A Participatory Design and Formal Study Investigation into Mobile Text Entry for Older Adults

Emma Nicol, University of Strathclyde, Glasgow, UK

Andreas Komninos, University of Strathclyde, Glasgow, UK

Mark D Dunlop, University of Strathclyde, Glasgow, UK

ABSTRACT

Text entry remains key to many tasks on touchscreen smartphones and is an important factor in the usability of such devices. The known problems of text entry can be particularly acute for older adults due to physical and cognitive issues associated with ageing. In a study of mobile text entry the authors employed a variety of participatory design and formal comparative study techniques in order to explore the requirements of this group of users and to discover the key differences in texting activity between them and younger users of mobile devices. They report on the findings of a lab study of texting behaviour of older adults. The authors' findings indicate differences in attitudes to texting styles and tasks between older and younger adults. They also identify some differences in typing behaviour and reflect on methods.

KEYWORDS

Older Adults, Participatory Design, Qualitative and Quantitative Studies, Text Entry

INTRODUCTION

The UK, in line with most of the EU and developed world, has an ageing population: in 1985 15% of the population was over 65 years old. This rose to 17% by 2010 with a predicted growth to 23% by 2035 (Office for National Statistics, 2012). In recent years, mobile technologies have had a massive impact on working life, for example it is estimated that over 25% of emails are now opened on mobile devices (Knotice, 2012). As the older working population rises, due to both ageing population demographics and increasing retirement age, a growing number of digital economy workers will require to continue to use mobile technologies for work into their mid/late 60s. Furthermore, many people will want to continue professional, social and lifestyle usage into their late retirement as the

technologies can support increased community involvement and personal independence. Unfortunately, the ageing process can interfere considerably with mobile technology usage. The normal ageing process typically involves a decline in visual and auditory abilities together with a decline in working memory, selective attention, and motor control (Fisk, Rogers, Charness, Czaja, & Sharit, 2012). For example, many people in their 40s start to have vision changes that affect their near focus while, as a rule-of-thumb, older people require 1.5 to 2.0 times longer to perform a motor action compared to younger people (Fisk et al., 2012).

Text entry is core to mobile interactions such as emailing, social networking, instant messaging and interacting with services such as web or map searching. The majority of smartphones now do not have any physical keyboard but rely on on-screen touch keyboards. These have been shown to be slower and more error-prone than traditional mini-physical keyboards in studies such as Hoggan, Brewster, & Johnston (2008), and Lee & Zhai (2009a), but are popular as they allow for larger mobile screens. While there have been numerous studies into text entry usage on touchscreens, there has been very little work studying the effects of ageing on text entry, particularly on modern touchscreen phones where reduced visual acuity, reduced motor control and reduced working memory are all likely to have an impact. The most notable work is that of Nicolau & Jorge (2012) who examined the role of hand tremor in input by older adults on touchscreen phones and tablets. Younger users have been shown to be faster in text entry studies since the early days of personal digital assistant devices (Wright et al., 2000). The main reasons for this are the effects of age on both our vision and on the accuracy and speed of movement. In an example checklist for design of push buttons for a stereo, (Fisk et al., 2012) included the following key recommendations when designing for older (65+) users:

- To accommodate older users with tremor and arthritic conditions, avoid combinations of low force (< 0.35 N) and low travel (< 0.2 mm);
- Ensure that there is adequate spacing between buttons to prevent inadvertent activation of controls;
- Ensure that surfaces have sufficient frictional resistance to prevent finger sliding;
- Ensure that there is proper illumination for activation under conditions of dim lighting.

Smartphone touchscreen keyboards provide an extreme counter-example of these principles:

- Soft keys require almost zero force to activate and have zero travel;
- Soft keyboards typically have no gap between keys (any visible gap is ignored by the underlying software that uses predictive technology to guess which key was intended);
- Surfaces are high gloss low resistance glass;
- Keyboards are often reduced contrast to aid styling with strong sunlight and high reflection levels further reducing contrast.

Although mobiles provide great support for communication and engagement across all ages, the natural ageing processes can make smartphone interaction particularly challenging for older adults.

BACKGROUND

While the mobile industry is currently focussed on targeting younger users, our aim is to develop innovative keyboard designs that are easier to use and less error-prone for older adults. During a 2-year investigation of mobile text entry for older adults we aimed to identify key design criteria for touchscreen keyboards with a view to building and evaluating prototypes of these and to quantitatively

measure the touchscreen behaviour of older adults and compare with that of younger adults. This paper reports some of our findings in these related areas of enquiry. When considering how to investigate text entry with older adults, we found from the literature that so far a complete methodology for use with this group has yet to be proposed or evaluated by researchers. In our studies we made use of a number of research techniques recommended for participatory design and we report on our experiences of using those here.

In an early lab study (Dunlop et al., 2014), we investigated the behaviour of older and younger adults and compared their text entry on a QWERTY keyboard with a half-qwerty keyboard prototype that employed a shake gesture to flick between sides of the keyboard. This prototype keyboard attempted to address the problem of small key sizes while maintaining an unambiguous keyboard. In studying this and other prototype keyboards, we identified some differences in the behaviours and attitudes of older and younger participants: Older users made fewer mistakes using the shake keyboard than the younger users did and were more willing to adopt the new keyboard layouts. These preliminary studies encouraged us to further investigate input behaviour by older users.

PARTICIPANTS

Taking our lead from that initial study, we began a participatory design process involving older adults recruited from the Centre for Lifelong Learning (CLL) at the University of Strathclyde with a view to generating ideas for novel interfaces and text entry techniques. Over the course of the project we worked with around 35 older adult participants in various groups and configurations, the typical age being 65, with many of our participants returning multiple times to take part in later stages of the research, dependent on their availability. We also conducted parallel research at times with cohorts of younger people, typically aged in their early to mid 20s, recruited mostly from our undergraduate and postgraduate body, all current users of smartphones. We first recruited for people aged 50+ chiefly by emails to the student body of the CLL and by putting up posters around the centre. A similar process occurred for subsequent studies with further recruitment also occurring by participants inviting friends, neighbours and colleagues to take part. We specifically asked for people who had some experience of using touchscreen smartphones, e-readers, tablets or any similar device. Every study conducted had been first approved by an ethics committee and participants were asked to give their consent to take part. Participants were thanked for their participation with shopping vouchers and refreshments at the research sessions and we held a special evening event near the close of the project to thank them for their participation and to share some of our findings with them. None of the participants had significant motor or visual impairments and all can be described as mobile, active people. In what follows we outline briefly the techniques used and the insights that our participatory design sessions with these participants allowed us to uncover, before moving on to describe a lab study in which we investigated and compared the text entry habits of older and younger adults.

PARTICIPATORY DESIGN SESSIONS

Mapping Exercise

Our first research session involved 13 participants aged 53-77yrs (average age= 64.8, sd=5.4, 7f, 6m). This session was designed to help us to begin to explore the context of text entry and mobile use with our participants and to begin to get a sense of their experiences of and attitudes to using mobile devices. We envisaged a lengthy participatory design process and thus we did not begin immediately with the specifics of keyboard layout. The session was planned for us to get to know better the needs and desires of the people with whom we were working and also as an icebreaker, for

them to get to know us and each other better. A short introductory questionnaire established that 10 of the 13 participants were daily users of touchscreen smartphones, 2 used Blackberries and one used a feature-phone. 8 participants, including the 3 using Blackberry and feature phones, reported daily use of tablets, and 6 used e-readers on a daily basis. All of participants met our recruitment criteria, with future sessions having a similar experience and mobile usage profile.

Working in groups of 4, participants were given a large map of a fictional town with images of locations where they might use text entry. Using stickers colour-coded for categories of mobile device activity such as emailing or web searching, groups identified places where they would enter text and annotated the map accordingly (see Figure 1 and Figure 2). Groups then gave a presentation on how they had made their decisions, which provided further rich data about their attitudes and experiences of using their mobile devices. The mapping exercise was successful in getting participants to talk about their experiences with and attitudes to mobile devices but it also highlighted the importance of good task description, agreeing terminology and keeping participants focused, as per Coleman, et al. (2010), Fisk, et al. (2009) and Lee & Zhai (2009b): all 3 groups interpreted the task rather differently making the results interesting but not directly comparable.

Topics that emerged from this session included the appropriateness and inappropriateness of locations for certain tasks, e.g. some participants felt it would not be appropriate under any circumstance to text from a theatre regardless of curtain up or curtain down and some felt that making calls while using public transport could never be justified. Some were bemused at the need for anyone to be emailing while on the move while others were frequent users of social media and other text-based applications in that situation. Appropriateness was an issue we had not anticipated nor specifically asked about, however it came to dominate much of the emerging discussion. A number of interesting novel uses of mobiles also emerged e.g. using a tablet during hospital visits to make notes of questions to ask before a consultation, then typing the answers or further notes when with the doctor or immediately following the appointment.

There were clear indications that our participants were generally making flexible use of multiple devices in different locations and contexts and often in ways that confounded the unfair stereotypes that abound, regarding the use of technology by older people. We believe this exercise was largely

Figure 1. A group of participants annotates a map

Figure 2. An annotated map. Stickers indicate the locations where members of this group decided they thought each of the colour-coded activities would be appropriate. The handwritten post-it notes represent locations they added by themselves.

successful but would recommend that future uses of such a task should start with careful task description and include discussion with the participants that allows opportunities for questions and clarification before the task commences.

Cultural Probes

At the close of the first session the 13 participants were given cultural probe packs (Gaver, Dunne, & Pacenti, 1999) to take home containing postcards to probe their mobile use and text entry activities and a disposable camera to take pictures related to their texting activities. In a later session, participants were asked to share their postcards and photos and, in groups, to present their most common and pressing issues. The postcards were successful in uncovering several mobile and texting issues. Many commented that on-screen keyboard and other buttons were too small to hit accurately and that screens were extremely sensitive. We got the impression that they had many problems with both text entry and accidentally tapping keyboard and other buttons (e.g. home, back and send buttons). The cameras were less successful as a research tool as many photos were of poor quality (particularly when attempting to capture screen images), and participants were less keen to share their photos than postcards. Participants were happy to present their issues to the others and some interesting dialogue developed as a result. Issues arising from these discussions included a frustration over the lack of support: in the absence of manuals, users struggled to understand their phones' capabilities and they were reluctant to use discussion forums etc. rather than manufacturers' publications. The cultural probes method used in this context was an effective way for us to find out about how our participants used their mobile devices with some keyboard related issues already beginning to emerge strongly.

Observational Studies

In a small group setting we asked 6 participants aged 58-70 years (5m, 1f) to complete a number of text entry tasks on their own mobile devices while we videoed and logged their interaction. We also asked them to try four different novel keyboards and gained feedback in short semi-structured interviews. The observational studies were effective in highlighting issues related to text entry itself: higher error rates with the first word in any message and the tendency of users to enter a "review mode" of what they had typed before committing/sending. Semi-structured interviews revealed an openness to trying new keyboard layouts with some showing a reluctance to rely on prediction. This style of session in which participants tried a range of keyboards also acted as an "eye opener" to design of soft input methods – many had not considered that the possibilities for design were so varied, as mainstream solutions are very similar to each other. Indeed after first exposure in this session a number of participants have begun, without prompting, to use gesture typing as their normal input method – a method now widely available on Android and iOS developed from the original work of Kristensson & Zhai (2004).

Metaphor and Paper Prototyping

Inspired by earlier comments about awareness of errors and review mode, we ran small group sessions to discuss the topic of error correction on mobile devices. Following the recommendations of Göllner et al. (2010), for this discussion we used a metaphor to establish a common language with participants. We used the cinematic term "blooper", to avoid the potentially blame-assignment language of "mistakes" or "errors". After introduction to the topic, participants initially described typical things that went wrong for them when entering text using post-it notes in the lab wall. We then demonstrated different approaches to error correction to motivate and gain feedback on attitudes to error correction (Traditional desktop style modal review dialogue, modern desktop red underline approach and a mobile solution using the word suggestion bar). Following on from these exercises, in a later session we asked groups of participants (16 participants working in groups of 4 or 5: 54-72yrs 11f, 5m) to draw ideal designs on paper mobile phone templates (e.g. Figure 3). A representative from each group presented the group's ideas to the others. Using metaphor and paper techniques to investigate error correction resulted in some rich feedback about participants' preferences for solutions to deal with this and was a good lead in to the prototyping exercise. The prototyping tasks resulted in detailed hand-drawn and labelled prototypes (e.g. Figure 3) for which participants were able to explain the design rationale to the others. Auto-correction approaches were viewed with concern, which is in-line with the views we had earlier encountered regarding predictive texting, while identification of mistakes for later review was widely suggested in the paper prototypes. We found that the metaphor approach as proposed by Göllner et al. (2010) was very helpful in establishing common language with the participants and that the selection of the neutral term "blooper" kept participants from placing blame on the software or themselves, allowing them to focus instead on designing solutions instead of discussing the causes of input errors.

LAB STUDIES AND LOGGING

One of the original aims of our research was to quantitatively measure the touchscreen behaviour of older adults and compare it with that of younger adults. Based on our experience of using logging in other user studies, we surmised that the capture of detailed key press data from both groups would be key to understanding this behaviour better. Over several months we conducted a series of individual lab studies and a longitudinal study using a keyboard designed to log tap events. In lab studies, users

Figure 3. Participant drawn paper prototypes of error correction solutions
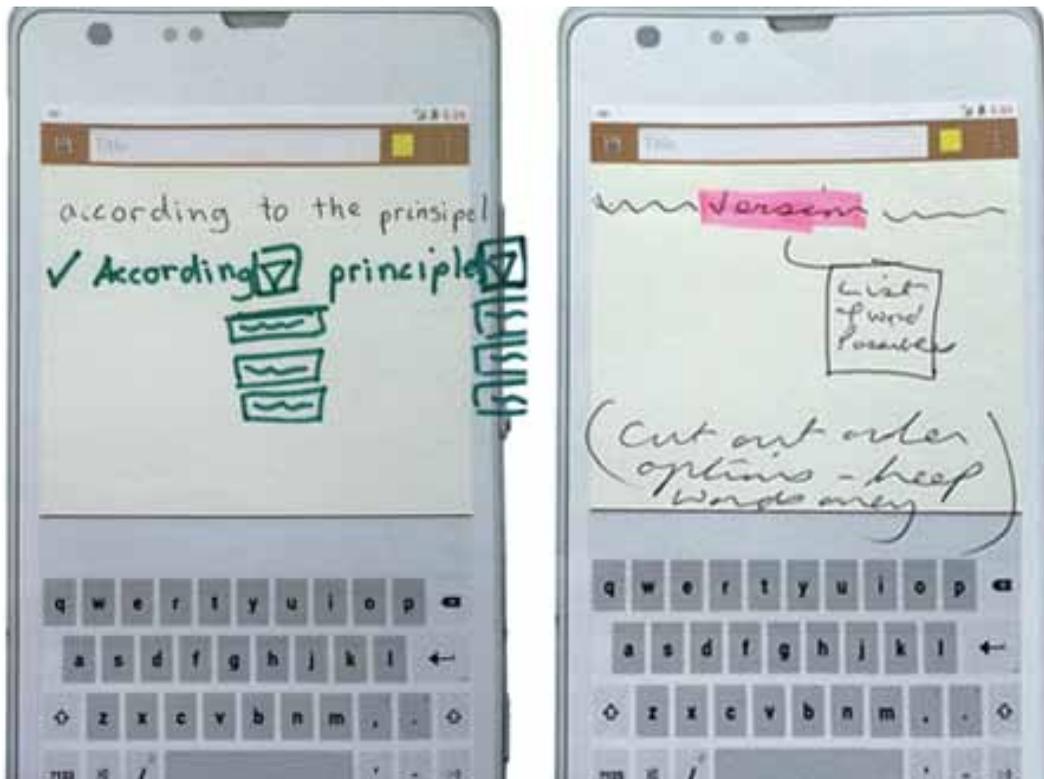
performed a fixed set of tasks of varying types (from copying short phrases to describing images in a fixed time limit), while our longitudinal studies recorded text entry interaction behaviour over 2 weeks. In what follows we explain our study design and the motivation for the use of particular tasks for our investigation.

Methodology

The design of formal text entry studies is primarily concerned with internal validity and external validity with the text entry community widely adopting adopted a standard approach to studies in which users are asked to transcribe a fixed set of phrases. While suggesting a standardized phrase set, MacKenzie & Soukoreff (2003a) focused on internal validity and stated that "Internal validity is attained if the effects observed are attributable to controlled variables". External validity means the results are generalizable to other subjects and situations. Kristensson & Vertanen (2012) also highlight reproducibility of studies and study-heterogeneity (cross-study comparison) as valuable aspects for user studies. They also state that "a phrase set that is not representative of the text that end-users are likely to write may potentially generate misleading results".

In controlled studies, the time it takes users to enter the phrases along with the number of errors made are used as metrics to compare text entry methods within the study (internal consistency). To ensure study heterogeneity and to allow for comparison across studies, standard phrase sets are now widely used, the two most widespread being the original standard phrase set (MacKenzie &

Soukoreff, 2003b) that is composed of 500 short phrases (e.g *have a good weekend*), and the Enron Mobile (Vertanen & Kristensson, 2011) collection of phrases that were written on mobiles (e.g. from the memorable set, *Can you help me here?*). While these two phrase sets now dominate text entry studies, there are various specific collections such as an SMS corpus (Chen & Kan, 2012), a child-oriented corpus (Kano, Read, & Dix, 2006) and one developed by detailed sampling of phrases from collections to ensure that they are representative of that collection (e.g. e-mail or Facebook) (Paek & Hsu, 2011).

While the approach of fixed phrase transcription gives strong internal consistency, reproducibility and heterogeneity advantages, the scenario of copying phrases and typing them is clearly not representative of most mobile text entry. The alternative to copy tasks is to ask users to generate text in composition tasks. In their work on speech recognition Karat et al. (1999) included a comparison in both speech and desktop typing, of copying sections of a novel to composing shorter, but detailed, replies to scenarios. They found that composition typing speed was 58% of that of copy typing. Vertanen & Kristensson (2014) investigated complementing copy tasks with composition tasks. Through a set of experiments emulating mobile entry style, they conducted a detailed comparison of copy tasks with different composition tasks by asking users to (a) reply to a message, (b) compose a message without scenario prompting and (c) compose with scenario prompting. They showed that composition tasks had an entry rate of 65-85% of the copy tasks depending on task type and that typed responses varied in length between 55% and 135% of copy tasks. They conclude that "providing participants with a simple instruction of creating a short message in the domain of interest was successful in getting participants to quickly invent and compose text. It does not appear necessary to provide participants with a specific situation or message in order to help them invent a message." However, the produced phrases were approximately the same length as copy tasks and still relatively short.

While the focus on much text entry experimentation is on speed of entry, usually reported as words-per-minute (WPM), accuracy of entry is also important. In experimental conditions users are normally asked to type "quickly but accurately" and this typically results in very low error rates in final submissions. However it is still prudent to check error rates. For transcription tasks, users are expected to copy a phrase exactly. As such, edit distance can be used a measure of accuracy of the final phrase (Soukoreff & MacKenzie, 2001) (Levenshtein, 1966). No comparative text exists for composition-based tasks, thus error rates have to be determined either by manual inspection either by the researchers or by crowd sourcing (Vertanen & Kristensson, 2014), or estimated by comparing out-of-dictionary word usage.

Logging Keyboard

We developed a keyboard to log a user's key presses and other touch interactions with mobile devices, working within a standard QWERTY layout based on Android source code. Figure 4 shows how our keyboard looked when installed on an Android device. The keyboard was designed to record all taps but did not log these sequentially or allow character streams to be regenerated. This allowed us to investigate considerable detail of each user's tapping behaviour while at the same time maintaining a degree of privacy and anonymity for the participant, a particular concern in longitudinal studies. The keyboard is designed to log each keystroke on the keyboard as a character, along with details of the tap itself (timing information, coordinates and other contact point data). As a heuristic we also tag keystrokes as "suspect" if they are later corrected. This simple algorithm monitors backspace use and identifies as suspect the last backspace-removed character from a sequence (e.g. if a user types *saisage*, followed by backspacing all characters up to and including the *i,* and then corrects to *sausage* we record *i* as a suspect corrected to *u*).

Figure 4. The logging keyboard interface
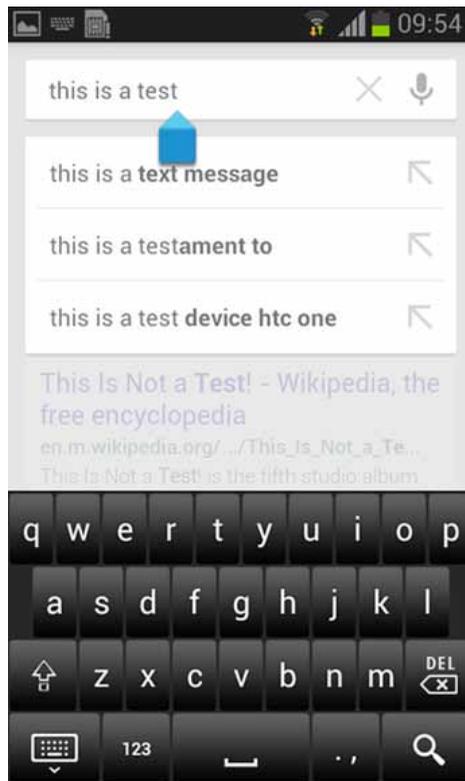
Study Design

As with the participatory design sessions, older adult participants for the lab studies were recruited in large part via the University of Strathclyde Centre for Lifelong Learning with younger adults recruited largely from our student body. For these studies we specifically recruited users of Android devices. Participants were invited to attend our offices and performed the tasks using our Text Entry Experiment System (TEES) client, running on Motorola MotoG Android 4.5" smartphones. TEES is a customisable system for conducting text input studies, which presents users with a configurable set of tasks and performs input performance logging and it is discussed later in the paper. The majority of research sessions were completed on an individual basis with a researcher present, though there were a couple of occasions when participants completed their tasks with other participants present. On those occasions, efforts were made to keep the participants separated within the room so that there was no conferring or collaboration on the performance of the tasks or on the feedback. A typical session lasted 40-45 minutes. We recruited 15 older adults (average age=60.5 yrs stdev=6.25, 5m, 10f) and 17 younger adults (average age=29.2 yrs, stdev=6.5, 10m, 7f).

Sessions started with participants completing a brief background questionnaire about themselves and their texting style, including which mobile devices they used and which keyboard layout and input style they preferred to use. Following this, and using a balanced study design with regard to the order in which tasks were performed they each completed 3 blocks of tasks: Copy, Compose and Describe each of which are detailed in what follows.
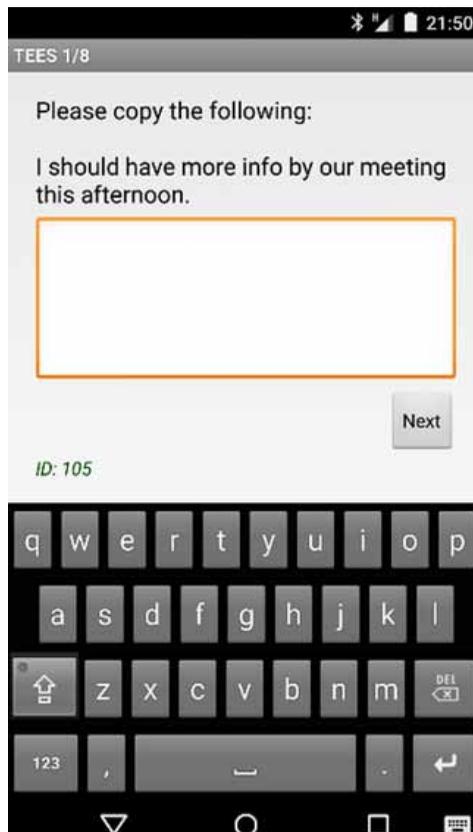
Copy Task

The use of copy or transcription tasks to examine text entry behaviour is widely reported in the literature on text entry studies as this type of task gives strong internal consistency between experimental conditions. In line with the most widespread practice, we showed the user a phrase on screen and timed their input from first character to last (we did not hide the phrase during typing). Figure 5 shows the TEES interface with a partially completed phrase copied. Phrases were sampled from the Enron memorable corpus (Vertanen & Kristensson, 2011) with two different phrase subsets in a balanced study design. Participants were presented with the following instruction on a sheet of A4 paper:

- You will be shown an English sentence. You just need to type it in including correct capitalisation and punctuation. Please proceed quickly and accurately.

Participants completed 8 such tasks in succession, with the onscreen TEES application displaying the next phrase to be copied after the previous copied phrase had been submitted. We did not impose a time limit for completion of this task.

Figure 5. TEES Copy task interface

Compose Task

As discussed above, Vertanen & Kristensson (2014) investigated complementing copy tasks with composition tasks. We replicated their suggested approach of asking people to simply create phrases. Participants were presented with the following instructions on a sheet of A4 paper:

- Imagine that you need to send a message using your mobile. We want you to invent and type in a fictitious (but plausible) message. Use your imagination. If you are struggling for ideas, think about things you often write about using your own mobile device (but don't just copy one of your messages). Please proceed quickly and accurately. Do NOT include any private information (such as real email addresses, phone numbers, or names). Invent a new message for each task of this type. Write as if you were actually typing the message for real. Please write complete sentences with proper grammar and spelling. Do NOT use texting abbreviations or slang.

Each participant completed 8 such tasks in succession, with the onscreen TEES application prompting participants for the next phrase after the first had been submitted. We did not impose a message length limit or time limit for this task.

Describe Task

Encouraged by pilot studies where we asked participants to describe images with a view to their generating large amounts of analysable text, we wanted to investigate using this approach with older adults. In our participatory design sessions it became clear that older adults were more aware of the formality expected for a message given the social context and the image description task might allow them to use appropriate language. Furthermore, some of the participants had expressed a desire to type full formal sentences (including correct capitalisation and punctuation) so this type of task seemed to fit well. The images were presented to our participants as A4 colour prints. Our images were sourced from the Internet and are all published with Creative Commons licenses that permit distribution. Users were instructed on the scenario and picture choice on screen and then typed their description into an on-screen text box (see Figure 6). A time-bar under the text area showed time completed and started on first key-press – all entry was blocked after 3 minutes. Participants were presented with sets of photos (Figure 7) along with the following instructions:

- You will be shown a set of three photos. We want you to describe any two of the photos using your mobile. We want you to write a message describing two of the photos as if you were describing them to a friend. Use your imagination: include a description of each scene and tell a story about the people in the scene. Please proceed quickly and accurately. Do NOT include any private information (such as real email addresses, phone numbers, or names). You will have 3 minutes to describe the two photos and can split your time as you wish between the photos;
- Write as if you were actually typing the message for real. Please write complete sentences with proper grammar and spelling. Do NOT use texting abbreviations or slang.

Each participant completed 4 Describe tasks, again using a balanced study design in terms of the image set that each participant received. The onscreen TEES application prompted participants to move on to beginning their description of the next image after the previous description had been submitted.

Figure 6. The TEES application as it appears during a Describe task

Text Entry Experiment System (TEES)

Developed over several years at the University of Strathclyde, TEES is a server-based system with clients available for web browsers (HTML5) and Android (Java), designed to prompt users to perform a configurable set of text input tasks, in order to capture data associated with that text entry into the server-based database. TEES records details such as time to enter a phrase, the phrase as typed, number of backspaces used and pauses in the text entry that last 1 second or more. For Copy and Compose tasks, a simple prompt is presented on screen along with a text input area and a "next" button to tap when finished – timer recording starts with the first keystroke and ends with the last. For Describe tasks, a small timer bar is explicitly displayed on screen to participants (see Figure 6) that starts filling on the first keystroke and blocks all entry then revealing a "next" button when the time limit has expired. Based on our pilot studies, we chose to impose a time limit and display a time progress bar in order to set expectations for how much to type and, in some way, to mimic the time pressure that often accompanies the need to send text messages in real life situations.

Post-Task Evaluation

Following each block of tasks, participants completed a NASA-TLX (Hart & Staveland, 1988) form about their experience of completing the tasks. At the conclusion of all three task blocks, participants completed a short exit questionnaire featuring a mixture of multiple choice and longer form questions.

Figure 7. Example image set used in the Describe task
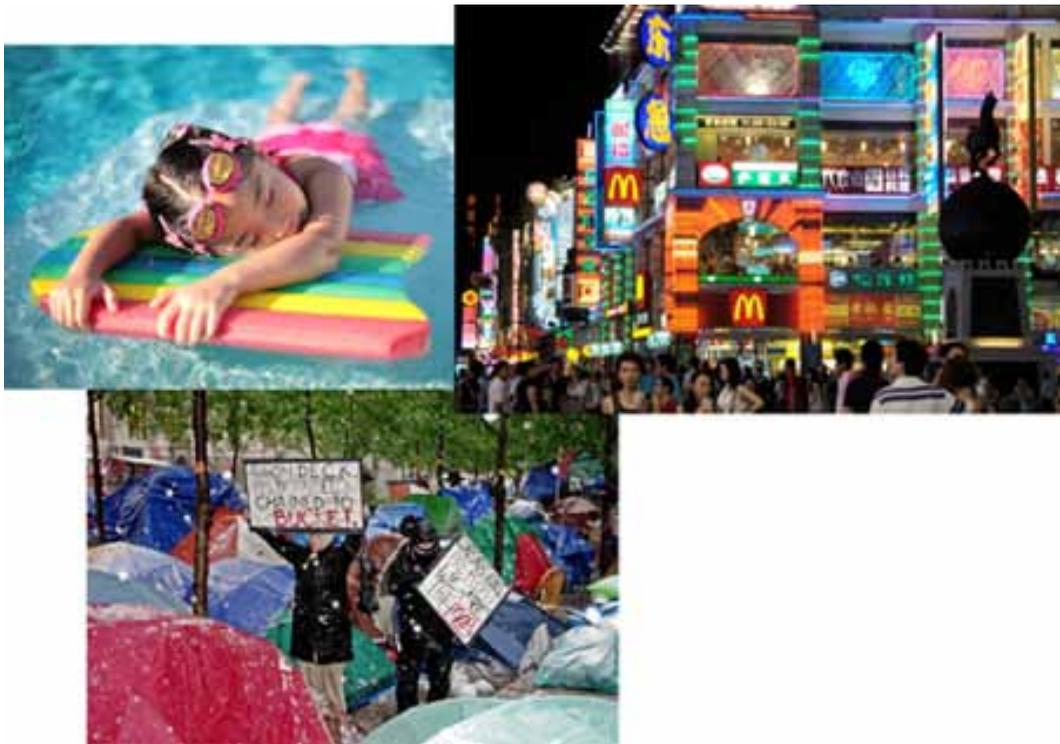
RESULTS AND DISCUSSION

In what follows, we discuss the results obtained in the study with the 15 older adults and 17 younger adults, beginning with the tap data from the logging keyboard and moving on to the data captured by the TEES system and the subjective feedback questionnaires.

Tap Data from Logging Keyboard

Analysis of the tap data revealed no statistically significant differences in the average duration of key-presses (independent sample T-test), which for both groups were around 90ms (M_Y=92.02ms, SD_Y=19.44ms, M_O=92.67ms, SD_O=16.75ms), indicating that neither group of users exhibits a habit of resting their fingers on the keys or dragging across the keyboard and so they are similar in this respect. A Mann-Whitney U test showed that where the older and younger adults differed was in the longer inter-key times (M_Y=465.92ms, SD_Y=108.97ms, M_O=768.31ms, SD_O=228.29ms, $p<0.01$). We speculate that this could perhaps be attributed to older adults' lower familiarity with the mini-QWERTY layout as we found in our earlier study (Dunlop et al., 2014), slower physical movement between targets or, simply, being more careful or reflective during input. However, slower movement times would contradict Murata et al. [10] who showed that the time required to point to a target with a touch interface is similar for both young and older adults, in contrast to other pointing devices. Younger adults exhibited a greater number of touch events per session on average than did the older adults. Additionally, younger adults took less time per session. Taken together, this indicates that younger adults favour fast typing and input more characters than older users. However, younger users

did not exhibit a statistically significant larger number of suspect characters, so their accuracy is not particularly compromised by this faster typing.

For older adults a paired-sample T-test showed that there existed a statistically significant x-axis difference ($p<0.05$) between taps on suspect ($M_S=1.88px$, $SD_S=1.46px$) and non-suspect keys ($M_{NS}=2.82px$, $SD_{NS}=1.69$). However, here we note that suspect keys exhibit a lower x-axis displacement during the touch – dismissing the notion that erroneous taps might be due to finger slippage between touch-down and touch-up locations. Both the position and displacement differences were, however, very small (approx. 2 vs 3 pixels) so could not be used to programmatically identify "suspect" keys at tap-time. For younger users, a difference was also observed ($M_S=2.27px$, $SD_S=1.47px$, $M_{NS}=2.55px$, $SD_{NS}=0.81px$) but it was marginally non-statistically significant (paired sample T-test, $p=0.06$). For older adults there was no statistically significant difference between vertical axis movement for suspect and non-suspect touches ($M_S=3.96px$, $SD_S=3.13px$, $M_{NS}=4.21px$, $SD_{NS}=2.29px$), although one exists for younger users ($M_S=3.66px$, $SD_S=2.16px$, $M_{NS}=4.00px$, $SD_{NS}=1.54px$, paired sample T-test, $p<0.01$). We note that in both cases, the vertical movement is considerably larger than the horizontal movement but is still quite small.

TEES Data

The data that we collected using the TEES system was unfortunately subject to significant data loss, and as such, of the 32 people who took part in our study, we were left with TEES data for only 18 of them: 12 older adults and 6 younger adults. It is on this reduced dataset that we report in the remainder of this section. The data collected by TEES revealed differences in the behaviour of the two groups of users. Examining the data by breaking it down by task type revealed that, not unexpectedly given the fixed nature of the task, both groups typed a roughly equal amount for the Copy tasks. For the Compose tasks, older adults were typing slightly more than the younger adults (Figure 8). In stark

Figure 8. Length of texts submitted for each task
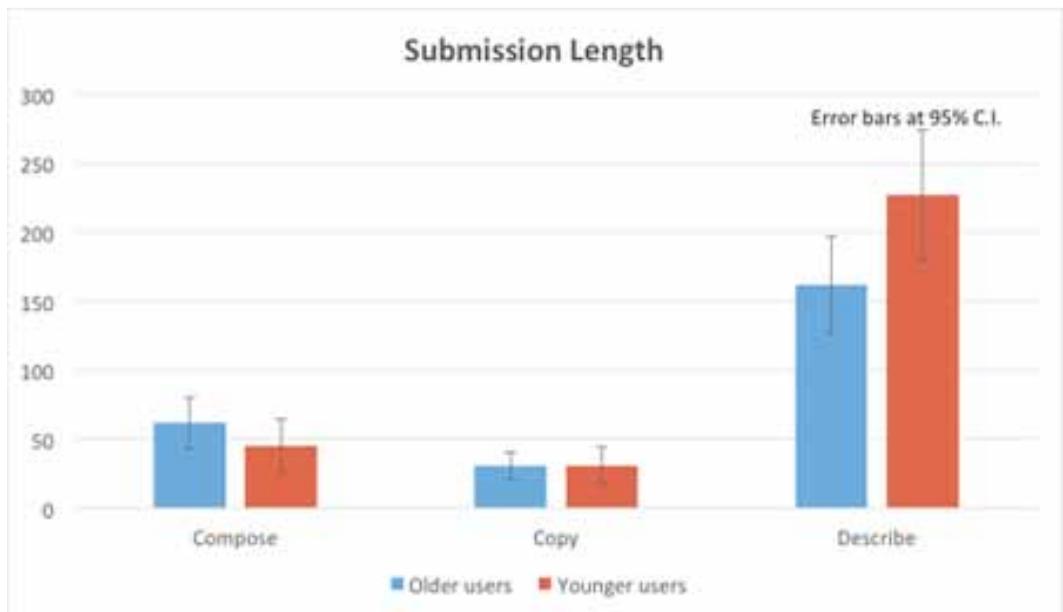
contrast, we found that the younger adults were typing a good deal more than older adults in the Describe tasks (Figure 8).

One of our key motivations in using a task such as Describe was to generate large amounts of text from our participants. When a comparison is made with the Compose tasks (Figure 8) it can be seen that for younger participants in particular this task generates around 5 times as much text as does the Compose task, and in the case of the older adults, the Describe task generates around 3.5 times the volume of text generated by the Compose tasks.

For all three tasks, the older adults spent longer typing than did the younger adults (Figure 9), albeit with a marginal difference in Describe tasks due to the system imposed time limit.

In all three tasks, younger adults typed at a word per minute (wpm) rate that was noticeably faster on average than the rate by older adults (Figure 10). It is also interesting to note that Copy tasks for younger adults are in line with many reported studies at around 20 wpm, while both older adults and composition-based tasks are slower than the more widely used copy tasks.

On average, older adults inputted a slightly higher number of backspaces than the younger adults as a proportion of the length of the final submission across all three task types. Copy tasks were most accurate on average (Figure 11) with a lower backspace rate but with a large variation of accuracy across all tasks and age groups, with in the range of 10-20% of keystrokes being backspaces, which equates to one backspace per one or two words.

The TEES system also recorded any pauses during typing that were more than a second in duration. Summing these per phrase showed no difference between groups for Copy and Describe tasks but a clear difference in Compose (Figure 12) – indicating that older adults may spend more time reviewing and thinking than our younger participants (we excluded pauses over 1 minute as outliers and did not include pauses before the first keypress per phrase).
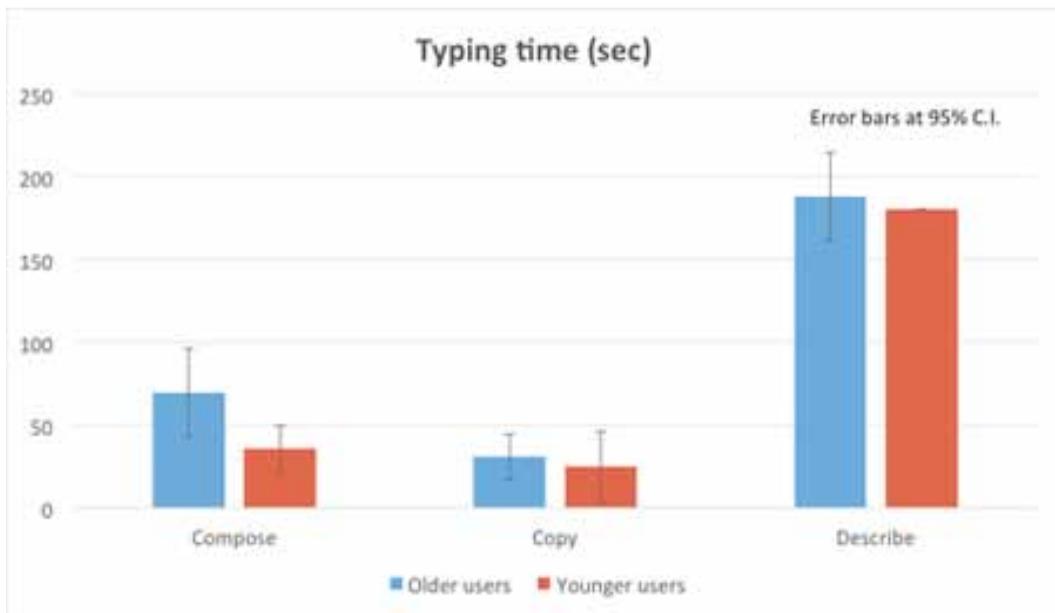
Figure 9. Typing time for each task

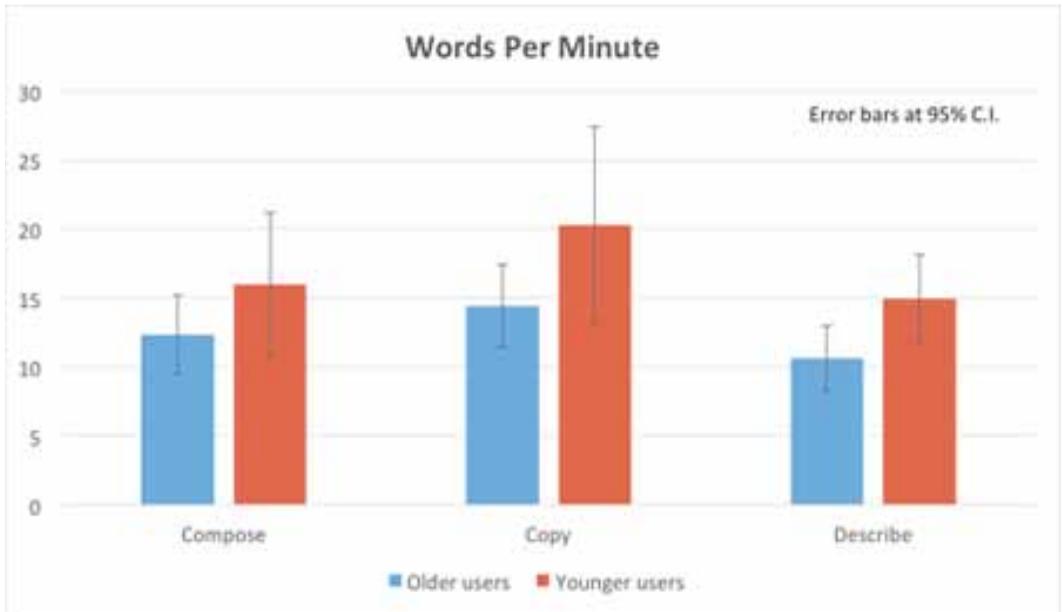**Figure 10. Words per minute (wpm) typed for each task**

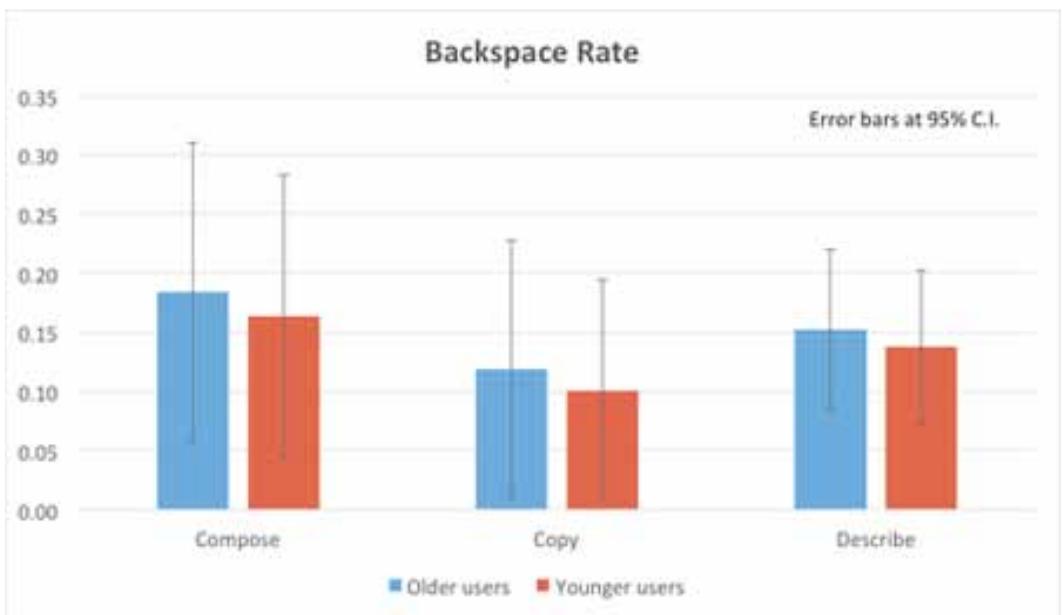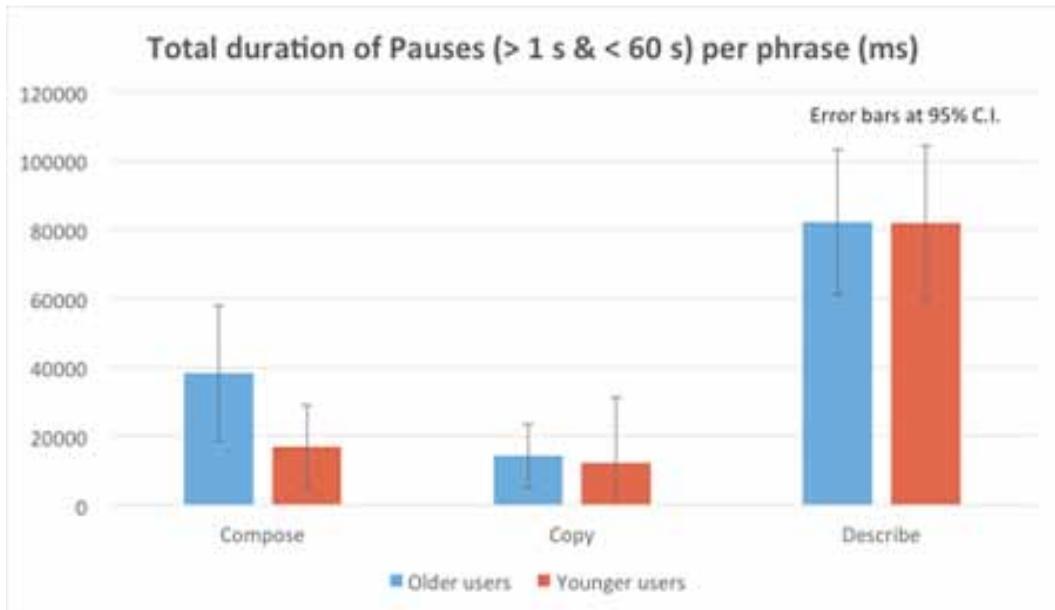**Figure 11. Backspace rate for each task**

Figure 12. Length of pauses per phrase typed

As expected, and in line with other studies, error rates on final submission were very small (using direct comparison for Copy tasks and spell-checking comparison for Compose and Describe tasks).

Subjective Feedback

The NASA-TLX questionnaires asked the participants to rate on a Likert scale how they felt following each block of a particular task type with regard to the following measures: Mental Demand; Physical Demand; Temporal Demand; Performance; Effort; Frustration. Figure 13, Figure 14, Figure 15 and Figure 16 illustrate how participants fared in each task according to each of these measures. In general, the Copy task was rated least demanding on all measures by both older and younger adults. The Compose task was rated as being slightly more demanding than Copy on all measures by both older and younger users, with the Describe task being rated as most demanding on all measures by both groups.

Only one statistically significant difference between the younger and older groups was found in the NASA feedback: on the Compose task, the older adults rated temporal demand significantly higher than the younger adults (M_O=8.20, SD_O=4.41, M_Y=4.59, SD_Y=3.91, independent sample T-test, $p<0.05$) (Figure 14).

In their answers to the exit questionnaire, older adults indicated that they found the logging keyboard easier to use than the younger users had. This could be related to the absence of auto-correction, suggestions and auto-completions. Our younger adult participants had reported in the pre-study questionnaire that they were slightly more reliant on these features on their own keyboard which may account for this difference, though it is worth noting that several of our older adults said in their post-task written feedback that they had noticed the effects of the absence of these features too. In designing our keyboard we based it on the stock open-source Android keyboard and kept the interface and functionality closely resembling the keyboards that our participants would have experienced already. Our aim here was that participant behaviour could be as close to their natural

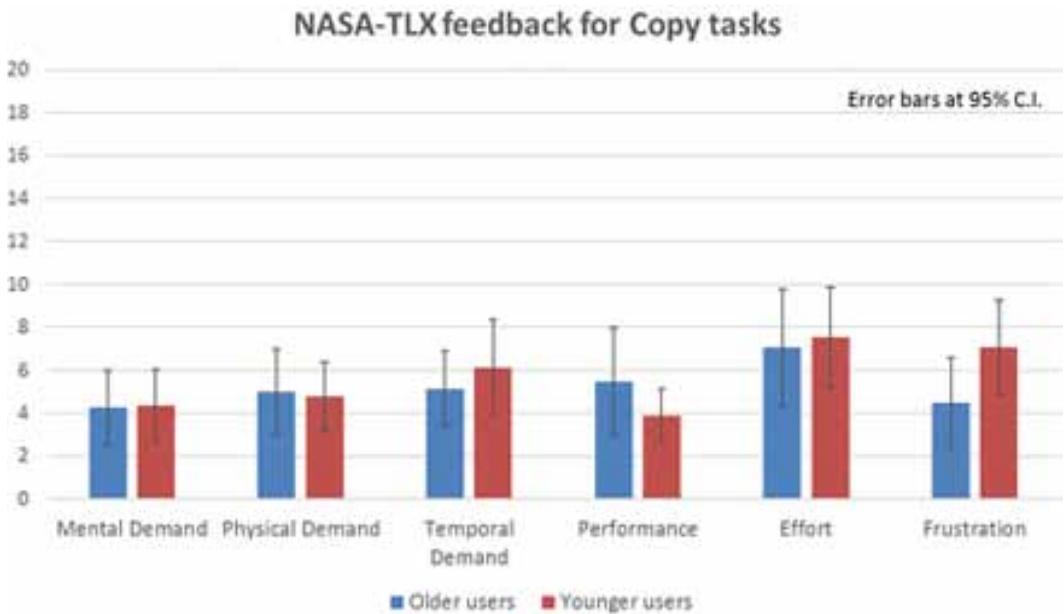**Figure 13. NASA-TLX mean scores for the Copy task**

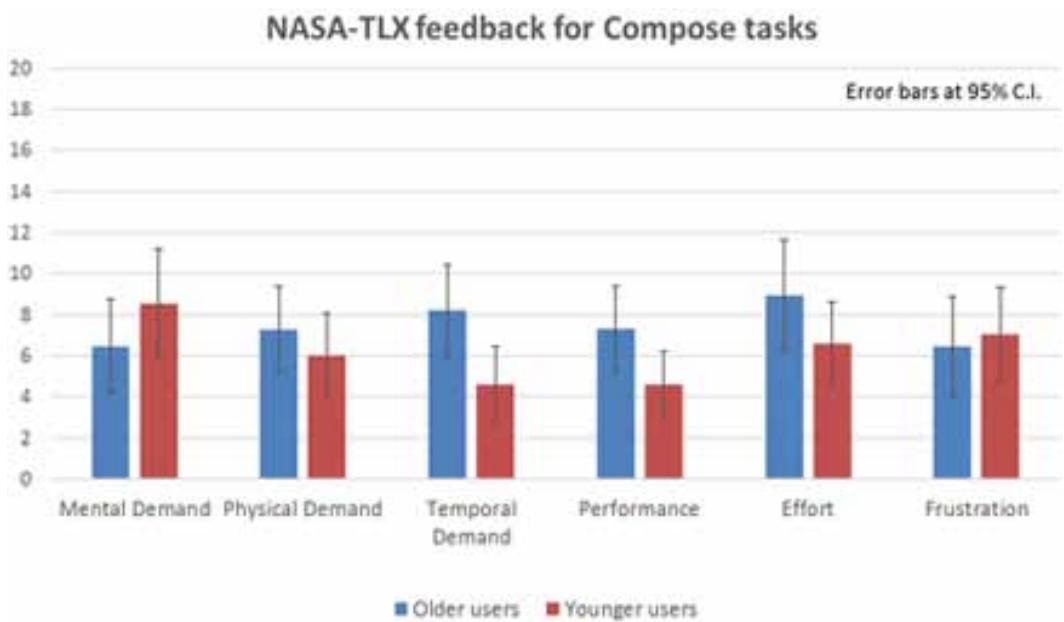**Figure 14. NASA-TLX mean scores for the Compose task**

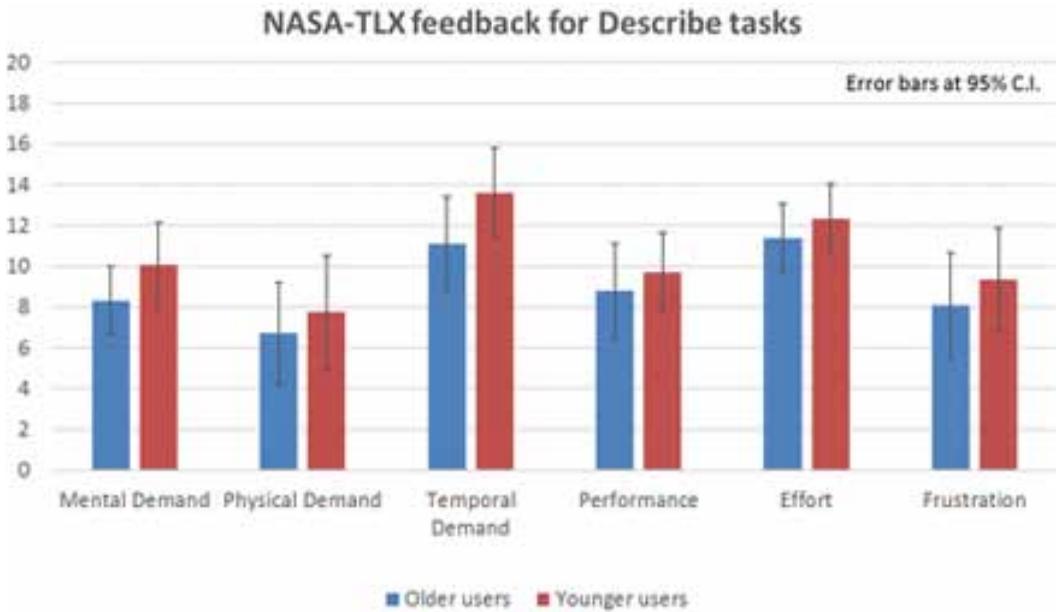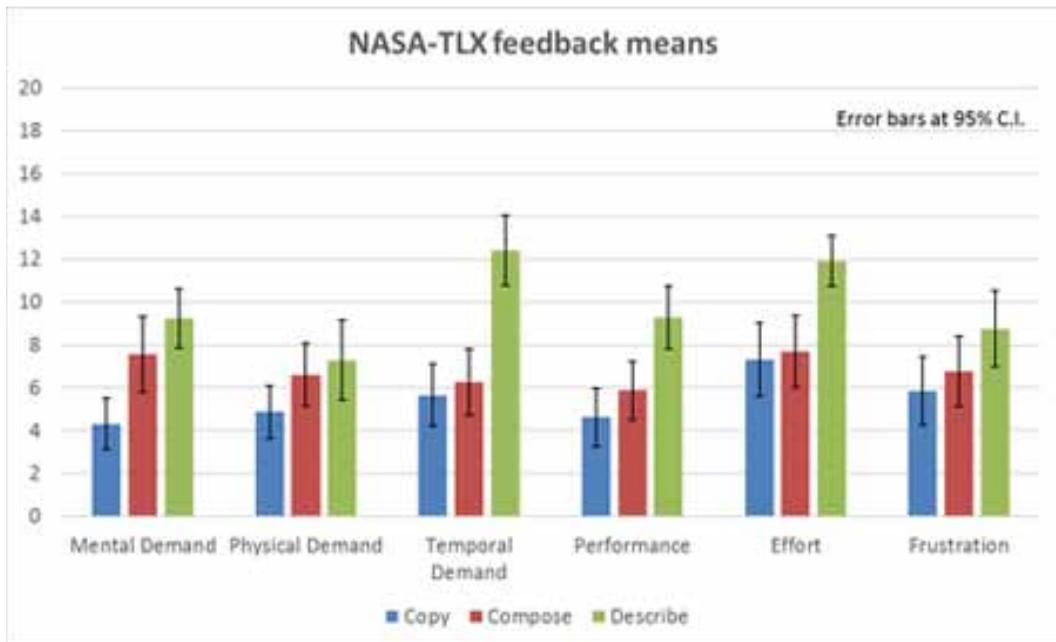**Figure 15. NASA-TLX mean scores for the Describe task**

Figure 16. Mean NASA-TLX scores across all participants (young and old) for each task

behaviour when using their own devices as possible. Accordingly, we asked how similar they thought the keyboard was to that which they normally used. The older adults reported finding the keyboard closer to their usual keyboard than did the younger adults. We also asked about how easy they had found each task type that we had given them to perform. On the Copy tasks, the older and younger adults reported a very similar level of ease. On the Compose task, the older adults reported finding the task more difficult than the younger adults, where the older adults reported finding the Describe task easier than the younger adults had. These results have some echoes in the results for favourite task reported later in this section.

We asked them how they thought they spent their time during each task in terms of thinking and typing. Older and younger adults differed in how they thought they split their time when performing tasks. This was the case for each of the 3 types of task to varying degrees (Figure 17, Figure 18, Figure 19), with older adults reporting more time spent thinking than younger adults reported. The most pronounced difference was reported in relation to the Describe task (Figure 19).

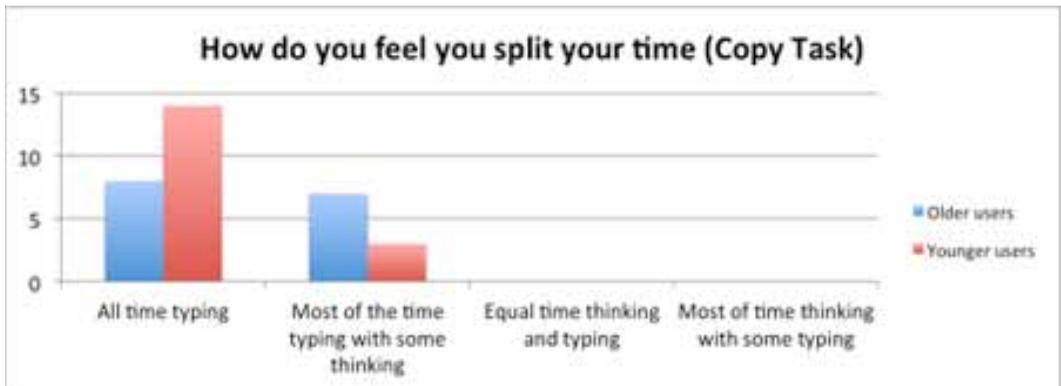**Figure 17. Time split between thinking and typing on Copy task**

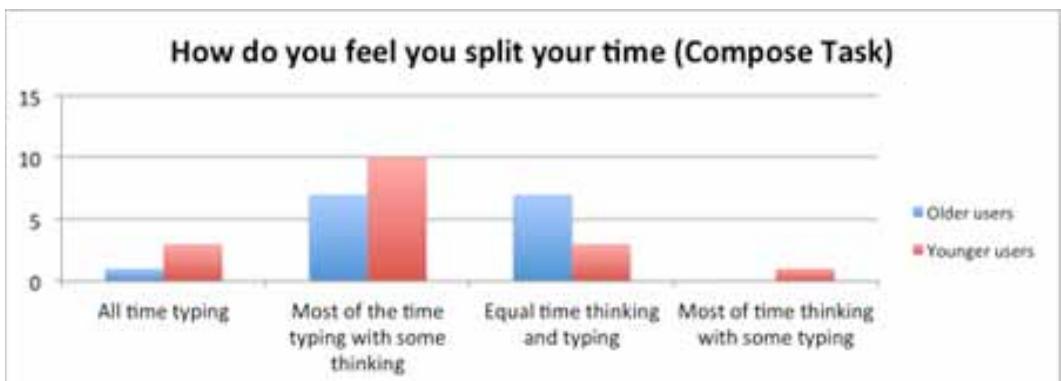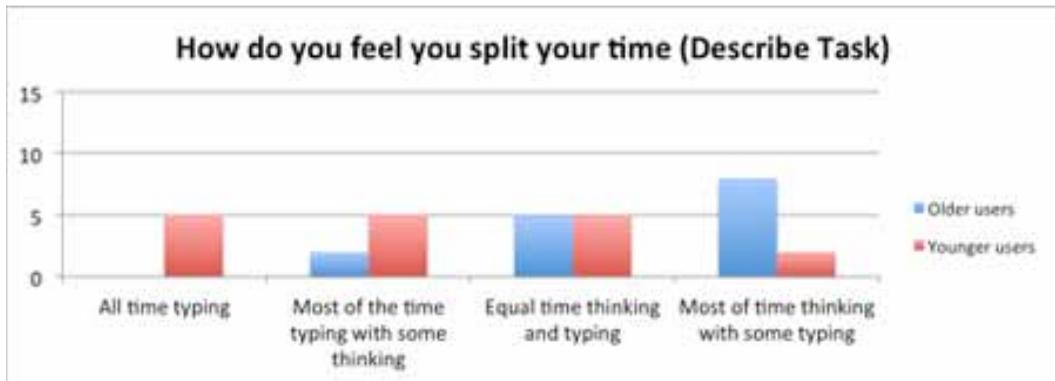**Figure 18. Time split between thinking and typing on Compose task**

Figure 19. Time split between thinking and typing on Describe task

When asked how easy it was to think of 8 phrases for the Compose task, the younger participants reported that they found it easier to think of the phrases than the older participants reported. When asked how easy it was to think of descriptions in the Describe task, the younger adults reported finding it easier than the older adults. As for how easy they found it to fill the 3 minutes allowed for the Describe task, the older adults found it easier to do this than the younger adults, which is interesting when considered together with their responses about the task they had found most enjoyable.

When asked about the most and least enjoyable tasks, the difference between older and younger adults was striking. Younger adults favoured the Describe task least of all, with half of the older adults choosing it as their favourite and the other half naming it as their least favourite. As one older participant said of the Describe task: "I don't think enjoyment came into it!" (See Figure 20, Figure 21, and Table 1).
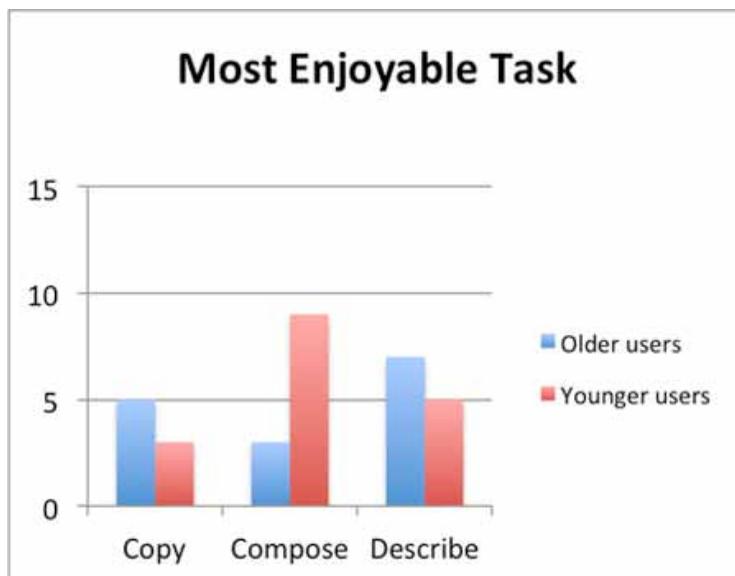
Figure 20. Most enjoyable task
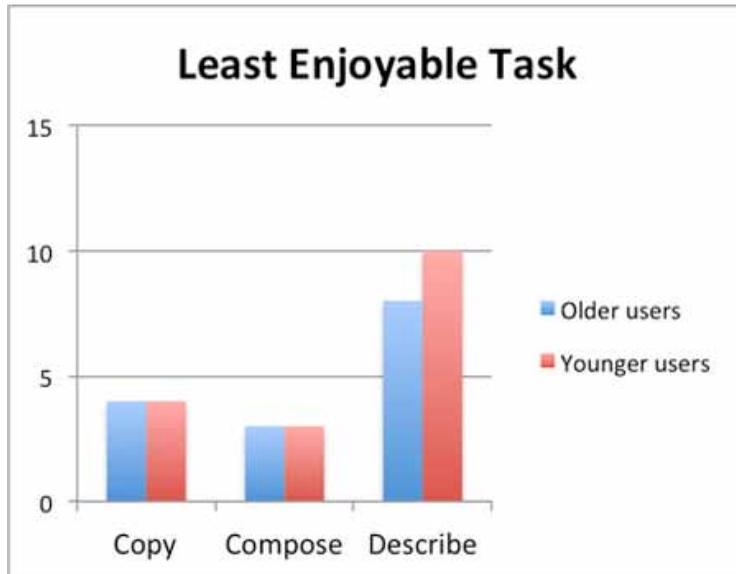
Figure 21. Least enjoyable task

Table 1. Summary of task enjoyment

	Most Enjoyable		Least Enjoyable	
	Older	**Younger**	**Older**	**Younger**
Copy	5	3	4	4
Compose	3	9	3	3
Describe	7	5	8	10

Other issues that arose from the subjective data: older adults made more frequent mentions of pressing incorrect keys than the younger adults did. Older adults also mentioned difficulties with punctuation more frequently. In summary, with the exception of the Copy task, which was regarded as having an equal level of difficulty by both older and younger adults and did not attract strong feelings in terms of enjoyability by either group, the tasks undertaken by participants revealed rather different behaviour and attitudes between the age groups with regard to the ease, enjoyability and time dimension of the tasks.

CONCLUSION AND FURTHER WORK

This paper reported the results from a set of studies of text entry for older adults. Firstly, in our research we saw evidence that supports our initial statement that people wish to use mobile text entry into their retirement in order to address diverse needs and desires. We saw this most directly in the map and cultural probe exercises but also in discussions throughout our design workshops. Our participants imparted a rich variety of contexts (spatial, temporal, activity and social) under which they might

like to perform text entry using their mobile devices. The discussions and comments from our older participants have, in addition, challenged the stereotype of simplistic use of technology by older adults: our users appear very socially aware of appropriate context of use and of messaging style, they are comfortable swapping between different devices in different situations and task contexts and have shown a strong interest in new ways of entering text. One user stated "I know when I am doing a Twitter post I have to be very careful to hand carve those 140 characters, with Facebook I used to be quite casual until some posts were re-posted and taken out of context so I am more careful now. With SMS to family or friends, I am much less careful as they know me well." while another stated "this is similar to the kind of description I might send from my allotment".

Our participatory design sessions confirmed a perceived concern over the button size and sensitivity of smartphone screens by older adults (in addition to issues unrelated to text entry such as battery life, lost signal and lack of support from handset manufacturers or operators). In our design and observation sessions a theme of "write and review" emerged, with older users tending to take a view that it is better to write with little support from the keyboard and then to review the message before sending. We saw support for this in our lab-based Compose tasks where, despite writing slightly more than younger participants, older participants also paused more frequently during the fixed time allowed for input. This finding was reflected in the self-judged statements of how much time they felt they were spending between thinking and typing, with older participants reporting spending more time thinking than younger ones. Confirming Wright et al. (2000) for finger-based touchscreens, our lab-based studies showed that older users did indeed type at a slower rate with lower average words-per-minute rates observed across all task types. More detailed analysis showed that this difference was caused by longer inter-key times and not the actual tap times.

Methodologically, we strongly support Coleman et al. (2010), Fisk et al. (2009) and Lee & Zhai (2009b) in their view that care must be taken, particularly with older adults, to negotiate terminology in order to establish common language between researchers and participants. The danger of not approaching the language issue with sensitivity and openness is that discussions can be side-tracked into clarification. We would also extend this to the study processes themselves – we found openness and clarity of the aim of the studies (e.g. "we wish you to describe a picture in 3 minutes as a way of capturing fairly long blocks of text") to be appreciated more by older adults than we are used to from younger adult studies. With regard to the use of metaphors as discussed by Göllner et al., (2010), our work has embraced the technique and utilised it to address the technical issues during participatory design – though we would caution that the use of metaphors can be considered patronising if not handled carefully.

Finally, we conducted a preliminary longitudinal study that showed that our users were willing to use a logging keyboard for daily activities. We are now planning a full longitudinal study in which we will be able to record text input behaviour over a longer period and get some insights into possible Hawthorn effects of our lab-based studies.

The work presented in this paper focuses on identifying the differences in input behaviour between older and younger users. This can be used as a basis on which to conduct future research into the keyboard designs that older users need and to produce specific design guidelines for such software. As a starting point towards framing design limitations, future efforts should consider that our older adult participants were particularly conscious of errors; they also expressed a dislike of predictive texting and a self-ownership attitude to texting quality believing it is their responsibility to type accurately. Together these desires are unachievable on small touchscreen phones where the key size is simply too small to allow accurate typing without predictive algorithm support. However, we did see strong interest in alternatives such as gesture input with some participants. For those who dislike prediction, one obvious solution is to encourage the use of larger phones, though this may not always

be practical or possible for all users. Further research is required to investigate the reasons behind the dislike of predictive input and concerns over error. Given the necessity for predictive input support on small screen devices, the underlying mechanisms should be designed to both improve predictive texting algorithms and tune the feedback provided by the input interface to users so as to support feelings of certainty and trust in use.

ACKNOWLEDGMENT

REFERENCES

Chen, T., & Kan, M.-Y. (2012). Creating a live, public short message service corpus: The NUS SMS corpus. *Language Resources and Evaluation*, *47*(2), 299–335. doi:10.1007/s10579-012-9197-9

Coleman, G. W., Gibson, L., Hanson, V. L., Bobrowicz, A., & McKay, A. (2010). Engaging the disengaged: How do we design technology for digitally excluded older adults? In Proceedings of the 8th ACM Conference on Designing Interactive Systems (pp. 175–178). ACM; Retrieved from http://dl.acm.org/citation.cfm?id=1858202 doi:10.1145/1858171.1858202

Dunlop, M., Komninos, A., Nicol, E., & Hamilton, I. (2014). Shake "n" Tap: A Gesture Enhanced Keyboard for Older Adults.

Fisk, A. D., Rogers, W. A., Charness, N., Czaja, S. J., & Sharit, J. (2009). *Designing for older adults: Principles and creative human factors approaches*. CRC press. Retrieved from http://books.google.co.uk/books?hl=en&lr =&id=uSXmGIHXyZUC&oi=fnd&pg=PP1&dq=%5B2%5D%09Fisk,+A.+D.,+Rogers,+W.+A.,+Charness, +N.,+Czaja,+S.+J.,+Sharit,+J.+Designing+for+Older+Adults:+Principles+and+Creative+Human+Factors +Approaches.+CRC+Press+(2009).&ots=8uoS_6k7cf&sig=MKsA128IaiNQASpRDAvZwUHQ1-I

Fisk, A. D., Rogers, W. A., Charness, N., Czaja, S. J., & Sharit, J. (2012). *Designing for older adults: Principles and creative human factors approaches*. CRC press.

Gaver, B., Dunne, T., & Pacenti, E. (1999). Design: Cultural probes. *Interaction*, *6*(1), 21–29. doi:10.1145/291224.291235

Göllner, S., Lindenberg, J., Conradie, P., Le, J., & Sametinger, F. (2010). The Enchanted Neighborhood: Using Metaphorical Devices for the Inclusion of Seniors in the Co-designing Process. Retrieved from http://www.researchgate.net/profile/Peter_Conradie/publication/236150346_The_Enchanted_ Neighborhood_Using_Metaphorical_Devices_for_the_Inclusion_of_Seniors_in_the_Co-designing_Process/ links/0c96052b1bcf8de9b6000000.pdf

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, *52*, 139–183. doi:10.1016/S0166-4115(08)62386-9

Hoggan, E., Brewster, S. A., & Johnston, J. (2008). Investigating the Effectiveness of Tactile Feedback for Mobile Touchscreens.*Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1573–1582). New York, NY, USA: ACM. doi:10.1145/1357054.1357300

Kano, A., Read, J. C., & Dix, A. (2006). Children's Phrase Set for Text Input Method Evaluations.*Proceedings of the 4th Nordic Conference on Human-computer Interaction: Changing Roles* (pp. 449–452). New York, NY, USA: ACM. doi:10.1145/1182475.1182534

Karat, C.-M., Halverson, C., Horn, D., & Karat, J. (1999). Patterns of Entry and Correction in Large Vocabulary Continuous Speech Recognition Systems.*Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 568–575). New York, NY, USA: ACM. doi:10.1145/302979.303160

Knotice (2012). Mobile Email Opens Report, 2nd Half 2011. Retrieved October 16, 2015 from http://www. knotice.com/reports/Knotice_Mobile_Email_Opens_Report_SecondHalf2011.pdf

Kristensson, P. O., & Vertanen, K. (2012). Performance Comparisons of Phrase Sets and Presentation Styles for Text Entry Evaluations.*Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces* (pp. 29–32). New York, NY, USA: ACM. doi:10.1145/2166966.2166972

Kristensson, P.-O., & Zhai, S. (2004). SHARK 2: a large vocabulary shorthand writing system for pen-based computers. *Proceedings of the 17th annual ACM symposium on User interface software and technology* (pp. 43–52). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1029640

Lee, S., & Zhai, S. (2009a). The performance of touch screen soft buttons.*Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 309–318). ACM. doi:10.1145/1518701.1518750

Lee, S., & Zhai, S. (2009b). The performance of touch screen soft buttons. . Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 309–318). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1518750 doi:10.1145/1518701.1518750

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics, Doklady*, *10*, 707.

MacKenzie, I. S., & Soukoreff, R. W. (2003a). Phrase sets for evaluating text entry techniques. *Proceedings of CHI'03 extended abstracts on Human factors in computing systems* (pp. 754–755). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=765971

MacKenzie, I. S., & Soukoreff, R. W. (2003b). Phrase Sets for Evaluating Text Entry Techniques. Proceedings of CHI '03 Extended Abstracts on Human Factors in Computing Systems (pp. 754–755). New York, NY, USA: ACM. doi:10.1145/765891.765971

Murata, A., & Iwase, H. (2005). Usability of Touch-Panel Interfaces for Older Adults. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *47*(4), 767–776. doi:10.1518/001872005775570952 PMID:16553065

Nicolau, H., & Jorge, J. (2012). Elderly Text-entry Performance on Touchscreens.*Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 127–134). New York, NY, USA: ACM. doi:10.1145/2384916.2384939

Office for National Statistics. (2012, March 2). Population Ageing in the United Kingdom, its Constituent Countries and the European Union. Retrieved from http://www.ons.gov.uk/ons/rel/mortality-ageing/focus-on-older-people/population-ageing-in-the-united-kingdom-and-europe/rpt-age-uk-eu.html

Paek, T., & Hsu, B.-J. (Paul). (2011). Sampling Representative Phrase Sets for Text Entry Experiments: A Procedure and Public Resource.*Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2477–2480). New York, NY, USA: ACM. doi:10.1145/1978942.1979304

Soukoreff, R. W., & MacKenzie, I. S. (2001). Measuring Errors in Text Entry Tasks: An Application of the Levenshtein String Distance Statistic. Proceedings of CHI '01 Extended Abstracts on Human Factors in Computing Systems (pp. 319–320). New York, NY, USA: ACM; doi:10.1145/634067.634256

Vertanen, K., & Kristensson, P. O. (2011). A Versatile Dataset for Text Entry Evaluations Based on Genuine Mobile Emails.*Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services* (pp. 295–298). New York, NY, USA: ACM. doi:10.1145/2037373.2037418

Vertanen, K., & Kristensson, P. O. (2014). Complementing Text Entry Evaluations with a Composition Task. *ACM Trans. Comput.-Hum. Interact., 21*(2), 8:1–8:33. doi:10.1145/2555691

Wright, P., Bartram, C., Rogers, N., Emslie, H., Evans, J., Wilson, B., & Belt, S. (2000). Text entry on handheld computers by older users. *Ergonomics*, *43*(6), 702–716. doi:10.1080/001401300404689 PMID:10902881

Emma Nicol is a researcher based in the Department of Computer and Information Sciences at the University of Strathclyde in Glasgow, Scotland and is a member of the i-lab and Mobiquitous lab research groups there. In recent times, she has been working on an EPSRC-funded project investigating mobile text entry for older adults (OATS) with Mark Dunlop from which a number of publications have resulted. She was co-organiser of workshops at the MobileHCI conferences of 2014 and 2015 on the topic of re-imagining mobile interfaces for older adults, the second of which focussed on methodologies for designing with this group of users. Her past research activities have encompassed both human computer interaction and information seeking behaviour, often with an educational focus. Emma has extensive experience of the design and evaluation of digital products for children and has published on in-the-wild evaluations of museum interactives, and on the use of digital simulations in education.

Andreas Komninos is a researcher at the Mobiquitous Lab, in the University of Strathclyde, Scotland. He is also an adjunct professor for the Engineering of Pervasive Computing Systems MSc at the Hellenic Open University, Greece. His research has focused on human interaction with mobile and ubiquitous computing systems, since 2001.

Mark D. Dunlop is a senior lecturer in Computer and Information Sciences at Strathclyde University, Glasgow. His research focuses on usability of mobile and wearable systems including mobile text entry, visualisation and user evaluation. He is currently leading the OATS EPSRC funded project on text entry for older adults and leads The Mobiquitous Lab at Strathclyde, which is researching user behaviour in the context of mobile devices, ubiquitous computing, and multimodal interaction. He is an Associate Editor for International Journal of Mobile Human Computer Interaction*, a Sub-Committee chair for ACM CHI and secretary to the MobileHCI Steering Committee. Prior to joining Strathclyde, Mark was a senior researcher at Risø Danish National Laboratory and a lecturer at Glasgow University.*