Evaluation and Context for In-car Speech Systems for Older Adults

Mary Zajicek
Department of Computing
Oxford Brookes University
Wheatley Campus
Oxford OX33 1HX, UK
+44 (0)1865 484579
mzajicek@brookes.ac.uk

Ing-Marie Jonsson
Department of Communication
Stanford University
Toyota Information Technology Centre
4009 Miranda Avenue
Palo Alto, CA 94025, US
+1 650 575 9906
ingmarie@csli.stanford.edu

ABSTRACT
Listeners to speech based information systems attribute qualities to voices which will make them receptive, or otherwise, to the information. This paper looks at the value of laboratory-based usability testing on voices for an in-car speech messaging system. It compares results concerning the users' perception of voice, in particular source credibility, performed first on voices presented out of context in a laboratory setting and then the same voices embedded in an in-car system in a driving simulator. In the case of older adults the results are significantly different raising issues of context and value in usability testing for this group.

Keywords
Speech systems, voice quality, in-car speech systems, usability testing.

INTRODUCTION
Testing the interfaces of potentially dangerous real-time systems poses a problem for interface designers. In-car information systems fall into this category and are rapidly becoming a common accessory in cars. Some of these systems rely on speech-based interactions and thus emphasize the importance of voices and characteristics of voices, which carry both linguistic and paralinguistic cues, and have the potential to influence listeners and communication partners.

These qualities of speech pose a number of questions for designers of speech-based in-car information systems. Should there be one voice conveying all types of information content in the car? Or should there be one voice per information category, or perhaps one voice per age group of drivers? Previous studies on specialization and generalization indicate that having one voice per information category in traditional media improves the listeners' trust in the quality of the information [12]. Will multiple voices have the same effect in the car? What characteristics of the voice impacts perception and trust of information? What is the impact on people's perception of information, system and car when the characteristic of the voice is mismatched with the content of the information? These and many more questions needs to be answered for successful introduction of speech based in-car information systems. In this paper we address the question of the optimum age of voice in an in-car information system for particular age group, 18 – 25 and 55 years of age and older.

This paper describes the experiments and sets out the results of trials carried out in the laboratory setting to gauge users' perception of voices to be used in in-car speech systems. The results are then compared with those for the same age groups carried out not in a laboratory setting but with the same voices in a driving simulator. The comparison demonstrates that for older adults in particular the perception of voices when heard out of context and when heard in the driving simulator is very different. This points to a difference in flexibility in dealing with computer systems between older and younger people which must be accommodated in interface design.

VOICE PERCEPTION
Humans are well tuned to detect characteristics in a voice and use that skill both when communicating with humans and with speech-based computer systems [10]. Humans also assess the characteristics of the voice and this affects their perception of the liking and credibility of what is said, i.e. the content of the information conveyed by the speech-based system. The psychological literature suggests that consistency is important. People expect consistency and prefer it to inconsistency. When inconsistency is encountered, people enter a state where they are motivated to adapt their perceptions in order to resolve inconsistency [4]. The need for consistency is well understood in traditional media but less clear for human-computer

interaction. Studies indicate that both synthesized and recorded voices [10] influence perception of 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 was mismatched. Credibility is generally positively associated with liking [1] but a plausible explanation is that in cases of voice characteristic and content mismatches, people draw 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 be seen as a sense of detachment and hence more objective and credible. Follow-up studies [11] show that credibility and persuasion is higher also for a mismatch between the personality of the voice of speech-based computer system and the personality of the person interacting with the systems. It has also been shown [9] that better human communication occurs between a source and a receiver who are alike, i.e., homophilous and have a common frame of reference. Communication is more effective [13] when source and receiver are similar. When two individuals share common meanings, belief, and mutual understandings, communication between them is more likely to be effective. 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. Communication between dissimilar individuals may also cause cognitive dissonance because an individual is exposed to messages that are inconsistent with existing beliefs, resulting in an uncomfortable psychological state.

Characteristics of voice can also influence people's attention, and affect performance, judgment and risk-taking. These properties are crucial for driving, and previous studies show that information provided by in-car systems have the potential to improve driving performance [6]. Studies also show that the linguistic and para-linguistic properties of the in-car voices influence driving performance [8]. Results from a study that considered older drivers show that the selection of voice had a significant impact on driving performance [7].

RATIONALE FOR EXPERIMENTS

Speech systems have been shown to help older adults accomplish tasks. They can also provide useful information about the environment and things happening around them, which older people may not readily absorb for themselves. In the domain of computing, voice prompt speech messages have been successfully used to provide reminders concerning previous interaction for those with poor memories [15] and enabled older people to get going on a computer system where they hadn't been able to before.

In studies with web browsing and speech systems for older adults [16],[18] and other work concerning explanation messages in computer interaction [16], the nature and quality of the voice messages was found to be very important.

Aging 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 ones actions and reluctance to try new things. Confidence boosting and affirmative speech prompts have proved to be very useful [15] for older adults when using a speech browser. We therefore have reason to expect that speech-based in-car information messages will also be able to instil confidence in older adults when driving.

The ability to absorb information also decreases with age. Older people were found to be less able to absorb long instructions than younger people [15]. They quickly forget them or were unable to remember them at the time that they were required. Speech based support provided exactly at the point when it will be useful can remove the need for long instructions to be given at the beginning of a task. In the experiments described in this paper particular care was taken to make speech messages as short as possible.

Speech messages therefore can compensate for memory loss by suggesting actions that have not been remembered, help with strategising by making contextually relevant suggestions, 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. The system 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.

It has also been shown [7] that drivers have strong preferences when presented with voices of different ages. In this study using speech messages in a driving simulator, older drivers definitely preferred a younger voice over an older voice.

LABORATORY EXERIMENTS

This goal of the experiments described in this paper was to identify characteristics of the voices used by the in-car information system that have the potential to influence drivers attitude to the voice and thus driving performance. The perception of the voices as well as the perception of the persons speaking was examined, in an attempt to identify characteristics of the voices that would explain the impact on driving performance for older adults as presented in [8] where older adults showed a definite preference for information delivered by a younger voice. The younger voice used was that of a 20 year old women and the older voice was that of a 73 year old woman.

Specifically answers were required to the following questions:

1. *Are there differences in the emotional coloring, the perceived trust, and credibility between the two voices?*
2. *Is there any difference in the voice quality between the two voices?*
3. *Similarity – 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. Participants attended one session listening to two voices while filling in a set of questionnaires.

Laptop and Earphones

The experiment was set up to rate the two different voices using a laptop with Microsoft Media Player and earphones. Participants controlled the laptop and could select and play any recording by clicking on them at any time. The recordings consisted of 26 short voice prompts, recorded in the two different voices. The prompts included the following:

- There is thick fog ahead.
- You are approaching an intersection.
- Warning there is a fallen tree in the road ahead.
- Beware of cyclists ahead.
- The current speed limit is 60 miles an hour.
- There are crosswinds in this area.
- Stop sign ahead.
- The police use radar here, you might need to slow down.
- There is heavy traffic ahead, turn left to avoid it.
- There is an accident ahead, turn right to avoid it.

Questionnaires were filled in using pen and paper.

Participants

The experiment consisted of a 3-day study at Oxford Brookes University, UK. There were 12 participants, 6 from the age group of 55 years of age and older, and 6 from the age group of 18 – 25 years of age. Both age groups were gender balanced. The age groups were selected to match the two age groups (18 – 25, 55 and older) from the driving simulator study that initiated this experiment.

All participants volunteered their time for their participation, gave informed consent and were debriefed at the end of the experiment

Procedure

All participants listened to the recordings of two voices, these were the actual recordings used in the driving simulator study mentioned above. The participants rated the voices using 4 questionnaires, one on emotional coloring of voice, one on trust and credibility of voice, one on quality of voice and finally one on how the participants perceived the persons speaking being similar to themselves in terms of background and attitude.

The participants 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.

All participants were informed that the experiment would take 30 minutes, and that they could play the recordings at any time and as often they liked while they filled in the questionnaires. After finishing the first questionnaire and hence the first voice, they notified the experimenter and were setup with recordings and questionnaire to rate the second voice.

Measures

Emotional Coloring of Voice

Positive emotional coloring of each voice was measured using a variation of the Differential Emotion Scale (DES) [5]. 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 (*alpha* = .77).

Negative emotional coloring 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 (*alpha* = .79).

Trust of Voice

Trust of voice was measured using a standard Individualized Trust questionnaire [14]. 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 [14]. 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 (Rubin et al 1994), *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, alpha = .79.

Participants were also asked to judge the age of the person speaking for both the young voice and for the older adult voice.

Homophily – Similarity

A standard questionnaire on Homophily [14] was used to identify measures of similarity, we created three indices, *attitudinal similarity, behavioral similarity* and *similarity* as a combination of attitude and behavior. 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.

RESULTS

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 Coloring of Voice

The older adults rated both voices as significantly more positive (Positive emotional coloring) 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 (see Table 1 for mean values and standard deviation, SD). There was no significant difference in how the age groups rated the negative emotional coloring of the voice for either the older adult voice or the young voice. When comparing the two voices, there was no significant difference in the positive emotional coloring, $F(1, 22) =2.5,\ p< .13,$ and no significant difference in negative emotional coloring, $F(1,22) = 1.8,\ p< .19$ (see 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 (see 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, $F(1,22) = 182.6,\ p<.001$ (see 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.

Table 1: Comparison: Rating of Voices by Age Group

	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 coloring	**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**

Table 2: Comparison: Young Voice and Older Adult Voice.

	Young Voice		**Older Adult Voice**	
	Mean	SD	Mean	SD
Positive emotional coloring	44.9	12.1	38.8	9.7
Negative emotional coloring	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

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, $F(1,10) = 5.23, p<.045$ (see Table 1).

The younger age group perceived the dynamism of the young voice to be higher than the older adult group, $F(1,10) = 5.92, p<.035$ (see 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 perceive 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, $F(1,22) = 8.7, p< .007$ (see Table 2). There was no overall difference in character between the two voices.

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, $F(1,22) = 9.21, p<.006$ (see Table 2).

Homophily – Similarity

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 behavior ($F(1,10) = 89.2, p< .001$), 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 than 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 behavioral and overall similarity.

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

Answering the questions

The results show clearly that there were no significant differences in the emotional coloring of the two voices (see Table 2). The older adult group rated the positive emotional coloring higher than the young group, but this was consistent for both voices (see Table 1). We can therefore answer NO to the question, *"Is there any difference in the emotional coloring between the two voices?"* This result reduced the likelihood of emotional coloring of voice as the characteristics of voice that positively or negatively influenced the driving performance in the driving performance study.

The older adult voice was perceived to be significantly more trustworthy by both age groups, so the answer to the question *"Is there any difference in the trustworthiness between the two voices?"* must therefore be YES.

Likewise, we must answer YES to the question, *"Is there any difference in the credibility between the two voices?"* 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, as described below. Drivers in the driving simulator study showed much better driving performance when driving with the young voice than when driving with the older adult voice, indicating more trust for the young voice.

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.

Answering the last question, "*What are the perceived age, background and attitude of the person speaking*?" 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. Please note that the voices were recorded by a 73 year old and 20 year old.

When listening to the older adult voice, the two groups polarized on similarity, the older adult groups perceived the speaker to be similar to themselves on both attitude and behavior, and the young groups perceived the speaker to be different from them on all similarity measures. When listening to the young voice, the two groups only polarized on attitude, where the young group perceived the speaker to have the same attitude as them. For behavioral 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.

From these results it is clear that people had no problem in judging the age of the person in the case of both the young voice and old voice and furthermore that the older adult voice triggered responses to attitudinal and behavioral similarity more than the young voice.

COMPARISON WTH DRIVING SIMULATOR RESULTS

The results discussed above which were gathered under laboratory conditions, provide interesting insights into users' perceptions of voices in voice systems however when we compare them with previous results with the same voices and the same subject age groups, but using a driving simulator [7], we find startling discrepancies.

The results from the driving simulator study show that driving performance of older adult drivers was significantly better when driving with the young voice, not when driving with older adult voice. This clearly contradicts the results from this study that indicates the older adult voice to be more trustworthy and more credible. Given this, it would be expected that the driving performance would be better with the older voice since it is more credible. Results, however, show that this is not the case, but rather the opposite. The young voice, that is less trustworthy and less credible, leads to better driving performance than the older adult voice. This can possibly be explained by perceived similarity. The similarity theory claims that communication is more effective when source and receiver are the same in that 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, for instance, be associated with better physical and attention abilities such as vision and reflexes.

Source credibility results in detail

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 between older adult users and younger ones. They are therefore covered in detail here.

As described above, credibility of voice was measured using a standard Source Credibility Scales (SCS) [14] which comprised 5 factors referring to criteria by which receivers evaluate sources; 3 factors from Berlo [14] safety (e.g., pleasant - unpleasant), qualification or expertise (e.g., experienced - inexperienced), dynamism (e.g., aggressive - meek). The remaining two factors for the SCS are from McCroskey [14] authoritativeness (e.g., reliable - unreliable) and character (e.g., trustworthy - non-trustworthy). The McCroskey and Berlo scales 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. The scale uses a series of bipolar adjectives that are randomly ordered when presented to respondents.

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 5 source credibility factors from Berlo's and McCroskey's credibility scales [14], Authoritativeness, Character, Safety, Qualification and Dynamism.

Results from the voice only study

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

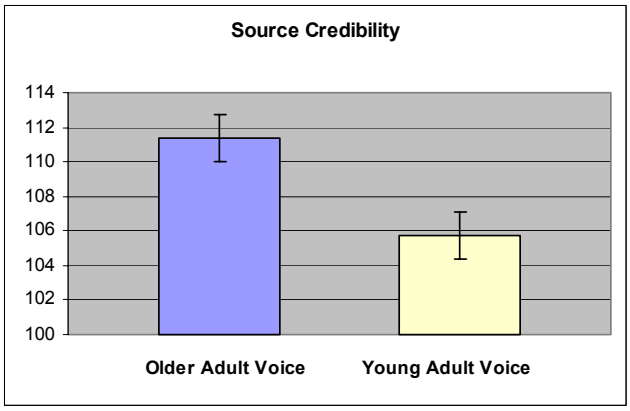

Figure 1. Overall source credibility of the two voices – voice only

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 (see Figure 2)

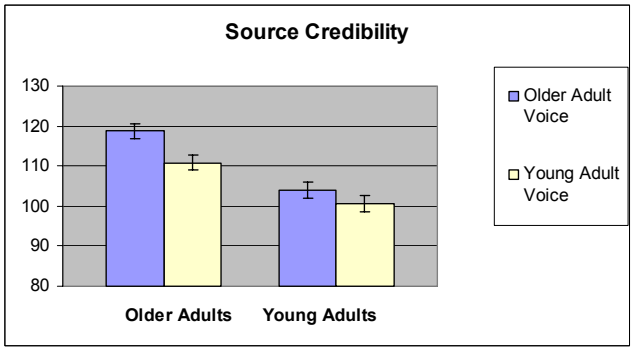

Figure 2. Source credibility of the two voices according to age of participants - voice only

Results from the 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 (see Figure 3).

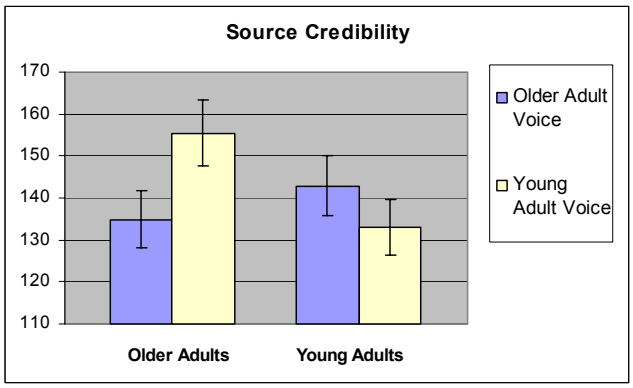

Figure 3 Source credibility of the two voices according to age of participants - driving simulator

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

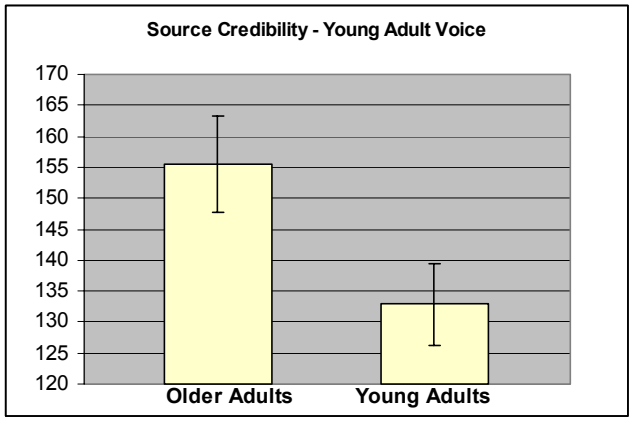

Figure 4. Source credibility of the young adult voice – driving simulator

Young adults rated the older adult voice as more credible than the older adults, but the difference was not significant (see 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.

VAUE IN INTERFACE DESIGN

These results 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 [2] 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

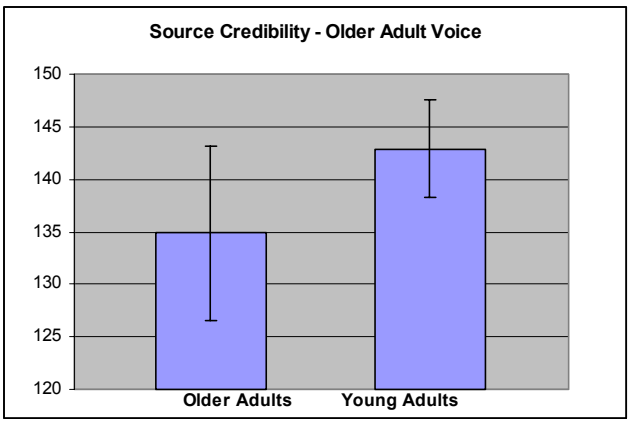

Figure 5. Source credibility of the older adult voice – driving simulator

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 produce or service in the world. It can be argued that the evaluation of speech hazard in-car systems as presented here goes some way to assessing *value* but that 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. Nevertheless the results reported here support a re-evaluation of the real value of an interface.

CONCLUSION

This study presents results that indicate that the selection of the appropriate voice for in-car system is crucial and most importantly that voices tested and selected for properties in a laboratory setting can be perceived differently in a driving simulator of car, and hence result in unexpected influence on driving performance particularly for older people. This opens an important discussion concerning the value of laboratory tests for different aspects of interface design for ubiquitous computing and systems which are part of a more general non-computing task such as driving, where older people are the proposed users.

The judgments of older participants in this study were far more affected by a change in context than those of younger people. Dulude [3] also demonstrated 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 that made their interaction difficult, whereas younger people were more flexible and able to work around them.

In the studies reported in this paper younger people appear yet again to be demonstrating their ability to be flexible, and perhaps subconsciously project the voices presented to them out of the driving simulator context into the driving simulator context or that the context does not matter to them. Their judgment of voices was more stable and independent of context compared with older people.

These results once again point to the older population as a very different user group from younger people and also to the need for us to be aware of these differences in every step of the design process.

The next step in our research is to investigate the relationship between older adults' responses to speech messages using a driving simulator compared to their responses when actually driving.

REFERENCES

1. Berlo, D. K., Lemert, J. B., & Mertz, R. J. Dimensions for evaluating the acceptability of message sources. Public Opinion Quarterly, 33 (1969), pp. 563-76.
2. Cockton, G., a development framework for value-centred design, Proceedings of CHI 2005, Portland, 2005
3. Dulude L., 2002, Automated telephone answering systems and aging. Behaviour Information Technology 21(3), (2002) 171-184.
4. Fiske, S., & Taylor, S., (1991), Social Cognition. New York, NY, McGraw-Hill Inc.
5. Izard,C., *Human Emotions*, (1977), New York, Plenum Press.
6. Jonsson, I-M., Nass, C., Endo, J., Reaves B., Harris, H., Le Ta, J., Chan, N., Knapp, S., (2005).Don't Blame me I am Only the Driver: Impact of Blame Attribution on Attitudes and Attention to Driving Task, Proceedings of CHI 2004
7. Jonsson, I-M., Zajicek, M., Harris, H., Nass, C., (2005). Thank you I did not see that: In-car Speech-Based Information Systems for Older Adults, Proceedings of CHI 2005, Portland, OR
8. Jonsson, I-M., Nass, C., Reaves, B., Endo, J. & Harris, H. (2005). Increasing Safety in Cars by Matching Driver Emotion and Car Voice Emotion, Proceedings of CHI 2005, Portland, OR.
9. Lazarsfeld, P Merton, R., . "Mass Communication, Popular Taste and Organized Social Action." 1964 (492-512).
10. Nass, C., Foehr, U., Brave, S., & Somoza, M. (2001) 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, North Falmouth, MA
11. Nass, C. & Lee, K. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. Journal of Experimental Psychology: Applied, 7(3), 171-181.
12. Reeves, B., Nass, C., The media equation, Cambridge University Press, 1996

13. Rogers, E, Bhowmik, D., Homophily-Heterophily: Relational Concepts for Communication Research, *Public Opinion Quarterly*, volume 34, page 523, 1970

14. Rubin, R. et al. (1994) *Communication Research Measures: A Sourcebook.* New York: Guilford Press, 1994

15. Zajicek M. & Hall, S., Solutions for elderly visually impaired people using the Internet, In S. McDonald, Y. Waern, G. Cockton (eds) People and Computers XIV – Usability or Else!, Proceedings of HCI 2000, (2000) pp 299 – 307

16. Zajicek M., and Morrissey W., Speech output for older visually impaired adults, In Blandford A., Vanderdonckt L,. Grat P (eds) Interaction without frontiers. Joint Proceedings of HCI 2001 and IHM 2001, (2001) pp. 503 – 513

17. Zajicek M., Wales, R., Lee, A., Towards VoiceXML Dialogue Design for Older Adults, In Palanque P., Johnson P., O'Neill E (eds) Design for Society. Proceedings of HCI 2003, (2003) pp 327 - 338

18. Zajicek, M., Wales, R., Lee, A., Speech interaction for older adults, Special Issue 'Design Principles to Support Older Adults' of Universal Access in the Information Society, (eds. M, Zajicek & S. Brewster) Springer, Vol 3/2 (2004) pp 122 - 13