Outbound Texting: Comparison of a Speech-Based Approach and a Handheld Touch-screen Equivalent

M. Lucas Neurauter[1], Jonathan M. Hankey[1], Thomas B. Schalk[2], and Gary Wallace[2]

[1]Virginia Tech Transportation Institute
[2]Agero

Submission Date: November 15, 2011

Word Count: 5,433 + 4 figures (1,000) and 2 tables (500) = 6,933 Total

M. Lucas Neurauter (Corresponding author)
Virginia Tech Transportation Institute
3500 Transportation Research Plaza
Blacksburg, VA 24061
P: (540) 231-1522
F: (540) 231-1555
lneurauter@vtti.vt.edu

Jonathan M. Hankey
Virginia Tech Transportation Institute
3500 Transportation Research Plaza
Blacksburg, VA 24061
P: (540) 231-1512
F: (540) 231-1555
jhankey@vtti.vt.edu

Thomas B. Schalk
Agero
8550 Freeport Parkway
Irving, TX 75063-2547
P: (972) 753-6278
F: (972) 753-6335
tschalk@agero.com

Gary Wallace
Agero
8550 Freeport Parkway
Irving, TX 75063-2547
P: (972) 753-6230
F: (972) 753-6335
gwallace@agero

Outbound Texting: Comparison of a Speech-Based Approach and a Handheld Touch-screen Equivalent

ABSTRACT

An evaluation of a speech-based system was conducted in an effort to assess general task and on-road performance across a range of vehicular control and subjective measures. The system under evaluation provided users the ability to send text messages and to obtain route guidance through destination entry while driving.

A brief introduction to the speech-based system and its capabilities were provided to participants before asking them to complete three practice tasks. General observations of these initial uninformed interactions (i.e., absent assistance by the experimenter) with the system indicated that the system was intuitive and easy to use. Errors, when observed, were commonly related to word recognition.

Participants were then instructed to navigate a closed test track at 72 km/h (45 mph) while engaging in nine secondary tasks: three manual modality texting tasks, three equivalent voice modality texting tasks, and three voice modality destination entry tasks. As expected, driving performance measures, glance behavior, and subjective ratings were all significantly degraded when comparing manual texting tasks to similar tasks completed using voice commands. Furthermore, performance under speech-based destination tasks was similar to that observed during speech-based texting.

Subjectively, participants generally felt that tasks requiring the use of the speech-based interface could be accomplished safely. The majority of participants (83%) expressed a desire to have this speech-based interface available on their next vehicle.

INTRODUCTION

The dangers of driving while texting are widely recognized due to the visual demand of this specific task type, thus increasing by more than 23 times that of normal driving a driver's likelihood of being involved in a "safety-critical event (i.e., crashes, near-crashes, crash-relevant conflicts, and unintentional lane deviations)" (*1*) . A speech-based texting interface offers the driver an alternative to tasks that are typically visually demanding. The underlying expectation is that by removing the manual component from any visually demanding secondary task drivers will be able to complete the task in a safer manner while driving.

Using the commonly accepted Lane Change Task (LCT) in a simulator environment, Maciej and Vollrath (*2*) compared effects related to lane maintenance, gaze, and subjective measures of distraction across manual and speech modalities. Song retrieval across various methods, finding a contact to call and selecting the appropriate number category (home, work, etc.), and point-of-interest (POI) destination-entry tasks were all evaluated using both a prototype speech-based interface and handheld devices dependent on the task type. As a whole, analysis suggested significant improvements in vehicular control and self-reported assessments for speech-based tasks when compared to those identical tasks executed manually. The authors did, however, identify vehicular control measures that, although significantly improved with speech compared to manual, were still worsened when contrasted to no secondary task involvement (baseline).

Tsimhoni et al. (*3*) evaluated two speech recognition approaches (word- versus character-based) that were compared to identical manual tasks conducted using a touch-screen keyboard. A simulator was also used to represent the driving environment, and secondary task involvement was evaluated under straight roadway sections and during curve negotiation. The authors found that tasks completed using the keyboard resulted in significantly degraded lateral control (defined in this report as a collective group of measures) when compared to results of the speech-based approaches, both aligned closely with performance observed during baseline driving. Tasks completed using the touch screen did result in significantly more frequent lane deviations compared to both speech approaches. However, while the character-based approach aligned with baseline driving, the number of deviations observed during tasks completed using the word-based approach was elevated, suggesting some degradation in vehicular control is to be expected with involvement in any type of secondary task.

Through a comprehensive review of speech versus manual literature available at the time, Baron and Green (*4*) were able to summarize their findings and conclusions across 15 separate studies. In all, the authors concluded that speech-based tasks typically allow for improved vehicular control (lane- and speed-related measures) compared to similar tasks completed manually. In many cases, however, a speech-based approach was still worse than driving with no secondary task involvement. As expected, a summary of the reviewed studies identified a common theme: a driver's eyes are more likely to be on the forward roadway when engaged in speech-based tasks as compared to manual tasks. To simplify, speech-based tasks allow for a driver's eyes to remain on the forward roadway, an obvious safety advantage that also impacts common vehicular control measures often assessed.

Similar to the existing literature, the primary objective of this effort was to assess on-road and general task performance while drivers engaged in secondary, speech-based tasks behind the wheel in a natural driving environment. This was accomplished through use of the speech-based interface for text-messaging tasks (compared to their manual equivalents and a baseline task) and for destination-entry tasks (compared to baseline and the other speech-based task types as there was no manual equivalent).

METHODS

Alternative implementations of text messaging while driving were evaluated using a within-subject design and comparing task performance across equivalent manual and speech-based tasks. Drivers participated in a two-hour experimental session during which they were asked to execute a series of text messaging and destination-entry tasks using either a handheld phone or a prototype speech-based interface. Also included within this effort was an observational analysis of driver interaction with the speech-based interface, representing generally uninformed users following only a brief introduction without specific training about task execution.

Instrumentation

Instrumentation comprising VTTI's data acquisition system (DAS; an updated version of what was implemented during the 100-Car Study [*5*]) was integrated into the test vehicle, which was a late model, full-size sport utility vehicle. The DAS captured full-time audio and video and driving performance measures (e.g., speed, steering wheel input obtained through the vehicle network) and inputs from an experimenter workstation (e.g., task number, trial). The video views (including driver-face view, forward view, views of the front tires [split within lower right quadrant], and an over-the-shoulder view) were captured continuously and allowed for post hoc analysis related to lane maintenance and eye-glance behavior.

In order to implement the speech-based interface, the test vehicle was also equipped with an aftermarket hands-free kit paired with a touch-screen smartphone. This hands-free kit allowed for a one-touch redial initiator through the use of a module installed at the top of the vehicle's center stack. Requiring a single manual input to initiate a call (through redialing) effectively simulated the intended production version of the speech-based interface under evaluation. Through this approach all tasks were in effect completed using the smartphone, either in a traditional manual method (handheld) or through use of a single button press of the hands-free system (speech-based).

Study Procedure

Twenty-four drivers from the New River Valley and surrounding localities (e.g., Roanoke, Salem) participated in this effort and represented two age groups balanced by gender: 1) younger (18 to 30 years old) and 2) middle-aged (45 to 55 years old). Beyond a standard phone screening across a number of eligibility requirements, qualified participants were required to text on a weekly basis (not necessarily behind the wheel); own a smartphone (e.g., Droid, iPhone); and, through self-report, identify themselves as users comfortable with an alphanumeric keyboard typically found on touch-screen phones. These selection criteria helped ensure that participants recruited would not have difficulty texting manually due to a lack of familiarity with the task or the device, thereby allowing for a robust comparison of task performance by users experienced with handheld tasks but inexperienced with a speech-based interface. A single, two-hour experimental session was all that was required on the part of the participants.

Vehicle and Smartphone Familiarization

Following completion of study-related paperwork, participants were escorted to the test vehicle where they received a brief orientation to the key driving-related controls. The experimenter then introduced the participant to the smartphone, noting that it would be used for all handheld texting tasks executed during the study. Although the participants were comfortable with touch-screen phone keyboards, few were familiar with the specific smartphone provided for this effort. A walk-through of the desired method was thereby provided, and participants were asked to use the demonstrated method for all subsequent handheld texting tasks. Participants were informed that common texting shortcuts and vernacular lingo were acceptable provided that the message would be accurately conveyed. The participants then completed a practice task following the desired method, conveying to the experimenter their comprehension of the demonstration they just received.

Introduction to the Speech-based Interface

Participants were subsequently introduced to the speech-based interface. Experimenters provided a brief overview of the system comprising services provided by the system (texting and destination entry) and how to initiate these services (hands-free module). Without any further instruction, the experimenter asked the participant to text the same message as they did during their smartphone practice task, this time using the speech-based interface. Participants were then asked to complete two more practice tasks, both destination-entry tasks: 1) obtaining and downloading directions for a given address and 2) searching for a categorical POI and obtaining and downloading directions for a specific option provided from the list of returned options. Once the participants finished all three practice tasks they were provided with the opportunity to repeat any of them, otherwise indicating they were ready to continue.

On-road Task Evaluation

Once they completed the practice tasks, participants were instructed to enter the Smart Road where the experimenter introduced them to the study protocol, the completion of which would satisfy the primary objective of this research effort. The Virginia Smart Road is an approximately 3.5 km (2.2 mi) closed test track with two traffic lanes and turnarounds at each end. For this effort, testing was limited to a 1.6 km (1 mi) section within which participants were allowed to complete the secondary task. Participants were first allowed to become accustomed to both the road and the test vehicle by driving a full lap while maintaining 72 km/h (45 mph). Following this orientation lap, participants were informed of the first task they were to perform. In total, 10 tasks were included with each of them performed two times independently of one another. The task sequences were randomly generated, and a balanced design was accomplished through use of mirrored orders. The experimenter used a place card to visually convey the task while the vehicle was parked, but beyond that no further instruction was provided about how to complete the task.

During the first trial the experimenter allowed the participant to achieve 72 km/h (45 mph) and a level of steady state before saying, “Begin.” Once the experimenter stated, “Begin,” the participant was free to initiate the task whenever he/she was comfortable doing so. There was no interaction between the experimenter and the participant once the task was underway, and the participants were instructed beforehand to say, “Done” when they felt they had completed the task successfully. It is important to note that the point at which the participant verbalized, “Done” was considered the point of task completion; data, where applicable as referenced within this document, are based on the window between initialization of the task by the participant through the point at which they stated, “Done.” Participants were then instructed to turn around and travel in the opposite direction, at which point they were verbally reminded of the task they were to perform for the second trial of that task, and the general protocol was repeated. For all manual tasks, the smartphone was placed in the center console cup holder located between the participant and the experimenter; participants were required to retrieve the smartphone from this location, returning it upon completion of the task.

Once parked at the starting point and following completion of both trials participants were asked to evaluate the task using three subjective metrics (*6*, *7*):

- Mental Demand: "…was the task easy or demanding, simple or complex?" on a scale from 1 (easy) through 100 (hard).
- Frustration Level: "How … stressed, annoyed, versus … relaxed and complacent did you feel during the task?" on a scale from 1 (low) through 100 (high).
- Situational Awareness: "How aware were you of surrounding traffic when you were performing the task?" on a scale from 1 (low) through 100 (high).

It is important to note that the Situational Awareness metric was presented as a hypothetical assessment of the participants' attentiveness to "surrounding traffic" had they been on a real road with other traffic while engaged in the task in question; no other traffic was present on the Smart Road during testing. Once these ratings were obtained, participants were asked to provide any open-ended feedback they might have before continuing to the next task.

Upon completion of the on-road assessment, participants were brought back to the building where they were asked to complete a Post-Drive Questionnaire to capture additional feedback specific to the speech-based interface.

Included Tasks The following tasks were included as part of the on-road assessment. Short, medium, and long versions of texting tasks were provided to represent a range of message types. TABLE 1 provides detailed analytical steps for tasks performed using the speech-based interface.

- Baseline (maintaining 72 km/h [45 mph]; no secondary task involvement)
- Texting to a specified contact (handheld and speech-based)
    - Short: "Testing 1 2 3"
    - Medium: "Have a nice day"
    - Long: "I'm driving to the grocery store"
- Destination-entry tasks (speech-based only)
    - Address
    - POI (closest location)
    - POI category (choose specific POI from list provided)

TABLE 1 Speech-Based Interface Analytical Task Steps

Task	Step	User Step	System Response
Text the message 'Testing 1 2 3' to Luke using the speech interface	1	Press Redial button	"Redial"; "Please say your text message"
	2	"testing 1 2 3"	"Please wait"; "Your message is 'testing 1 2 3’"; "Who would you like to text?"
	3	"Luke"	"Your message has been sent to Luke"; "Would you like to try again?"
	4	"No"	"Thank you for calling text by voice"
Text the message 'Have a nice day' to Dad using the speech interface	1	Press Redial button	"Redial"; "Please say your text message"
	2	"have a nice day"	"Please wait"; "Your message is 'have a nice day"; "Who would you like to text?"
	3	"Dad"	"Your message has been sent to Dad"; "Would you like to try again?"
	4	"No"	"Thank you for calling text by voice"
Text the message 'I'm driving to the grocery store' to Mary using the speech interface	1	Press Redial button	"Redial"; "Please say your text message"
	2	"I'm driving to the grocery store"	"Please wait"; "Your message is 'I'm driving to the grocery store"; "Who would you like to text?"
	3	"Mary"	"Your message has been sent to Mary"; "Would you like to try again?"
	4	"No"	"Thank you for calling text by voice"
Obtain directions for '100 Clay Street, Blacksburg, VA' and download them to the vehicle	1	Press Redial button	"Redial"; "Please say a destination"
	2	"100 Clay Street, Blacksburg, Virginia"	"You said '100 Clay Street, Blacksburg, Virginia'"; "Would you like to search for this destination?"
	3	"Yes"	"Please wait"; "100 Clay Street, Blacksburg, Virginia; Would you like to download this destination?"
	4	"Yes"	"Your destination is being downloaded to your vehicle"; "Thank you for calling"
Obtain and download directions to the nearest Starbucks using the speech interface	1	Press Redial button	"Redial"; "Please say a destination"
	2	"Starbucks"	"You said 'Starbucks'"; "Would you like to search for this destination?"
	3	"Yes"	"Please wait"; "Starbucks, located at 2475 Franklin Street, Christiansburg, Virginia; Would you like to download this destination?"
	4	"Yes"	"Your destination is being downloaded to your vehicle"; "Thank you for calling"
Using a basic search for Gas Station, obtain and download directions for Campus Exxon.	1	Press Redial button	"Redial"; "Please say a destination"
	2	"Gas Station"	"You said 'Gas Station'"; "Would you like to search for this destination?"
	3	"Yes"	"Please wait"; "Valero, located at 3620 South Main Street, Blacksburg, Virginia; Would you like to download this destination?"
	4	"No"	"Campus Exxon, located at 415 National Drive, Blacksburg, Virginia; Would you like to download this destination?"
	5	"Yes"	"Your destination is being downloaded to your vehicle"; "Thank you for calling"

RESULTS

A variety of measures were examined as participants engaged in secondary tasks while driving on the Smart Road, including task success rates; vehicle control measures (e.g., speed, lane maintenance); frame-by-frame, eye-glance analysis; and subjective ratings. All nine secondary tasks, not including baseline, were collapsed into the following task types: 1) handheld texting, 2) speech-based texting, and 3) speech-based destination entry. An analysis of variance (ANOVA) was conducted on these variables using a 2 (age) x 3 (task type) mixed-factorial design. Furthermore, a Duncan's post hoc test identified significant differences between the individual levels of main effect.

Task Performance

Task success rates of approximately 90% were observed for each of the three task types collapsed across both independent trials. As a reminder, a successful outcome for the handheld and speech-based texting tasks was defined as a message that was accurately conveyed as intended in addition to having been sent to the intended contact. This could also include cases where multiple messages were sent, even if inadvertent, as long as the final message was accurately communicated and sent to the correct recipient. Differences in task success rates when comparing trials one and two were nonexistent.

Figure 1 identifies task outcome by participant age group. Younger participants had very high success rates for the handheld tasks as compared to the middle-aged group (96% and 82%, respectively). However, task success rates between the two age groups for the speech-based tasks were similar, thus suggesting that using the speech-based interface equalizes the outcome or at the very least largely reduces the performance gap apparent for handheld task performance between younger and middle-aged users.

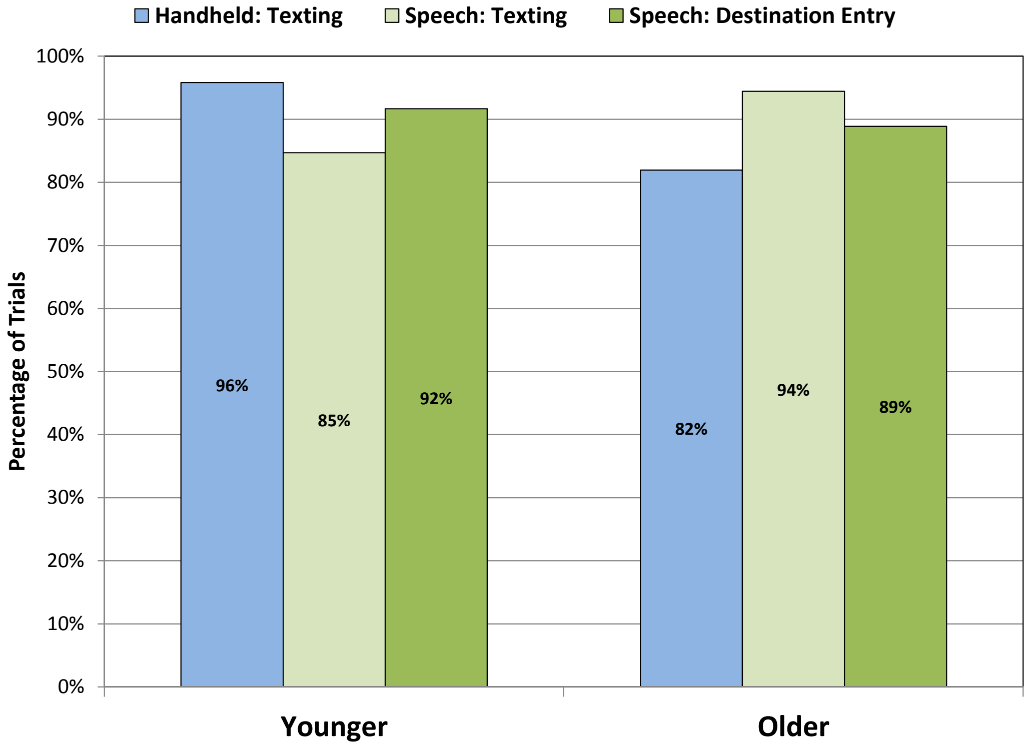

FIGURE 1 Task outcomes by age and gender collapsed.

Vehicle Performance Measures

Vehicle performance measures of lateral and longitudinal control were examined for each task and trial. Table 2 provides a summary of significant findings for each of the vehicle control measures across the tested model.

TABLE 2 Summary of ANOVA Results across Vehicle Control Measures

		Task Type	Age	Task Type x Age
Longitudinal Control Measures	Mean Speed		***	
	Speed Variance	*		**
Lateral Control Measures	Steering Variance	***	*	***
	Frequency of Lane Deviations	***	**	***
	Time out of Lane	***	*	**

*p<.05, **p<.01, ***p<.0001

Longitudinal Control Measures

Longitudinal control was assessed by each participant's ability to maintain 72 km/h (45 mph) while engaged in a secondary task. As a whole, participants were able to maintain a steady average speed, independent of task type. Differences in average speed were found only with respect to age, where younger participants were observed to drive approximately 1.6 km/h (1 mph) faster than the middle-aged participant group. Speed variance was the more discriminating measure, with significant differences found across both task type and task type by age (F [2,431] = 4.2, p = 0.0156 and F [2,431] = 5.32, p = 0.0052, respectively). As a direct comparison, texting tasks accomplished using the speech-based interface showed less speed variance than when engaging in the same task manually. Specific to the speech-based interface, differences in task type (texting versus destination entry) do not appear to influence speed maintenance. Surprisingly, younger participants appear unaffected by task type when examining speed variance across task type by age: between 2.16 km/h (1.34 mph) and 2.29 km/h (1.42 mph) across the three task types. The interaction uncovered that specific to the handheld texting tasks speed variance for the middle-aged group was higher than it was for its younger counterpart (Figure 2): 2.45 km/h (1.52 mph) compared to 2.19 km/h (1.36 mph), respectively. Conversely, this finding was reversed relative to both speech-based task types, where speed variance for the middle-aged group dropped below that of the younger drivers: 1.79 km/h (1.11 mph) and 1.80 km/h (1.12 mph) compared to 2.29 km/h (1.42 mph) and 2.16 km/h (1.34 mph) for the speech-based texting and destination-entry tasks, respectively. This finding suggests that middle-aged drivers may benefit more from speech-based tasks over manual equivalents when it comes to general speed maintenance.

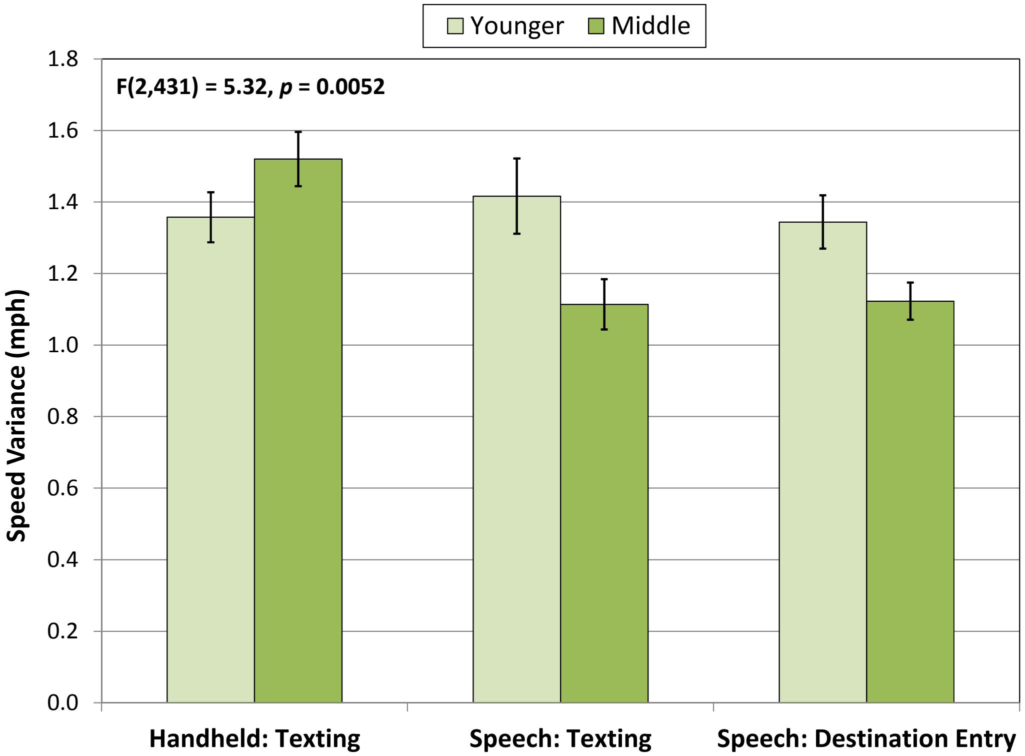

FIGURE 2 Average number of lane deviations across task type and age.

Lateral Control Measures

The ability of drivers to maintain their lane is a well-recognized safety surrogate and should receive greater emphasis over steering wheel variance when considering safety implications. Significant differences were found across task type, with lane deviations more likely to occur during handheld texting as compared to either speech-based task type (F [2,431] = 23.85, $p <$ 0.0001): 0.35 on average compared to 0.03 and 0.05 across all trials for handheld texting, speech-based texting, and speech-based destination entry, respectively. The interaction (Figure 3) indicated that the middle-aged group had significantly degraded lane maintenance while engaged in handheld texting compared to that of the younger group. Not shown but of equal importance is the number of participants observed to deviate the lane by task type. More than half of the participants (54%, or 13 out of 24) were observed to deviate the lane at least once during handheld texting tasks compared to only 13% (3 out of 24) and 21% (5 out of 24) for both speech-based texting and destination-entry tasks, respectively.

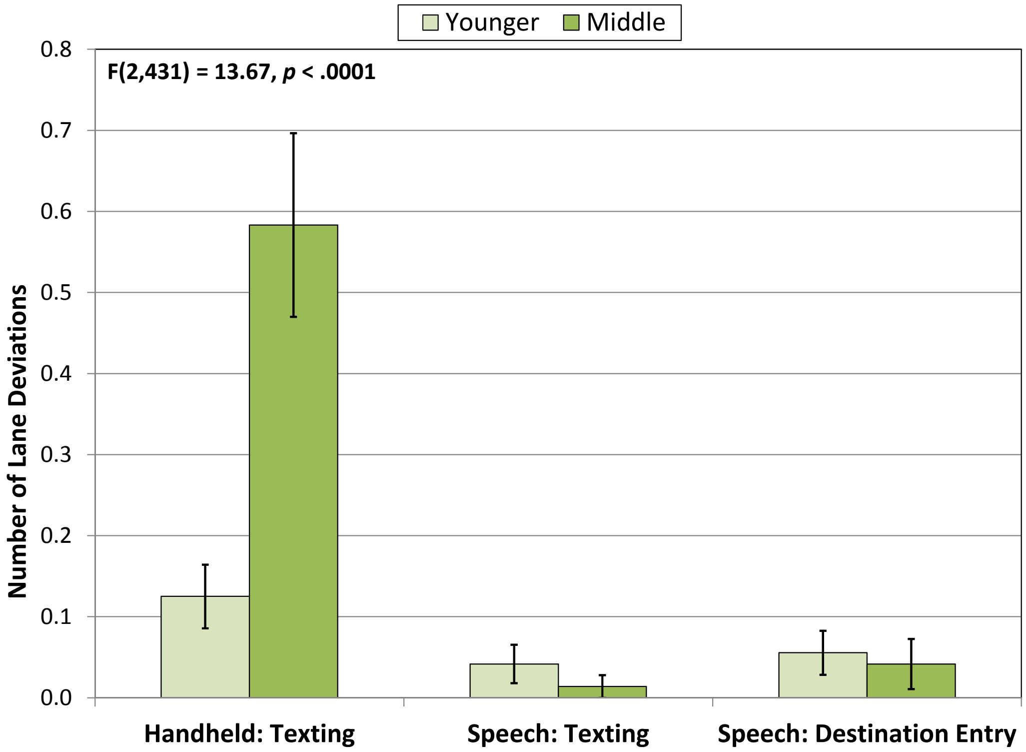

FIGURE 3 Average number of lane deviations across task type and age.

Similar trends also existed when examining the average time spent out of the lane during observed lane deviations. Handheld texting tasks, due in large part to the higher frequency of observed deviations, accumulated a significantly higher per-task average of 1.25 seconds spent out of lane compared to only 0.06 seconds for both speech-based task types (F [2,431] = 13.37, *p* < 0.0001). Differences in age groups again showed a higher average value for middle-aged participants, with an average time spent out of lane of 0.73 seconds across all tasks compared to only 0.18 for younger participants (F [1,431] = 6.31, *p* = 0.0124). This finding between the age groups was overwhelmingly due to the differences specific to the handheld texting task, an interaction between task type and age that was also significant (F [2,431] = 7.75, *p* = 0.0005). Middle-aged participants spent on average 2.12 seconds out of the lane per handheld texting task compared to only 0.37 seconds for the younger age group. Moreover, differences between the two age groups relative to the speech-based task types were nonexistent, dropping under an observed maximum of 0.11 seconds across both age groups and speech-based task types. Although younger drivers do see benefit, vehicular control is again dramatically improved during speech-based tasks for the middle-aged drivers when compared to performing similar tasks manually.

Eye-glance Analysis

Eye-glance reduction of participants while engaged in each secondary task was conducted by trained analysts who reviewed the video frame by frame, recording a glance location every 100ms. Glance location was particularly important with respect to the source of interest for each task type. A source of interest refers to the specific location where glances are required in order to complete the task. For handheld tasks the source of interest was the smartphone, whereas the source of interest for the speech-based tasks was the center stack area where the hands-free module was located.

As a point of comparison, glances made to these sources of interest were collapsed across a cumulative frequency distribution for each task type (Figure 4). Glance durations to the center stack for speech-based tasks, as a group, were noticeably shorter compared to glance durations to the smartphone during handheld tasks, as witnessed by a complete separation between the speech-based and handheld trends. When limiting focus to the texting tasks between handheld and speech-based task types, the large discrepancy in the overall number of observed glances (3,453 versus 170, respectively) suggests that to complete the same task across these two methods the average user would require more than 20 glances to the smartphone for each glance to the initiating button for the speech-based interface (hands-free module). The analysis also discovered 357 glances to the smartphone (approximately 10% of the sample) with durations greater than 2 seconds, which is considered a safety-critical cutoff at which point crash risk begins to dramatically increase (*5*, *8*). Conversely, no glances to the center stack more than 2 seconds in duration were observed using the speech-based interface while texting, and only one glance lasting more than 2 seconds occurred while engaged in a destination-entry task. Differences in glance duration between the two speech-based task types were not observed.

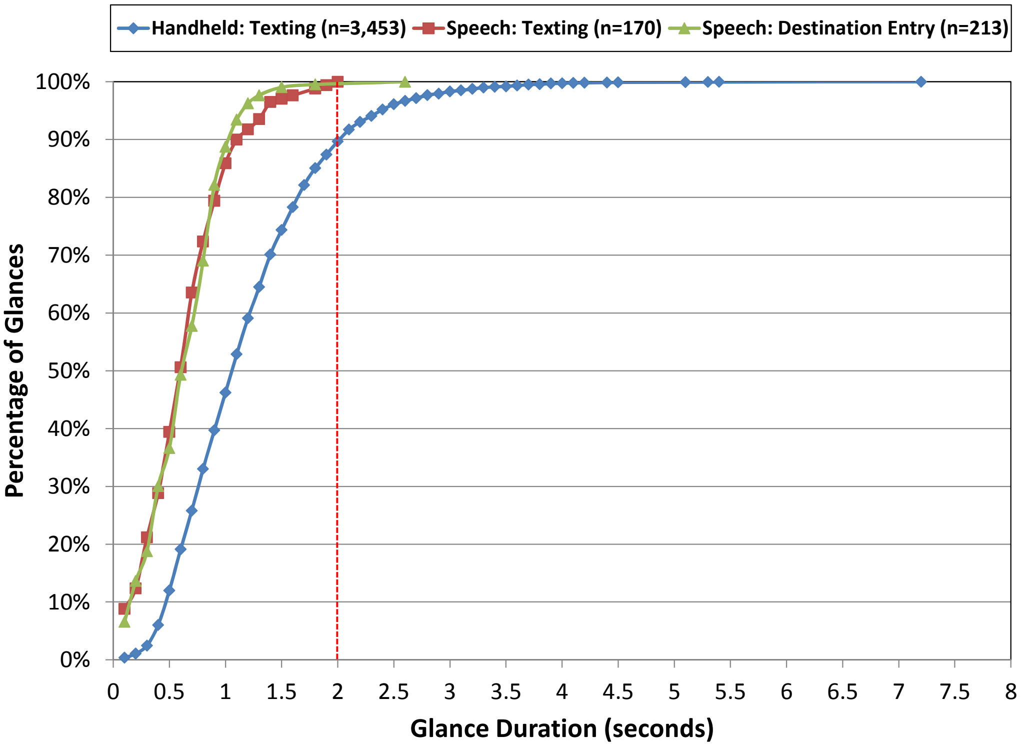

FIGURE 4 Distribution of glance durations to system of interest by task type.

Beyond glance durations to specific locations, total eyes-off-road time (EORT) as a function of task duration allows for comparison of the time the user is looking away from the forward roadway while engaged in a secondary task. Eyes were considered “off road” whenever the participant was not looking forward. Not surprisingly, the speech-based task types (percentage of EORT at 26%) required significantly less EORT (F [2,431] = 753.37, $p < 0.0001$) compared to handheld texting (percentage of EORT at 68%). This is very telling in that, regardless of task length or complexity, task type is the dictating factor in influencing the percentage of time a driver is “required” to have his/her eyes off of the forward roadway to complete a given task. Significance in percentage of EORT was also observed across the age groups (F [1,431] = 15.81, $p < 0.0001$), but a significant interaction across task type and age did not emerge. Middle-aged participants were observed to spend, on average, approximately 42% of task time looking away from the forward roadway across all tasks, compared to 38% for the younger participants.

Subjective Ratings

Participants were asked to rate their mental demand, frustration level, and situational awareness once both iterations of each task were completed. As shown in Table 3, significant differences were observed across task type, age, and task type by age for ratings of both mental demand (0:easy through 100:hard) and frustration level (0:low through 100:high) and by task type for situational awareness (0:low through 100:high).

TABLE 3 Summary of ANOVA Results across Workload Ratings

	Task Type	Age	Task Type x Age
Mental Demand	***	*	***
Frustration Level	***	**	***
Situational Awareness	***		

**p<.05, **p<.01, ***p<.0001*

On average, manual texting tasks received mental demand ratings of more than twice their speech-based counterparts: 57.7 versus 24.9, respectively (F [2,215] = 103.59, $p < 0.0001$). Both speech-based task types received comparable ratings, suggesting that participants felt that both task types were equally demanding. For comparison, the average rating for mental demand following the baseline trial was 13.3. With that, engaging in any secondary task, even a speech-based task with relatively low ratings, still increases the self-reported level of mental demand required by the driver to complete the task. Significance was observed for the interaction between task type and age (F [2,215] = 10.26, $p < 0.0001$). Again, as with many of the objective measures, this difference is accounted for by responses provided following the handheld tasks. Middle-aged drivers, on average, reported mental demand ratings of 67.2 compared to 48.3 for the younger driver group following handheld texting tasks. Ratings of mental demand were comparable between the two age groups for both speech-based task types: between 21.6 and 25.6 on average.

Similar differences and trends observed for the mental demand ratings were also evident for ratings of frustration level. Manual texting, as suggested by the participants, again received a rating of a frustration level approximately twice that of its speech-based counterpart: 49.2 for handheld texting versus 25.5 and 24.3 for speech-based texting and speech-based destination entry, respectively (F [2,215] = 33.03, $p < 0.0001$). Both speech-based task types were again comparable, suggesting that resulting frustration exists as part of using the speech-based interface and not due to engagement in texting versus destination-entry tasks. For reference, ratings following the baseline task were on average 7.8 across the driver sample. As with the ratings of mental demand, the interaction between task type and age was significant (F [2,215] = 14.51, $p < 0.0001$). Differences were specific to the handheld texting tasks where middle-aged drivers, on average, reported frustration levels of 63.9 compared to only 34.4 for the younger driver group. Frustration level ratings were comparable between the two age groups for both speech-based task types: between 23.0 and 28.0, on average.

As expected based on the EORT results discussed earlier, awareness of the surrounding environment suffers most when engaged in handheld texting tasks. As a group, these ratings were significantly lower than for all speech-based texting and destination tasks (F [2,215] = 111.39, $p < 0.0001$), with observed average responses of 49.3, 82.2, and 82.1 for handheld texting, speech-based texting, and speech-based destination-entry tasks, respectively. For reference, ratings of situational awareness following the baseline task were, on average, 92.4 across the driver sample.

DISCUSSION

A speech-based interface capable of sending text messages and obtaining route guidance through destination-entry tasks was evaluated as part of this effort. The primary objective was to assess general driving performance across a range of measures, comparing speech-based tasks to manual equivalents. Twenty-four participants comprising both younger (18 to 30 years of age) and middle-aged (45 to 55 years of age) groups and both balanced by gender were recruited for this effort.

Each participant, once familiar with the vehicle and the protocol, was asked to complete 10 separate tasks (two trials each) while maintaining 72 km/h (45 mph) on the Virginia Smart Road. These tasks included three manual texting tasks using a smartphone (handheld), three equivalent texting tasks using the speech-based interface, three destination-entry tasks using the speech-based interface, and a baseline task where participants simply maintained 72 km/h (45 mph). Workload ratings were obtained following the second trial to measure mental demand, frustration level, and situational awareness.

Perhaps the most important, yet not surprising, finding based on the existing literature discussed beforehand is that of drivers showing greater difficulty texting with a handheld device as compared to using the speech-based interface while behind the wheel. Drivers exhibited a significantly degraded ability in controlling the vehicle laterally while engaged in a handheld secondary task and subjectively thought that the task was more mentally demanding, more frustrating, and led to lower self-assessed situational awareness. The preeminent safety surrogate during this study was lane maintenance (i.e., deviations and time out of lane). Participants were much more likely to deviate from the lane while engaged in manual texting (e.g., lane deviations were significantly higher for handheld texting when compared to both speech-based texting and speech-based destination entry). Average time out of lane also demonstrated the clear advantages of a speech-based alternative over manual texting.

Driving, by nature, is a visually intensive task, so minimizing a driver's EORT and number of 2 second or longer glances to the source of interest is an important design consideration relative to these types of 'convenience' features. Dramatic differences were observed during this study when examining the analysis of frame-by-frame, eye-glance reduction. As expected, glance durations to the source of interest (center stack, where the initiating button was located) during speech-based tasks were noticeably shorter as compared to glances to the smartphone during handheld task involvement. Moreover, the frequency of observed glances suggested that for completion of equivalent texting tasks, manual tasks required, on average, 20 glances to the smartphone for every one glance to the center stack for speech-based tasks. Of similar concern, approximately 10% of the glances to the smartphone lasted more than 2 seconds, widely considered a safety-critical cutoff (*5*). The percentage of EORT also proved to be a clear separating measure between manual and speech-based tasks, revealing significant differences by task type with magnitudes beyond double for manual texting tasks compared to their speech-based counterparts.

Workload ratings showed similar divisions. Significant differences were found across all three measures, with clear divisions between speech-based and manual task types. As a group, speech-based tasks resulted in low mental demands, low frustration levels, and high situational awareness, typically only marginally different from baseline.

Age-Related Differences

Interestingly, when assessing the impact of age on the assigned tasks, the speech-based interface showed a propensity to equalize performance in terms of outcome between the two age groups as opposed to a clear disconnect during handheld task performance. Collectively, findings suggest that middle-aged participants would benefit the most from a speech-based texting alternative. In general, the middle-aged group’s driving performance was worse during the manual texting tasks as compared to their younger counterparts, yet they performed comparably and even at times better than their younger counterparts when using the speech-based interface. This finding was consistent for both lateral and longitudinal measures of vehicle control.

Subjectively, ratings of mental demand and frustration level following handheld texting tasks were higher for middle-aged participants but were similar to those of the younger participants following the speech-based tasks. This finding is in accordance with the increased difficulty exhibited by the middle-aged age group across the various vehicle performance metrics.

Study Limitations

This effort serves as a standard validation study comparing a speech-based interface with the ability to perform tasks that, traditionally, are manual only. Results should be tempered to the fact that these data were collected on a controlled test track with no other traffic present. Therefore, the driver’s workload was lower than would typically be found when driving on public roads. The study also does not attempt to address issues of cognition other than subjective estimations of workload when engaged in any type of secondary task. A suggestion for future research is to assess if speech-based interfaces in general increase secondary task involvement behind the wheel and, furthermore, what impact this may have on driving performance. For example, due to its ease of use, a user might be inclined to increase his/her texting behind the wheel to a level of high involvement. Effects of this increased interaction should be examined and appropriately understood.

SUMMARY

The speech-based modality provides a viable alternative for conducting secondary tasks in the vehicle. The speech-based system assessed during this study appeared easy to use and intuitive, which is critical for implementation in an automotive environment. The system resulted in a significant improvement over handheld texting in terms of driver performance, and no differences were found between speech-based texting and speech-based navigation tasks evaluated during this study.

ACKNOWLEDGMENT

This study was supported by a contract from Agero, Inc..

REFERENCES

1. Olson, R. L., R.J. Hanowski, J.S. Hickman, and J. Bocanegra. *Driver Distraction in Commercial Vehicle Operations.* Publication DTMC75-07-D-00006, FMCSA, U.S. Department of Transportation, 2009.
2. Maciej, J., and M. Vollrath. Comparison of manual vs. speech-based interaction with in-vehicle information systems. *Accident Analysis & Prevention,* Vol. 41, 2009, pp. 924-930.
3. Tsimhoni, O., D. Smith, and P. Green. Address Entry While Driving: Speech Recognition Versus a Touch-Screen Keyboard. *Human Factors: The Journal of the Human Factors and Ergonomics Society,* Vol. 46, No. 4, 2004, pp. 600-610.
4. Baron, A., and P. Green. *Safety and Usability of Speech Interfaces for In-Vehicle Tasks while Driving: A Brief Literature Review.* Publication UMTRI-2006-5, University of Michigan Transportation Institute, 2006.
5. Klauer, S. G., T.A. Dingus, V.L. Neale, J.D. Sudweeks, and D.J. Ramsey. *The Impact of Driver Inattention on near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data.* Publication DTNH22-00-C-07007 Task Order 23, NHTSA, U.S. Department of Transportation, 2006.
6. Blanco, M. Effects of in-vehicle information systems (IVIS) tasks on the information processing demands of a commercial vehicle operations (CVO) driver. Unpublished Master’s Thesis, Virginia Polytechnic Institute and State University, Blacksburg, 1999.
7. Angell, L. S., R.A. Young, J.M. Hankey, and T.A. Dingus. An Evaluation of Alternative Methods for Assessing Driver Workload in the Early Development of In-Vehicle Information Systems. Presented at the Government/Industry Meeting, Washington, D.C., 2002.
8. Zwahlen, H.T., C.C. Adams, and D.P. DeBald. Safety aspects of CRT touch panel controls in automobiles. In A.G. Gale et al. (eds.), *Vision in Vehicles – II*, Elsevier Science, Amsterdam, 1988, pp. 335-344.