

On the Road and on the Web?

Comprehension of synthetic and human speech while driving

Jennifer Lai, Karen Cheng
IBM Corporation/
T.J. Watson Research Center
30 Saw Mill River Road
Hawthorne, NY 10532
jlai@us.ibm.com
kgcheng@ucla.edu

Paul Green, Omer Tsimhoni
University of Michigan
Transportation Research Institute
2901 Baxter Road
Ann Arbor, MI 48109-2150
pagreen@umich.edu
omert@umich.edu

ABSTRACT

In this study 24 participants drove a simulator while listening to three types of messages in both synthesized speech and recorded human speech. The messages consisted of short navigation messages, medium length (approximately 100 words) email messages, and longer news stories (approximately 200 words). After each message the participant was presented with a series of multiple choice questions to measure comprehension of the message. Driving performance was recorded. Findings show that for the low driving workload conditions in the study, (cruise control, predictable two-lane road with no intersections, invariant lead car) driving performance was not affected by listening to messages. This was true for both the synthesized speech and natural speech. Comprehension of messages in synthetic speech was significantly lower than for recorded human speech for all message types.

Keywords

Speech synthesis, text-to-speech, comprehension, driving performance, driving simulator

INTRODUCTION

The term “mobile worker” no longer describes a transient state for a person, but represents a category of corporate employees who work from a multitude of locations. Employees are being encouraged to “work smarter” since it often does not seem possible to work harder. As a result, system designers are stepping up to the challenge of giving users access to the same functionality in a mobile setting that they would have if seated in their offices.

Car manufacturers and computer companies alike view the time workers spend in a car as an opportunity to deliver

email and other individualized messages. The in-vehicle information systems market is projected to exceed \$5 billion by the year 2003 [3]. The U.S. National Highway Traffic Safety Administration estimates that currently 44% of drivers have phones in their vehicles, 7% have email access and 3% can receive faxes [12].

Given the difficulties and the high cost of delivering large amounts of dynamic data with pre-recorded speech, it is most likely that these applications will use synthesized speech for the reading of the email messages or profile specific news stories. Synthesized speech is speech produced by a computer. It is often referred to as “rule-based” speech because the computer uses a series of rules to convert text to the sounds that are generated. Compared to human speech, synthetic speech is generally not as well liked and takes longer for people to become accustomed to [5].

Using synthetic speech in the car, as discussed above, may be easier from a system implementation viewpoint. However, we have wondered whether there are any costs to attention and driving performance compared to using recorded human speech. The goals of this study were 1) to examine the impact of listening to messages in synthetic and human speech on driving performance and 2) to measure the comprehension of those messages. This work was conducted jointly by IBM Research and the University of Michigan Transportation Research Institute (UMTRI). Complete details will appear in an upcoming technical report.

PRIOR WORK

The safety of introducing potential distracters into the driving environment continues to be debated in both the research and the public arenas. When radios were first added to cars, they were heavily criticized as distracting to drivers. However, research suggests that it is not the listening per se that disrupts driving [2] but rather the additional tasks such as putting in tapes or changing stations, that deteriorate driving performance [7].

Recent research [e.g., 6] focuses on the effect of using cell phones while driving. Most evidence indicates that using phones has the potential to distract drivers and even cause

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SIGCHI'01, March 31-April 4, 2001, Seattle, WA, USA.
Copyright 2001 ACM 1-58113-327-8/01/0003...\$5.00.

accidents. A statistical analysis of traffic accidents suggests that drivers using cellular phones are four times more likely to be involved in a crash [10,11]. Studies from both on-road and simulator driving reveal substantial increases in driver responsiveness (i.e. length of time to react to an event) during telephone conversations [1]. These findings hold for both hands-free cellular phone usage and traditional usage.

Research on driver response to in-vehicle auditory messages is still fairly limited. Fleming, et al. [4] had 32 licensed drivers drive an instrumented car on an expressway. While driving, they were presented with traffic messages containing 6 to 14 items. Participants were asked to identify whether the message was relevant to a particular route and to recall as much of the message as possible. Typically only four items were recalled regardless of message length, with the road and crossroad being most common. They found that listening to a message increased speed variance. Other driving performance measures such as lane variability and lane excursions were generally unaffected. As drivers became more loaded, they tended to slow down.

Concurrent with our research, Lee, et al. [9] had 24 young people (ages 18-24) drive in a simulator while verbally interacting with a speech-based email system. Findings show that participants experienced a 30% increase in response time to a periodically braking lead vehicle when using the email system. The drivers generally reported that the speech-based interaction caused a greater cognitive load for them and they perceived it to be distracting.

The present study was motivated by findings in Lai, et al. [8], which suggest that perception of synthetic speech requires greater cognitive resources than perception of human speech. Such demand of cognitive resources could divert attention away and negatively impact performance from concurrent tasks, such as driving.

Our hypothesis was that an increase in cognitive workload would lead to a decrease in both comprehension of messages and driving performance. In this study, cognitive workload increased with the use of synthetic speech, the increased complexity of the messages presented to the driver, and the increased difficulty of the driving course. All of these factors were predicted to negatively impact comprehension and driving performance.

METHODOLOGY

Experimental Design

This study utilized a 2 (voice type: human, synthetic) by 3 (message type: news, email, navigation) by 3 (driving workload: parked, straight, curved) within-subjects design. All participants participated in all conditions. Half of the navigation messages were preceded by an earcon and half were not. The earcon consisted of a brief chime sound. Participants were randomly assigned to one of two groups, with the restriction that each group had an equal number of men and women, and young and old. Messages that were presented in human voice in one group were presented in synthetic voice in the other group. Each message was only presented once.

The dependent variables in this study were accuracy of message comprehension, response time to questions, standard deviation of the vehicle's lateral position within the lane, and standard deviation of the steering wheel angle. We also assessed participants' subjective rating of the perceived difficulty of the message.

Participants

A total of 24 licensed drivers were recruited from the Ann Arbor, Michigan area. Half of the participants were between the ages of 21 and 28 and half were older (between 65 and 71). These two age groups were selected since they represent two extreme segments of the driving population in terms of driving performance. All participants were paid \$35 for the participation. Participants were invited to complete the study only if they had no hearing problems. Only 37.5% of the participants owned cell phones (4 of the younger subjects and 5 older subjects).

Previous exposure to synthetic speech was rated on a six point scale (1-"never heard of it", 3-"heard it once or twice", 6-"work with it"). Except for two young participants, which reported their exposure at a level 4, none of the participants had significant prior exposure to synthesized speech. Most reported that they had heard it once or twice. Any participant reporting a level of 5 or 6 would have been excused.

Simulator

Participants drove the UMTRI driving simulator version 7.2.2 (www.umich.edu/~driving/sim.html), a medium-fidelity driving simulator based on a network of Macintosh computers. The simulator consists of a mockup of a car, a projection screen, a torque motor connected to the steering wheel, a sound system, a sub-bass sound system, a computer system to project images of an instrument panel, and other hardware. The sound system provides various noises (engine, drive train, tire, and wind) to enhance the realism of the experience. The sub-bass sound system provides vibration. The projection screen, offering a horizontal field of view of 33 degrees and a vertical field of view of 23 degrees, is 6 m (20 ft) in front of the driver, effectively at optical infinity.

Simulated two-lane roads had 3.6 meters (12 feet) wide lanes. Oncoming vehicles were positioned statically in the left lane to provide a more realistic feeling of a busy road. A lead vehicle drove in front of the simulator vehicle at a fixed speed of 72.5 km/hr (45 mi/hr). A sample of the view of the road is shown in Figure 1.

The simulator has both a functioning brake and gas pedal. For this study neither was required since the car was used in cruise control and automatically accelerated from a stopped position to the preset cruise speed. Participants were only required to use the steering wheel to control the position of the car within the lane. If a lane deviation occurs, a loud bump and grinding noise is heard as the car moves onto the shoulder. If the car crosses the lane into other cars there is no negative impact on the driver. One is able to "drive through" any object.



Figure 1: Sample view of road and lead vehicle

Task

Every participant heard three types of messages of varying lengths in both synthetic and recorded human speech. The three message types were navigation, email, or news.

All navigation messages were four seconds long and consisted of the direction of the turn, name of road, and distance until turn. For example “*turn right in 1.5 miles onto Radiance Drive.*” These three pieces of information appeared in random order to reduce any training effect. The participants were then presented with two questions, which asked for two of the three pieces of information. Thus the purpose of listening to the messages was to attempt to understand them and not to direct the driving. The road in the simulator had no intersections and did not require turning off onto other roads.

The email messages were around 100 words in length and discussed a variety of topics, many geared towards a corporate recipient. The participants were asked three questions about each message, the first of which was a general question about the purpose of the message or the tone in which it was spoken. The other two questions probed for comprehension about details contained in the message.

Sample email message:

Colleagues,
Safe driving at our site is everyone's responsibility. In the past few months, Security has received a number of complaints about speeding cars in the parking areas. Excessive speed, combined with decreased visibility and wet roadways during the winter months, increases the risk of accidents. In addition, a number of employees have been parking in the roadway next to the grassy islands, and therefore obstructing the flow of traffic. All employees are asked to observe the following guidelines:

- Observe the posted speed limits;
- Observe traffic signs and pavement markings;
- Respect the two-hour limit on visitor parking;
- Do not park in areas marked “No Parking” or “Fire Zone.”

Sample general comprehension question:

1) The tone of this message is best described as:

- a) friendly;
- b) informative;
- c) persuasive;
- d) encouraging;

Sample specific comprehension question:

2) In this message which factor was NOT attributed to increased risk of accidents:

- a) irresponsible drivers;
- b) excessive speed;
- c) low visibility;
- d) wet roads;

News messages were taken from the CNN web site and shortened to fit the word limit. The participants were asked four questions about each news message, one general and three specific comprehension questions. See Table 1 for characteristics of the email and news messages.

The IBM TTS engine Via Voice Outloud, for embedded platforms, was used to record the synthetic messages at a rate of 180 words per minute (wpm). With the exception of the wpm rate, none of the other default settings (e.g. pitch and breathiness) were altered. A voice talent recorded the natural speech messages. As seen in Table 1, the messages recorded in TTS were slightly longer than those recorded with natural speech, which is due to the lower number of words per minute. The recordings for both voices were at a sampling rate of 11.025 KHz, 16 bit, mono.

Type	Number of Words	Length in seconds		Reading Level	
		TTS	Human	Ease	Grade
Email	106 (7)	43 (3)	40 (3)	50	10.3
News	223 (8)	85 (5)	79 (4)	51	10.7

Table 1. Mean values and standard deviations. Length of messages in seconds.

In Table 1, the reading ease was measured with the Flesch scoring method where 0 is very hard and 100 is very easy. Flesch-Kincaid was used for grade reading level.

Procedure

The participants began by completing a biographical form and a consent form. They then underwent both a hearing test and a visual acuity test. All were well qualified to drive and could easily hear the messages. After moving to the driving simulator the experimenter played a sample message and verified that the participant found the sound level comfortable. Then the participant read a few sentences from the in-vehicle display, to verify that the participant would be able to read both the questions and multiple-choice answers about the spoken messages.

The experiment started with a 5-minute practice of driving the simulator on a road that consisted of a few straight sections and several curves. The participant then drove on a similar road to collect baseline-driving data. Thus comfortable with how to operate the driving simulator, the participant was given a series of practice messages in both the synthetic and human voices with the simulator running but parked. After listening to each passage the participant was presented with several questions with multiple choice answers on the in-car display. Thus questions and their possible answers were read by participants and not spoken with either TTS or human speech. Questions were answered by the participant by speaking the selection (a, b, c or d) out loud. Each question was presented one at a time, with the participant controlling advancement to the next question by means of a finger switch fitted to the right index finger.

A similar practice session was conducted with the participant driving in both straight sections and curves. After the participants pulled away from the side of the road the car automatically accelerated to a preset cruise control speed of 72.5 km/hr (45 mi/hr). Cruise control was selected for this study to eliminate the variability caused by speed reductions when the driver listened to the messages. This simulates a very easy driving scenario. In order for the participant to read and answer the questions, the car went into an “auto-pilot” mode after the message was played. To gauge self-perception of performance, the participants were asked to rate their understanding of the message on a 10-point scale (with 10 being the highest), after each message.

Participants had a fixed amount of time to answer the set of questions. After that fixed period of time, the auto-pilot was turned off and the driver needed to resume control. The participants were made aware of the time remaining for the set of questions by a progress bar at the bottom of the screen that moved towards 0 as they ran out of time. Also, when the auto-pilot was turned off, there was a road sign by the side of the road that read “Auto-pilot is now disengaging”.

Evaluation Measures

Comprehension accuracy was measured by scoring the number of multiple choice questions that were answered correctly. Time on task was measured for each question and included the time from when the question was presented until the participant clicked the switch to move to the next question. Driving performance measurements included standard deviation of the steering wheel angle, and standard deviation of the vehicle’s lateral position in the lane. The number of lane excursions was also measured. In general as standard deviations of these measures increase, driving performance decreases.

Baseline Studies

Two baseline studies were conducted to determine if there could be vast differences in the comprehension of messages that was due solely to the message itself (i.e. were certain messages much harder than others), or to the nature of the questions.

In the first baseline, 10 participants were asked to answer all the questions used in the study without ever having heard the text. The subjects were presented with a paper copy of the questions and instructed to answer them in the order presented based on: common sense, prior knowledge of the subject of matter, or hints and clues in the structure of the questions and answers. As in the main experiment, they said the number of their choice for the answer out loud. The mean comprehension level was 23% correct, which is close to random guessing (25%). Therefore, the difference in comprehension of different messages could not be attributed to clues in the questions themselves.

In the second baseline, the main experiment was replicated without driving. Twelve young subjects, whom had not previously participated, were asked to respond to the messages while seated in the simulator room. The apparatus for this baseline was the same as used for the main driving experiment, except for the factors associated with driving the simulator (i.e., there was no motor sound and the subject was not seated in the simulator). The purpose of this experiment was to measure the difficulty of messages without the effect of driving workload. The second baseline study revealed differences in comprehension between the different messages. Such a difference between messages is somewhat predictable and is extremely hard to control for. In real life situations some messages are harder to follow than others, just as all messages are not perceived as equally interesting by all listeners.

Statistical analyses

Repeated measures Analysis of Variance tests were performed on all dependent variables. Tests of 3-level factors were followed by pairwise comparisons, with Bonferonni corrections made to the alpha levels to account for multiple comparisons.

FINDINGS

Overview

Consistent with our hypotheses, participants rated the synthetic voice more difficult to understand than the human voice and were less accurate in remembering the content of the messages read by synthetic voice. However, voice type had no effect on driving performance.

We also had several findings that were contrary to our expectations. While we expected that comprehension accuracy would degrade as the messages got longer and more detailed, we found that email messages, our medium length message, had the lowest accuracy scores. Also comprehension accuracy appeared to increase with a more difficult driving workload. These findings and others are detailed in the sections below.

No participants committed any lane excursions, therefore no results are reported for that measure of driving performance.

Subjective Rating

After each message was played, participants were asked to rate how much of the message they felt they understood on a

scale of 1 to 10, with 10 being ‘understood all of message.’ Consistent with prior research, we found that participants rated the human voice ($M=8.65$) easier to understand than the synthetic one ($M=7.70$), $F(1,16)=39.85$, $p<.001$.

Participants reported that they understood the email messages ($M=8.44$) more than they understood news messages ($M=7.91$), $F(1,16)=17.49$, $p<.01$. This was expected, given that news messages were longer and contained more detail. However the participants’ perception of performance and ease of the email messages was contrary to actual accuracy scores. Accuracy scores by message type and voice type can be seen in Figure 2.

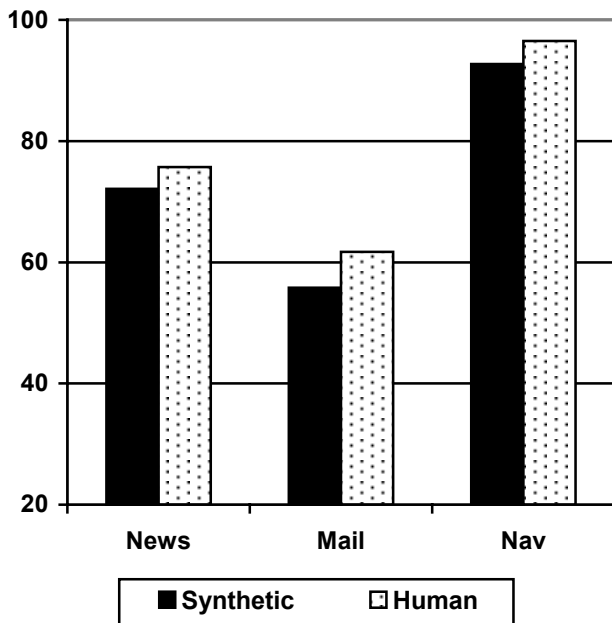


Figure 2. Comprehension by message type and voice type.

Participants also felt they understood more while involved in the task of driving (straight $M=8.19$; curved $M=8.44$) than while parked in the simulator ($M=7.91$), $F(2,32)=13.33$, $p<.001$. We had predicted that an increase in driving workload would divert attention away from one’s ability to comprehend text messages and would have expected them to perceive the task as harder.

Effect of Voice Type

The results suggest that listening to synthetic speech requires more attention than listening to human speech. Participants were less accurate in answering questions following synthetic speech messages ($M=73.5\%$) than questions following human speech messages ($M=78\%$), $F(1,16)=11.39$, $p<.01$. The time needed to answer the comprehension questions was not affected by voice type. Both measures of driving performance were also unaffected by voice type.

Effects of Message Type

No differences were found between navigation messages preceded by an earcon and those not preceded by an earcon. We speculate that this is due primarily to the fact that participants were seated in the simulator waiting for the next message to play. In real driving situations there would be more distractions and the earcon would be needed to alert the driver to broaden his attentional focus. However since there were no differences in the findings for the two types of navigation messages (with and without earcon) we collapsed the results and present only one type of navigation message.

Perhaps not surprisingly, participants were most accurate in answering questions about the least complex messages: navigation messages ($M=94.6\%$), $F(2,32)=155.84$, $p<.001$. However, longer news messages ($M=73.9\%$) were more accurately remembered than the medium-length email messages ($M=58.8\%$). Pairwise comparisons revealed that all means were significantly different from each other.

Though news messages are longer than email messages, they may be easier to understand because most people already have developed a model for listening to news stories in the car by listening to the radio. One participant commented, “Yeah, I hear news messages on the radio. With the email, it seemed like important stuff that you had to think about. You couldn’t just passively listen to it.” In addition, because the content of the email messages was somewhat business-related, it may have been more unfamiliar to the participants than the content in the news messages. One of the participants, when asked to explain why she thought news messages were easier to understand than email messages, remarked, “It was the subject matter that made the difference. Some of them were funny... I could relate to them... If I understood the message, the voice didn’t make a difference. It was the message itself.”

More complex messages required more time to answer the comprehension questions. Questions about the navigation messages were answered very quickly ($M=4.0$ sec), while news ($M=10.3$ sec) took longer, and email ($M=10.5$ sec) messages took the longest to respond to, $F(2,32)=335.08$, $p<.001$.

The type of message the participant heard appeared to affect driving performance. Lateral position within the lane was not affected. However, there was a greater deviation in steering wheel angle during navigation messages ($M=.041$) compared to news ($M=.033$) and email messages ($M=.031$), $F(2,32)=11.81$, $p<.001$. This could have been due to the smaller sampling of time.

Effects of Driving Workload

The driving performance results affirm that curved sections were more difficult to drive than straight sections. There were greater deviations in lateral position within the lane and deviations in steering wheel angle for the curved sections compared to the straight sections, $p<.001$ for both dependent variables.

More interestingly, we found that an increase in driving workload was accompanied by increased accuracy in answering comprehension questions. Messages heard during curved sections ($M=80.4\%$) were more accurately remembered than those heard during straight sections ($M=77.5\%$) and parked sections ($M=69.4\%$), $F(2,32)=9.942$, $p<.001$. All pairwise comparisons were significantly different from one another, $p<.001$.

It is possible that the slight increase in workload caused subjects to focus more on the experiment, thus increasing attention to the message as well as to the driving task. A different explanation might stem from the order of messages. Due to the experiment setup, the first parked condition always occurred before the first two driving conditions. Straight and curved sections were not presented in a fixed order. Some task training could have occurred after the parked position. However the subjects would also be more fatigued as the study progressed.

Alternatively, these results may have been an artifact of the particular messages presented in the different driving workload conditions. Because we balanced the order of TTS and human voice messages within each age and gender group, we had 3 participants in a given condition (age, gender, order of voice type). As a result of the small number and the additional programming requirement, we did not randomize messages across driving workload conditions.

Response time to answer questions did not increase linearly with driving workload. We found instead that participants were quickest to respond in the straight section ($M=7.4$ sec), compared to the parked ($M=8.7$ sec) and curved ($M=8.6$ sec) sections, $F(2,32)=47.95$, $p<.001$.

Interactions with Driving Workload

A two-way interaction revealed that accuracy in comprehending email messages was greatest in the straight sections, but accuracy of comprehending news messages was greatest in the curved sections, $F(4,64)=8.45$, $p<.001$ as seen in Table 2. It was surprising to find that the most complex messages were understood best in the most difficult driving condition.

	Parked	Straight	Curved
News	67.7%	72.7%	81.3%
Email	43.1%	68.1%	65%
Navigation	97.4%	91.7%	94.8%

Table 2. Mean values for accuracy comprehension by driving workload and message type.

A two-way interaction also revealed that comprehension of synthetic speech was most accurate in the straight sections but comprehension of human speech was most accurate in the curved sections, $F(2,32)=10.09$, $p<.001$ as seen in Figure 3. This finding supports our hypothesis that perceiving synthetic speech requires more cognitive resources than perceiving human speech. The increased workload of driving the curved sections made it more difficult to perceive

the synthetic speech, thus the lower accuracy in remembering synthetic voiced messages in those sections.

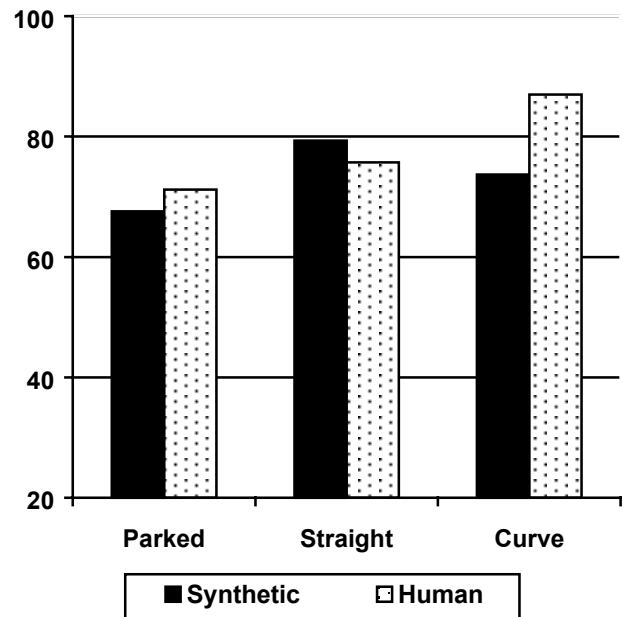


Figure 3. Mean values for accuracy comprehension by driving workload and voice type.

Effects of Age and Gender

As seen in Figure 4, younger participants were more accurate in answering comprehension questions than older participants, $F(1,16) = 9.07$, $p<.01$. There were no significant differences in comprehension between men and women and no interaction between age and gender.

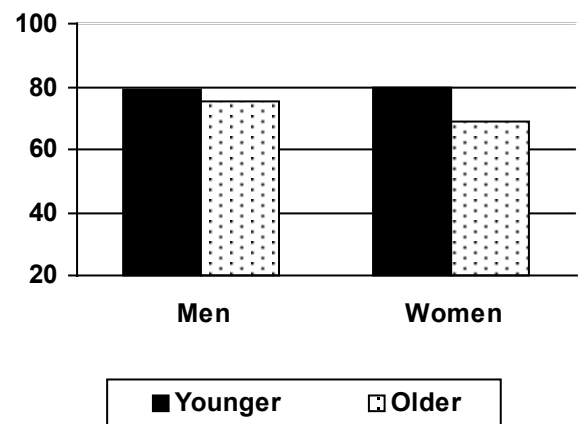


Figure 4. Mean accuracy scores displayed by gender and age group.

Women ($M=.0399$) displayed greater deviation in steering wheel angle than men ($M=.030$), $F(1,16) = 4.82$, $p<.05$. Older participants ($M=.039$) showed greater deviation in steering wheel angle than younger participants ($M=.030$), however the difference did not reach statistical significance p

= .065. There was no interaction between age and gender. Age and gender did not affect lateral position within the lane.

DISCUSSION

Though findings from this study did not show a significant change in driving performance while listening to messages produced with either synthesized or human speech, the authors caution that this result should not be generalized until higher levels of driving workload can be examined. As one of the first studies in this field, we wanted to keep the initial parameters fairly basic. Thus driving conditions were simple: cruise control, a lead vehicle that never changed speed or braked, and very predictable two-lane roads with minimal traffic and no intersections or unexpected events. This is perhaps one of the easiest driving scenarios that drivers could find themselves in. However, while easy, it is still reflective of driving conditions on open highways in cruise control. Future work should involve measurements of the effect of listening to email messages with greater levels of complexity in the driving conditions, such as random unexpected events (e.g., pedestrians, swerving vehicles, road hazards) and more traffic.

Comprehension of messages was lower with synthesized speech than for recorded human speech and this was true across all message types. This finding confirms previous research in the area. Subjective evaluations showed that perception of the increased difficulty of the synthetic speech messages was correlated with lower accuracy rates. Participants reported feeling that synthetic speech was more difficult to understand than human speech. One participant commented, "The computer voice seemed annoying and not easy to understand."

Perhaps one of the most interesting findings was that comprehension accuracy for synthesized speech did not degrade linearly with the length of the message, as it had in an earlier in-lab experiment [8]. When driving, the navigational messages were easiest to understand, followed by the news stories, with email messages coming in last even though the news stories were longer with a greater level of detail. This may be because many of us have a well-established model for driving and listening to news on the radio. News stories are general and usually do not affect or involve us directly. In contrast email messages are usually addressed directly to us, involve our lives and often require us to take an action (i.e. attend a meeting, place a phone call, or get in touch with a person). Also news stories are written by professional reporters and if reading the entire article, may have more redundancy than email messages. However, since these articles were modified to fit the word length restriction most, if not all, of the redundancy was removed. With regard to the professional writer argument, this was not a benefit that aided comprehension in the in-lab experiment referenced earlier.

We believe that examining technology as a potential distraction for drivers is an important area in need of additional study. Future work not only needs to address higher levels of driving workload, it also needs to address

more complex levels of interaction with in-car speech systems. While in our study participants merely listened to messages, in a real speech-based system they would need to interact with it. This would involve both trying to remember the correct phrasing or command for the system as well as keeping the context of the interaction in mind. These additional complexities would further tax the cognitive resources of the driver. We expect that analyses of higher driving and interaction workloads will reveal a negative impact on driving performance.

REFERENCES

1. Alm, H., Nilsson, L. (1995) The Effects of a mobile telephone task on driver behavior in a car following situation. *Accident Analysis and Prevention*, 27(5), pp.707-715
2. Brown, I. D. (1965). Effect of a car radio on driving in traffic. *Ergonomics*, 8(4), 475-479.
3. Bruno, A. (1999) Auto Industry Drives Telematic Services. RCR Radio Communications Report, v18, n37, September 13, 1999
4. Fleming, J., Green, P., and Katz, S. (1998). Driver Performance and Memory for Traffic Messages: Effects of the Number of Messages, Audio Quality, and Relevance *Technical Report UMTRI-98-22*, Ann Arbor, MI: The University of Michigan Transportation Research Institute.
5. Francis, A.L., Nusbaum, H.C. (1999). Evaluating the quality of Synthetic Speech. In Gardner-Bonneau, D. (Ed.), *Human Factors and Voice Interactive Systems* (pp. 63 - 97)
6. Goodman, M., Bents, F.D., Tijerina, L., Wierwille, W., Lerner, N., and Benel, D. (1997). An Investigation of the Safety Implications of Wireless Communication in Vehicles (Technical Report DOT HS 808 635), Washington, D.C.: U.S. Department of Transportation
7. Jaencke, L., Musial, F., Wogt, J., & Kalveram, K. T. (1994). Monitoring radio programs and time of day affect simulated car-driving performance. *Perceptual & Motor Skills*, 79(2), 484-486.
8. Lai, J., Wood, D., Considine, M., The Effect of Task Conditions on the Comprehensibility of Synthetic Speech. *ACM SIGCHI CHI 2000 Proceedings* pp. 321-328
9. Lee, J.D., Caven, B., Haanke, S., Brown, T. (under review). Speech-based Interaction with In-vehicle Computers: The Effect of Speech-based E-mail on Drivers Attention to the Roadway. *Human Factors*
10. Redelmeier, D.A., Tibshirani, R.J. (1997) Association between cellular-telephone calls and motor vehicle collisions. *New England Journal of Medicine* 336(7) pp.453-458
11. Violanti, J.M., (1997) Cellular phones and Traffic Accidents. *Public Health*, 111(6) pp.423-428
12. www.driverdistraction.org

