input methods, for example, by modifying the key sizes on the screen [39], as well as organizing the layout of the keys to a more effective order (e.g., [40, 41, 42, 43]). Adding audio or tactile feedback to visual feedback is shown to improve performance in several studies. For example, the presentation of visual icons combined with audio and tactile feedbacks is investigated in [44, 45, 46]. Novel feedback methods include earcons and tactons, which are aural and tactile icons [47]. Earcons and tactons recognition and discrimination performance has shown the possibility for further development and implementation [34]. Compensating touch screen's lack of tactile feedback compared to physical key's edges and button presses by providing artificial tactile feedback has shown promising results as well [48]. In addition, novel audio presentation methods include 3D audio-spaces utilized for instance, for presenting menu options [37, 49].

Studies on context recognition (e.g. [35, 50]) have shown the technical potential to develop interfaces that automatically adapt the utilized modalities according to contextual attributes. In addition, environmental thresholds have been suggested, in which the utilization of modalities should be changed.

Hoggan [51], for example, studied the exact environmental levels where audio and tactile feedbacks become ineffective, implying that performance decreased significantly for audio feedback when the context's noise level exceeded 94dB, and for tactile feedback when vibration levels exceeded 9.18g/s. However, studies examining modality integrations to adaptive interfaces are rare.

2.3.3 Novel Interfaces

One motivation in developing novel multimodal interaction methods is to provide information cross-modally to make the interface more accessible. The benefit of cross-modality is, that as mobile interaction occurs in varying contexts, it is more likely that the information is communicated successfully when it is transmitted via various modalities instead of one. Similarly users with impairments are more likely to be able to use multimodal interface than a unimodal one [52]. As vision is known to be an important modality for interaction with the environment in mobile contexts, an "eyes-free" approach has been suggested [36] to let vision free from the interface to be allocated elsewhere. Similarly, hands are assumed to be often utilized in context-related tasks, and to avoid manual multitasking, some "hands-free" input methods are developed [53, 54] in addition to more traditional methods, such as speech input in common hands-free devices.

To mention a few, gestures [55, 56], foot tapping [53], wrist rotation [57], and head tilting [54] are novel input methods developed to provide "eyes-free" and "hands-free" interaction. However, these novel input methods are often very visible to other people in the environment, and social acceptability might set limitations to their usability [56, 58]. Moreover, the "eyes-free" and "hands-free" approach makes assumptions on the contextual attributes – it assumes both, that replaced modality (e.g. vision) is reserved by the context and, that utilized modality (e.g. foot) is not reserved or distracted by the context. This paradigm faces again the fact, that there are an infinite number of unpredictable contexts in the real world, and designing to eliminate one problem does not necessarily contribute to others.

2.3.4 Interaction Research on Multimodal Flexibility

Most studies evaluating multimodal flexibility are focused on measuring effects that two interactions (e.g. typing and walking) cause to each other. In other words, the effect that interaction with the context has on the interaction with the device, or the other way around. Contextual tasks have been demonstrated to hamper the performance with an interface, for example in studies focusing on walking [59], bumpy contexts [51], or lack of vision [36]. On the other hand, studies have shown the interaction with the interface to hamper task performance in the context. For example, walking speed [39, 60, 61] and driving [62, 63] performances have been shown to decrease as a result of simultaneous interaction with the system. In addition, the dynamicity of the mobile context has been in focus; as attention is required for the safety reasons when navigating in dynamic environments, avoidance is also needed. Lumsden [64] has studied avoidance cues while interacting in the dynamic laboratory test environment the authors developed observing both, the context's effects on the performance with the interface, and the effects on contextual task the interaction with the interfaces causes.

Some studies have given effort on attempts to cover multiple contexts in the real world to assess the effects of differing attributes. For example, the previously mentioned study [34] utilized four environments; home, office, commuting, and restaurant, and was able to distinguish these contexts by the logged attributes (vibration level and noise). On the other hand, some studies attempt to recognize and categorize limitations of contexts. Lemmela et al. [65] approached multimodal interaction design by first identifying the interaction limitations of different mobile situations by observations. They identified contexts and estimated aural, visual, physical and cognitive load in them as well as the cause of the load (i.e. the type of the stimuli, such as traffic noise or speech in aural load). However, these measures are subjective and can only cover a limited selection of possible contexts. Furthermore, the problem of measuring the effects of interactions on each other lies in the utilization of dual-tasking. If the interface's performance is measured while conducting another task (e.g. walking, attending to context, hearing noise), dual-task interference exists. Not only the resources are withdrawn from one interaction, but also allocated to another. As a result, the measured effect of utilization of some users resource includes the effect caused by cognitive load from dual-tasking. This suggests, that measuring an interface's ability to let a user's modalities free to be allocated to the context is not measured, as the effect was not isolated from other effects such as the dual-tasking inference.

3 Experiment Method in this Thesis

The previous chapters presented research in the field of multimodal interaction. In the cooperation of modalities, examples on the synergy and interference of different sensory cues were given, as well as the comparison between unimodal, dual-modal and multimodal utilization. In addition, differences on simultaneous and sequential utilization, as well as the utilization method (input/output) were discussed. Approaches on multimodal flexibility were then introduced. By examples from previous research, the focus of developing multimodal flexibility was shown to be on interfaces that adapt to the contexts of use. Additionally, the research on evaluating multimodal interfaces presented controlled laboratory tasks, in situ and field experiments, and observations on effects between the interaction with the context and the interaction with the interface.

However, assessing the context's effects in a limited number of conditions does not generalize to other possible conditions (which are infinite in the real world). The problem of measuring the effects of interactions on each other results from dual-tasking, as the effects cannot be controlled. Then how to measure the utilization of modalities without setting a condition where modalities are allocated away from the interface to the interaction with another task? The answer is to isolate the effect of modality utilization in the interaction with the interface; instead of using a secondary task, the modalities have to be withdrawn from the interaction.

3.1 Importance of this Study

The approach on multimodal flexibility applied in this thesis focuses on how flexible the *interaction* is to adapt to the contexts of use. The difference to previous research is to evaluate interaction that the interface enables, not the interface itself. The purpose is to study, how free sensory modalities are from one interaction to be allocated to another. This approach leads to a more generic measure of multimodal flexibility, as it does not depend on the context, but only on the interface's abilities for flexible interactions. Furthermore, the utilized method [1] allows a highly controlled experiment, which is still applicable to any modalities and generalizable to any interactions' with an interface. The procedure is to withdraw sensory modalities and combinations thereof by "blocking" those from the interaction. This contributes to the real world modality allocations by blocking two-way information transmission, not just an input or a feedback one.

3.2 HCI Research Methodology

Behavioral research in human-computer interaction (HCI) can be divided into three (however overlapping) types: Descriptive, Relational and Experimental research [66]. This thesis uses a controlled experiment method, utilizing both, relational and experimental research. The study includes two independent variables, the mobile text input interface, and the sensory modalities (and combinations thereof), which are hypothesized to affect the text input performance. Relational research identifies the relations between the variables, and experimental research identifies the causes of events [lazar]. This experiment seeks to identify which modality is utilized and how much withdrawal of one affects the performance with a certain interface, thus applying both types of research.

3.2.1 Design

Experimental design usually starts from hypothesis. Hypothesis reveals the variables that are examined in the experiment, then the significance of the relations between variables should be tested, and finally the limitations and potentiality of results considered. Typical variables in HCI are an interface (independent variable) and the performance time or performance errors (dependent variables) in the interaction with the interface. In addition to interfaces, this experiment has a second independent variable, the sensory modalities.

Experimental design must be decided considering the type of independent variables and time use. Using within-subjects design, it has to be possible to conduct all the conditions with every subject (a time issue), and naturally, the type of independent variable must allow this (e.g. in examining the effects of gender, a between-group design has to be used, as subject cannot belong to both genders) [66]. This experiment is possible to design to last only 1-1.5 hours per subject conducting all the conditions. Furthermore, between-group design is not necessary as the experiment task is simple and do not require grouping of the subjects.

When there is more than one independent variable, a factorial design is used. Factorial design can include both or either of between-group and within-group design [66]. Having the interface and modality condition as independent variables, this study uses factorial design. The number of conditions needed in this experiment is the number of interfaces multiplied by the number of sensory modality conditions. The number of modality conditions can be defined with combinations. There are the single-modality conditions and bimodality combinations, as well as conditions where none of the three modalities or all of the three modalities are blocked, resulting in eight combinations in total. Thus, the number of conditions in this experiment is three interfaces multiplied by eight modality combinations resulting in 24 conditions.

To minimize the effects on performance caused by fatigue and learning, the randomization or counterbalancing of the order of experiment conditions is important, especially in within-subject designs. In this study, the order of the interfaces and modality conditions is rotated and reversed. In addition, training the task can be used in simple experiment tasks. The subject should be made to feel comfortable and given enough time to adapt to the condition to avoid effects caused by the context (if that is not one of the examined variables).

3.2.2 Measures

Words per minute (WPM) is a common performance variable used in interaction experiments to describe the transcribing or typing speed with an interface. To address the errors, a common variable is the Keystrokes per character (KSPC), calculating how many keystrokes were needed to produce one (correct) character. To simplify the analysis in this study, the errors and the performance time are not analyzed separately, but together forming just one dependent variable. As the design allows, the variable can be the number of correctly transcribed words during a certain time. Thus the errors can be discarded, and only the typing speed of producing correct words is counted. This design also keeps the experiment time under control, as the task time is limited.

The variable in this experiment is chosen to be 80% correctly transcribed letters in half a minute. This is due to the consideration that 80% correct words are still readable and understandable – real-life text messages include some errors as well.

The data often needs coding before the analysis. To discard the bias caused by personal factors (as experience in fast typing in general, or with certain interfaces), and to code the data to more comparable form, the performances are normalized within the subjects within every condition. This is conducted by dividing the WPM with subject's "baseline" WPM (the condition where all modalities are in use). This results in scores that indicates the performance change (percents) in conditions in relation to the baseline performance.

3.2.3 Statistical Tests

Statistical tests are used to assess the significance of the results. Paired-samples t-tests are used to compare mean values, when the means are contributed within group [66]. The t-test returns a t-value, high value implying the means to differ significantly. T-tests are used in this thesis to define 95% confidence intervals (CIs) when comparing the means of modality conditions.

The F-test is analysis of variance (ANOVA), also comparing means. However, ANOVA is used to compare the means of two or more groups (whereas t-test within one or between two). Multiple-level, repeated measures ANOVA is needed for this within group study having two independent variables to determine the effects of the interfaces. In addition, Fisher's least significant difference (LSD) is used as a post-hoc analysis to define the significant differences between the means. Fisher's LSD requires the rejection of null-hypothesis (from the F-test) before it can be utilized.

3.3 Multimodal Flexibility Method and Indices

The purpose of this study is to compare, how three common text input interfaces utilize three sensory modalities in interaction by applying the multimodal flexibility method [1]. The idea of the method is to assess (1) how flexibly the interaction with an interface adapts to the lack of modalities, (2) how modalities cooperate in the interaction and (3) how dependent the interaction with the interface is on each modality. Moreover, the method generates quantitative and comparable information on these modality utilizations. The procedure of the method is to measure the effect of blocking sensory modalities on the task performance with the interface. The blocking conditions include the baseline (none-blocked) condition, the single-modality conditions, and all blocking combinations of modalities. From the performance scores in the conditions, the purpose is to calculate an overall index to each interface's multimodal flexibility.

The index is calculated by first normalizing the performance scores in every condition. This results in scores where the baseline performance has the highest score (1) indicating 100% performance. The conditions are hypothesized to affect performance of an interface. Thus the scores in other conditions are supposed to vary between 0 and 1, indicating the percentage of the baseline performance. The multimodal flexibility index (MFI) is simply the average over the conditions (*S*) where modalities (*b*) are blocked from the interaction (Equation 1).

$$MFI = \frac{\sum_{b \subset B} s_b}{2^n - 1} \qquad (1)$$

In addition, the interface's dependence (*D*) on a single modality (*m*) is calculated as in equation 2:

$$D_m = \frac{\sum_{b \subset B} (s_{b \cup \{m\}} - s_b)}{2^{n-1}} \tag{2}$$

Dependence value (D-value) is the average decrease in performance caused by the withdrawal of a single modality from other present ones over every tested condition. The D-value can be interpreted as the interface's percentual dependence of a modality.

Finally, the cooperation of modalities is calculated to determine the synergies and interferences existing in the experimental setup. Bimodal values are simply the performance scores in bimodal conditions, and unimodal in single modality conditions. The unimodal values are summed to indicate if the bimodal performance exceeds the sum of its parts, and thus an occurring synergy.

4 Study

4.1 Introduction

In the real world, the mobile device user typing the text usually generates the words, but in this study, transcribing was utilized in the experiment task. Salthouse [67] reviewed the research on transcription typing, integrating the found phenomena into a four-component heuristic model. The model consists of the phases of transcription typing; first, the verbal material is registered and perceived, and next partitioned to appropriate chunks and discrete characters, then the material is translated into physical movements, and finally movements are executed as key presses. Salthouse's findings include effects related to typing speed, such as the positioning of the movements, the interkey intervals, eye-movements, and error types occurring in different phases of typing. The author notes, that fast or experienced typists' interkey intervals are only a fraction of normal choice reaction time, suggesting that processes in typing are overlapping in time. However, in a choice reaction task, the stimulus appears only after the previous response, preventing simultaneous processes. According to Salthouse's review, the rate of typing is nearly the same for random words as it is for meaningful text. Instead, the typing speed decreases if the view to the material is restricted. These three findings suggest the transcribing method in the text input task applied in this thesis to be relatively natural as (1) the vision was not restricted to the material (allowing parallel processing), (2) meaningfulness of text does not effect the speed and, and (3) pre-view to material was not restricted, thus not decreasing the typing speed.

Salthouse [67] also reviews the copying span and the stopping span in typing. Copying span is the number of words that can be typed with a single inspection of the material, and it ranges from about two to eight words. Stopping span is the number of characters the typist types after stop notice (whether it is a given signal, or the typist himself perceiving an error), ranging about from one to two keystrokes. Copying span might affect on the performance of experiment task in this thesis, especially in conditions where vision is not blocked, as the attention is then shifted between the material and the keyboard or the transcribed text on the screen. The greater the copying span is, the more the vision can be attended to the interface.

Errors also relate closely to the mobile text input task when sensory modalities are distracted. Salthouse noted that typist detects only 40% to 70% of typing errors without reference to the transcribed text [67]. In this study, the transcription was visible all the time, but error corrections were not permitted. Salthouse [67] mentioned four different error types: substitution (e.g., modal for model), intrusion (e.g., moddel for model), omission (e.g., mdel for model) and transposition (e.g., moedl for model). If the error occurs in the reading phase of the transcribing, it is always a perceptual confusion. If the error occurs in execution phase, it is either (1) a misplaced finger position or inaccurate movement trajectory, (2) a simultaneous depression of two adjacent keys, (3) an inadequate force or reach on keystroke, or (4) a keystroke preparation out of sequence, according to the four error types (in that order) [67]. All these error types were counted to determine the 80% correctly transcribed words in this study.

The three text input interfaces compared in this experiment utilize the same three sensory modalities (audio, visual, tactile) in interaction, but in diverging manners. The keys are virtual on touchscreen, and physical buttons in keypad and keyboard. For example, in a text input, vision is utilized first to locate the key of a letter. Then tactition lets the user know when finger touches, and then he presses the key. From the physical keys, the user can feel the edges of the key, and the button pressing down. Touch screen also supports artificial tactile feedback (vibration). The interface might play a sound indicating that the key press is registered, and finally user sees the letter on the screen. So what happens to this complex sequence of processes when typing is conducted in the real world?

Typing can be learnt to perform faster without looking at the keyboard. With the desktop devices, a common way to type fast is the 10-digit typing system, where the writer only looks at the feedback (the typed text) on the display. Predictive and corrective text entry mode is optional but integrated in most mobiles in the market to enable faster typing speed. However, in mobile contexts the vision might be totally allocated to the context, and in addition to the keyboard, the display cannot be attended to either. The alphabetical 12-digit keypad in a mobile text input is fairly easy to learn to use eyes-free. Instead, a qwerty-keyboard in mobile phones has smaller keys than the computer keyboard, and thus more error prone input. Qwerty, however, generally has a faster typing speed resulting from requiring less keystrokes per character than the 12-digit keypad. Typing with tabletop Qwerty-keyboard is measured to be approximately 50 words per minute (WPM) [68], decreasing to about 25 WPM with mobile Qwerty-keyboard [69, 70]. 12-digit keypad text input speed has been measured to perform even slower speeds [69, 70, 71].

4.2 Experiment

The purpose of the experiment was to compare the multimodal flexibility of three mobile interfaces in a text input task. The interfaces were:

1. Nokia Xpress Music 5800 touch screen Qwerty keyboard
2. Nokia E75 physical Qwerty keyboard
3. Nokia E75 12-digit, ITU-12 keypad

Each interfaces' multimodal flexibility was assessed based on the utilization of three sensory modalities; vision, audition and tactition. The study was conducted in a controlled laboratory experiment applying the multimodal flexibility method [1] to the text input task performance.

4.3 Method

4.3.1 Subjects

Twelve students were recruited for the experiment from Helsinki University of Technology. Their mean age was 22.8, with an age range of 21 to 26 years (SD = 1.6 years). Seven of the subjects were male. As for usage experience, 11 were currently using an ITU keypad, seven with predictive text entry and four without. One subject was using a physical qwerty-keyboard but was also experienced in using a 12-digit keypad. Eleven subjects had experience of typing with a physical qwerty-keyboard, and five with touchscreen. Two subjects reported that they send fewer than 10 text messages per month, five reported sending 10–50, four between 50 and 100, and one over 100 text messages.

4.3.2 Text Typing as an Experiment Task

The task was to type words as correctly as possible for 30 seconds. For every task, there were 5 sentences presented on the computer screen at the same time. After 30 seconds had passed, the sentences disappeared, the screen turned red, and a sound mark was played. The sentences were real including real words, to represent normal interaction with a mobile, such as writing a text message. However, the task differed from common real text input task in that there was no text generation, but rather transcribing. The material was kept visible during the whole task conduction time, as 30 seconds was considered to be too long for memory based transcribing. Real sentences were used, as copying pseudo text would require even more attention to the source displaying the text than real words. The sentences were from a set of 500 sentences from Soukoreff & MacKenzie [72], translated into Finnish by Isokoski [73]. There were no special characters, punctuation marks, uppercase letters, or Scandinavian letters.

4.3.3 Apparatus

With the Nokia XpressMusic 5800, the touchscreen virtual Qwerty keyboard ("Touch-Qwerty") was used holding the device horizontally. Nokia E75's both text input interfaces, physical Qwerty keyboard ("Physical-Qwerty") and the 12-digit ITU keypad (ITU-12) were used (Figure 3). The default tactile (with Touch-Qwerty) and audio feedback of these devices was set to "high". Predictive text entry was turned off.

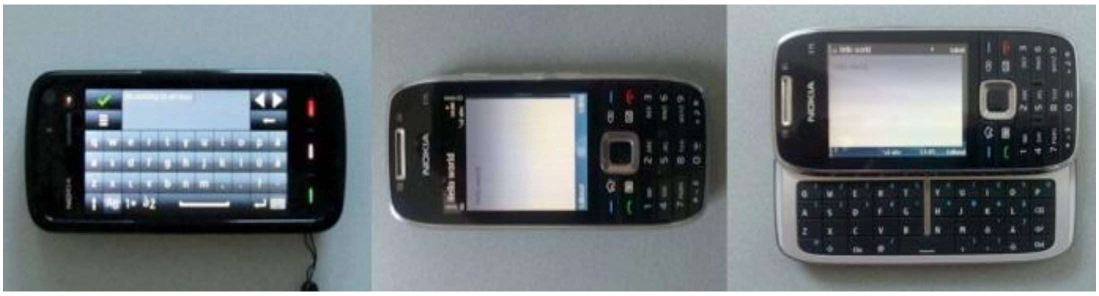

Figure 3. Text input interfaces compared in the study. From left: Touch-Qwerty keyboard, ITU-12 keypad and Physical-Qwerty keyboard.

4.3.4 Blocking of Modalities

The vision was blocked with a cardboard placed under the participants' chin (Figure 4), so that the subject was still able to maintain a natural sitting position in the chair, and hold the mobile in a similar, natural way as in other conditions. In addition, the vision to the computer screen could be maintained free with this blocking solution. The keyboard of the computer was covered with cardboard so that the user could not see the Qwerty layout from it. Hearing was blocked by turning the key-press sound off from the mobiles, and by hearing protectors (Peltor Optime H520) (Figure 4) to block the mechanical sounds caused by key presses with physical keys.

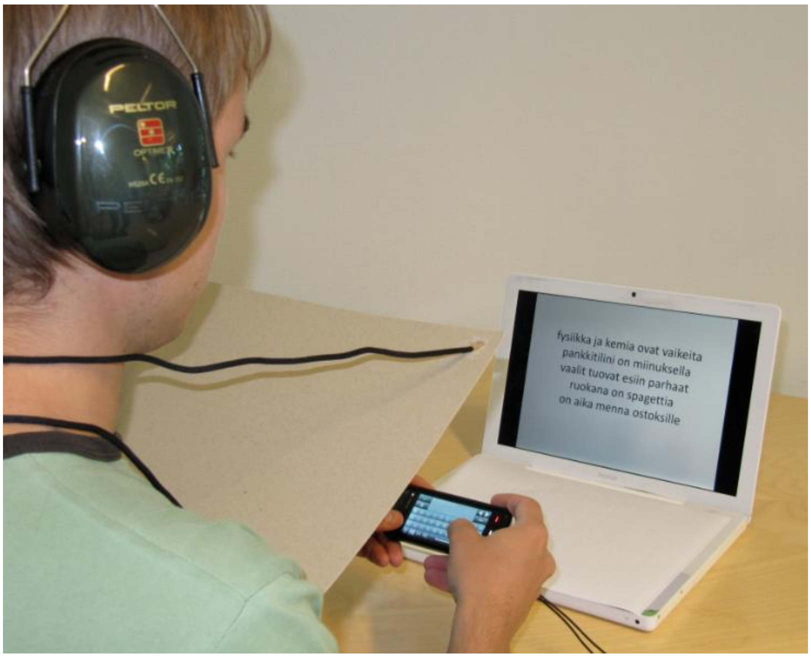
PELTOR
OPTIME
fysiikka ja kemia ovat vaikeita
pankkitilini on miinuksella
vaalit tuovat esiin parhaat
ruokana on spagettia
on aika menna ostoksille

Figure 4. Tactile feedback unblocked and vision (cardboard) and audition (ear protection) blocked.

The physical keyboard provides natural tactile feedback, such as button edges to separate keys, and key press to feel that the input is given. A thin plastic layer was placed on the keyboard to block these tactile feedbacks (Figure 5). In addition, the tactile feedback -feature was turned off from Touch-Qwerty. However, this solution could not prevent feeling the whole keyboard edges. The key layouts were printed on the plastic layer to not to block the vision. Nevertheless, the layer occluded visual feedback from Touch-Qwerty's text input interface, which flashes the key when pressed.

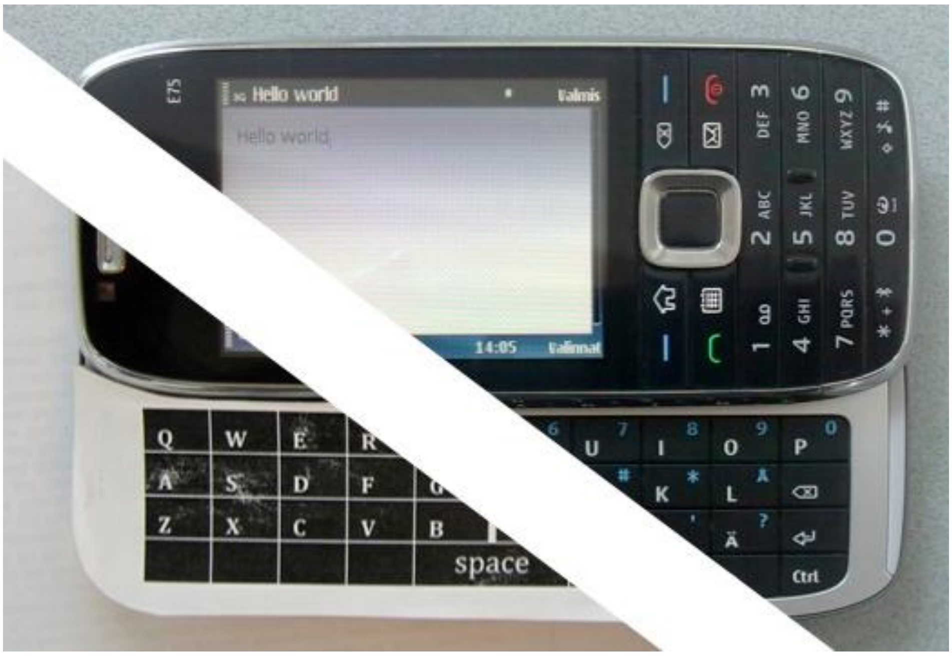

Figure 5. A thin layer of plastic with printed key layouts on the keyboard (left) was used to block feedback from the button edges and key releases of the original keyboard (right).

4.3.5 Design

The experimental design was an eight-by-three within-subjects design with blocking combinations as the first factor and an input interface as the second. In total, there were eight modality conditions (Table 1*): Ø, a, t, v, ta, av, vt,* and *atv,* with two trials performed in each. Every subject thus completed 48 trials, and the experiment time was around 1-1.5 hours. The order of the two factors was counterbalanced by reversing and by rotating.

Table 1. Modality conditions.

	Free modalities	Condition
1	tactition-audition-vision	atv
2	audition	a
3	tactition	t
4	vision	v
5	tactition-audition	ta
6	audition-vision	av
7	vision-tactition	vt
8	Ø (none)	Ø

4.3.6 Procedure

The participants were first trained to use each keypad with a three-task training set. They were instructed to write the words as fast and as correctly as possible, and to separate words and sentences with space characters. Correction was forbidden to minimize variance due to strategic differences and to ensure comparability of blocking conditions, as correction would be hard to use when vision is blocked.

Before every blocking combination, the subject had a chance to practice the typing with the next blocking. When the subject was ready, the moderator made the set of sentences visible. After 30 seconds, a red indicator flashed to mark the end of the time, when subject was instructed to stop the typing and hand the device to the moderator, who saved the transcription. All trials were videotaped and a brief demographic questionnaire was filled.

The instructions for the subjects were:

- Type the words as correctly as possible, and as many as you can in the 30 seconds the sentences are visible.
- Do not correct if you type a wrong letter, just proceed to the next one.
- Do not mind if the letters are upper case or lower case.
- We don’t use Scandinavian letters, so type with a’s and o’s, but do not mind if you unintentionally type ‘ä’ or ‘ö’.
- Even if the sentences are presented in different rows, just separate them by space ‘ ‘ as you do separate words.
- If you unintentionally exit the typing state, we will move on to the next task.
- In this experiment we don’t use predictive typing.

4.3.7 Measurement

The 80% correct words transcribed in 30 seconds was chosen to performance variable with the idea that 80% correct words would still be mostly understandable and because when blocking the vision, 100% correct typing is not realistic. Furthermore, analogous results were obtained with alternative variables such as 100% correct words, or number of correct letters. The value was calculated by first subtracting the number of letter substitutions, intrusions, and transpositions from each transcription’s length, and then dividing the result by the length of the presented word.

4.4 Results

4.4.1 Absolute Performance

There were two trials conducted in every condition. First, the average was taken from these two performances to get the absolute performance scores. Absolute performance refers here to the actual number of 80% correct words transcribed in 30 seconds. Physical-Qwerty was best in terms of absolute performance, with a mean of 3.42 (95% CI ± 0.51) words on average (Figure 6). It was significantly best among the three interfaces. Touch-Qwerty also performed significantly better with a mean of 2.77 (95% CI ± 0.45) words, than ITU-12 with 2.43 (95% CI ± 0.25) words.

The performance of ITU-12 probably results partly from the design of the input method, as it naturally takes more keystrokes per character than the Qwerty-keyboards because there are three to four letters in every key. The need for normalizing the performance scores results from bias caused by individual differences, but it also compensates for the effect caused by the required keystrokes.

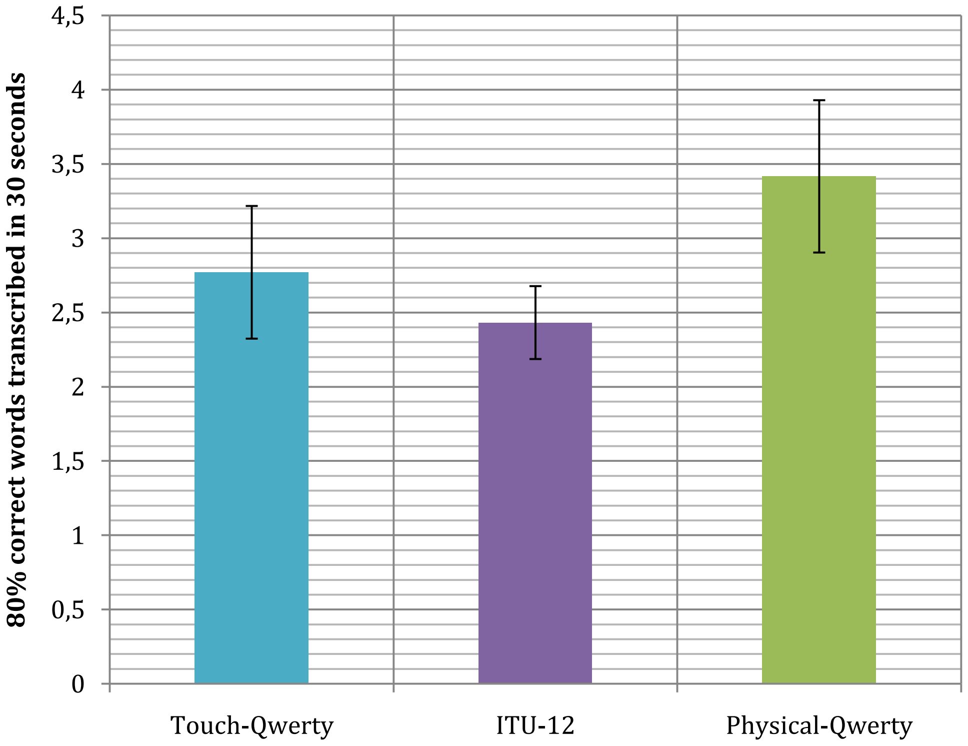

Figure 6. Absolute (the number of 80% correct words transcribed in 30 seconds) performance of interfaces. Vertical bars denote 95% confidence intervals (CIs) calculated from Student's t-distribution.

4.4.2 Multimodal Flexibility Index

A multiple-level, repeated measures ANOVA was ran for the interfaces, showing a significant effect for the multimodal flexibility indices, F(2,22) = 5.885, p < .01 (Figure 7).

ITU-12 performed best in terms of multimodal flexibility. The flexibility index for ITU-12 is better (0.517, 95% CI ± 0.12) than the Qwerty-interfaces' indices. The mean difference of ITU-12 was post-hoc analyzed (with Fisher's LSD) to be significant with Touch-Qwerty (p = .007) and with Physical-Qwerty (p = .041). There was no significant difference between Touch-Qwerty (0.349, 95% CI ± 0.07) and Physical-Qwerty (0.364, 95% CI ± 0.05) with p = .742.

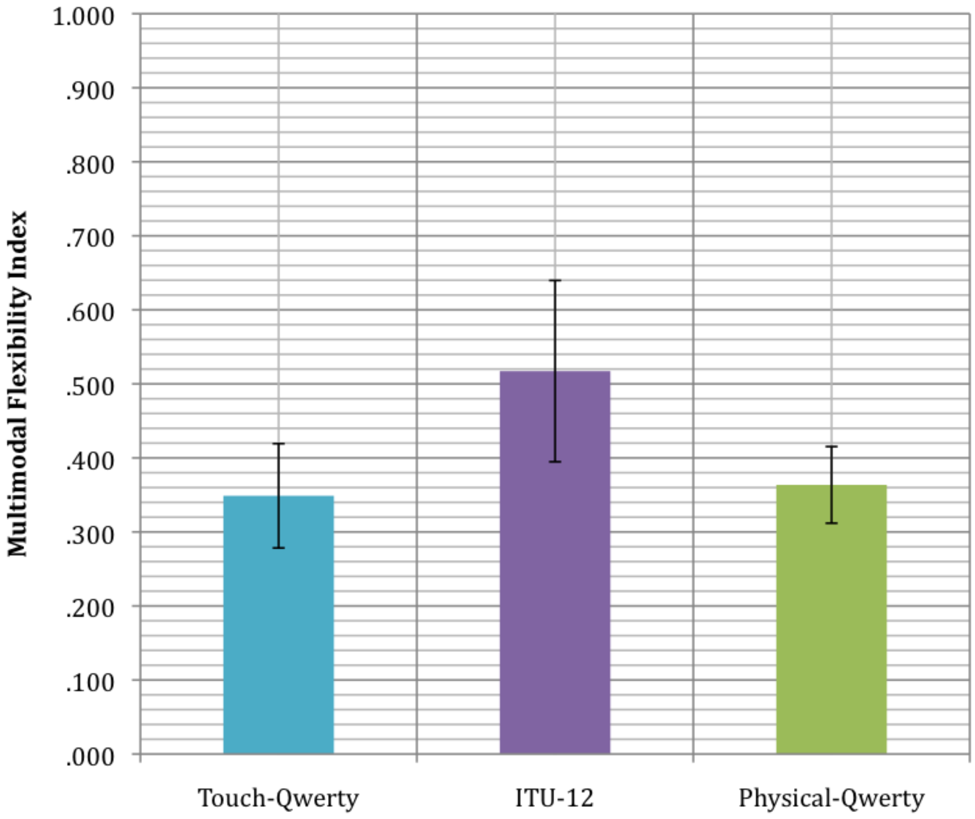

Figure 7. Multimodal Flexibility Indices for the interfaces. Vertical bars denote 95% confidence intervals (CIs).

The results indicate, that ITU-12's performance decreased 48% on average over the modality conditions. The Qwerty keyboards suffered 64-65% on average as a result of modality withdrawals.

4.4.3 Modality Conditions

ITU-12's flexible performance in modality conditions is clearly visible in Figure 8. Compared to the baseline (score 1.0), all other conditions significantly hampered the performance with these interfaces, except "audition-vision" with Touch-Qwerty (0.84, 95% CI ± 0.23), and "vision-tactition" with ITU-12 (1.05, 95% CI ± 0.10) and Physical-Qwerty (0.97, 95% CI ± 0.09).

The only condition where the Qwerty-keyboards performed better than ITU-12, was when audition and vision were free, but tactition blocked. In this condition, performances with Touch-Qwerty and Physical-Qwerty (0.78, 95% CI ± 0.07) were significantly better than ITU-12's performance (0.62, 95% CI ± 0.10). Moreover, ITU-12 performed significantly better than Touch-Qwerty in all other conditions except "audition-vision". In addition, this was the only condition where Touch-Qwerty was better than Physical-Qwerty, however, the difference is not significant. Physical-Qwerty was significantly better than Touch-Qwerty when all the modalities were blocked from interaction, the Touch-Qwerty performed 0.00 (95% CI ± 0.00) whereas Physical-Qwerty scored 0.04 (95% CI ± 0.03). When vision was blocked, ITU-12 always performed significantly better than either Qwerty, which were equally and devastatingly hampered (more than 95% decrease in performance) by the withdrawal of vision.

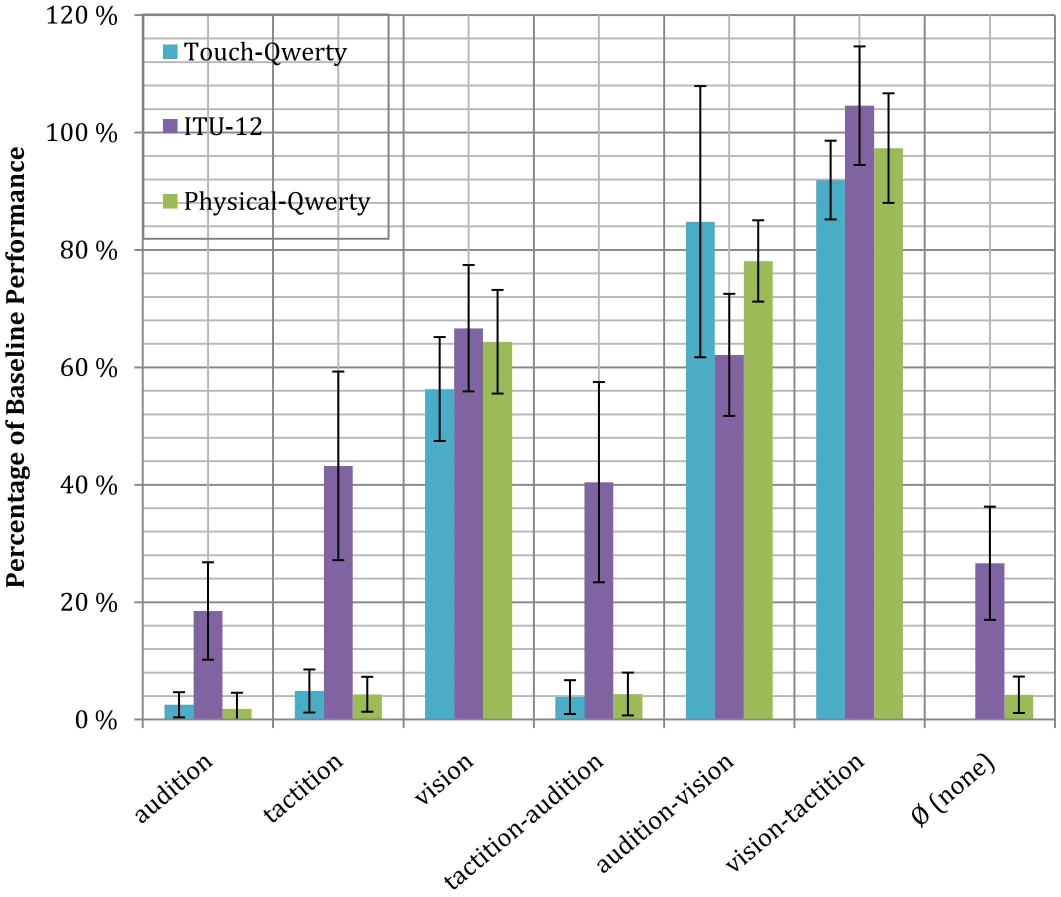

Figure 8. Normalized performance scores in modality conditions. Vertical bars denote 95% confidence intervals (CIs) calculated from Student's t-distribution.

4.4.4 Modality Dependence

Modality dependencies were calculated according to Equation 2. The performances of the Qwerty-keyboards were significantly hampered by withdrawal of vision from the interaction (Figure 9). Their vision dependence was quite similar, the Touch-Qwerty having dependence value 0.80 (95% CI ± 0.27) and the Physical-Qwerty 0.81 (95% CI ± 0.26). ITU-12 was significantly less dependent on vision, the performance decreasing approximately to 51% (95% CI ± 17%) of that in the baseline.

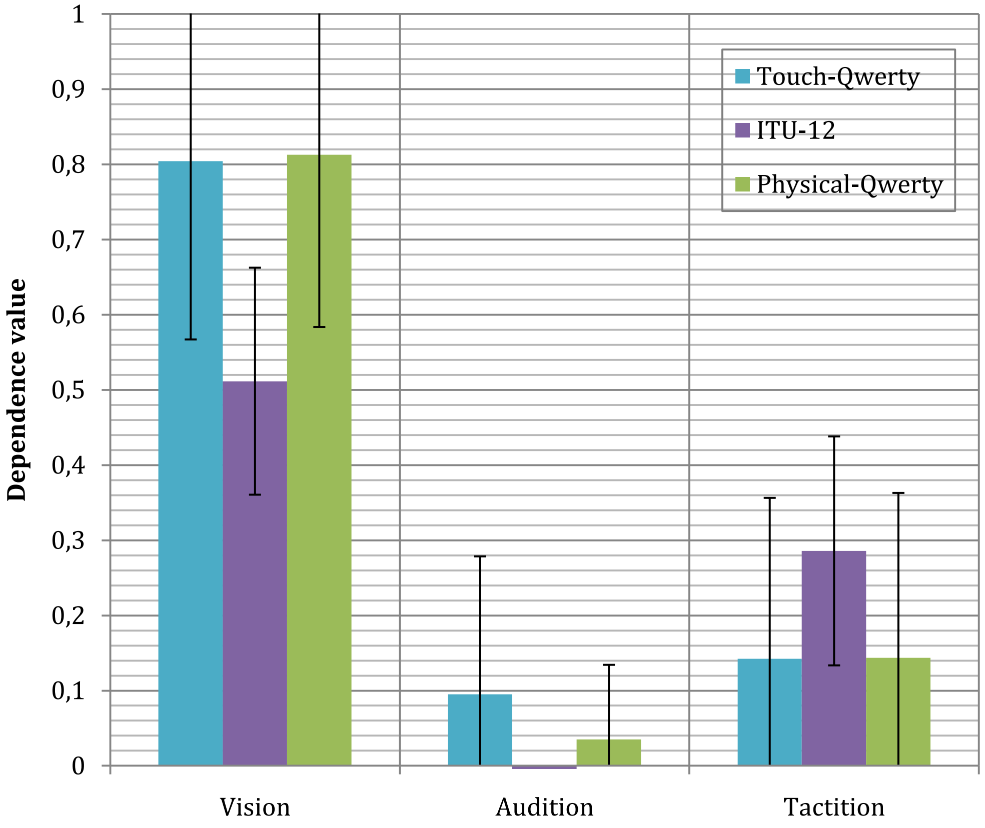

Figure 9. Modality dependence values for the interfaces. Vertical bars denote 95% confidence intervals (CIs) calculated from Student's t-distribution.

4.4.5 Bimodality Indices

The synergy of modalities was examined averaging the normalized performance scores in modality conditions over all interfaces (Table 2). The mean performance score in the "vision-tactition" condition is better than the sum of the scores in the unimodal "vision" and "tactition" conditions (vt = 0.98 > 0.80 = v+t). This indicates, that vision and tactition work in synergy for the text input task with the tested interfaces. Similar, although smaller, effect exists with vision and audition. Vision and audition work at least complementary and additively for each other (av = 0.75 ≈ 0.70 = a+v), if not in synergy. However, audition seems to be almost distractive when utilized simultaneously with tactition to these interfaces (ta = 0.16 < 0.17 = t), or tactition is strongly dominant over audition (ta = 0.16 ≈ 0.17 = t > a)

Table 2. Normalized performance scores in the modality conditions for the three interfaces.

Condition	Touch-Qwerty	ITU-12	Physical-Qwerty	Average
audition	0.025	0.185	0.019	0.076
tactition	0.049	0.432	0.043	0.175
vision	0.563	0.667	0.643	0.624
tactition-audition	0.038	0.404	0.043	0.162
audition-vision	0.848	0.621	0.781	0.750
vision-tactition	0.919	1.046	0.973	0.979
Ø (none)	0.000	0.266	0.042	0.103

4.4.6 Individual Differences

Almost all subjects were currently using a mobile with an ITU-12 keypad. Moreover, eleven subjects had some experience with Physical-Qwerty, so these could not be used as a predictive factors for individual differences. Five subjects had experience with Touch-Qwerty, but their performance did not differ from those who lacked the experience. The only heavy (> 100 text messages per month) user had the best mean MFI with all three interfaces, however the difference was not significant.

Predictive text entry was turned off in the experiment, and four subjects who were not using predictive text in their mobile phones performed significantly better in terms of their personal mean MFI = 0.45 (student's t 95% CI ± 0.10) compared to those using it (MFI = 0.39, 95% CI ± 0.06).

5 Conclusion and Discussion

5.1 Research Questions

The primary goal of the study was to determine the *multimodal flexibility of the three common mobile text input interfaces.* Results imply alphabetical 12-digit keypad to be multimodally the most flexible of the three compared interfaces. Although the 12-digit keypad was slowest to type when all modalities are free to be allocated to the interaction (baseline condition), it was the most flexible in performing under constraints that the real world might set on sensory modalities. The performance of the physical and touch Qwerty-keyboards' did not differ in terms of MFI despite the fact that physical Qwerty performed better than touch in absolute performance (baseline).

The efficiency of modalities was studied to determine the values of single modalities, and contributing the second goal on addressing *how much each modality is utilized with each interface in a text input task.* All the interfaces were shown to be highly dependent on vision in the text input task. The Qwerty-keyboards' performances dropped by more than 95% in conditions where vision was blocked. In addition, the vision-dependence of these interfaces was suggested to be approximately 80%. ITU-12 was least vision dependent, the performance being approximately 50% of that in the baseline. Lack of audition was not affecting the text input performance significantly in any interface. Withdrawal of tactition hampered the performance for 10-30%, but significant differences were not discovered between the interfaces. Furthermore, Qwerty-keyboards were not shown to be significantly dependent on tactition at all, Touch-Qwerty's performance being almost at the baseline level in "audition-vision" condition.

Finally, effectiveness was studied by examining *the cooperation of modalities in a mobile text input task*. Vision was shown to work in synergy with tactition and with audition, suggesting, that the modalities added value to each other when utilized simultaneously in the interaction. However, audition and tactition were not significantly providing extra value working bi-modally, compared to unimodal conditions, as suggested already observing the performances of individual interfaces.

5.2 Validity of the Results

The study was designed carefully applying experimental methods and variables utilized in the field of human-computer interaction. The participants presented a typical mobile phone user group, and there were enough subjects for within-group design applied in the experiment. To eliminate bias, the obtained performance scores were normalized within the subjects. This further ensured control on the affects of the tested variables. The validity and significance of the results was analyzed with appropriate statistical tests.

The multimodal flexibility method [1] utilizes modality “blocking” to measure the effects of withdrawal of a modality. As noted, the blocking is in some cases difficult to conduct purely. Obviously, sensory modalities cannot be fully withdrawn from the interaction. Only way to study total lack of sensory modality would be to use impaired subjects lacking perceptions from sensory modalities. However, it was not an option as this within-subject experiment design compares the performances in modality blocking conditions to the baseline performance where modalities are free to be allocated to the interface. Furthermore, as the design was comparative, all interfaces were subjected to the same conditions and same text input task, allowing within-comparison of MFI, but not between other studies. In other words, the results are comparable only when the same blocking method is used in the same experiment task. However, as text input represents also other interactions, for example target selection speed and accuracy when both the errors and speed of key presses are considered, the results might suggest performance of interfaces also beyond this particular experiment task. Lastly, the cooperation of modalities in the mobile text input task was calculated over the interfaces. The cooperation results would be generalizable to all mobile text input tasks only if all mobile text input interfaces would be included in the study. As so, the results in this study suggest the modalities’ cooperation only with *common* mobile text input interfaces.

5.3 Discussion

This thesis evaluated the multimodal flexibility of three interfaces with a novel approach. Instead of investigating the effects of interactions on each other, the study indicated each interfaces' abilities to let sensory modalities free to be utilized elsewhere. The major advantage of the method is that the results contribute to the real world modality allocations by controlling the modalities (instead of measuring effects of dual- or multi-tasking) and withdrawing the two-way information transmission of modalities (instead of an input or a feedback one) from the interaction. The mobile usability of an interface depends on the amount of information a user is able to retrieve or transmit while on the move. Moreover, the information transmission capacity depends on how flexibly usable the interface is across varying real world contexts. The flexibility of mobile systems is one of the least studied mobile usability attributes [2]. This thesis took a novel approach on multimodal flexibility, conducting an empirical and objective study and delivering valid and general results. Future work on mobile usability in terms of multimodal flexibility can apply a similar method to measure the effect of other user's resources utilized in interaction, such as other interaction modalities or physical resources.

6 References

1. Oulasvirta, A. and Bergstrom-Lehtovirta, J. A simple index for multimodal flexibility. In *Proceedings of the 28th international conference on Human factors in computing systems.* (2010), 1475--1484.

2. Coursaris, C. K. and Kim, D. J. A qualitative review of empirical mobile usability studies. In *Proceedings of the Twelfth Americas Conference on Information Systems.* (2006), 1--14.

3. Oulasvirta, A., Tamminen, S., Roto, V., and Kuorelahti, J. 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), 919--928.

4. Lindenberger, U., Marsiske, M., and Baltes, P. B. Memorizing while walking: Increase in dual-task costs from young adulthood to old age. *Psychology and Aging, 15,* 3 (2000), 417.

5. Pashler, H. and Yantis, S. Stevens' Handbook of Experimental Psychology, Sensation and Perception. Wiley, 2004.

6. Sweller, J. Cognitive load during problem solving: Effects on learning. *Cognitive science, 12,* 2 (1988), 257--285.

7. Sweller, J., Chandler, P., Tierney, P., and Cooper, M. Cognitive load as a factor in the structuring of technical material. *Journal of Experimental Psychology: General, 119,* 2 (1990), 176--192.

8. Sweller, J. Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction, 4,* 4 (1994), 295--312.

9. Wickens, C. D. Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 2,* 3 (2002), 159--177.

10. Wickens, C. D. and Liu, Y. Codes and modalities in multiple resources: A success and a qualification. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 30,* 5 (1988), 599--616.

11. Cassell, J. Speech, Action and Gestures as Context for Ongoing Task-oriented Talk. In *AAAI Fall Symposium Working Notes: Embodied Language.* (1995).

12. Allwood, J. Cooperation and flexibility in multimodal communication. In *Cooperative Multimodal Communication.* (2001), 113--124.

13. Tindall-Ford, S., Chandler, P., and Sweller, J. When Two Sensory Modes Are Better Than One. *Journal of experimental psychology: Applied, 3,* 4 (1997), 257--287.

14. Mayer, R. E. and Anderson, R. B. The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of educational Psychology, 84* (1992), 444--444.

15. Mayer, R. E. and Sims, V. K. For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of educational psychology, 86* (1994), 389--389.

16. Blake, R., Sobel, K. V., and James, T. W. Neural synergy between kinetic vision and touch. *Psychological Science, 15,* 6 (2004), 397.

17. McGurk, H. and MacDonald, J. Hearing lips and seeing voices (1976).

18. Howard I. P. Human Spatial Orientation. *John Wiley & Sons Ltd; First Edition* (1966).

19. Klatzky, R., Lederman, S., and Matula, D. Haptic exploration in the presence of vision. *Journal of Experimental Psychology, 19,* 4 (1993), 726--743.

20. Rock, I. and Victor, J. Vision and touch: An experimentally created conflict between the two senses.. *Science (New York, NY), 143* (1964), 594.

21. Botvinick, M. and Cohen, J. Rubber hands' feel'touch that eyes see. *Nature, 391,* 6669 (1998), 756--756.

22. Zangaladze, A., Epstein, C. M., Grafton, S. T., and Sathian, K. Involvement of visual cortex in tactile discrimination of orientation. *Nature, 401,* 6753 (1999), 587--590.

23. Posner, M. *Chronometric Explorations of Mind*. Erlbaum (1978), Hillsdale, NJ.

24. Duncan, S. On the structure of speaker-auditor interaction during speaking turns. *Language in society, 3,* 02 (1974), 161--180.

25. Duncan, S. and others On signalling that it's your turn to speak. *Journal of Experimental Social Psychology, 10,* 3 (1974), 234--247.

26. Perakakis, M. and Potamianos, A. Multimodal system evaluation using modality efficiency and synergy metrics. In *Proceedings of the 10th international conference on Multimodal interfaces.* (2008), 9--16.

27. Brewster, S. A., Wright, P. C., and Edwards, A. D. N. The design and evaluation of an auditory-enhanced scrollbar. In *Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence.* (1994), 173--179.

28. Oakley, I., Brewster, S., and Gray, P. Solving multi-target haptic problems in menu interaction. In *CHI'01 extended abstracts on Human factors in computing systems.* (2001), 357--358.

29. Prewett, M. S., Yang, L., Stilson, F. R. B., Gray, A. A., Coovert, M. D., Burke, J., Redden, E., and Elliot, L. R. The benefits of multimodal information: a meta-analysis comparing visual and visual-tactile feedback. In *Proceedings of the 8th international conference on Multimodal interfaces.* (2006), 333--338.

30. Jacko, J. A., Barnard, L., Kongnakorn, T., Moloney, K. P., Edwards, P. J., Emery, V. K., and Sainfort, F. Isolating the effects of visual impairment: exploring the effect of AMD on the utility of multimodal feedback. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* (2004), 311--318.

31. Nigay, L. and Coutaz, J. A design space for multimodal systems: concurrent processing and data fusion. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems.* (1993), 172--178.

32. Oviatt, S., Coulston, R., and Lunsford, R. When do we interact multimodally?: cognitive load and multimodal communication patterns. In *Proceedings of the 6th international conference on Multimodal interfaces.* (2004), 129--136.

33. Oviatt, S. Ten myths of multimodal interaction. *Communications of the ACM, 42*, 11 (1999), 74--81.

34. Hoggan, E. and Brewster, S. A. Crosstrainer: testing the use of multimodal interfaces in situ. In *Proceedings of the 28th international conference on Human factors in computing systems.* (2010), 333--342.

35. Siewiorek, D., Smailagic, A., Furukawa, J., Krause, A., Moraveji, N., Reiger, K., Shaffer, J., and Wong, F. L. Sensay: A context-aware mobile phone. In *Wearable Computers, 2003. Proceedings. Seventh IEEE International Symposium on.* (2003), 248--249.

36. Brewster, S., Lumsden, J., Bell, M., Hall, M., and Tasker, S. Multimodal'eyes-free'interaction techniques for wearable devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* (2003), 473--480.

37. Marentakis, G. and Brewster, S. A. A study on gestural interaction with a 3d audio display. *Mobile Human-Computer Interaction--MobileHCI 2004* (2004), 529--529.

38. Oviatt, S. Multimodal interactive maps: designing for human performance. *Human-Computer Interaction, 12*, 1 (1997), 93--129.

39. Mizobuchi, S., Chignell, M., and Newton, D. Mobile text entry: relationship between walking speed and text input task difficulty. In *Proceedings of the 7th*

international conference on Human computer interaction with mobile devices \& services. (2005), 122--128.

40. Lyons, K., Starner, T., Plaisted, D., Fusia, J., Lyons, A., Drew, A., and Looney, E. Twiddler typing: One-handed chording text entry for mobile phones. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* (2004), 678.

41. MacKenzie, I. S. The one-key challenge: searching for a fast one-key text entry method. In *Proceeding of the eleventh international ACM SIGACCESS conference on Computers and accessibility.* (2009), 91--98.

42. Mackenzie, I. S. and Felzer, T. SAK: Scanning ambiguous keyboard for efficient one-key text entry. *ACM Transactions on Computer-Human Interaction (TOCHI), 17,* 3 (2010), 1--39.

43. Ryu, H. and Cruz, K. LetterEase: Improving text entry on a handheld device via letter reassignment. In Proceedings of the 17th Australia conference on Computer-Human Interaction: Citizens Online: Considerations for Today and the Future. (2005), 1--10.

44. Hoggan, E. E. and Brewster, S. A. Crossmodal icons for information display. In *CHI'06 extended abstracts on Human factors in computing systems.* (2006), 857--862.

45. Hoggan, E. and Brewster, S. Mobile crossmodal auditory and tactile displays. In *Proceedings of HAID 2006: First International Workshop on Haptic and Audio Interaction Design.* (2006), 9--12.

46. Hoggan, E., Kaaresoja, T., Laitinen, P., and Brewster, S. Crossmodal congruence: the look, feel and sound of touchscreen widgets. In *Proceedings of the 10th international conference on Multimodal interfaces.* (2008), 157--164.

47. Hoggan, E. and Brewster, S. Designing audio and tactile crossmodal icons for mobile devices. In *Proceedings of the 9th international conference on Multimodal interfaces.* (2007), 162--169.

48. Hoggan, E., Brewster, S. A., and Johnston, J. Investigating the effectiveness of tactile feedback for mobile touchscreens. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems.* (2008), 1573--1582.

49. Vazquez-Alvarez, Y. and Brewster, S.A. *Audio Minimization: Applying 3D Audio Techniques to Multi-Stream Audio Interfaces*. In Proceedings of HAID 2009 (Dresden, DE), Springer LNCS Vol 5763.

50. Schmidt, A., Aidoo, K., Takaluoma, A., Tuomela, U., Van Laerhoven, K., and Van de Velde, W. Advanced interaction in context. In *HandHeld and Ubiquitous Computing.* (1999), 89--101.

51. Hoggan, E., Crossan, A., Brewster, S. A., and Kaaresoja, T. Audio or tactile feedback: which modality when. In *in ACM CHI.* (2009).

52. McGookin, D., Brewster, S., and Jiang, W. W. Investigating touchscreen accessibility for people with visual impairments. In *Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges.* (2008), 298--307.

53. Crossan, A., Brewster, S., and Ng, A. Foot Tapping for Mobile Interaction. In *Proceedings of BCS HCI*, (2010)

54. Crossan, A., McGill, M., Brewster, S., and Murray-Smith, R. Head tilting for interaction in mobile contexts. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services.* (2009), 1--10.

55. Rico, J. and Brewster, S. Gestures all around us: user differences in social acceptability perceptions of gesture based interfaces. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services.* (2009), 1--2.

56. Rico, J. and Brewster, S. Gesture and voice prototyping for early evaluations of social acceptability in multimodal interfaces. In *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction.* (2010), 16.

57. Crossan, A., Williamson, J., Brewster, S., and Murray-Smith, R. Wrist rotation for interaction in mobile contexts. In *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services.* (2008), 435--438.

58. Rico, J. and Brewster, S. Usable gestures for mobile interfaces: evaluating social acceptability. In *Proc. CHI.* (2010).

59. Crossan, A., Murray-Smith, R., Brewster, S., Kelly, J., and Musizza, B. Gait phase effects in mobile interaction. In *CHI'05 extended abstracts on Human factors in computing systems.* (2005), 1312--1315.

60. Kane, S. K., Wobbrock, J. O., and Smith, I. E. 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.* (2008), 109--118.

61. Mustonen, T., Olkkonen, M., and Hakkinen, J. Examining mobile phone text legibility while walking. In *CHI'04 extended abstracts on Human factors in computing systems.* (2004), 1243--1246.

62. Kujala, T. Efficiency of visual time-sharing behavior: the effects of menu structure on POI search tasks while driving. In *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications.* (2009), 63--70.

63. Bernsen, N. O. and Dybkjaer, L. Exploring natural interaction in the car. In CLASS Workshop on Natural Interactivity and Intelligent Interactive Information Representation. (2001), 75--79.

64. Lumsden, J., Durling, S., and Kondratova, I. A Comparison of Microphone and Speech Recognition Engine Efficacy for Mobile Data Entry. In *The International Workshop on MObile and NEtworking Technologies for social applications (MONET'2008), part of the LNCS OnTheMove (OTM) Federated Conferences and Workshops.* (2010).

65. Lemmelä, S. and Vetek, A. and Mäkelä, K., and Trendafilov, D. Designing and evaluating multimodal interaction for mobile contexts. *In Proceedings of the 10th international conference on Multimodal interfaces,* (2008), 265--272

66. Lazar, J., Feng, J. H., and Hochheiser, H. *Research methods in human-computer interaction*. Wiley, 2010.

67. Salthouse, T. A. Perceptual, cognitive, and motoric aspects of transcription typing. *Psychological bulletin, 99,* 3 (1986), 303--319.

68. Lewis, J., Potosnak, K., and Magyar, R. Keys and Keyboards. Handbook of Human-Computer Interaction. M. Helander, T. Landauer and P. Prabhu. *Amsterdam, North-Holland, 1285* (1997), 1316.

69. Silfverberg, M., MacKenzie, I. S., and Korhonen, P. Predicting text entry speed on mobile phones. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* (2000), 9--16.

70. Green, N., Kruger, J., Faldu, C., and St Amant, R. A reduced QWERTY keyboard for mobile text entry. In *CHI'04 extended abstracts on Human factors in computing systems.* (2004), 1429--1432.

71. James, C. L. and Reischel, K. M. Text input for mobile devices: comparing model prediction to actual performance. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* (2001), 365--371.

72. MacKenzie, I. S. and Soukoreff, R. W. Phrase sets for evaluating text entry techniques. In *CHI'03 extended abstracts on Human factors in computing systems.* (2003), 754--755.

73. Isokoski, P. and Linden, T. Effect of foreign language on text transcription performance: Finns writing English. In *Proceedings of the third Nordic conference on Human-computer interaction.* (2004), 109--112.