Behaviour & Information Technology
2009, 1–14, i*First* article

Age-related differences in the initial usability of mobile device icons

Rock Leung[a]*, Joanna McGrenere[a] and Peter Graf[b]

[a]*Department of Computer Science, University of British Columbia, 2366 Main Mall, Vancouver, British Columbia, V6T IZ4, Canada;* [b]*Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, British Columbia, V6T IZ4, Canada*

(*Received 29 March 2008; final version received 27 June 2009*)

Mobile devices offer much potential to support older adults (age 65+). However, older adults have been relatively slow to adopt mobile devices. Although much ongoing HCI research has examined usability problems to address this issue, little work has looked at whether existing graphical icons are harder to use for this population compared with younger adults. We conducted a qualitative exploratory study and a follow-up experimental study to determine which icon characteristics help initial icon usability for older adults. We found that our older participants did have more problems using existing mobile device icons, but that particular icon characteristics – semantically close meaning (i.e. natural, close link between depicted objects and associated function), familiar, labelled and concrete (i.e. those depicting real-world objects) – improved icon usability for them. We discuss how these findings can help icon designers to create mobile device icons that are more suited to the abilities and technology experience of older adults.

Keywords: aging; graphical icons; initial usability; mobile devices; icon characteristics; user interface design

1. Introduction

The proportion of older adults in developing countries is growing and their mobility has increased because of better general health and financial resources (Goodman and Gray 2003). However, as people age, they experience a normal decline in a number of abilities, such as perceptual, motor and cognitive abilities, which limit mobility and their independence. Thus, to remain independent older adults often require more support. As the proportion of older adults increases, the possibility of relying on human care givers to provide this support decreases, increasing the need for other types of support such as mobile computer devices (Goodman *et al.* 2004). Mobile devices can support older adults in many ways; for example mobile phones can help older adults stay connected, innovative memory aids can help them to remember important information (e.g. Inglis *et al.* 2003), and portable game systems can even provide them with fun and stimulating mental exercises (e.g. Nintendo Brain Age™).

Older adults, however, have been slower to adopt mobile computer technologies and find them more difficult to use (Kurniawan *et al.* 2006). For example, in a UK survey conducted by Ofcom (2006), only 49% of seniors (age 65+), compared with 82% of all adults surveyed, reported owning a mobile phone. Of these seniors, 44% reported being able to listen to voicemail messages and 29% reported being able to send a text message (compared with 83% and 81% of all adults, respectively). Recent HCI research has examined many different usability issues that might explain the reluctance of older adults to adopt mobile devices e.g. decline in users' vision (Jacko *et al.* 2002) or motor skills and coordination (Moor *et al.* 2004), interface complexity (Ziefle and Bay 2005).

Although icons are an integral part of most user interfaces, as they are the gateway to access the myriad of functional capabilities for such devices, little work has investigated the influence of graphical icons on younger and older adults' use of traditional computers and mobile devices. It seems likely that the decline in perceptual and cognitive abilities which accompanies normal aging has some effect on older adults' ability to interpret graphical icons. First, an icon has a number of graphical properties to be processed (e.g. visual details such as lines and dots, its resemblance to a real-world object) in order to identify the visual objects being presented in the icon. This step may be more difficult for older adults with decreased perceptual abilities. Second, an icon is a pictographic item that represents one or more objects, a concept or a function (Peirce 1932), and interpreting the icon requires understanding the link between that item and the meaning for which it stands (Heim 2007), which may be more difficult for someone with declining cognitive

*Corresponding author. Email: rockl@cs.ubc.ca

doi:10.1080/01449290903171308 (http://dx.doi.org/10.1080/01449290903171308)

abilities. Interpreting the icon also depends on other factors, such as the context provided by the software application in which the icon is used, text labels that might be displayed together with an icon and the user's familiarity with the icon as well as with its application context (Horton 1994). Compared with young adults, older adults are likely to have less experience with contemporary handheld devices and be less familiar with a device's icons and applications, making the icons more difficult to interpret.

Initial icon usability is particularly important for learning to use mobile technology because of the many features and interfaces available on mobile devices. Mobile devices have a wide variety of functionality that is not commonly available on laptop/desktop computers, such as a number of text entry methods and data connectivity options (e.g. wi-fi, Bluetooth, infrared and cellular). Many of these functions can only be accessed through unique icons and generic buttons (as opposed to dedicated buttons and other controls). In addition, there are many more operating systems for mobile devices (e.g. Symbian, Windows Mobile, PalmOS, Blackberry OS, iPhone OS and embedded Linux) than desktop operating systems (e.g. Windows, MacOS), and each system typically uses its own set of icons. Finally, initial icon usability may be more critical for mobile devices because mobile devices appear to evolve more quickly and users tend to replace them more frequently than laptop or desktop computers.

Many icon design guidelines have been published, for example: 'Design icons to identify clearly the objects or concepts they represent' (Sun Microsystems 1999), 'Icons should be suggestive of the functionality with which they are associated.' (Benson *et al.* 2005), 'Icons should be familiar to the user' (Heim 2007, p. 430). However, these guidelines are not age specific and the implied assumption is that they apply universally to all ages. We do not know whether existing guidelines are appropriate for older adults because, to our knowledge, they are based on research involving only younger adults. Thus, a further motivation for the present research was to investigate whether existing icon design guidelines can be generalised to older adults, particularly those aged 65 and older.

The design of the icon, and its resulting characteristics, greatly affects its usability, and also the usability of the application in which the icon is found. Take for example the following sets of icons (see Figure 1) from two handheld computer calendar applications that are both used to view schedule information for 1 day, week or month. The icons on the left seem to have a greater resemblance to traditional paper calendars than the icons on the right. Users are generally familiar with paper calendars and would likely find it easier to identify the icon objects on the left (assuming icons are the same size), and may find it much harder to identify and interpret the icons on the right without additional information or prior experience.

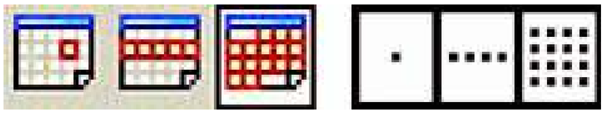

Figure 1. Example calendar application icons for viewing a day, week and month; icons are from the HP iPaq rx3715 (left) and the Palm® Treo™ 650 (right).

Our goal for this research was to find empirical evidence for age-related effects of salient icon characteristics on icon usability, with the overall aim of understanding how to design icons more inclusively. We sought to identify which icon characteristics help or hinder usability, and to determine experimentally whether the effects of those characteristics differ across age groups. Little has been reported in the research literature about icon usability issues that seniors experience (beyond size (Siek *et al.* 2005), and colour and contrast (Hawthorn 2000)) and our work seeks to take steps to fill that gap. With a view on technology adoption, we focused on the initial usability of existing mobile device icons, specifically on icons employed on smart phones, handheld computers and personal digital assistants (PDAs).

To ground our understanding of age-related differences in icon usability and to identify icon characteristics which might cause usability problems for older adults, we began with a qualitative exploratory study with 10 participants of varying ages. The results highlighted the importance of familiarity of the visual stimuli used in icons and of their *concreteness*, that is how closely the object(s) depicted in an icon resembles real-world items, places or people (McDougall *et al.* 2000). Concrete icons show objects that closely resemble items in the real world, whereas abstract icons 'represent information using graphic features such as shapes, arrows and so on' (McDougall *et al.* 1999, p. 488). In Figure 1, the icons on the left are more concrete than those on the right. The results also suggested that icon usability might be related to *semantic distance*, that is the closeness of the relationship between the objects depicted in the icon and the function being represented (McDougall *et al.* 1999). In Figure 1, icons on the left have semantically closer meanings than those on the right.

To investigate in depth the specific influence of each of these factors on the usability of icons, we conducted a controlled experiment with 36 participants, half of them young adults and the other half seniors. We also investigated age-related differences in the effect of labels on icon usability. Our results showed that mobile device icons were generally harder for older adults to use, compared with younger adults. In addition, we found that icons with semantically closer meanings had a larger positive impact on interpretation accuracy for seniors than younger adults. We also found that concrete icons helped older adults, more than younger ones, to identify objects in icons with semantically far meanings. Furthermore, we found that labels improved initial icon usability but did not help older adults more than younger ones. We discuss implications of these findings for designing mobile device icons that are better suited to older adults. Our findings constitute a much needed empirical foundation on age-related differences in icon usability and for icon design guidelines targeting older users.

2. Related work

There has been much emerging work on designing computer interfaces for older users (e.g. Czaja 1997, Docampo Rama *et al.* 2001, Gregor *et al.* 2002, Fisk *et al.* 2004), but little work has looked specifically at computer icon usability. However, past research has identified user characteristics that affect the usability of computer technology for this group, and we expect many of these characteristics also affect the usability of icons. These user characteristics include visual and verbal abilities, attention, and the capacity to learn and remember new information and associations. In addition, lack of experience with software interfaces by this cohort may also affect icon usability. Although our research did not focus on particular age-related user characteristics and their effects on icon usability, we were conscious of them so that we could control as much as possible these characteristics.

Past work has identified many icon characteristics that affect icon usability, such as physical characteristics (e.g. visual detail, colour and size), choice of icon object(s) associated with the intended icon meaning and how that object is depicted (e.g. dots vs. calendar) (Heim 2007). Researchers have also identified various user-related characteristics (e.g. intelligence, experience and culture) and the context in which the icon is found (e.g. mobile device capabilities, task and software application interface) as factors that influence icon usability (Horton 1994). A number of studies have looked closely at the effects on icon usability of some of these characteristics, such as animation (Baecker *et al.* 1991), and spacing and size (Lindberg and Nasanen 2003).

More closely related to our work, McDougall *et al.* (2000) investigated the effects of icon concreteness and visual complexity on tasks involving visual search and matching icons with labels. Visual complexity refers to the amount of visual detail or intricacy (e.g. lines and shading) in the icon. The icons used in that study were taken from a corpus of 239 icons that had been rated on a number of icon characteristics. These icons included graphics from road signs and electronic symbols as well as computer icons. Participants were recruited from a local university; age was not reported and was not a factor in the study. McDougall *et al.* found that participants, upon first use, did worse on the tasks with abstract icons than concrete ones. The researchers suggested that 'concrete icons are likely to be most useful when icon learning needs to occur quickly or instantly' (McDougall *et al.* 2000, p. 304).

There has also been work looking at the effect of labels on initial icon usability. Wiedenbeck (1999) conducted an experimental study in which 60 undergraduate students with little computer experience were asked to operate a desktop email program using buttons that had icons, labels or a redundant combination of icons and labels. Usability was measured by correctness of the tasks performed, time to perform tasks and number of times the help facility was accessed. It was found that participants performed better with text labels than unlabelled icons during initial use. Participants performed similarly with labelled icons and labels alone on each of the three usability measures, but significantly better than with icons alone. Further, participants reported finding the labelled icons the easiest to use.

Our work built on McDougall *et al.*'s (2000) and Wiedenbeck's (1999) work by exploring age-related differences in the effects of concreteness and labels, plus semantic distance, on initial icon usability. Further, we extend their research by including in our study actual icons from existing commercially available mobile devices and looking at the icons' usability in context.

3. Qualitative exploratory study

To ground our experimental investigation of age-related factors that influence the usability of mobile device icons, we first carried out a qualitative exploratory study with 10 participants from three age groups (20s, 60s and 70s). All participants reported good or corrected vision, basic but not extensive computer experience, basic cell phone experience and little to no experience with more advanced mobile devices. For the study, we used a laptop computer to display enlarged (approximately two times) screen captures of icons selected from two commercial mobile devices, an HP iPaq rx3715 and a Palm® Treo™ 650. Participants were first required to look at each unlabelled icon, to identify the object(s) they saw in the icon and to indicate which device function they assumed to be associated with it. In addition, participants were asked to complete a series of tasks on each device, such as finding icons for a particular application (e.g. camera program) or function (e.g. 'the help function'). Finally, participants were shown the icons used for the same function on the iPaq and on the Treo™; they were asked to choose the one that they found more usable and to explain their choice.

The most important findings from this exploration were, first, that the older participants were less accurate in identifying icon objects and in interpreting icon meanings. To illustrate, the older participants were unable to correctly identify the rolodex card in the Contacts icon or interpret the meaning of the Tasks icon (see Figure 2). Second, when required to choose between paired icons and to explain their choices, participants tended to prefer and focus on icons depicting something concrete or familiar, or icons with an obvious link between the depicted object and associated device function. Although such a preference for concrete and familiar icons with semantically close meanings is not surprising, it seemed that this preference was more pronounced in older participants.

4. Experimental study – methods

On the basis of the findings of the qualitative exploratory study, we sought to understand more precisely through a controlled experiment the extent to which concreteness and semantic distance affect icon usability in young versus older adults. In addition, we aimed to understand the effect of labels on initial icon usability.

We hypothesised that older adults would find existing mobile device icons less usable than younger adults (H1). More importantly, we hypothesised that older adults would find the following types of icons significantly harder to use than younger adults: icons showing abstract objects compared with those showing concrete objects (H2); icons with semantically distant compared with close meanings (H3); and unlabelled icons compared with labelled ones (H4). In other words, we hypothesised that compared with younger adults, older adults would find it relatively easier to use concrete icons, icons with semantically close meanings and labelled icons.

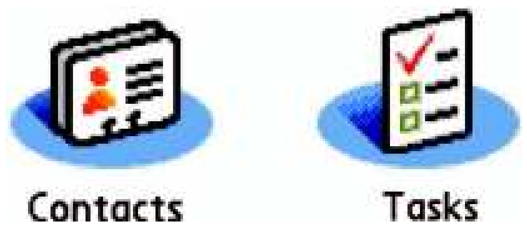

Figure 2. Example icons – 'Contacts' (left) and 'Tasks' (right) – from the Palm® Treo™ 650 which older adults found more difficult to use.

4.1. Experimental design

A 4-factor mixed design was used: age group (20–39 or 65+) was a between-subjects factor, whereas concreteness (concrete or abstract), semantic distance (close or far) and icon-label condition (icon-only, icon + label or label-only) were all within-subject factors. Presentation order of the icon-label conditions was fully counterbalanced whereas the presentation order of the three icon sets (described later) followed a Latin-square design.

4.2. Participants

Two groups of 18 participants, 20–37 years old (mean = 30.7) and 65 years old and older (mean = 71.5), were recruited (referred hereafter as our younger and older participants). One of the younger participants was replaced because her responses suggested that she did not understand the study tasks. Participants were recruited through advertisements placed in a free local newspaper and online classifieds. Interested potential participants were pre-screened over the phone to have at least basic computer experience (e.g. regular use of a computer, experience with internet browsers, email and word processing), functional eyesight, no colour blindness and fluency in English. In addition, interested participants were excluded if they had some experience using handheld computers, PDAs or advanced smart phone functions (e.g. messaging, browsing internet pages and taking photos), as we wanted our participants to be as unfamiliar as possible with the icons we used in the study. To confirm the screening results, eyesight and verbal fluency were tested at the study session using a reduced Snellen eye test and the FAS test (Benton and Hamsher 1978), respectively. The results from these tests showed participants had normal, age-appropriate eyesight and verbal fluency levels.

4.3. Measures

To test all four hypotheses, data were collected, one score per icon/label, for the following dependent variables:

- Accuracy in identifying icon object(s) (values: 0 or 1, no scores recorded for the label-only condition)
- Accuracy in interpreting icon/label meaning (values: 0 or 1)
- Confidence in interpreting icon/label meaning (self-reported on a scale of 1–4; 1 = 'not sure at all', 4 = 'very sure')

- Icon/label familiarity (self-reported responses to questions described in the Procedures section and mapped to values from 1–10; 1 = 'never seen before', 10 = 'seen and used on a similar mobile device')

In order to focus on initial usability, we aimed to recruit participants who were unfamiliar with the icons presented in the experimental study. Realistically, however, it was not feasible to find people who were completely unfamiliar with our icons or anything similar, so participants' familiarity with icons/labels was collected to assess, and possibly control, its influence on the usability scores.

4.4. Materials

Three sets of 20 icons were used in the study; one set was needed for each of the three icon-label conditions presented in a study session. Icons were selected from 149 icons of existing popular mobile devices (Sony Ericsson W610i, W850i Blackberry® 7730, Nokia N95, HP iPaq rx3715, Palm® Treo™ 650, Apple iPhone). The icons were obtained from high-quality screen captures posted on the Internet.

Independent raters, three HCI graduate students, rated the concreteness and semantic distance of each of the 149 icons. Using these ratings, 60 icons were chosen to represent the four concreteness–semantic-distance combinations, which are depicted in Figure 3. Figure 4 shows a sample of these icons. We arranged the icons into three sets, with each set consisting of six concrete + close, four abstract + close, four concrete + far and six abstract + far icons (there were fewer choices for icons that could be categorised as either abstract + close or concrete + far, hence the unequal number within each set). In addition, for each set of 20 icons, we controlled for the number of icons that were selected from an application interface (i.e. icons on buttons to operate an application; see Figure 5) and from a menu list (e.g. list of preference settings, list of programs that could be launched from the main menu; see Figure 6). To provide context, the application interface icons were presented in a screen capture that displayed all other icons used in the interface, and the

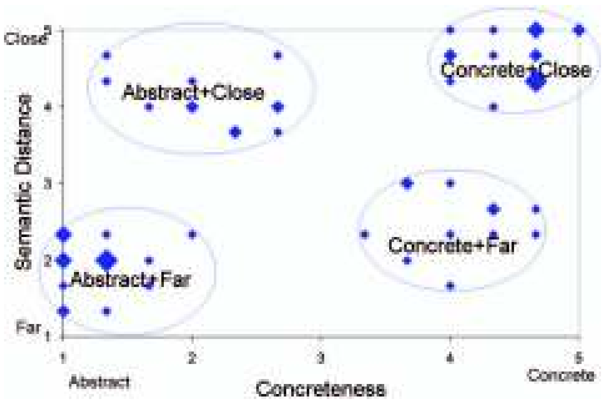

Figure 3. Average concreteness and semantic distance ratings, and associated groupings, for the 60 icons used in experiment; data point size reflects the number of icons characterised with a particular rating.

Figure 4. Example icons representing the four concreteness–semantic-distance combinations.

Figure 5. Screen capture of the HP iPaq rx3715 camera program showing example application interface icons.

Figure 6. Example pages for: icon + label (left), icon-only (inset top-right), and label-only conditions (inset bottom-right); icons are from the HP iPaq rx3715 program menu and are examples of menu item icons.

other menu icons were usually presented with at least one other icon from the menu.

For the label conditions, we used existing labels unless they included a product name (e.g. HP Image Zone, Quickoffice) or abbreviated words (e.g. Prefs). To avoid such labels, we replaced 5 out of the 60 icon labels with a more generic, unabbreviated word or phrase (e.g. Image Viewer, Document Editor and Preferences).

Most icons were enlarged to $1'' \times 1''$ (300 dpi) and smaller icons were enlarged to $0.5'' \times 0.5''$ (300 dpi). This was done to minimise effects of icon size because of individual differences in eye sight. Icons were presented to users on paper instead of a computer screen to eliminate any possibility of glare, to which many older adults are sensitive (Marmor 1998). Furthermore, icons were presented on paper instead of actual mobile devices so that our results would not be influenced by participants' varying abilities to operate mobile devices. This removed one more source of cognitive load that probably would have been heavier for the older participants, which would have made the overall icon interpretation task more difficult for them.

Icons were grouped by mobile device brand and by menu/application, and there were six pages of icons (1–8 icons/page) per icon set. Pages were created for each of the three icon-label conditions for each icon set (e.g. see Figure 6). The six icon set pages were randomly ordered for each session.

4.5. *Procedures*

All study sessions took place in a usability lab on the university campus, and were recorded using a video camera. After giving consent, ~5 min was spent familiarising participants with the functional capabilities of existing mobile devices (e.g. wireless connectivity, taking and viewing digital photos, and contact information management). Participants were asked to list the capabilities that they knew of, and were informed afterwards by the experimenter of other existing capabilities. Participants were also given a reference sheet, which listed general capabilities of mobile devices, that they could use if desired.

For the rest of the study session, participants were shown the three sets of icons and were asked a series of questions for each icon/label that was presented (60 total). Icons were shown in three blocks, one block per icon set paired with a different icon-label condition. The specific purpose/function of each application and menu list was described to participants, as the associated icons were presented, to help participants interpret the icons in context (e.g. for application shown in Figure 5: 'This camera application is used to capture photos. This is the viewfinder and these are the buttons used to operate this application.'; for menu shown in Figure 6: 'This menu lists a number of programs that a user can run on the device').

In the icon-only and icon + label conditions, participants were asked:

- To identify the icon objects ('What is shown in the icon?');
- To interpret the icon's function ('How might the icon be used?' or 'What would happen if you "clicked" on the icon?');
- How familiar they were with the icon ('Have you seen this particular icon before? If so, have you used this icon before?'); and,
- How confident they were about their interpretation ('How sure are you of the icon's function?').

In the label-only condition, participants were asked:

- To describe the function associated with the label;
- How familiar they were with the label ('Have you seen and used this exact label before, say on a computer, mobile device, etc.?'); and,
- How confident they were about their interpretation ('How sure are you of the label's function?').

4.6. Scoring

Participants' accuracy in identifying icon objects and in interpreting icon meanings were scored during the study session by the experimenter (first author), and then separately by an independent rater using the session video recordings. The agreement percentages for the two sets of icon object identification and icon interpretation scores were 88% and 92%, respectively, corresponding to Cohen's kappa scores of 0.47 ($p < 0.001$) and 0.81 ($p < 0.001$). All scores on which the two raters disagreed were discussed and resolved by consensus.

To assess whether the icon object(s) were correctly identified and whether the icon meaning was correctly interpreted, the raters focused more on the expressed concepts and ideas, rather than on whether the right technical words or phrases were used. For example, for the icon used to exit the camera program, responding with either to get out of program or to close program would be scored as correct. For icons showing abstract objects, participants could identify the objects by naming the shapes (e.g. dots and arrows) and graphical features (e.g. lines). Care was taken during the study session to not let participants know whether or not they had given the correct answer.

5. Experimental results

We used ANOVAs to test our hypotheses. For icon object identification accuracy, we ran a 2(age group) × 2(concreteness) × 2(semantic distance) × 2(icon-label condition: icon-only, icon + label) ANOVA. For both icon interpretation accuracy and interpretation confidence, we ran a 2(age group) × 2(concreteness) × 2(semantic distance) × 3(icon-label condition) ANOVA.

Whenever a statistically significant interaction was found, we followed up with post hoc pair wise comparisons, using Bonferroni correction to protect against Type I error. In addition, Greenhouse–Geisser corrections were used when sphericity was an issue. Using this correction can result in degrees of freedom that are not whole numbers. We also report the eta-squared (η^2) statistic, a measure of effect size, which is often more informative than statistical significance in applied human–computer interaction research (Landauer 1997). To interpret this value, 0.01, 0.06 and 0.14 are considered small, medium and large effect sizes, respectively (Cohen 1988).

5.1. Confidence and familiarity

As expected, participants' confidence was significantly higher when they correctly interpreted an icon's meaning (mean = 3.14, standard deviation = 0.7) than when they gave an incorrect interpretation (mean = 2.05, standard deviation = 0.9) (paired $t_{(285)} = 16.8$, $p < 0.001$). When reporting interpretation confidence scores, we only use the scores for icons whose meanings were correctly interpreted, as we were less interested in participants' confidence when they incorrectly interpreted an icon. Some participants did not correctly interpret any icons for a particular experimental condition and thus did not have confidence scores for each experimental condition; this is reflected in the lower degrees of freedom in the related ANOVA results.

We asked participants to report their familiarity with the icons/labels in order to assess the influence of familiarity on icon usability scores. An ANOVA of the familiarity scores showed that our younger participants gave significantly higher familiarity scores (average of 4.8 out of 10) than our older participants (average of 2.6 out of 10) ($F_{(1,34)} = 21.5$, $p < 0.001$, $\eta^2 = 0.39$). Because of this difference, we examined the influence of familiarity on (i.e. correlation with) each of our primary dependent usability measures. We found that familiarity was significantly correlated with each of our three usability measures, but the amount of variance that could be attributed to familiarity (r^2) was relatively small, 16% or less, on the three dependent measures (see Table 1). Nevertheless, because familiarity had a significant effect on our usability scores, we reran all our ANOVAs controlling for its effects by treating it as a covariate. We found that even with familiarity controlled, most of the significant effects of the other independent variables remained. Thus we present the ANOVA results based on our unadjusted usability scores, and note when significant effects were different in the analyses where the scores were controlled for familiarity.

5.2. Overall usability

As predicted by H1, compared with our younger participants our older participants were significantly less accurate in identifying icon objects and interpreting icon meanings (i.e. significant and very large main effects of age). Interpretation confidence scores did not differ significantly between the two age groups, suggesting that when icons were correctly interpreted, both groups of participants felt equally confident in

Table 1. Correlation between perceived familiarity and icon usability measures, r (Pearson correlation) and r^2.

	r	r^2	p	N
Identification	0.27	0.07	<0.001	288
Interpretation	0.40	0.16	<0.001	432
Confidence	0.35	0.12	<0.001	398

their interpretation. The results are summarised in Table 2.

5.3. *Effect of concreteness, semantic distance and labels*

5.3.1. *Identification accuracy*

We found significant age-related effects of concreteness and semantic distance on icon identification accuracy. Specifically, a significant three-way interaction among age group, concreteness and semantic distance was found ($F_{(1,34)} = 8.44, p = 0.006, \eta^2 = 0.02$), as shown in Figure 7. Both age groups performed worse on abstract-far icons than concrete-far and abstract-close icons ($p < 0.001$ for both groups), but the decline was greater for our older participants, supporting hypotheses H2 and H3. No significant age-related effects of concreteness were found for icons that were semantically close, but this may be because of ceiling effects in our younger participants' results. In addition, no significant age-related effects of labelling on icon object identification was found, offering no support to H4. Mean scores and standard deviations for concreteness, semantic distance and icon-label conditions are summarised in Table 3.

5.3.2. *Interpretation accuracy*

An analysis of the interpretation scores did not find a significant interaction between age group and concreteness ($F_{(1,34)} = 0.53$, $p = 0.47$, $\eta^2 < 0.001$), as shown in the left panel of Figure 8. This lack of interaction offers no support to H2.

As hypothesised in H3, a significant age-related effect of semantic distance on icon interpretation was

Table 2. Means, standard errors, *F*-values and effect sizes on dependent usability measures ($N = 36$).

	Mean scores (Std error)					
Effect of age	20–39	65+		F	p	η^2
Identification	96% (1.3%)	86% (1.3%)		$F_{(1,34)} = 29.4$	<0.001	0.47
Interpretation	75% (1.9%)	60% (1.9%)		$F_{(1,34)} = 27.1$	<0.001	0.44
Confidence	3.14 (0.125)	2.95 (0.158)		$F_{(1,11)} = 0.90$	0.36	0.08
Effect of concreteness	Concrete	Abstract		F	p	η^2
Identification	93% (1.1%)	89% (1.3%)		$F_{(1,34)} = 6.99$	0.01	0.02
Interpretation	66% (1.7%)	69% (1.4%)		$F_{(1,34)} = 5.51$	0.03	0.004
Confidence	3.05 (0.102)	3.03 (0.108)		$F_{(1,11)} = 0.10$	0.76	<0.001
Effect of sem. distance	Close	Far		F	p	η^2
Identification	95% (0.8%)	87% (1.4%)		$F_{(1,34)} = 31.77$	<0.001	0.08
Interpretation	84% (1.0%)	51% (2.1%)		$F_{(1,34)} = 279.6$	<0.001	0.31
Confidence	3.34 (0.084)	2.75 (0.125)		$F_{(1,11)} = 70.38$	<0.001	0.21
Effect of labels	Icon-only	Icon + label	Label-only	F	p	η^2
Identification	88% (1.5%)	94% (0.8%)	–	$F_{(1,34)} = 15.4$	<0.001	0.04
Interpretation	45% (2.0%)	78% (1.5%)	79% (1.8%)	$F_{(2,68)} = 192.9$	<0.001	0.32
Confidence	2.58 (0.086)	3.29 (0.146)	3.26 (0.117)	$F_{(2,22)} = 26.1$	<0.001	0.26

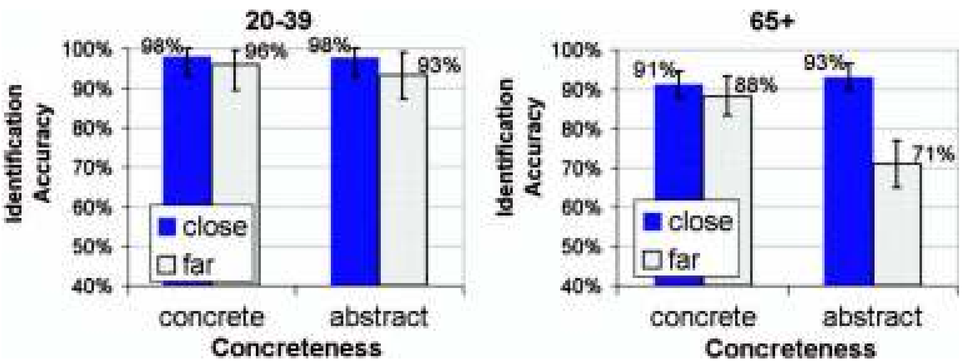

Figure 7. Mean icon object identification accuracy scores and 95% confidence intervals by age group, concreteness and semantic distance ($N = 36$).

Table 3. Mean icon usability scores, and standard deviations in brackets, for concrete and semantic distance conditions, and icon-label conditions ($N = 36$).

	Concrete + close	Concrete + far	Abstract + close	Abstract + far	Icon-only	Icon + label	Label-only
Identification accuracy							
20–39	98% (8%)	96% (12%)	98% (7%)	93% (12%)	95% (12%)	98% (7%)	
65+	91% (13%)	88% (18%)	93% (13%)	71% (22%)	82% (21%)	90% (16%)	
Interpretation accuracy							
20–39	84% (17%)	63% (34%)	93% (14%)	62% (28%)	52% (32%)	86% (16%)	87% (15%)
65+	75% (18%)	43% (33%)	84% (21%)	39% (25%)	38% (30%)	71% (27%)	73% (25%)
Confidence							
20–39	3.46 (0.43)	2.96 (0.67)	3.43 (0.47)	2.71 (0.77)	2.71 (0.77)	3.27 (0.61)	3.39 (0.47)
65+	3.25 (0.46)	2.75 (0.86)	3.24 (0.61)	2.90 (0.77)	2.61 (0.76)	3.27 (0.67)	3.22 (0.56)

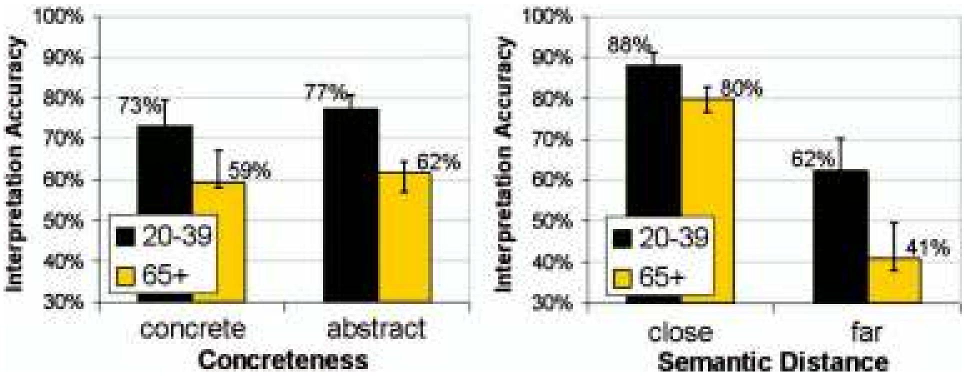

Figure 8. Mean icon interpretation accuracy scores and 95% confidence intervals by age group and concreteness (left) and age group and semantic distance (right) ($N = 36$).

found (i.e. a significant two-way interaction between age group and semantic distance, $F_{(1,34)} = 11.9$, $p = 0.002$, $\eta^2 = 0.013$). Both age groups performed significantly worse on semantically far than close icons ($p < 0.001$ for both groups), as shown in the right panel of Figure 8. However, the decline was greater for our older participants, revealing that they correctly interpreted significantly fewer semantically far icons.

Although both age groups correctly interpreted significantly more icons when labels were present ($p < 0.001$ for both groups), no significant age-related effect of labels was found. A significant three-way interaction of age group, icon-label condition, and semantic distance was found on icon interpretation scores ($F_{(2,68)} = 3.9$, $p = 0.025$, $\eta^2 = 0.006$), but this interaction was not significant when familiarity was controlled ($F_{(2,68)} = 1.1$, $p = 0.36$, $\eta^2 < 0.001$). This is our only reported result where a significant effect was no longer significant after we controlled for familiarity.

5.3.3. *Confidence*

A significant four-way interaction of age group, icon-label condition, semantic distance and concreteness on interpretation confidence was found ($F_{(2,22)} = 3.5$, $p = 0.047$, $\eta^2 = 0.03$). To facilitate interpretation of this interaction, we ran follow-up three- and two-way ANOVAs, but did not find significant age-related effects of concreteness and/or semantic distance on interpretation confidence, which offers no clear support to H2, H3 or H4.

5.4. *Limitations*

One limitation to our study design was its potential for ceiling effects (i.e. correctly identifying or interpreting 100% of the icons). Our younger participants had perfect identification/interpretation scores 6% of the time while our older participants scored perfectly 0% of the time. The presence of some ceiling effects for younger participants' results may have reduced our study's power to find some age-related effects (e.g. Section 5.3.1). However, these ceiling effects do not negatively impact the validity of the significant effects that were found.

In our study we showed to participants enlarged icons and text labels on paper, and this might have impacted the ecological validity of the results. Enlarging the icons and text and presenting them on paper allowed us to ensure that they were as readable as possible and to reduce the influence of individual and age-group differences in eyesight, which we felt was an important first step in understanding icon characteristics. Future work will need to look at the extent to which our findings change, if at all, when different icon and label font sizes are considered. When incorporating icons into a mobile interface, the designer should consider the results of past research on effect of icon and text size on older users (e.g. Siek *et al.* 2005).

5.5. *Summary*

This study revealed the following critical findings concerning each of our hypotheses:

H1 Supported: Older participants had more difficulty than younger participants identifying icon objects and interpreting icon meaning.

H2 Partially supported: Concrete icons helped older participants, more than younger participants, to identify objects in semantically far icons, but there was no effect of concreteness on interpretation accuracy.

H3 Supported: Icons with semantically closer meanings helped older participants, more than younger ones, to better identify the objects in abstract icons, and to interpret more icon meanings.

H4 Not supported: Labels did not appear to help older participants, more than younger ones, to use icons.

6. Discussion

6.1. *Existing icons harder for older adults to use*

We found empirical evidence that older adults do have significantly more difficulty than younger adults with the initial usability of a variety of existing mobile device icons. Having similar mobile device experience and at least basic computer experience, our older participants could only correctly interpret 38% of our unlabelled icons and 71% of our labelled ones, whereas our younger participants did substantially better, interpreting 52% and 86%, respectively. Although both age groups were found to have trouble with similar types of icons, our results suggest that older adults may get stuck more often because they are not able to interpret many icons in an existing mobile device interface, whereas a younger person may be able to 'get by'. Difficulties with using icons, which leads to difficulties using the entire interface, may partly explain why older adults find mobile devices difficult to use (Kurniawan *et al.* 2006) and why they have been relatively slower to adopt mobile devices. This suggests that there is a strong need to redesign some existing icons in order to make mobile device interfaces easier for older adults to use, especially for devices that can help improve their quality of life.

We found that abstract icons, icons with semantically far meanings and unlabelled icons were especially difficult for seniors to use. We next discuss our findings and the implications for designing better icons for seniors.

6.2. *Concrete icons had little effect on icon interpretation*

We found evidence that concrete icons help older adults to identify more icon objects with semantically far meanings, but did not have an age-related effect on icon interpretation. In fact, when looking at the two age groups together, participants correctly interpreted significantly more abstract icons than concrete icons (as shown in Table 2 and Figure 8). Given the strong effect of concreteness on increasing initial icon usability reported in McDougall *et al.* (2000), we were interested in understanding why our concrete icons were not always easier than our abstract ones to use (e.g. icon interpretation by our younger participants).

We believe that this discrepancy may be due in part to the difference between the way icons were rated on concreteness in McDougall *et al.* (2000) and in the research we present here in this paper. According to our definition of concreteness, our concrete icons included representations of real-world objects, places or people, but icons with symbols, even those commonly used in the real world that had precise well-established meanings (e.g. math symbols and musical notes), were classified as abstract icons. The icons used in McDougall *et al.* (2000) also depicted commonly used symbols, but they were often rated as being more concrete than abstract. As a result, familiarity with McDougall *et al.*'s set of icons (unlabelled) was reported to be highly correlated with concreteness ($r = 0.78$) (McDougall *et al.* 1999), whereas familiarity with our set of icons (unlabelled) was found to be much less correlated with concreteness ($r = 0.15$). This shows that the way one defines icon concreteness impacts how concreteness relates to familiarity and its effect on icon usability. More generally, this shows that defining icon characteristics is non-trivial and that slight variations in definitions can result in large differences in study outcomes. Further research aimed at achieving standardised definitions for icon characteristics is required to address this issue.

6.3. *Icons with semantically close meanings are much easier for older adults to use*

We found that icons with semantically further meanings were generally much harder for our older participants to use. As presented earlier, our older participants correctly interpreted 80% of our semantically close icons, compared with only 41% of our semantically far icons. Further, our older participants had significantly more difficulty than our younger participants in interpreting icons with semantically far meanings. On the basis of our findings, semantic distance appears to have a larger effect on initial icon usability than concreteness.

Why would older adults have more difficulty with the meaning of semantically far icons (and icons in general)? We argue that their difficulties may be related to older adults' difficulty in forming and using mental models, which has been reported in the literature (Freudenthal 1998, Ziefle and Bay 2004). Users often rely on their understanding of the system and how it operates (i.e. their mental model) to help 'cross' the semantic distance from the icon object to its meaning. In other words, icons, especially those with semantically far meanings, generally require an accurate mental model of the system that the user can apply in interpreting the icon. For example, one needs to know that a device can perform calculations in order to know how to interpret a calculator icon. One also needs to know that a device can and is sometimes used to compress files, in order to correctly interpret the 'Zip' icon in Figure 4 showing a clamp. Although we spent time during the experiment reviewing common mobile device capabilities and chose participants with similar levels of computer and mobile device experience, our older participants may have had more difficulties remembering many of the new functions that we introduced to them, or had difficulties applying their mental model of the device to interpreting the icons. A comment by one of our older participants that one needed 'umpteen degrees to keep up with technology' suggests that he did have difficulty developing accurate mental models of new technology.

We have found evidence that icons with semantically far meanings are particularly difficult for older adults to learn to use, and should be redesigned. One method to reduce semantic distance is to choose icon objects with semantically close associations to the icon meaning. We describe in this section other ways for reducing semantic distance.

6.3.1. *Use familiar metaphors*

We observed in this research that familiarity with an icon often plays a large part in being able to use it. Existing icon design guidelines suggest using images and metaphors that are familiar to the target user. However, an icon's familiarity depends on an individual's experience. Commonly used computer metaphors (e.g. disk for saving and wrench for device options) may not always be known to older adults, who generally have less experience with computers. Instead, when designing icons for older users we suggest using everyday metaphors with which they are familiar. If it is not feasible to only use familiar metaphors, one should ensure that the metaphors used in the interface are taught to the user, perhaps through documentation (reference card) or by someone (e.g. care giver or customer support). One of our older participants commented that she was very interested in learning the commonly used metaphors used in computer icons.

Some may argue that older adults' lack of familiarity with commonly used computer metaphors will no longer be an issue in the future when the upcoming generations, who generally have relatively more computer experience, become older adults themselves. Hawthorn (2002) counters that there are still many younger adults who have jobs or home situations where they do not interact much with computers. In addition, he points out that even if older working adults have opportunities to keep up to date with new computer technology, they may be less able to adjust to the changes. Further, once retired, older adults may have less need to keep up to date with new technology. Computer metaphors will continue to evolve and we therefore expect future generations of older adults to continue to have more difficulty than younger adults staying current with these metaphors.

6.3.2. *Label icons*

We found that labels greatly help both young and old to initially use icons. Icon object identification and interpretation performance in the icon-only condition was generally worse than the two label conditions (i.e. icon + label and label-only), whereas performance in the two label conditions was similar, which is consistent with Wiedenbeck's findings (1999). Although we did not find that labels helped our older participants significantly more than our younger ones to interpret icons, we expect that labels provide greater benefits to the general older adult population (that likely has on average less computer experience than the older participants in our study). In fact, three of our older participants commented that they usually interpreted the label before the icon. On the basis of our findings, we suggest that all mobile device icons be labelled for older users at least initially. It would be interesting to explore whether older adults continue to rely on labels over continued use of the icons.

Although the presence of labels greatly improves icon usability, it may not always be feasible to label all mobile device icons because of the device's limited screen real-estate. However, there are a number of alternative labelling techniques that may be suitable. For example, some mobile interfaces show only one label at a time (usually for the selected icon) in a designated area in the interface, or have 'tool tip' labels pop up above the highlighted/selected icon. Past research has found that older adults have poorer visual spatial abilities (Czaja 1997), and thus research is needed to see which labelling technique works best with this population.

6.3.3. *Allow user to select icons*

Our results highlight the fact that an icon that is easy for one person to use may be difficult for another. We propose that a mobile interface not just provide one icon for the different functions in the interface, but allow users to choose, from a set of icons for each function, an icon that they each feel is most suitable. A mobile device interface could provide, for example, a variety of icons associated with the 'device options' function portraying different objects such as a wrench, control knobs, form radio buttons or a graphic of the device itself. The semantic distance between an icon and its meaning can be different for each individual as it depends in part on the user's familiarity with the objects depicted and metaphor used in the icon. A feature to allow users to select icons is more suitable for a mobile device interface, compared with a desktop computer interface, because only a relatively small set of icons would need to be selected. The process of choosing suitable icons for an application could be supported by a software wizard to minimise the time and effort required from the user. The selection of icons could also be performed by a loved one or care giver. A number of existing commercial desktop and mobile assistive technologies for people with communication disorders, many who are older, currently help users to personalise the interface's icons (e.g. Pocket Communicator (Gus Communications 2009), Lingraphica (Lingraphicare 2009)). We expect that giving older adults the option to choose their icons would improve both initial icon usability and usability over time, and younger users might benefit from this feature as well.

6.4. Long-term icon usability

Having explored initial icon usability, future work should look at the effects of icon characteristics on the usability of these icons over time or over multiple exposures. Older users probably interact less frequently with their devices, decreasing the frequency of contact with icons, making it more important to design icons whose meanings are easy to remember as well as learn. McDougall *et al.* (2000) found that, for their university student participants, abstract icons tended to become as usable as concrete icons over multiple exposures. Wiedenbeck (1999) also found that undergraduate students had more problems retaining the meaning of unlabelled icons over a 1-week period, compared with labelled icons and labels without icons (i.e. label-only), but that these difficulties diminished over further exposures. We are interested in seeing whether such learning effects can also be found in older populations.

On the basis of the literature and our findings, we speculate that there are a number of age-related differences in icon usability over time. Older adults have been found to have more difficulty learning new associations (Chalfonte and Johnson 1996, Naveh-Benjamin *et al.* 2004), which may make it more difficult to learn unfamiliar abstract icons or those with semantically far meanings. In addition, the recall of associations learned in old age has been found to be harder, especially in multitasking situations (Hawthorn 2000), which are common in mobile contexts. Older adults may also have more difficulty than younger adults in retaining learned meanings of unlabelled icons and may continue to rely on icon labels longer than younger adults. We also speculate that concrete icons, which were found to help older adults to identify more icon objects, may be more effective as cues for recalling associations and be easier for older adults to use over time. Research is needed to validate these speculations.

7. Conclusions

In this research, we sought to find empirical evidence for age-related effects of salient icon characteristics on icon usability, in order to learn how to better design icons for older adults and develop icon design guidelines that account for a user's age. Through a qualitative exploratory study, we observed older adults having more difficulty using unfamiliar mobile device icons, and identified, based on their comments, some icon characteristics that helped or hindered their icon usability.

A follow-up controlled experiment was conducted to better understand the effects of these characteristics, specifically concreteness, semantic distance and the presence of labels. We found that these characteristics did impact initial icon usability, but some more than others. From the experiment results, we found that older adults were able to interpret significantly fewer existing mobile device icons than young adults, particularly those with semantically far meanings. We also found that older adults found it more difficult to identify objects in abstract icons with semantically far meanings. Further, we found that unlabelled icons were more difficult for both age groups to use.

We concluded that many icons need to be redesigned in order for older adults to be able to use the icons and their interfaces, and presented a number of reasons why good icon and interface usability is particularly important on mobile devices used by this population. We suggested, based on our results, that icons incorporate concrete objects or commonly used symbols. More importantly, we suggested reducing semantic distance by choosing icon objects semantically close to the icon meaning, using familiar metaphors, using labels and allowing users to choose an icon from a set of potentially suitable icons. Although our empirical results are consistent with many existing icon design guidelines, the results highlight related guidelines that, when followed, should improve an older user's initial usability of icons on mobile devices and other computer interfaces.

Existing and future mobile devices offer older adults many opportunities to remain active and increase their independence. By making mobile device icons easier for older adults to use, we expect that the overall device will be more usable and will have a better chance of being adopted.

Acknowledgements

This research was supported in part by the Canadian Institutes of Health Research and the Natural Sciences and Engineering Research Council of Canada.

References

Baecker, R., Small, I., and Mander, R., 1991. Bringing icons to life. *In*: S.P. Robertson, G.M. Olson, and J.S. Olson, eds. *Proceedings of SIGCHI conference on human factors in computing systems (CHI)*, 27 April – 2 May 1991. New Orleans, Louisiana, USA. New York, NY, USA: ACM Press, 1–6.

Benson, C., *et al*., 2005. *Designing effective icons* [online]. The GNOME Usability Project. Available from: http://developer.gnome.org/projects/gup/hig/hig-diff/icons-design.html [Accessed 24 February 2008].

Benton, A.L. and Hamsher, K., 1978. *Multilingual aphasia examination manual*. Iowa City: University of Iowa.

Chalfonte, B.L. and Johnson, M.K., 1996. Feature memory and binding in young and older adults. *Memory and Cognition*, 24 (4), 403–416.

Cohen, J., 1988. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale, NJ: Erlbaum.

Czaja, S.J., 1997. Computer technology and the older adult. *In*: M. Helander, T.K. Landauer, and P. Prabhu, eds. *Handbook of human-computer interaction*. 2nd ed. Amsterdam: Elsevier, 797–812.

Docampo Rama, M., de Ridder, H., and Bouma, H., 2001. Technology generation and age in using layered user interfaces. *Gerontechnology*, 1 (1), 25–40.

Fisk, A.D., *et al*., 2004. *Designing for older adults: principles and creative human factors*. London: CRC Press.

Freudenthal, T.D., 1998. *Learning to use interactive devices: age differences in the reasoning process*, Thesis (PhD). Technische Universiteit Eindhoven.

Goodman, J., Brewster, S., and Gray, P., 2004. Older people, mobile devices and navigation. *In*: J. Goodman and S. Brewster, eds. *HCI and the older population, workshop at the HCI 2004*, Leeds, UK, 13–14. Available from: http://www.dcs.gla.ac.uk/~stephen/research/utopia/workshop/goodman.pdf.

Goodman, J. and Gray, P., 2003. A design space for location-sensitive aids for older users. *In*: *Proceedings of workshop on HCI in mobile guides*, 8 September 2003, Udine, Italy, 12–16. Available from: http://www.comp.lanes.ac.uk/~kc/mguides03/Goodman-final.doc.pdf.

Gregor, P., Newell, A.F., and Zajicek, M., 2002. Designing for dynamic diversity: interfaces for older people. *In*: *Proceedings of the fifth international ACM conference on assistive technologies (ASSETS)*, 8–10 July 2002, Edinburgh, Scotland. New York: ACM Press, 151–156.

Gus Communications, 2009. *Gus! Pocket Communicator* [online]. Gus Communications. Available from: http://www.gusinc.com/pocketcomm.html [Accessed 1 June 2009].

Hawthorn, D., 2000. Possible implications of aging for interface designers. *Interacting with Computers*, 12 (5), 507–528.

Hawthorn, D., 2002. How universal is good design for older users? *SIGCAPH Computers and the Physically Handicapped*, (73–74), 38–45.

Heim, S., 2007. Icons. *In*: *The resonant interface: HCI foundations for interaction design*. 1st ed. Boston, MA: Addison Wesley, 407–451.

Horton, W., 1994. *The icon book: visual symbols for computer systems and documentation*. New York: John Wiley & Sons.

Inglis, E.A., *et al*., 2003. Issues surrounding the user-centred development of a new interactive memory aid. *Universal Access in the Information Society*, 2 (3), 226–234.

Jacko, J.A., *et al*., 2002. Macular degeneration and visual icon use: deriving guidelines for improved access. *Universal Access in the Information Society*, 1 (3), 197–206.

Kurniawan, S., Mahmud, M., and Nugroho, Y., 2006. A study of the use of mobile phones by older persons. *In*: *CHI '06 extended abstracts on Human factors in computing systems*, 22–27 April 2006, Montréal, Québec. New York: ACM Press, 989–994.

Landauer, T.K., 1997. Behavioral research methods in human-computer interaction. *In*: M.G. Helander, T.K. Landauer, and P.V. Prabhu, eds. *Handbook of human-computer interaction*. 2nd ed. Amsterdam: Elsevier Science B.V., 203–227.

Lindberg, T. and Nasanen, R., 2003. The effect of icon spacing and size on the speed of icon processing in the human visual system. *Displays*, 24 (3), 111–120.

Lingraphicare America, 2009. *Lingraphica product overview* [online]. Lingraphicare America. Available from: http://www.aphasia.com/patients/product.aspx [Accessed 1 June 2009].

Marmor, M.F., 1998. Normal age-related vision changes and their effects on vision. *Aging & Vision News* [online]. Available from: http://www.lighthouse.org/education-services/professional-education/patient-management/managing-the-patient-with-low-vision/normal/ [Accessed 27 March 2008].

McDougall, S.J., Curry, M.B., and de Bruijn, O., 1999. Measuring symbol and icon characteristics: norms for concreteness, complexity, meaningfulness, familiarity, and semantic distance for 239 symbols. *Behavior Research Methods, Instruments, and Computers*, 31 (3), 487–519.

McDougall, S.J., de Bruijn, O., and Curry, M.B., 2000. Exploring the effects of icon characteristics on user performance: the role of icon concreteness, complexity, and distinctiveness. *Journal of Experimental Psychology applied*, 6 (4), 291–306.

Moor, K.A., Connelly, K.H., and Rogers, Y., 2004. *A comparative study of elderly, younger, and chronically ill novice PDA users*. Bloomington, IN: Indiana University. Technical Report TR 595.

Naveh-Benjamin, M., Guez, J., and Shulman, S., 2004. Older adults' associative deficit in episodic memory: assessing the role of decline in attentional resources. *Psychonomic Bulletin and Review*, 11 (6), 1067–1073.

Ofcom, 2006. *Media literacy audit: report on media literacy amongst older people*. London: Ofcom. Available from: http://www.ofcom.org.uk/advice/media_literacy/medlitpub/medlitpubrss/older/older.pdf.

Peirce, C.S., 1932. *The collected papers of Charles Sanders Peirce*. Cambridge, MA: University Press.

Siek, K.A., Rogers, Y., and Connelly, K.H., 2005. Fat finger worries: how older and younger users physically interact with PDAs. *In*: *Proceedings of tenth IFIP TC13 international conference on human-computer interaction (INTERACT)*, 12–16 September 2005, Rome, Italy. Berlin/Heidelberg: Springer, 267–280.

Sun Microsystems, 1999. *Design guidelines: Application graphics* [online]. Available from: http://java.sun.com/products/jlf/ed1/dg/high.htm [Accessed 24 February 2008].

Wiedenbeck, S., 1999. The use of icons and labels in an end user application program: an empirical study of learning and retention. *Behaviour & Information Technology*, 18 (2), 68–82.

Ziefle, M. and Bay, S., 2004. Mental models of a cellular phone menu. Comparing older and younger novice users. *In*: *Proceedings of 6th international conference on human computer interaction with mobile devices (MobileHCI)*, 13–16 September 2004, Glasgow, Scotland. Berlin/Heidelberg: Springer-Verlag, 25–37.

Ziefle, M. and Bay, S., 2005. How older adults meet complexity: aging effects on the usability of different mobile phones. *Behaviour & Information Technology*, 24 (5), 375–389.