The Effects of Different Voices for Speech-Based In-Vehicle Interfaces: Impact of Young and Old Voices on Driving Performance and Attitude

Ing-Marie Jonsson, Nils Dahlbäck

Department of Computer and Information Science, Linköping University, Linköping, Sweden

`ingmarie@ansima.com, nilda@ida.liu.se`

Abstract

This paper investigates how matching age of driver with age of voice in a conversational in-vehicle information system affects attitudes and performance. 36 participants from age groups, 55 -75 and 18 - 25, interacted with a conversational system with young or old voice in a driving simulator. Results show that all drivers rather communicated with a young than old voice in the car. This willingness to communicate had a detrimental effect on driving performance. It is hence important to carefully select voices, since voice properties can have enormous effects on driving safety. Clearly, one voice doesn't fit all.

Index Terms: Spoken dialogue systems, Applications in other areas – cars,

1. Introduction

Today's cars are often fitted with interactive information systems that handle high quality audio/video, satellite navigation, hands-free telephony, and control over climate and car behaviour. Most in-vehicle information systems use visual displays for interaction, with a recent shift to speech.

Wickens [1] suggests that using speech-based output is less distracting than interactions with a visual display because the driving task is primarily visual. Interactions with a voice-based system still demand the driver's attention, McCarley et al. [2] demonstrate that conversation can disrupt attentive scanning and representation of a traffic scene. However, Baron and Green [3] reviewed and summarized papers on the use of speech interfaces for tasks such as music selection, email processing, dialling, and destination entry while driving. They concluded that "People generally drove at least as well, if not better (less lane variation, speed was steadier), when using speech interfaces than visual graphical interfaces". Speech interfaces led to less workload than graphical interfaces and reduced eyes-off-the-road times, all pro-safety findings.

The research reviewed above is about general aspects of speech based systems. Very little attention has been paid in these and other similar studies on the impact of different voices and tone of voice. In other areas such as radio, TV, movies, however, choice of voice has long been an important factor to consider, but this seems often overlooked in much research on speech based interfaces in vehicles.

One exception is the studies where Jonsson et al [4, 5] compared the effect of young and old adult voices in an in-vehicle warning and hazard system. The study was run using a driving simulator [6]. The hazard and warning system was designed as a purely informational system. It generated a speech-prompt in a timely fashion before an upcoming complex traffic situation or hazard. It was found that a young adult voice was preferred by both young (18-25) and old (55-75) volunteers. The research also measured significantly fewer driver mistakes when drivers drove with the in-vehicle hazard and warning system using the young adult voice than the system with the older adult voice [4, 5].

The study presented here is a follow-up study where a conversational interface using different age voices is investigated. The rationale for studying a conversational system is that they are introduced by car manufacturers (as well as insurance companies, and rental car companies) to provide drivers with personalized services, and assess driver abilities. The "in-vehicle information system" learns about its driver(s), by either a) (automatically) collecting information about driving habits and ability, or b) interacting with driver(s). The focus of this paper is not on the morality or legality of such activities, but on the conversational interfaces used as part of the interactive assessment of a driver.

When combined with the improved safety attributes of speech based interfaces [7] rather than GUIs, conversational interfaces will likely become more prevalent for future in-car information systems. The quality of the conversation with the vehicle is critical to user acceptability. A distracting voice or a badly designed dialogue may result in the customer returning the vehicle to the dealership - asking to get the "annoying" voice disabled. This paper focus on voice selection for in-vehicle conversational systems and how the voice and the system affect performance and attitude.

2. Experimental Method

2.1. Goal

The experiment was designed to investigate the effect of a matching age of voice with age of driver in a conversational information system on driving behaviour and attitude.

The conversational system was designed to ask questions with self-disclosures to learn more about the driver. Two versions were implemented, one system asking questions using an "old" adult voice, and one system asking questions using a "young" adult voice.

The experiment was conducted in a driving simulator that was instrumented to collect and save data on driving behaviour. Driving performance data included obedience to traffic laws (e.g., stopping at stop signs, and not speeding), ability to stay in lane and accidents. Attitudinal data was gathered by a set of both standard and newly-designed questionnaires [8]. This included willingness to communicate with the in-vehicle information system, trust, reliability and liking of the system and how the system affected the trust and liking of the car. The experiment was a 2 (age group: 18-25 years of age, and 55 years of age and older) x 3 (Voice system: 20 year old voice, 76 year old voice, and no voice) gender balanced between-participants design setup to answer the following questions:

1. *Are there differences in drivers' willingness to communicate based on age of voice?*

2. *Are there differences in the perceived affection, trust and dominance between the two voices?*

3. *Are there differences in how the voices of the in-vehicle system affect driving behaviour?*

2.2. Apparatus

The STISim driving simulator [6] was used in the experiment. Users sat in a real car seat and used a Microsoft Sidewinder steering wheel and pedals. The simulation was projected on a wall in front of the participants. A driving scenario was specifically designed to be as varied and realistic as possible. It included villages, scenic drives and small towns with traffic varying from sparse to dense. All drivers completed exactly the same driving scenario with the same settings. The simulator was configured to automatically record driving performance for each participant.

Conversational speech prompts were inserted at 18 points in the course. The prompts were all questions, some of them scripted with self-disclosures to elicit responses. As an example, a question related to driving:
"Do you find it stressful talking to people while driving?"
and a question with self-disclosure related to driving:
"I like driving on mountain roads, where is your favourite place to drive? "

Literature shows that similarity attraction is important when selecting interaction partners [9]. Lydon, Jamieson and Zanna [10] furthermore suggest that similarity and attraction is multidimensional and people are attracted to others similar to them for instance in attitudes, social and cultural background, personality and physical appearance. Hence two voices were selected for the conversational system, an older adult voice (76 year old woman) and a young adult voice (20 year old woman). The speech talents were instructed to read the sentences in a calm and neutral voice; the recorded voices were assessed and compared for age cues, amplitude and intelligibility [11]. Results show no specific emotional content in either voice or any difference between the styles except the age of voice.

2.3. Participants

There were 36 participants, 18 from the age group 55 – 75, and 18 from the age group of 18 – 25 years of age. These are two groups of drivers that warrant extra interest based on accident statistics.

The age-span of the older adult group, over 55 (i.e. 55 – 75), was based on a frequently used evaluation form and report, Drivers 55 Plus: Check your own performance, published by the AAA Foundation of Traffic Safety [12]. None of the participants were diagnosed with mental illness, and all were active drivers. Both age groups were gender balanced. All participants volunteered their time for their participation, gave informed consent and were debriefed at the end of the experiment

2.4. Procedure

Participants started by completing a pre-driving web-based online questionnaire for general information and prior driving experience. Each participant then drove a short training course with verbal guidance from the experimenter. This course was 8000 feet and took approximately five minutes to finish. The purpose this was to familiarize participants with the control and feedback from the driving simulator and to screen participants for simulator sickness [13]. None of the 36 participants felt nauseous or discomfort during or after the training course.

Both young and old participants were randomly placed in one of three gender-balanced groups, creating a total of six conditions; 1) older adults driving with in-vehicle system with young voice. 2) older adults driving with the in-vehicle system with old voice. 3) older adults driving without voice system, 4) young adults driving with the in-vehicle system with young voice, 5) young adults driving with the in-vehicle system with old voice, and 6) young adults driving without a voice system. They then drove the 60000 feet long driving scenario that took on average 27 minutes to complete.

After the driving session all participants completed a set of online questionnaires. Included in the questionnaires were the Relational Communication Scale [8], the Interpersonal Communication Motives Scale [8], and a questionnaire on the properties and influence of the in-vehicle information system. Participants used a desktop computer to complete and submit the web-based questionnaires. All participants were debriefed at the end of the experimental session.

2.5. Measures

Driving is a complex activity that continually tests drivers' abilities to react to traffic and road conditions. Measures of numbers of driving accidents and violations serve the same function in both real and simulated driving. They indicate critical breakdowns in driver attention, judgment, and vehicle management; it is these failings that predict future driving problems. Focus for driving performance is therefore on three measures of the most dangerous behaviours: number of accidents, and obedience to the most important traffic laws (adherence to traffic lights and adherence to stop signs), and lane deviations.

Driving Performance: Three indices for were created based on data from the driving simulator. *Accidents*, was based combined collisions and off-road accidents. *Breaking Traffic Regulations*, combined speeding with running stop signs and red-lights. *Swerving*, was the sum of the drivers lane deviations. The indices were very reliable, (*alpha .65, alpha .62 and alpha .66 respectively).* Also reported is time to finish the driving course.

Relational Communication: Drivers' interpersonal relationship with the in-vehicle system was measured using a standard Relational Communication Scale [8]. The items affection, similarity, trust, dominance, equality and task orientation, were combined into one index.

Interpersonal Communication: Drivers' willingness to communicate with the in-vehicle system was measured using a standard Interpersonal Communication Motives Scale [8]. The items pleasure, affection, inclusion and relaxation were combined into one index.

Quality and Influence of In-Vehicle system: Participants assessed the quality of the in-vehicle system. The index, *Voice System Quality*, was comprised of five items, "fun to use", high quality, use again, "want to have" and "discourage others to use" reverse coded. The index was very reliable (*alpha .80*).

Drivers also rated the *influence of the in-vehicle system* on driving performance. The index was comprised of three items; Confident driver, Careful driver and Inattentive driver reverse coded. The index was reliable (*alpha .72*).

Prior Driving Experience: Participants self reported on driving habits, such as miles driven per week, and normal driving scene (city, urban, rural) and past two years driving incidents (accidents and tickets).

3. Results

The effects of the in-vehicle system on driver performance were measured by a two (driver age) by two (in-vehicle system, no system) between-participants ANOVA. The effects of age of voice of in-vehicle conversational system on the age of the driver were measured by a two (driver age) by two (age of voice) between-participants ANOVA.

Prior Driving Experience: There was no significant difference between the groups on self reported driving style, and driving performance (tickets and accidents).

Relational Communication: Drivers' interpersonal relationship with the in-vehicle system was measured in terms of affection, similarity and trust. The data shows an interaction effect between age of driver and age of voice. Old drivers have less affection and trust for the old voice system than for the young voice system ($M= 76.9$ and $M=93.5$), whereas young drivers have equal affection and trust systems with both old and young voice ($M=91.8$ and $M=96.3$), $F(3, 20) =5.43$, $p < .03$. There were main effects for both age of driver and age of voice. Young drivers showed more affection and trust for the in-vehicle system than older drivers ($M=94.0$ and $M=85.2$), $F(3,20) = 11.5$, $p < .003$. All drivers showed more affection and trust for the in-vehicle system with the young voice, than for the system with the old voice ($M=94.9$ and $M=84.4$), $F(3,20) = 16.6$, $p <.001$, see Table 1.

Table 1: Attitudinals, Quality, and Voice Influence

		Affection, Similarity, Trust	Willingness to Comm.	Quality of System	Positive Influence of System
Older Adults					
Old voice	Mean	76.9	20.0	10.1	14.7
	SD	6.5	4.0	2.6	4.4
Young voice	Mean	93.5	25.1	23.8	14.8
	SD	4.4	6.1	4.4	3.1
Young Adults					
Old voice	Mean	91.8	26.8	32.0	16.0
	SD	7.3	1.5	5.7	3.9
Young voice	Mean	96.3	31.0	28.9	9.9
	SD	6.9	4.0	1.6	0.8

Interpersonal Communication: There are main effects in drivers' willingness to communicate with the in-vehicle system. Young drivers were more willing to communicate with the in-vehicle system than the older drivers ($M=28.9$ and $M=22.6$), $F(3,20)=13.6$, $p <.001$. Both young and older drivers were also more willing to communicate with the in-vehicle system with the young voice ($M=28.0$ and $M=23.4$), $F(3,20)=7.4$, $p < .01$, see Table 1.

Quality and Influence of In-Vehicle system: The data show a clear interaction effect age of voice with age of driver. Young drivers rate the systems equally, and older drivers rated the in-vehicle system with the young voice higher than the system with the old voice $(M=23.8$ and $M=10.1)$, $F(3,20)=28.0$, $p < .001$. There were main effects for both age of driver and age of voice. Older drivers rated the quality of the system lower than young drivers ($M=17.0$ and $M=30.4$), $F(3,20)=71.2$, $p <.001$. Even though younger drivers showed no preference of voice, older drivers preference ensure that on average, all drivers rated the in-vehicle system with the young voice higher than the system with the old voice ($M=26.4$ and M=21.0), F=$(3,20)=10.0$, $p < .004$, see Table 1.

There is an interaction age of voice and driver age for positive system influence. Young drivers report that the system with the young voice had a bad influence on their driving ($M=9.9$ and $M=16.0$ for positive influence), while older drivers report no difference in influence based on age of voice ($M=14.7$ and $M=14.8$), $F(3,20)=5.1$, $p < .04$. There was

a main effect for age of voice. Overall, all drivers reported more negative influence from the in-vehicle system with the young voice than from the system with old voice ($M=12.4$ and $M=15.4$ for positive influence), $F(3,20) = 4.8$, $p <.04$. There was no main effect from age of driver, see Table 1.

Driving Performance - Accidents: The data show no interaction effects but two main effects. Young drivers drive much worse than older adults ($M=3.4$ and $M=1.9$, $F(5, 30) = 4.5$, $p < .04$), and all drivers drove worse with the young voice system ($M=3.8$ and $M=1.5$, $F(3,20) = 10.9$, $p <.004$). Note that the data show the fewest accidents when driving with the old voice system, see Table 2.

Table 2: Driving Performance

		Driving time	Swerving	Accidents	Breaking Traffic rules
Older Adults					
Old voice	Mean	1560	10.5	1.0	2.3
	SD	221	2.5	1.0	1.6
Young voice	Mean	1814	15.5	2.5	4.7
	SD	704	8.5	2.0	3.2
No Voice	Mean	1809	27.8	2.3	2.5
	SD	144	11.27	1.9	2.7
Young Adults					
Old voice	Mean	989	16.8	2	14.7
	SD	243	3.8	1.4	2.9
Young voice	Mean	1053	23.0	5	14.5
	SD	191	8.0	1.9	2.7
No Voice	Mean	1201	41.0	3.3	12.8
	SD	267	26.2	3.4	7.9

Driving Performance – Breaking Traffic Regulations: Young drivers broke traffic regulations significantly more often than older adult drivers ($M=14.0$ and $M=3.2$), $F(5,30)=64.7$, $p<.001$ (see Table 2). There was no effect from age of voice or from driving with or without the in-vehicle system.

Driving Performance – Swerving: Young drivers deviated more from their lane than older adult drivers ($M=26.9$ and $M=17.9$), $F(5,30)=4.5$, $p<.04$. There was a positive effect of the in-vehicle system, drivers without a voice system swerved the most ($M=68.8$ and $M=49.8$), $F(5,30)=8.6$, $p<.001$. Drivers with the young voice system swerved significantly more than drivers with the older adult voice system ($M=62.3$ and $M=37.4$), $F(3,20)=4.7$, $p<.04$, see Table 2.

Driving Performance – Driving Time: The data show that young drivers drove significantly faster than older drivers ($M=1081$, $M=1727$ respectively with $F(5,30)=30.8$, $p<.001$), see Table 2. There was no effect for age of voice or driving with or without a system.

4. Discussion and Conclusion

The goals of this experiment were investigate the effects on attitudes and driving performance of matching age of drivers to age of voice in a conversational in-vehicle information system. The results from the experiment (see table 1 and 2) provide answers to the questions posed in section 2.1.

Are there differences in drivers' willingness to communicate based on age of voice? Results show that all drivers, regardless of age, prefer to communicate with a younger voice in the vehicle. Post driving questionnaires show that young people were inclined to "forget" driving (primary task) in favour of answering questions from the system. Note that young drivers were acutely aware of the fact that they preferred to talk rather than drive. They also noted that the young voice had a bad influence on their driving performance.

Are there differences in the perceived affection, trust and dominance between the two voices? Older adults found the young voice to be more affectionate, trustworthy and less dominant than the old voice, see Table 1. Young drivers did not report a difference between voices. This result contradicts similarity attraction [9], but confirm previous studies with in-vehicle systems [4, 5, 11]. In these studies, older adults preferred the young voice over the older adult voice in vehicles, but not in an office setting.

Are there differences in how the voices of the in-vehicle system affect driving behaviour? The data show that both age groups drove worse with the young voice conversational in-vehicle system. There was however also a significant difference in the breadth of the effect. The young voice affected the driving performance of young drivers negatively on two out of three driving measures (accidents and swerving). Older drivers, on the other hand, were less affected; they performed worse on one of three measures (number of accidents). This confirm results from a previous study matching age of voice with age of driver [4, 5]. A purely informational Hazard and Warning system was used in this previous study, and data show that older adults drive better (listen more) to a young voice than to an old voice. The young voice was better at focusing older drivers' attention to road hazard, and their driving improved. Similarly, in the study reported here, the older drivers also paid more attention to the young voice, but the nature of the system (conversational) resulted in decreased driving performance.

Date from these experiments suggests that using young voices for in-vehicle conversational systems may have a detrimental effect on driving performance for all drivers, but seems particularly bad for young drivers. Since this is a group of drivers that often cause concern, and with in-vehicle systems becoming more prevalent, the effect found here should be further investigated. Furthermore, the reason for older adult drivers being better focused on the task of driving (even though they like to communicate with the young voice system, see Table 1 and 2), also warrants further investigation.

Voice interfaces and dialogue systems in the vehicle are special; they are secondary task applications, while driving is the main task. Attention should be paid to drivers' state and driving conditions to ensure that interactions with the dialogue system are safe, especially when the dialogue is initiated by the system and not the driver. The system needs to "know" when to cease interactions and when it is safe to resume interactions again.

Data from this study furthermore confirm previous findings [4, 5, 11], and emphasizes that - The car is different! Solid findings of similarity attraction no longer hold. There are some possible explanations for this behaviour in cognitive science. Norman [14] makes a distinction between experiential and reflective cognition, where experiential cognition is associated with expert/skilled and reactive behaviour and reflective cognition with slow/planned and reconsidered behaviour. Driving is primarily a data-driven reactive activity, similar to other critical control rooms such as for instance cockpits. Here the voice of the expert is not necessarily similar to oneself, but instead associated with properties necessary for the task at hand. For cars, younger eyes, voices and cognition seems to be preferred even for older drivers.

The results presented here also highlight a trade-off for secondary task voice applications: attractiveness of voice interface versus performance/task orientation. In the case presented here, a conversational system in the vehicle needs to be appealing so that drivers are willing to communicate with it. However, increasing the attractiveness of the interface should not be done at the expense of driving safety. This result can most likely be generalized to other application areas outside the vehicle where a conversational/dialogue system is the secondary task, such as other control tasks where speech-based system are mainly used for help, information and warning. Results from this study indicate that voices for speech systems in cars needs to be carefully selected.

5. References

[1] C. Wickens, *Processing Resources and Attention, Varieties of Attention*. New York: Academic Press, 1984.

[2] J. S. McCarley, M. Vais, H. Pringle, A. F. Kramer, D. E. Irwin, and D. L. Strayer, "Conversation disrupts visual scanning of Traffic scenes," presented at Ninth International Conference on Vision in Vehicle, 2001.

[3] A. Barón, and Green, P., "Safety and Usability of Speech Interfaces for In-Vehicle Tasks while Driving: A Brief Literature Review," Transportation Research Institute (UMTRI), The University of Michigan, 2006.

[4] I.-M. Jonsson, M. Zajicek, H. Harris, and C. I. Nass, "Thank you I did not see that: In-car Speech-Based Information Systems for Older Adults," presented at SIGCHI Conference on Human Factors in Computing Systems, Portland, OR, 2005.

[5] M. Zajicek and I. Jonsson, "A Complex Relationship, Older People and In-Car Message System Evaluation," *Journal of Gerontology*, vol. 6, pp. 66-78, 2007.

[6] Systems Technologies Inc, "STISim Drive," Systems Technologies Inc, 1990.

[7] H. Lunenfeld, "Human factor considerations of motorist navigation and information systems," presented at Vehicle Navigation and Information Systems Conference, Toronto, Ontario, 1989.

[8] R. B. Rubin, P. Palmgreen, and H. E. Sypher, *Communication Research Measures: A Sourcebook*. New York: Guilford Press, 1994.

[9] D. Byrne, *The Attraction Paradigm*. New York, NY: Academic Press, 1971.

[10] J. E. Lydon, D. W. Jamieson, and M. P. Zanna, "Interpersonal similarity and the social and intellectual dimensions of first impressions," *Social Cognition*, vol. 6, pp. 269 - 286, 1988.

[11] I.-M. Jonsson and M. Zajicek, "Selecting the voice for an In-car information system for older adults," presented at Human Computer Interaction International Las Vegas, Nevada, USA, (ISBN 0-8058-5807-5), 2005.

[12] American-Automobile-Association-Foundation-for-Traffic-Safety, "Drivers 55 Plus.," Washington DC 20005: 202/638-5944, 1994.

[13] R. J. V. Bertin, A. Guillot, C. Collet, F. Vienne, S. Espié, and W. Graf, "Objective measurement of simulator sickness and the role of visual-vestibular conflict situations: a study with vestibular-loss (a-reflexive) subjects," presented at Neuroscience, San Diego, California, 2004.

[14] D. Norman, *Things that make us smart*. Boston, MA, USA: Addison-Wesley Publishing Company, 1993.