A complex relationship: Older people and in-car message systems

Mary Zajicek
Department of Computing, Oxford Brookes University,
Wheatley Campus, Oxford OX33 1HX, United Kingdom
E: mzajicek@brookes.ac.uk

Ing-Marie Jonsson
Department of Communication, Stanford University,
450 Serra Mall, Stanford, CA 94305, USA
E: ingmarie@csli.stanford.edu

M. Zajicek, I-M. Jonsson. A complex relationship: Older people and in-car message systems. Gerontechnology 2007; 6(2):66-78. The motivations and preferences of different user groups are important when designing for all and accessibility, in the sense of access to computer systems, is only one aspect of the enterprise. Ubiquitous computing by its very nature is difficult to evaluate in a laboratory setting and if an application is safety critical it is also difficult to evaluate in a real life scenario. This paper examines the part that context plays in the evaluation of a speech-based in-car messaging system when tested with older adults. The results show that older people are significantly influenced by the context in which evaluation takes place and in fact produced different evaluations of speech messages when heard on a laptop compared with within a simulated driving scenario. Younger people were far more consistent in their evaluation under both conditions, and less influenced by the change in context. These results reveal the complexity entailed when designing for all with in-car speech systems which include older people, and point to the need for extra care when testing ubiquitous computer systems for older people.

Keywords: in-car speech system, driving simulator, driving safety

Technology developers should be aware of the preferences of different user groups when designing for all. Many instances of technology exist these days which could be of great benefit to older people where unfortunately uptake is mainly by the young. In-car message systems, for example, under discussion in this paper, are available in many cars and could be of value to older people for orientation and help with driving. However, they are under-used by this age group.

This paper explores how age-related changes affect those facilities required for successful technology use and discusses why user-centred design and evaluation of computer-based systems for older adults creates new challenges. This holds especially for ubiquitous computing such as in-car message systems, rather than for desktop office based computing. In particular the paper looks at the way in which older people have different perceptions from younger people and are more sensitive to change of context when evaluating ubiquitous computer systems.

Experiments carried out with older adults and in-car speech message systems illustrate the complexity of the evaluation of ubiquitous computing with older adults. Previous experiments using a driving simulator to represent a driving scenario with in-car speech messages showed considerable difference in the preference of voice type between younger and older people[1]. This paper describes experiments using a laptop computer rather than a driving simulator. They are aimed at finding out more about what affects the acceptance of the voice that is used. The results concerning voice perception by older adults derived from the laptop experiments contradict those derived from the driving simulator experiments and serve to illustrate that design and evaluation of ubiquitous systems for older people is complex, and should be treated with care. The diversity of users' previous experience and perception can significantly affect their acceptance of a system.

Design for All therefore not only means that everybody should have access to computer systems but that the design is such that all users benefit equally from the use of the system once it has been made accessible.

Interface design

Age and ageing

Chronological age alone cannot be used to categorise a group of older people, as there exists considerable diversity in the rate at which older people experience the effects of ageing. Researchers have investigated the effects of ageing with differing results. Dulude[2] found that older people with whom she worked either experienced considerable effects of ageing or experienced none at all, with bi-polar results, whereas Hawthorn[3] found that the effects of ageing start to become noticeable from as early as the mid-40s, thus indicating the difficulty of even defining older people. The sample older subject groups in this study was selected not only by age, but also by attributes considered particular to the application. Thus in-car speech messages were tested with professional people aged 55 and over, who also held a valid driving licence. In contrasting previous work on memory support systems for older people[4], these were evaluated with older subjects who attended a day care centre for those who are just about able to live independently in their own home, and tended to be over 65.

The older adult subjects in the experiments described in this paper were over 55, currently held a driving licence and were from the higher socio-economic level more likely to purchase cars with in-car speech systems. This selection of subjects ensured that diversity was minimised. The younger group in these experiments was aged between 18 and 35 years.

Adults as they get older experience a wide range of age related impairments including loss of vision, hearing, memory and mobility, the combined effects of which contribute to loss of confidence and difficulties in orientation and absorption of information and, most significantly, reduced ability to build strategies at the computer interface.

Interaction

There exists significant legislative pressure for the development of systems that are accessible to older and disabled people. The 1990 Americans with Disabilities Act[5] asserts the individual's right to use products and services on an equal access basis and in 1995, the United Kingdom implemented the Disability Discrimination Act[6]. Although there is therefore currently increased awareness of the need for universal access, i.e., access for everybody to technology, many questions concerning suitable methodologies for design for this group remain unresolved. Designing for this group of users is not easy, and the cultural and experiential gap between designers and older people can be especially large when developing new technology[7].

User-centred design requires that the user is involved in each step of the design process, starting with gathering requirements for the system and then participating in the testing and evaluating of prototypes which are used to inform the design of the finished system. Younger people, through familiarity with the technology, can more easily participate as users in user-focused activities. Older people are commonly unaware of the possibilities of new technologies, and this can severely limit their

April 2007, Vol 6, No 2

www.gerontechjournal.net

ability to contribute actively to a discussion about their requirements.

Initial requirements for a system are commonly elicited using a focus group. This has proved challenging when working with older people as they have a tendency, in general, to stray from the point, experience tiredness and have difficulty hearing. There exist instances of successful use of focus groups with older people. Kirakowski[8], for example, reports instances where standard focus group procedures were used successfully for requirements elicitation with older people, and that no adjustments for this user group were required. However, more recent work has demonstrated that focus groups must be adapted for older people and that their organisation requires considerable interpersonal skills. For example, when gathering requirements for an interactive memory aid, researchers at Dundee University[9] reported difficulties in managing focus groups comprising more than three older people. They reported that auditory impairment was affecting older people's attention and the ability to follow a discussion, and that where depth and volume of information are important smaller groups or individual interviews were required.

Lines and Hone[10] also found that older people are inclined to 'wander' from the topic under discussion, providing unrelated anecdotes and chatting amongst themselves. They reported that it was difficult to keep the participants' attention focused on the task and felt that smaller numbers in sessions were preferable, allowing everybody time to contribute and those who appeared nervous to be drawn into the discussion more easily by the moderators. These comments highlight the challenges involved in defining systems requirements from older people's experience and perspectives.

Gathering interface requirements from older people therefore requires considerable skill and understanding of the user group. Newell and Gregor[11] also proposed that standard user-centred design techniques, which rely on relatively homogeneous user groups for user testing, should be replaced by User Sensitive Inclusive Design, which seeks out diversity, in order to ensure that systems are truly usable by older people.

In-car messages

The experiments described below were carried out as part of a larger project aimed at investigating the potential for older people to use in-car hazard warning messages to help them drive more safely and enjoyably. The focus is user acceptance, in this case acceptance of the voice to be used by the system. Driving a car is of course safety critical and a driving simulator with simulated in-car speech messages was used to gather information concerning the safety of the driving and reactions to the voice rather than using a real car which could be dangerous.

One limitation of the current research is that we used a driving simulator rather than an actual car. However, we believe that the results are nonetheless relevant because results obtained in driving simulator studies are indicative of likely patterns in actual cars, and are often used as the first screening of various car assistance systems and information systems. It is less costly to test systems in driving simulators during the development cycles, and as seen from this study, there are many more parameters that can be fine-tuned before testing the system in real cars. We cannot extrapolate to the actual levels of poor driving in real life, but we are confident that the patterns associated with the presence or absence of the in-vehicle systems do extrapolate to less difficult driving situations. Furthermore, it would be impossible to do this study with actual cars: There is no current system that can accurately or reliably judge road conditions or road hazards in real-time and then format and present the information through an in-car information system in an accurate and timely manner.

The authors were also aware that speech-based interactions, where voices and characteristics of voices carry both linguistic and paralinguistic cues, have potential to influence the behaviour of listeners. Different characteristics of voice are discussed below and are shown to be important contributors to success in in-car speech systems.

Voice perception

People easily detect characteristics in a voice when communicating with humans and also with speech-based computer systems[12] and this affects their perception of the credibility of the content of the speech and their enjoyment when listening. Studies indicate that both synthesized and recorded voices[13] influence content so that a happy voice makes content seem happier and a sad voice makes content seem less happy. Results also show that people prefer the content when voice characteristics match the content, but interestingly rated information as more credible when voice characteristics and content were mismatched. Credibility is generally associated with liking[14] and where voice characteristics and content were mismatched; people drew on their experience of communication with people to understand the mismatch. When interacting with people, the mismatch occurs when emotional content is read in a neutral voice to reflect objectivity. The mismatch might create a sense of detachment and hence appear more objective and credible. Follow-up studies[12] show that credibility and persuasion is higher also for a mismatch between the personality of the voice of a speech-based computer system and the personality of the person interacting with the systems. It has also been shown by Lazarsfeld and Merton[15] and Rogers and Bhowmik[16], that better human communication occurs between a source and a receiver who are alike, i.e., homophilous and have a common frame of reference. Individuals enjoy the comfort of interacting with others who are similar. Talking with those who are markedly different from us requires more effort to make communication effective.

Characteristics of voice can also influence people's attention, and affect performance, judgment and risk-taking. Jonsson et al.[1] have shown that the selection of voice has considerable impact on older drivers and previous studies show that information provided by in-car systems have the potential to improve driving performance[17] and that the linguistic and para-linguistic properties of the in-car voices influence driving performance[18].

Rationale

Speech systems can provide useful information for older people about the environment and things happening around them, which they may not readily absorb for themselves. Voice prompt speech messages have been used successfully to provide reminders concerning previous interaction for those with poor memories[19] and enabled older people to get going on a computer system where they had failed before. In studies with Web browsing and speech systems for older adults[4] and other work concerning explanation messages in computer interaction[20] the nature and quality of the voice messages was found to be very important.

Ageing affects short-term memory and the ability to absorb general background contextual information, together with the ability to multi-task. Loss of memory and general awareness can cause a decrease in confidence in one's actions and reluctance to try new things. Confidence boosting and affirmative speech prompts have proven to be very useful for older adults when using a speech browser[19]. Therefore there is reason to expect that speech-based in-car information messages will instil confidence in older adults when driving and contribute to driving safety and enjoyment.

The ability to absorb information also decreases with age. Older people were found to be less able to absorb long instructions than younger people[19]. They completely forget the messages or were unable to remember them at the required time. Speech based support that is invoked exactly at the point when it will be useful removes the need for long instructions to be given at the beginning of a task.

Speech messages can therefore compensate for memory loss by suggesting actions that have not been remembered, help with strategising by making contextually relevant suggestions and provide contextually relevant advice, i.e., advice about road conditions and provide warnings in safety-critical situations. The user's perception of the speech-based support system is in many cases as important as its functionality. It must be attractive to the user and engender feelings of trust and confidence in the information provided; otherwise it will not be accepted or used.

Driving simulator

The paper by Jonsson et al.[1] reports in full the experiments on the effectiveness of voice messages carried out using a driving simulator used to simulate a drive with speech messages warning of upcoming hazards. The results clearly demonstrated that driving safety in the driving simulator was enhanced with the use of in-car systems, with the younger voice providing better performance.

Older adults actually perceived that the in-car system with the younger voice made them drive better, found the young voice to be more credible than the old voice and felt more at ease after driving with the young voice than the old voice. These experiments demonstrate that there is significant potential for increasing the safety of drivers (over 55 years of age) by providing information concerning road hazards, and that these notifications are well-received by the drivers. The experiments also demonstrate that the choice of voice for the system is very important. It was therefore necessary to investigate further the most important aspects of the voice and what makes it successful for in-car speech systems. Further investigations of voice were carried out using a laptop computer in isolation, without the driving simulator context.

Voice only

The goal of the voice only experiments described in this paper was to further identify the characteristics of the voices, used by the in-car information system that can influence drivers' attitude to the voice and thus their driving performance. The perception of the voices as well as the perception of the persons speaking was examined, to identify characteristics of voice which explain the significant impact on driving performance found in the experiments using a driving simulator where older drivers showed a significant preference for a 20 year old woman's voice to that of a 73 year old woman.

Specifically answers were required to the following questions concerning the younger and older women's voices used in the driving simulator experiment: (i) Are there differences in the emotional colouring, the perceived trust, and credibility between the two voices?; (ii) Is there any difference in the voice quality between the two voices?; (iii) What are the perceived age, background and attitude of the persons speaking?

The experiment was a 2 (age group: 18-25 years of age, and 55 years of age and older) x 1 (two voices), balanced (for gender and order) between-participants, design. The younger and older women's voices were chosen to eliminate external factors.

They were both middle class and spoke well enunciated Queen's English. They also spoke at the same speed and with the same inflection.

Participants
The subject groups were selected to match the two age groups (18 - 25, 55 and older) who participated in the driving simulator study that initiated this experiment. Neither the younger nor the older subjects had previously driven the driving simulator and had therefore not heard the voices before. The younger group comprised 18 students of a variety of levels of IT competence. The older group comprised of 18 older people who currently held a driving licence and were from the higher socio-economic level where people were more likely to purchase cars with in-car speech systems. All participants volunteered their time for their participation, gave informed consent and were debriefed at the end of the experiment.

Procedure
Participants listened to the same two younger and older women's voices used in the driving simulator experiments while filling in a set of questionnaires using pen and paper. They were randomly divided into two groups, one group that listened to and rated the young voice first, and one group that listened to and rated the older adult voice first. They used a laptop with Microsoft Mediaplayer and earphones where they could select and play 26 short voice prompts, recorded in the two different voices. These included: (i) There is thick fog ahead, (ii) You are approaching an intersection, (iii) Warning! There is a fallen tree in the road ahead, (iv) Beware of cyclists ahead, (v) The current speed limit is 60 miles an hour, (vi) There are crosswinds in this area, (vii) Stop sign ahead, (viii) The police use radar here, you might need to slow down, (ix) There is heavy traffic ahead, turn left to avoid it, and (x) There is an accident ahead, turn right to avoid it.

Participants rated the voices using questionnaires investigating different aspects of voice: emotional colouring; trust, credibility; quality and homophily, i.e., how much participants perceived the persons speaking as similar to themselves in terms of background and attitude. All participants were informed that the experiment would take 30 minutes, and that they could play the recordings at any time and as often as they liked while they filled in the questionnaires.

Emotional colouring
The positive emotional colouring of each voice was measured using a variation of the Differential Emotion Scale (DES)[21]. A positive emotion index was used based on a questionnaire with a 20-term DES, using the terms happy, delighted, enthusiastic, amused, curious, attentive, alert and interested in a 10-point Likert scale (1 = describes very poorly to 10 = describes very well). The index was very reliable (Cronbach's α = .77).

The negative emotional colouring of each voice was measured using the same questionnaire with the 20-term DES. A negative emotion index was created using the terms angry, aggressive, hostile, mad, distressed, sad, upset, and unhappy in a 10-point Likert scale (1 = describes very poorly to 10 = describes very well). The index was very reliable (Cronbach's α = .79).

Trust of voice
Trust of voice was measured using a standard Individualized Trust questionnaire[22]. Participants were asked to rate a number of adjectives based on the question "How well does each of the following words describe the voice you just heard?" Contrasting adjectives were paired on opposite sides of a 10-point scale such that, for example, reliable and unreliable would appear at different ends.

Credibility of voice
Credibility of voice was measured using

a standard Source Credibility questionnaire[22]. Participants were asked to rate a set of adjectives based on the question "How well does each of the following words describe the voice you just heard?" Contrasting adjectives were paired on opposite sides of a 10-point scale such that, for example, qualified and unqualified would appear at different ends. Four standard measures from Berlo's and McCroskey's credibility scales[22] Authoritativeness, Character, Qualification and Dynamism, were created.

Quality of voice

Quality of voice was measured using a questionnaire where participants were asked to rate adjectives based on the question "How well does each of the following words describe the voice you just heard?" Contrasting adjectives were paired on opposite sides of a 10-point scale such that, intelligible and inarticulate would appear at different ends. The questionnaire was used to create one index, clarity of voice comprised of the terms intelligible, clear, non-breathy, fluent and enunciated. The index was very reliable (Cronbach's α = .79). Participants were also asked to judge the age of the person speaking for both the young voice and for the older adult voice.

Similarity of voice

A standard questionnaire on homophily[22] was used to identify measures of similarity. We created three indices, attitudinal similarity, behavioural similarity and similarity as a combination of attitude and behaviour. Participants were asked to rate the statements based on the questions "On the scales below, please indicate your feelings about the person speaking?" Contrasting statements were paired on opposite sides of a 10-point scale such that, 'similar to me' and 'different from me' would appear at different ends.

Table 1. Comparison of rating of voices by age group; bold indicates statistically different means at a confidence level of 0.05; SD = standard deviation

Parameter	Young Voice				Older Adult Voice			
	Age group 18-25		Age group 55 and over		Age group 18-25		Age group 55 and over	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Positive emotional colouring	**35.0**	**6.6**	**54.8**	**6.7**	**32.3**	**6.4**	**45.4**	**7.8**
Trust of voice	28.2	2.3	28.9	3.5	46.4	4.5	48.2	3.4
Character of voice (credibility)	18.3	3.2	18.5	2.6	**14.8**	**1.9**	**18.6**	**2.1**
Qualification of voice (credibility)	14.6	2.4	17.7	2.3	20.2	2.1	20.8	2.0
Dynamism of voice (credibility)	**29.9**	**4.7**	**24.8**	**2.2**	22.4	5.1	22.3	2.8
Clarity of voice (quality)	35.17	6.2	37.7	4.3	42.7	5.8	42	2.4
Age of persons speaking	18.33	2.7	23.2	4.8	62.7	4.4	63.0	4.6
Attitudinal similarity	**24.3**	**6.7**	**15.5**	**6.1**	**10.8**	**3.7**	**27.4**	**4.8**
Behavioral similarity	23.3	5.2	23.3	2.0	**9.6**	**3.2**	**27.7**	**3.5**
Similarity (combined)	47.7	11.4	41.0	4.5	**19.4**	**5.4**	**53.1**	**4.3**

Results of voice only experiments

The ratings of the two voices were measured by a one-way ANOVA with age group of participants (18-25, 55 and over) as the between participant factor. The ratings of the two voices, Young voice, Older Adult voice, were also directly compared by a one-way ANOVA.

Emotional colouring

The older adults rated both voices as significantly more positive than the younger group, with $F(1,10) = 10.2$, $p<.01$ for the older adult voice, and $F(1,10) = 26.8$, $p<.001$ for the young voice (*Table 1*). There was no significant difference in how the age groups rated the negative emotional colouring of the voice for neither the older adult voice nor the young voice. When comparing the two voices, there was no significant difference in the positive emotional colouring, $F(1,22) = 2.5$, $p<.13$, and no significant difference in negative emotional colouring, $F(1,22) = 1.8$, $p<.19$ (*Table 2*).

Trust of voice

There was no significant difference in how the two age groups rated the trustworthiness of the two voices, $F(1,10) = .57$, $p<.47$ for the older adult voice, and $F(1,10) = .16$, $p<.7$ for the young voice (*Table 1*). There was, however, a significant difference in the overall trust when the two voices where compared, with the older adult voice being perceived to be significantly more trustworthy than the young voice, $F(1,22) = 182.6$, $p<.001$ (*Table 2*).

Credibility of voice

There was no significant difference in how the two age groups rated the credibility as authoritativeness, qualification and dynamism for the older adult voice, and as authoritativeness and qualification for the young voice. The older adult group, however, perceived the older adult voice to have significantly more character than did the young group, $F(1,10) = 10.8$, $p<.008$. The older adult group also perceived the young voice to be more qualified than the young group, $F(1,10) = 5.23$, $p<.045$ (*Table 1*). The younger age group perceived the dynamism of the young voice to be higher than the older adult group, $F(1,10) = 5.92$, $p<.035$ (*Table 1*).

There were significant differences in the overall perceived credibility of the two voices, the older adult voice was perceived to be more authoritative than the young voice, $F(1,22) = 8.1$, $p<.01$, and it was perceived to be more qualified than the young voice, $F(1,22) = 19.2$, $p<.001$. The young voice was perceived to have significantly more dynamism than the old voice, $F(1,22) = 8.7$, $p<.007$ (*Table 2*). There was no overall difference in character between the two voices.

Table 2. Comparison of young voice and older adult voice; bold indicates statistically different means at a confidence level of 0.05; SD = standard deviation

Parameter	Young Voice		Older Adult Voice	
	Mean	SD	Mean	SD
Positive emotional colouring	44.9	12.1	38.8	9.7
Negative emotional colouring	16.2	7.2	21.8	10
Trust of voice	**28.5**	**2.8**	**47.3**	**3.9**
Authoritativeness of voice (credibility)	**29.2**	**6.4**	**36.2**	**5.6**
Qualification of voice (credibility)	**16.2**	**2.8**	**20.5**	**2.0**
Dynamism of voice (credibility)	**27.4**	**4.4**	**22.3**	**3.9**
Clarity of voice (quality)	**36.4**	**5.3**	**42.3**	**4.2**
Similarity (combined)	44.3	9.0	36.2	18.2

Quality of voice

There was no significant difference in how the age groups rated the clarity of both the young voice and the older adult voice. When comparing the two voices, however, there was a significant difference in the overall perceived clarity of the voices. The older adult voice was perceived to have more clarity than the young voice, $F(1,22) = 9.21$, $p<.006$ (*Table 2*).

Similarity of voice

There were significant differences in how the age groups rated the voices, or rather the persons speaking on the similarity measures. The older adult group perceived the older adult voice (person speaking) to be more similar to them both in attitude ($F(1,10) = 45.9$, $p<.001$), and behaviour ($F(1,10) = 89.2$, $p<.001$), than the young group. The combined similarity measure was highly significantly different, $F(1,10) = 141.2$, $p<.00$, with the older adult group feeling more similar to the older adult voice (person speaking) than the young group. Likewise, the young group perceived the older adult voice (person speaking) to be different from them on all similarity measures.

Similar significant differences do not show up when rating the young voice. The young group perceived the young voice (person speaking) to be more similar to them in attitude ($F(1,10) = 5.6$, $p<.04$) than the older group. There were, however, no significant differences in behavioural and overall similarity.

There was no significant difference in the perceived similarity rating for the two voices (persons speaking).

Answering the questions

Are there differences in the emotional colouring, the perceived trust, and credibility between the two voices? - Mixed response:

(i) Emotional colouring differences absent

The results show clearly that there were no significant differences in the emotional colouring of the two voices (*Table 2*). The older adult group rated the positive emotional colouring higher than the young group, but this was consistent for both voices (*Table 1*). This result reduced the likelihood of emotional colouring of voice as the characteristic of voice that positively or negatively influenced the driving performance in the driving performance study.

(ii) Trust differences present

The older adult voice was perceived to be significantly more trustworthy by both age groups.

(iii) Credibility differences present

The older adult voice was rated as more authoritative and qualified than the young voice by both age groups.

These two results contradict the results from the driving simulator study where drivers showed much better driving performance when driving with the young voice than when driving with the older adult voice, indicating more trust in the young voice.

Is there any difference in the voice quality between the two voices? - Yes.

Both age groups rated the clarity (intelligible, clear, non-breathy, fluent and enunciated) of the older adult voice higher than the clarity of the young voice. So the answer to the question "Is there any difference in the quality between the two voices?" must be yes. Given that drivers with the young voice showed better driving performance, this reduced the likelihood that the quality of voice was the characteristic of voice that influenced the driving performance.

What are the perceived age, background and attitude of the persons speaking?

Here we see the most interesting results. Both the older adult group and the young group placed the two voices (or rather the persons speaking) in the correct age groups, the older adult voice is perceived to be spoken by a 63 year old, and the young voice is perceived to be spoken by a 21 year old, where the actual ages were

73 and 20.

The two groups differed in similarity for the older adult voice, the older adult groups perceived the speaker to be similar to themselves on both attitude and behaviour, and the young groups perceived the speaker to be different from them on all similarity measures. The two groups only differed on attitude for the younger voice, where the young group perceived the speaker to have the same attitude as them. For behavioural and combined similarity, there were no differences between the age groups, so apart from attitude, both the young group and the older adult group felt equally similar to the speaker with the young voice.

In summary, both groups had no problem judging the age of the voice; furthermore, the older adult voice triggered more responses to attitudinal and behavioural similarity than the young voice.

Driving simulator results

The out of context, voice only, laptop results provide useful insight into users' perceptions of voices. However, when we compare them with results with the same voices and the same subject age groups, but using a driving simulator[1] we find significant discrepancies.

The driving simulator study shows that driving performance of older adult drivers was significantly better when driving with the young voice, which contradicts the results from the voice only study which indicates the older adult voice to be more trustworthy and more credible. It would be expected that driving performance would be better with the more credible older voice. However, the young voice which is less trustworthy and less credible, leads to better driving performance. This phenomenon can be explained in terms of perceived similarity. Similarity theory claims that communication is more effective when source and receiver are the same as they share common beliefs and have a common frame of reference. In this case, the older adult drivers are aware of their declining physical and attention abilities, and would therefore trust an older adult voice less than a young voice in the car. The young voice would be associated, for instance, with better physical and attention abilities such as vision and reflexes.

Source credibility

The difference in the source credibility results for voice only, and voice in the driving simulator, were most significant for older adults, and demonstrate clearly the difference in the perception of older and younger people. As described above, credibility of voice was measured using standard Source Credibility Scales (SCS)[22] which comprised five factors referring to criteria by which receivers evaluate sources; three factors from Berlo safety (for instance, pleasant - unpleasant), qualification or expertise (for instance, experienced - inexperienced), dynamism (for instance, aggressive - meek) and the remaining two factors from McCroskey authoritativeness (for instance, reliable - unreliable) and character (e.g., trustworthy - non-trustworthy). The McCroskey and Berlo scales, which use a series of bipolar adjectives that are randomly ordered when presented to respondents, have been used to confirm a wide variety of perceived properties; high and low credibility speakers, to assess credibility of trial witnesses, rate of speech and gender, non-verbal cues, agreeing with a message, and social status and dialect. The scales are often used to assess credibility of people such as speakers, peers, and teachers. Participants were asked to rate a set of adjectives based on the question "How well does each of the following words describe the voice you just heard?" Contrasting adjectives were paired on opposite sides of a 10-point scale such that, for example, qualified and unqualified would appear at different ends. We computed all five source credibility factors from Berlo's and McCroskey's credibility scales, Authoritativeness, Character, Safety, Qualification and Dynamism.

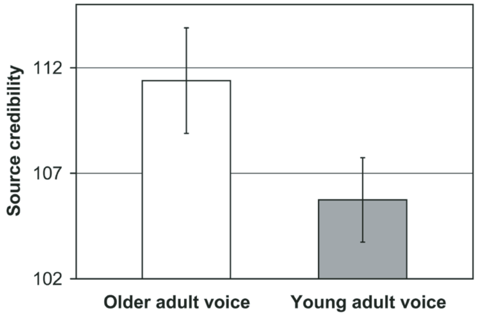

Figure 1. Overall source credibility of the two voices - voice only

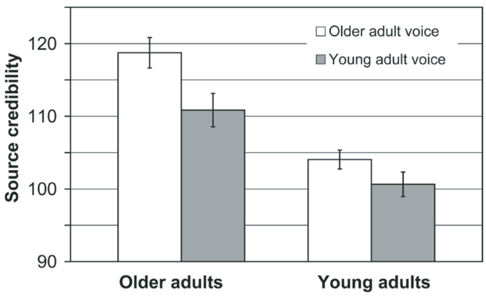

Figure 2. Source credibility of the two voices according to age of participants as tested with voice only

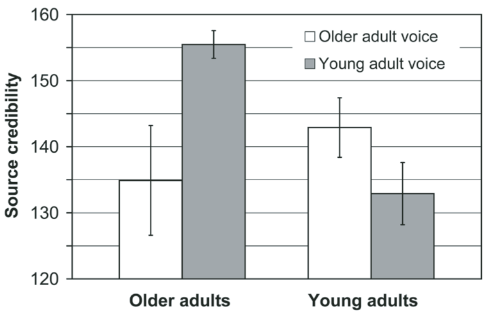

Figure 3. Source credibility of the two voices according to age of participants as tested in a driving simulator

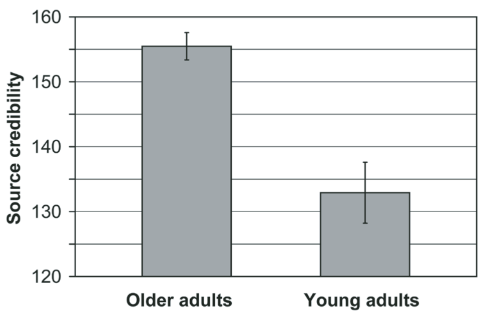

Figure 4. Source credibility of the young adult voice in the driving simulator

Credibility in voice only

Figure 1 shows the overall source credibility of the two voices, young and old, as seen by both groups. A two way ANOVA shows a significant difference between how the two voices were perceived $F(3,20) = 8.6$, $p<.008$.

The same ANOVA also showed that older adults rated both voices higher than the young adults $F(3,20) = 41.8$, $p<.001$. Even though there were clear differences between how the young adults rated the voices with more trust in the older adult than the young voice, the difference was significant only for the older adults with the older adult voice being more credible than the young voice (*Figure 2*).

Driving simulator study

While driving, the older adults, contrary to the results while testing the voice in the lab, found the young adult voice to be more credible than the older adult voice, $F(3,20) = 3.6$, $p<.03$, while the young adults still rated the older adult voice as more credible than the young adult voice (*Figure 3*).

The older adults rated the young adult voice as significantly more credible than did the young adults. $F(1,10) = 18.74$, $p<.001$ (*Figure 4*).

Young adults rated the older adult voice as more credible than the older adults, but the difference was not significant (*Figure 5*). The young adults preferred the older adult voice, but there was no significant difference. Here we see that older adults were far more affected than younger people by difference in context. Younger people were able to judge the credibility of a voice relatively consistently in different contexts, whereas older people did not.

Discussion

This study shows that voices tested and selected for properties in a laboratory setting can be perceived differently in a driving simulator of a car, and hence result in unexpected influence on driving perform-

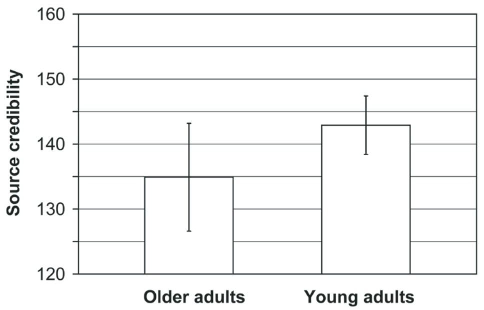

Figure 5. Source credibility of the older adult voice in the driving simulator

ance, particularly for older people. Thus laboratory tests for different aspects of interface design for ubiquitous computing and systems that are part of a more general non-computing task such as driving, where older people are the proposed users, should be treated with care.

The judgments of older participants in this study were far more affected by a change in context than those of younger people. Dulude[2] also found more flexibility in younger people where performance with interactive voice response systems was worse for older people than younger users because older people were responding more negatively to design problems, whereas younger people were more flexible and able to work around them.

The judgment of voices by younger people was more stable and independent of context compared with older people. Yet again they appear more able to adapt, and perhaps to be able to project the voices presented to them on a laptop into the driving simulator context. Context appears to have less impact on younger people and yet these results indicate that context was playing a major role in voice perception for older adults.

These results also point to the need for a contextual focus in interface evaluation and indeed a recognition of the value of the design in a wider context. Cocton[23] argues that the focus of interface design has shifted over the years from the system via the user to the context of use and that all are necessary but not sufficient for effective interactive systems design which requires a new value-centred focus. In his view 'the value-centred framework involves opportunity identification as well as design, evaluation and iteration where opportunity identification has the goal of stating the intended value for a digital product or service in the world'. The evaluation of speech hazard in-car systems as presented in this paper goes some way to assessing value but in safety critical systems such as these, value must be simply perceived value or be evaluated using a form of simulation of the safety critical scenario. The results reported here support a re-evaluation of the real value of an interface especially for older people.

The older population therefore emerges as a very different user group from younger people, who cannot be expected to participate in user evaluation tasks in the user centred design process in the same way. Evaluators should be particularly aware of these differences and take extra care when investigating older people's perceptions within systems. Design for all in the context of these studies with older people demands a 'design process for all' that can accommodate all user types effectively. It also means designing a system that will be 'perceived to be usable by all' which addresses the existent perceptions and previous experiences of all.

References

1. Jonsson I-M, Zajicek M, Harris H, Nass C. Thank you I did not see that: In-car Speech-Based Information Systems for Older Adults. Proceedings of Conference on Human Factors in Computing Systems, 2005, April 2-7. Portland: ACM Press; 2005; pp 1953-1956
2. Dulude L. Automated telephone answering systems and aging. Behaviour & Information Technology 2002;21(3):171-184

3. Hawthorn D. Possible Implications of Ageing for Interface Designers. Interacting with Computers 2000;12:507-528
4. Zajicek M, Wales R, Lee A. Towards VoiceXML Dialogue Design for Older Adults. In: Palanque P, Johnson P, O'Neill E, editors. Design for Society: Proceedings of Human Computer Interaction, 2003, September 8-12, Bath, UK; pp 327–338
5. ADA, Americans with Disabilities Act of 1990, US Public Law; 1990; pp 101-336
6. DDA, The Disability Discrimination Act, 1995, Department for Education and Employment, UK;1995, Chapter 50
7. Eisma R, Dickinson A, Goodman J, Mival O, Syme A, Tiwari L. Mutual Inspiration in the Development of New Technology for Older People. Proceedings of Include 2003, March 25-28, London, UK
8. Kirakowski J, Corbett M, Effective Methodology for the Study of HCI. Amsterdam: Elsevier; 1990
9. Inglis E, Szymkowiak A, Gregor P, Newell AF, Hine N, Wilson BA, Evans J. Issues surrounding the user centred development of a new interactive memory aid. Proceedings of the Cambridge Workshop Series on Universal Access & Assistive Technology (CWUAAT), 2002, March 25-27, Cambridge; pp 171–178
10. Lines L, Hone K, Research Methods for Older Adults. Proceedings of Human Computer Interaction 2002, September 2-6, London
11. Newell A, Gregor P. User sensitive inclusive design – in search of a new paradigm. Proceedings of the ACM Conference on Universal Usability, 2000, Washington; pp 39-44
12. Nass C, Lee K. Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. Journal of Experimental Psychology: Applied 2001;7(3):171-181
13. Nass C, Foehr U, Brave S, Somoza M. The Effects of Emotion of Voice in Synthesized and Recorded Speech. Proceedings of the AAAI Symposium Emotional and Intelligent II: The Tangled Knot of Social Cognition, 2001, North Falmouth
14. Berlo DK, Lemert JB, Mertz RJ. Dimensions for evaluating the acceptability of message sources. Public Opinion Quarterly 1969;33: 563-576
15. Lazarsfeld P, Merton R. Mass Communication, Popular Taste and Organized Social Action, in The Communication of Ideas. Bryson L, editor, New York: Harper; 1948; pp 95-118
16. Rogers E, Bhowmik D. Homophily-Heterophily: Relational Concepts for Communication Research. Public Opinion Quarterly 1970;34: 523
17. Jonsson I-M, Nass C, Endo J, Reaves B, Harris H, Le Ta J, Chan N, Knapp S. Don't Blame me I am Only the Driver: Impact of Blame Attribution on Attitudes and Attention to Driving Task. Proceedings of Human Factors in Computing Systems, 2004, April 24 – 29, Portland; pp 1219-1222
18. Jonsson I-M, Nass C, Reaves B, Endo J. Harris H. Increasing Safety in Cars by Matching Driver Emotion and Car Voice Emotion. Proceedings of Human Factors in Computing Systems, 2005, April 2-7, Portland; pp 1973-1976
19. Zajicek M, Hall S. Solutions for elderly visually impaired people using the Internet. In: McDonald S, Waern Y, Cockton G, editors. People and Computers XIV – Usability or Else! Proceedings of Human Computer Interaction 2000, September 5-8, Sunderland; pp 299–307
20. Zajicek M, Morrissey W. Speech output for older visually impaired adults. Proceedings of Human Computer Interaction 2001, September 10-14, Lille; pp 503–513
21. Izard C. Human Emotions. New York: Plenum; 1977
22. Rubin R, Palmgreen P, Sypher H, editors. Communication Research Measures: A Sourcebook. New York: Guilford; 1994
23. Cockton G. A development framework for value-centred design. Proceedings of Human Factors in Computing Systems, 2005, Portland, April 2-7; pp 1292-1295