

Usability Evaluation of a Volkswagen Group In-Vehicle Speech System

Jackie C. Chang, Annie Lien, Brian Lathrop, Holger Hees

Electronics Research Lab

Volkswagen Group of America, Inc.

4005 Miranda Avenue, Suite 100

Palo Alto, CA 94304

+1 (650) 496-7000

{jackie.chang,annie.lien,brian.lathrop,holger.hees}@vw.com

ABSTRACT

Driving behavior has been trending towards more time in the car and longer commutes. This has fueled the demand for an increasing number of in-vehicle infotainment features, at the cost of the driver splitting attention between the primary task of driving and other secondary tasks. To demonstrate one process we use for generating continuous improvements to the usability of our infotainment systems, we discuss a study where 30 participants were asked to interact with the speech dialogue system of a Volkswagen Group in-vehicle speech system. Participants performed tasks in telephone, navigation, and map contexts. Tasks were timed and videotaped for analysis of three performance measures: 1) Task Completion, 2) Task Time, and 3) participant rating of Task Difficulty. From this analysis, we identified issues that are especially important to the interaction between the system and the driver, which we categorized into a few broad areas: System Organization, Push-To-Talk Functionality, Data Entry, and Speech Commands. Analysis of the issues specific to each category and usability recommendations for each are discussed.

Categories and Subject Descriptors

H.5.2 [Information Systems]: Information interfaces and presentation – *user interfaces, evaluation/methodology*.

General Terms

Measurement, Performance, Design, Experimentation, Human Factors.

Author Keywords

Driver user interfaces, driver safety, voice user interfaces, speech interface, speech technology, speech dialogue systems.

Copyright held by author(s)
AutomotiveUI'09, September 21-22, 2009, Essen, Germany
ACM 978-1-60558-571-0/09/009

1. INTRODUCTION

With the ever-increasing features available in today's in-vehicle infotainment systems, the need for a simple, easy-to-use interface has become a necessity [2]. In the last 40 years, automotive infotainment displays and control elements have more than doubled [6]. As drivers spend more time in their vehicles, and with the trend towards longer commutes [2], the demand for infotainment features will not subside. Especially with the focus of current events and legislation on hands-free devices and the impact of these secondary tasks on driver distraction, the need for a usable interface that does not distract from the primary task of driving is ever more important. Of the control methods currently available, hands-free speech recognition is one of the most promising methods, resulting in better driving performance, less mental taxation, and less glances off the road compared to manual data entry [1], and better driving quality especially in more complex tasks such as navigation and phone dialing [3]. In fact, it has been said that the level of distraction involved in entering a voice command and listening to the vehicle's subsequent response is so low that it is comparable to that of listening to the car radio [12].

Although there have been attempts at developing natural language speech systems [4, 9, 10, 11], there is yet to be a viable product for the mass market, due to a number of challenges that must still be overcome when communicating between human and machine. These challenges range from issues with the sophistication of the speech technology to the user interaction schemes used to guide task performance. With respect to the speech technology itself, issues such as car noise interference and a limited vocabulary of speech commands that the system can recognize [5] constrain the ways that humans can interact with the system. On the usability side, many speech interfaces do not have a clear and transparent menu structure [5], which leads to confusion about why a certain command is misrecognized in certain contexts but not others. Also, the pace of speech system dialogues is far from the natural pace of human conversation [8]. Oftentimes, information needs to be entered in pieces, such as entering an address with house number, street, and city as separate utterances [4].

In light of the issues mentioned above, the goal of this paper is to illustrate a process of evaluating the usability of a speech interface system, and discuss the method of analysis that led to suggestions and rationale for future improvements.

2. EXPERIMENT DESIGN

We conducted an in-vehicle user study to evaluate the usability performance of a Volkswagen Group speech dialogue system. The goals of the user study were to: 1) Observe participants' interactions with the voice-activated phone, navigation, and map contexts of the system, and 2) Generate system specific recommendations as well as general suggestions and principles for how such a system can best interact with a user. All tasks were timed and performed only through voice operation.

2.1 Participants

Thirty native English speaking adults (17 male and 13 female), between the ages of 21 – 56 years old ($M=37$, $SD=9$), were recruited to participate in the study. Of the 30 participants, all were computer literate and all possessed a valid driver's license with 27 participants having had a license for 10 or more years; 24 participants drove 10,000 or more miles per year; and 26 participants had previous experience with navigation systems. In terms of level of familiarity with navigation systems, on a scale from 1 to 5 (1=Not Familiar, 5=Very Familiar), 25 participants possessed a familiarity level of 3 or more. Participants were recruited from the local community and compensated with a payment of an \$80 check for experiment participation. Each experiment session lasted approximately 1.5 hours.

2.2 Apparatus

Throughout the experiment, participants were seated in the driver's seat of a stationary Volkswagen Group vehicle equipped with a speech recognition system for infotainment control, which allowed users to operate contexts such as navigation, phone, media, and setup. Voice commands were available for operating a subset of the functions of the system.

2.3 Procedure

Participants began by signing a consent form and filling out a pre-experiment questionnaire in order to collect their demographic information and driving habits. After completing the pre-experiment questionnaire, each participant received a brief training session so that they could become familiar with the basic operation of the speech user interface and know how to access the system's help feature. Training included instruction and practice on operating the Push-To-Talk button on the steering wheel, listening for the beep as a cue to speak to the system, and accessing the help menu by saying the command "Help." Additionally, participants practiced interrupting the system while the system was speaking and canceling an action by saying "Cancel," or holding down the Push-To-Talk button (long press). No additional training was provided, as for the purposes of this experiment, it was important to keep participants naïve about which commands to use.

After training, participants were given a paper packet that listed 21 tasks in phone, navigation, and map contexts of the speech system (see below, Tasks). All tasks were performed by voice operation only. Tasks were videotaped and timed, with a maximum allowed time of 3 minutes to complete each task. After 3 minutes had passed, if the participant was not done with the task, the experimenter stepped in to finish the task. This was necessary from a practical and logistical standpoint.

Participants were encouraged to use the vehicle's "Help" menu if they were stuck and needed help. Beyond answering any questions the participant had prior to timing began, once the stop

watch had started, the experiment administrator only stepped in to clarify a task. Clarification was required if participants seemed confused about a task to the point where they were obviously off task and unaware of it, or thought they had completed a task but in actuality had not. Experimenter intervention occurred in approximately 19% of the tasks.

2.4 Tasks

The following is an example of the sequence of a task. The participant reads the task aloud, for example, "You need to find a house located at *NNN* S Blaney Ave in Cupertino, CA. Please enter this house address as a destination into the navigation system." The participant then asks the experimenter for any clarification if the task was not understood. Once any and all questions have been answered, the participant begins performing the task, and the experimenter starts timing. In the case of correct completion of this task, the participant begins by pressing the Push-To-Talk button and, upon hearing the subsequent beep of the microphone turning on, says the command "Enter destination." The system would then prompt the participant by asking "Please enter the city" and beep when it is ready to receive the next command. The participant states the city name. The system then repeats the city name that it registered, prompts the participant for the street name, and beeps when ready to receive the response. The task is complete once the participant has finished entering a destination in its entirety and activated route guidance. At this point, task timing stops.

The 21 tasks (in order of appearance in the experiment) were as follows:

2.4.1 Telephone Tasks

- (1) "Please navigate to the directory using voice command."
- (2) "Please call Harrison by finding his phone number in the directory through voice command and dialing the number."
- (3) "Please find Jackie's contact information, this time by using voice command to scroll down the list of contacts in the directory until you find Jackie."
- (4) "Please navigate to the telephone menu."
- (5) "Please call the number (*NNN NNN-NNNN*)." (Note: a real phone number was used; it is masked here for security reasons.)
- (6) "Please call the international phone number 011 *NN NNN NNNNNNN*." (Note: a real phone number was used; it is masked here for security reasons.)
- (7) "Please redial the last number that you have just called."

2.4.2 Navigation Tasks

- (8) "Please switch over to the Navigation menu."
- (9) "You need to find a house located at *NNN* S Blaney Ave in Cupertino, CA. Please enter this house address as a destination into the navigation system." (Note: a real address was used; it is masked here for security reasons.)
- (10) "You have just realized that the house is actually located at *NNNN* Alma St. in Palo Alto, CA, and you need to correct the address. Please change the destination." (Note: a real address was used; it is masked here for security reasons.)
- (11) "You no longer need to go to this house. Cancel the route guidance of the navigation system."
- (12) "Instead, you would like to go to Daniel Jones' address in Palo Alto, which has already been saved in your address

book. Find this address from your address book and navigate to it.”

- (13) “You need to go to the previously entered destination of NNNN Castro Dr in San Jose. Please select this address from a list of previous destinations.” (Note: a real address was used; it is masked here for security reasons.)
- (14) “You would also like to visit the corner of Prospect Rd and Miller Ave in Saratoga. Enter this intersection of Prospect Rd and Miller Ave as a new destination.”
- (15) “Select Gas Stations as a Point of Interest and use the navigation system to locate and display the addresses of the nearest gas stations. Select the 3rd gas station listed on the second page of the list of gas stations as your Point of Interest.”

2.4.3 Map Tasks

- (16) “You want to view a map of your vehicle and the surrounding area where you are currently located. Bring up a map of your current location onto the display screen.”
- (17) “Proceed to zoom out from your vehicle’s location on the navigation map display.”
- (18) “Zoom in on the navigation map to a scale of 50 yards.”
- (19) “Change the orientation of the map so that the display is oriented northward.”
- (20) “Change the map display from daytime to nighttime display.”
- (21) “Switch from a 2D map image to a 3D map display.”

In choosing the wording to describe each task, emphasis was placed on colloquial usage, in other words, tasks were not worded to provide clues to the participant as to which speech command to use. Any similarities between task wording and the system’s actual speech commands were not intentional.

2.5 Performance Measures

For each task, the following performance measures were recorded:

2.5.1 Task Completion

A task was considered “Complete” if the participant was able to finish the task successfully within the 3-minute time limit. If at 3 minutes the participant was not able to finish the task, the task was considered “Not Complete” and the experiment administrator intervened and finished the task for the participant.

2.5.2 Task Time

Total task time needed to complete a task was recorded with a stop watch. Both the task time including experimenter intervention (which means total task time recorded was more than 3 minutes), and excluding experimenter intervention (which means the maximum task time was cut off at 3 minutes), were recorded.

2.5.3 Task Difficulty

After performing each task, on the same paper packet containing the list of tasks, the participants rated task difficulty on a scale from 1 to 10, with 1 indicating “Very Easy” and 10 indicating “Very Difficult.”

3. RESULTS

We utilize the results of our analysis in two main ways: 1) to make system specific recommendations, and 2) to provide general guidelines for enhancing the usability of the voice or graphical interface.

System Specific Recommendations. Through our analysis of individual tasks (see Results by Task), we use the most problematic tasks to identify areas for system specific recommendations. These recommendations can be as detailed as examining the exact steps required to complete that particular task.

General Guidelines for Voice and Graphical Interface. Additionally, we analyze the data for more general problem areas (see Results by Problem Area) to provide guidelines in the overall design of the voice and graphical interface. These guidelines can be applied system wide, and are oftentimes general enough to be relevant to any similar speech interface system.

3.1 Results by Task

The following graphs show results for each task in terms of the three performance measures described above. (Two graphs are shown for the performance measure of Task Time, one which includes experimenter intervention time, and one which excludes experimenter intervention time).

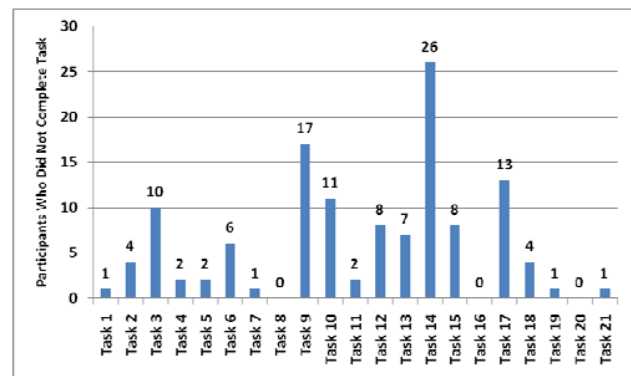


Figure 1. Task Completion (number of participants who could not complete the task)

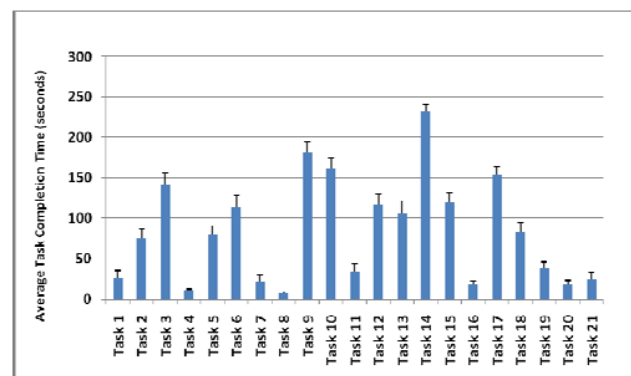


Figure 2a. Average Task Time, Including Experimenter Intervention (across 30 participants)

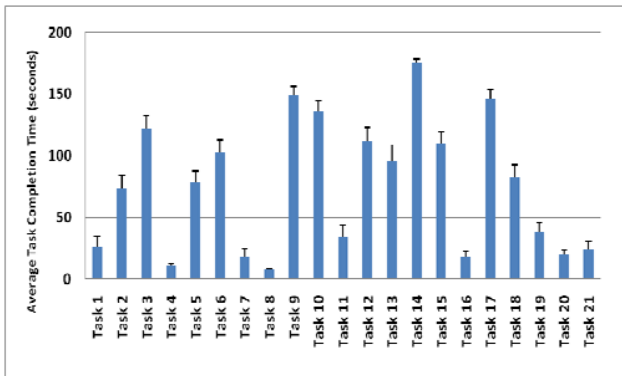


Figure 2b. Average Task Time, *Excluding* Experimenter Intervention (across 30 participants)

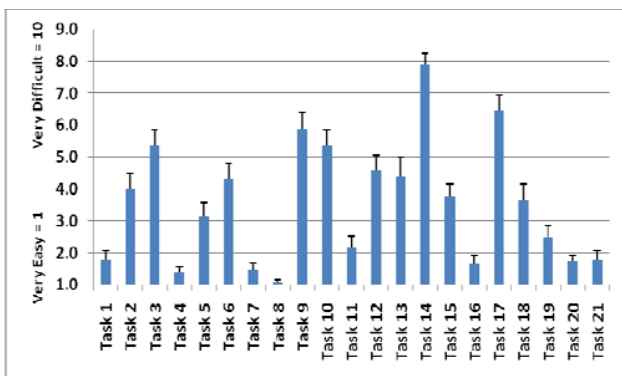


Figure 3. Average Participants' Reported Difficulty Rating (across 30 participants)

The table below (see Table 1) shows the correlations between Task Completion and the other dependent variables.

Table 1. Correlations between Task Completion and Other Dependent Variables

Dependent Variable	Correlation Coefficient
Task Time (<i>Incl.</i> Experimenter Intervention)	$r = 0.95$
Task Time (<i>Excl.</i> Experimenter Intervention)	$r = 0.91$
Difficulty Rating	$r = 0.92$

Since Task Completion was highly correlated ($r \geq 0.90$) with all other dependent variables, we used Task Completion as the marker in identifying areas for making system specific improvements. In particular, we focused on those tasks where one third or more of the participants could not complete the task. Additional conclusions relating to the other dependent variables will not be discussed here due to their high correlation to Task Completion. Given these criteria, the problematic tasks are listed below (see Table 2).

Table 2. Top 5 Problematic Tasks for Participants

Task 3: "Please find Jackie's contact information, this time by using voice command to scroll down the list of contacts in the directory until you find Jackie."

Task 9: "You need to find a house located at *NNN* S Blaney Ave in Cupertino, CA. Please enter this house address as a destination into the navigation system."

Task 10: "You have just realized that the house is actually located at *NNN* Alma St. in Palo Alto, CA, and you need to correct the address. Please change the destination."

Task 14: "You would also like to visit the corner of Prospect Rd and Miller Ave in Saratoga. Enter this intersection of Prospect Rd and Miller Ave as a new destination."

Task 17: "Proceed to zoom out from your vehicle's location on the navigation map display."

3.2 Results by Problem Area

After tabulating the task performance data, we did a more thorough analysis of all the videotaped sessions for all tasks to understand the causes of confusion and find improvements that are not necessarily task-specific. Through the experimenter's observation of the participants, the assumptions made from these observations, and the participants' self-reported comments, we identified a number of repetitive and consistent problem areas which disrupted the interaction between the user and the system and hindered participants' ability to perform tasks. The problem areas are system-wide and not specific to any particular task, and often spanned across multiple tasks. Problem areas we identified are comparable to those reported in other studies of speech system usability [3, 4, 5]. Those problem areas that were observed in 5 or more participants are listed in the table below (see Table 3).

Table 3. Problem Areas Identified (≥ 5 participants with problem)

Problem Area	# Participants with problem
System Organization	
- Global vs. Local Commands	21 out of 30
- Undo or Back	21 out of 30
Push-To-Talk Functionality	
- System Playback Interruption	13 out of 30
- System Misrecognition Not Conveyed	10 out of 30
- Microphone On/Off Not Apparent	23 out of 30
- Timing of Microphone On Indication	15 out of 30
Data Entry	
- Pace of Data Entry	5 out of 30
- Order of Data Entry	17 out of 30
- Format of Data Entry	19 out of 30
Speech Commands	
- Misleading Help Commands	22 out of 30
- Selection of Wrong Command	28 out of 30

To illustrate how problem areas were identified through observation of a task, we use the example of entering an intersection (Task 14). During the course of performing this task, there are a number of problems a participant may encounter.

Many of these problems (for example, “Microphone On/Off Not Apparent” or “Global vs. Local Commands”) can probably be overcome in the 3-minute time frame and successful completion of the task is still achievable. Other problems may be more difficult to overcome and thus will ultimately be the main factor in preventing a participant from successfully completing the task under 3 minutes. In our analysis (see Table 3), we have provided a general overview of how many participants experienced the problem area, which could have occurred in any of the 21 tasks.

Though one of the purposes of this experiment was to find system-specific improvements, we feel the problem areas identified could be generalized to be applicable to any similar speech system design. We will now discuss each problem area in more detail, using examples from the particular system evaluated.

3.2.1 System Organization

In using any system, users feel more comfortable when they have a mental map of how the system is organized. Having an unclear or overly complicated structure leaves the user feeling lost and unable to repeat an action that they have just completed. Having an unclear idea of the system’s structure creates further usability issues that manifest themselves in different ways.

3.2.1.1 Global vs. Local Commands

21 out of 30 participants had issues identifying which commands could be used globally (throughout the entire system) vs. locally (in certain contexts, such as navigation only or radio only). Global commands included commands such as “Navigate to [destination],” “Call [name],” or “Enter Destination.” Local commands included such commands as “Enter Street” which is only functional when the user is already in the navigation context. A related issue that exacerbates the problem is not having a clear delineation between the various contexts. Many systems, understandably, try to link “Map” and “Navigation” contexts, or link “Radio” and “CD/Media” contexts. The problem arises when the system attempts to partially link as well as partially keep these contexts separate. As with the test system used in this case study, while “Map” and “Navigation” were separate screens, they both shared the same color in terms of text font and graphics. This creates problems due to the global and local commands discussed earlier. Users are confused as to which command they can use and when. Contexts, therefore, should be clearly differentiated both in terms of the speech and graphical user interface to minimize user confusion.

3.2.1.2 Undo or Back

21 out of 30 participants had issues making a correction to their task. This happened in two main instances: 1) in making a correction during data entry, and 2) in returning to a previous state or menu. The system we evaluated employed two different commands that were similar but varied slightly in functionality. “Cancel” allowed the user to stop the current dialogue. (For example, when the user says “Navigation”, the system repeats “Navigation” once it arrives at the navigation context, then opens the microphone. At this point, saying “Cancel” would stop the dialogue and close the microphone). “Correction” allowed the user to make a correction to a data entry field. To complicate the matter, however, saying “Cancel” during a data entry field (instance 1) does not work because the system attempts to match “Cancel” to the closest sounding phonetic representation, which could be a street name, city name, or person name. For example, during street address entry, saying “Cancel” when the system is expecting a street name causes the system to find the street name that sounds most like “Cancel.” In this particular system, there

was no way to return to a previous state or menu (instance 2).

An example of an interface where this is not a problem is the computer. Word processing applications have one command (CTRL-Z) which works all the time, either in instance 1 or 2, and simply undoes the last user action. It is a feature that almost all users are familiar with and is easy to understand.

3.2.2 Push-To-Talk Functionality

Although this case study is an evaluation of a speech system interface, we would like to emphasize that we are not evaluating the accuracy of the speech recognition technology. Rather, we are evaluating the usability of the speech interface, in terms of interaction with the user, pace and timing of dialogue, etc.

3.2.2.1 System Playback Interruption

The particular system used in this evaluation repeated the command that it heard back to participants, as a method of feedback and verification. Many speech systems incorporate some form of this. In the example of the system used for this evaluation, when the user states “Navigation,” the system repeats the command “Navigation” as confirmation before changing to the navigation context. This worked for users most of the time, but this method of confirmation did not work very well in the telephone context. 13 out of 30 participants expressed confusion when the system interrupted them by repeating back the telephone number digits that they had just entered, prior to their completion of the entire telephone number entry. While the user can continue with the number entry where they left off once this happens, 13 out of 30 users did not know this, and instead thought they had to delete the entire number and start over from the beginning.

A few minor adjustments can be made to this interface in order to provide more clarity to the user. First, instead of repeatedly telling the participant to enter a number when a number has already been partially entered in the telephone number field, the system could give a command such as “Please *finish* the number entry” or “Please *continue* the number entry” which gives the user a cue that the previously entered digits are still acknowledged. A visual cue that can accompany the vocal dialogue is having the phone number field split into three separate smaller data entry fields, rather than having one large data field, which would subtly suggest to the user that the verbal entry can also be broken up into three separate parts. The system can also wait longer before interrupting with the playback, as most people cannot verbally recite an entire phone number without at least one short pause.

3.2.2.2 System Misrecognition Not Conveyed

With any technology that receives input and interaction from a user, such as speech recognition, it is necessary to convey to the user whether a misrecognized command is the fault of the system (speech recognition) or the fault of the user (wrong command used). The system sometimes asks for verification or asks the user to repeat the command when it does not recognize a command. However, above a certain threshold of certainty, the system does not ask the user to repeat, but rather simply executes the command. We noticed that in 10 out of 30 cases, the participant had in fact used the correct command, but due to system misrecognition and the fact that the system did not ask for verification, the user thought the wrong command was used and thus never tried that particular command again. One way to alleviate this problem is for the system to repeat what it thought it heard every time, while executing the command which allows the user to go back and repeat the command if necessary.

3.2.2.3 Microphone On/Off Not Apparent

Given the noisy environment of the vehicle and the fact that

speech recognition systems cannot yet detect the difference between human-to-human speech and commands directed at the system, it necessitates that there is a state when the microphone of the system is turned on in order to receive commands from the user, and turned off in order to ignore speech not intended for the system. In the system that we evaluated, the differentiation between the microphone on and off states was not always apparent to the user (in 23 out of 30 cases). The system indicates that it is ready to receive a command by beeping as well as showing a microphone on icon on the screen. In order to improve upon the current interface, we suggest that the icons for microphone on and off be more visually differentiated. Some systems make it apparent for the user when they can talk because the user needs to hold down a button the entire time that speech is being inputted. However due to the fact that both hands are occupied while driving, instead we suggest that when the microphone closes, the system can say "microphone off" to make it obvious to the user.

3.2.2.4 Timing of Microphone On Indication

As stated previously, the system evaluated uses a short beep to indicate to the user every time the microphone is on and ready to accept speech commands. The timing of the beep is such that for very short commands in rapid succession, the user's command is in many cases partially not registered by the system because the user spoke before the beep has occurred. In such instances, which happened in 15 out of 30 cases, the pace of the dialogue can be improved such that it more closely mimics the pace of natural conversation. Additionally, there can be a buffer so that the system is listening for a command already, shortly before the beep occurs.

3.2.3 Data Entry

Given the linear nature of speech input, data entry is a special consideration. The main use cases of data entry that were tested are: 1) telephone number entry, and 2) destination entry for navigation.

3.2.3.1 Pace of Data Entry

In the telephone context, when dialing a phone number, 5 out of 30 participants did not know that digits can be said in groups, for example, the user saying "650" [wait for system to register] versus the user saying "6" [wait for system to register], "5" [wait for system to register], then "0" [wait for the system to register]. This caused considerable frustration on the users' part because data entry could easily take twice as long when it is done digit by digit. One easy way that this could be solved is by providing a visual cue to users, for example (as stated previously), breaking up the telephone number field into three separate smaller fields as opposed to having one long field.

3.2.3.2 Order of Data Entry

Another unique aspect of navigation systems is that sometimes data entry must occur in a specific order before the system can move forward. This can be difficult for non-experienced users to understand, because it is counterintuitive to how a computer works. It is especially problematic when entering a destination for navigation. Given the nature of navigation systems, oftentimes the database of addresses needs to be narrowed down before the system can find the proper data. For instance, most systems require the user to specify the state in the U.S. where the destination is located before it can find a street. This is so that the system can narrow down the possible matching streets, since it does not have the vast memory and processing power of a desktop computer. When users are not aware of this mannerism of in-

vehicle navigation systems, this can cause some problems. In 17 out of 30 cases, participants did not know that in order to find the particular street that they are interested in, they need to have the correct city first. For example, when the city field is prefilled to "San Francisco" (from the previous destination entry), unless the user changes the city to "San Jose" prior to entering the street name, the system will only look for streets in "San Francisco."

To alleviate this problem, systems should clearly gray out unavailable fields, in order to guide the user into the correct order of entry. The order of the destination entry fields on the screen itself can also be rearranged. The other option is for the street entry prompt to clue the user in to the fact that only streets in the displayed city will be found. For example, changing the prompt from "Please enter a street" to "Please enter a street in the displayed city" will alert the user to this fact. If this is not enough, the prompt can say something more discrete such as "The city field needs to be updated first before a street can be entered."

3.2.3.3 Format of Data Entry

Because the system does not possess human understanding of speech, in addition to the data input that is needed, another piece of information that needs to be conveyed to the user is the format of the data input that is needed. This is most apparent in destination entry tasks. Many participants will enter an address by saying street and house number all as one string, for example, "NNN South Blaney Avenue" all as one string for the street field, rather than saying "South Blaney Avenue" for the street field and "NNN" for the house number field. Another issue occurs when entering an intersection. The system asks "Please enter the intersection." It is unclear what format the data for an intersection is supposed to be. Is it the first street, then the second street? Or both street names at once, separated by an "and"? When designing a system, these considerations need to be put in place, because when speaking to another human being, the particular format does not matter—all formats can be understood.

3.2.4 Speech Commands

The particular wording of the speech commands themselves has a large effect on the usability of the system. First, many real life users might not consult the system's instruction manual, and second, commands that more closely match users' natural predilection will be more memorable, and thus perceived as easier to use.

3.2.4.1 Misleading Help Commands

In using any unfamiliar system, the user looks to the instructions given by the system as guidance. When the instructions are misleading, it is almost impossible for the user to disregard the misleading information and do what is intuitive to him or her. 22 out of 30 participants came across this problem in the particular instance of attempting to enter in a destination as an intersection into the system. In the destination entry process, once participants specify the city and street, they have the option to enter either a house number or a second, intersecting street. The voice guidance at this point tells the participant to enter a house number. It does not mention that an intersecting street can also be entered at this time. While a minority of participants were able to intuitively guess that an intersecting street might also be acceptable by the system at this juncture, the majority did not figure this out.

3.2.4.2 Selection of Wrong Commands

Commands that most naturally correspond to local dialect are those that will be most accessible and memorable to the user. In a few instances of our evaluation, having a poorly chosen command

for a particular feature in the system rendered that feature almost impossible for the user to access, even with the aid of the help menu list of possible commands. The most poignant example was the task of zooming out on the navigation display map. While users are familiar with the terms “zoom in” and “zoom out,” the system’s commands for this feature were “Map Smaller” and “Map Larger.” Since this is not the vocabulary that users commonly use to refer to this feature, even after they had viewed a list of possible commands, of which “Map Larger/Smaller” was one of them, 28 out of 30 participants still could not activate the function. To further confound the issue was the fact that another, unrelated feature used similar language as what was thought to be appropriate for the zoom in/zoom out feature. The feature of “Intersection Zoom,” where the vehicle zooms in on the map display whenever the vehicle approaches the intersection, was commonly mistaken for zooming in and out of the map.

4. DISCUSSION

The current study discusses the methodology used and the results found in our evaluation of the Volkswagen Group speech system. Questionnaire and videotape data were collected across 30 participants. Statistics were compiled for task performance and by problem area. Major problem areas in System Organization, Push-To-Talk Functionality, Data Entry, and Speech Commands were identified. System specific suggestions as well as general recommendations for addressing these common speech interface usability issues were discussed.

An obvious limitation of this study was the fact that it was conducted while the vehicle was stationary. (Though it has been reported that speed of performing a speech task is relatively unaffected by whether the participant was driving or stationary [7]). However, in order to gain a more complete picture of the impact of driving on speech system use performance, the experiment should be replicated in a closed driving course during real driving, or during simulated driving using a driving simulator.

Although we touched on some visual cues in the context of how they could have supported the speech interface, a much deeper analysis into the system in its entirety, speech and visual cues (and maybe even tactile cues) in conjunction, could provide some more elegant solutions to improving the usability of the system.

Additionally, the task list could be broadened to cover some other areas of functionality. The current task list was developed to research the most common use cases according to our own personal experiences.

Also, to have a more accurate evaluation of different implementation details, the same experiment could be replicated with different vehicle infotainment systems of various brands.

The current study discusses findings and results in terms of usability from a human perspective. Although current limitations to speech recognition technology (difficulty distinguishing between similar sounding words, processing power, and limited database of commands) constrains much of what the system can do, the goal of this paper is not to offer a definitive solution that can be technically realized, but rather to explain why some current ways of implementation can be confusing to a user. In this way, we are influencing the future design of our systems by offering some explanations for the mismatch between the mapping of the system design and the human brain.

5. ACKNOWLEDGMENTS

We thank Kristin Dettmer and Dr. Stefan Christ for their support of this project. The statements contained in this document represent claims of the authors only, and are not necessarily representative of the views of Volkswagen AG, or their associates. We would also like to thank Jessica Yuan for her assistance in data collection, and Jerry LeBlanc of Volkswagen Group of America, Inc. for sharing his expertise regarding the system’s functionality.

6. REFERENCES

- [1] Barón, A., Green, P. Safety and Usability of Speech Interfaces for In-Vehicle Tasks while Driving: A Brief Literature Review. Technical Report UMTRI-2006-5. University of Michigan Transportation Research Institute (2006).
- [2] Driver Distraction: trends and issues. Computing & Control Engineering, IEE (2005), 28-30.
- [3] Gärtner, U., König, W., Wittig, T. Evaluation of Manual vs. Speech Input When Using a Driver Information System in Real Traffic. SENECA Evaluation. DOI=http://ppc.uiowa.edu/driving-assessment/2001/Summaries/Driving%20Assessment%20Papers/02_Gartner_Wittig.htm
- [4] Gruenstein, A., Reimer, B., Orszulak, J., Mehler, B., Liu, S., Seneff, S., Roberts, S., Glass, J., Zabel, J., Coughlin, J. City Browser: Developing a Conversational Automotive HMI. In Proceedings of CHI 2009, ACM Press (2009), 4291-4296.
- [5] Hataoka, N., Araki, M., Matsuda, T., Takahashi, M., Ohtaki, R., Obuchi, Y. Evaluation of Interface and In-Car Speech – Many Undesirable Utterances and Sever Noisy Speech on Car Navigation Application. In Proceedings of IEEE Workshop on MMSP (2008), 956-959.
- [6] Reissner, U. Gestures and Speech in Cars. DOI=http://www14.informatik.tu-muenchen.de/konferenzen/Jass07/courses/3/3_Gestures-and-Speech-in-Cars_Report.pdf
- [7] Tsimhoni, O., Smith, D., Green, P. Address Entry While Driving: Speech Recognition Versus a Touch-Screen Keyboard. Human Factors. Human Factors and Ergonomics Society (2004), 600-610.
- [8] Villing, J., Holtelius, C., Larsson, S., Lindström, A., Seward, A., Åberg, N. Interruption, Resumption and Domain Switching in In-Vehicle Dialogue. In Proceedings of GoTAL, Springer LNCS/LNAI, 5221 (2008), 488-499.
- [9] Weng, F., Varges, S., Raghunathan, B., Ratiu, F., Pon-Barry, H., Lathrop, B., Zhang, Q., Bratt, H., Scheideck, T., Xu, K., Purver, M., Mishra, R., Lien, A., Raya, M., Peters, S., Meng, Y., Russell, J., Cavedon, L., Shriberg, E., Schmidt, H., & Prieto, R. CHAT: A Conversational Helper for Automotive Tasks. In Proceedings of Interspeech: ICSLP (Pittsburgh, Pennsylvania, September 17-21, 2006).
- [10] Weng, F., Yan, B., Feng, Z., Ratiu, F., Raya, M., Lathrop, B., Lien, A., Mishra, R., Varges, S., Lin, F., Purver, M., Meng, Y., Bratt, H., Scheideck, T., Zhang, Z., Raghunathan, B., Peters, S. CHAT To Your Destination. In Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue (Antwerp, Belgium, September 2007).

- [11] Yan, B., Weng, F., Feng, Z., Ratiu, F., Raya, M., Meng, Y., Varges, S., Purver, M., Lin, F., Lien, A., Scheideck, T., Raghunathan, B., Mishra, R., Lathrop, B., Bratt, H., Peters, S. A Conversational In-car Dialog System. In Proceedings of NAACL/HLT (Rochester, New York, April 22-27, 2007).
- [12] Yoshitsugu, N., Miki, Y., Ito, T., Matsunaga, M. Study of Driver Distraction Due to Voice Interaction. SAE International (2003), 71-75.