Speech Recognition Interface Design for In-Vehicle System

Zhang Hua
Continental Automotive Singapore Pte
#08-02/08, Blk28 Ayer Rajah Crescent
Singapore 139959
+65 6779 9652
Hua.2.zhang@continental-corporation.com

Wei Lieh Ng
Continental Automotive Singapore Pte
1st line of address
2nd line of address
+65 6779 9678
ng.weilieh@continental-corporation.com

ABSTRACT

This paper aims to provide a framework of guidelines for the design of an in-vehicle speech recognition interface. In the first section, a background of speech recognition technology is introduced to explain why it is necessary to provide specific guidelines for in-vehicle speech recognition interfaces. The second session reviews two parts of previous research work; existing guidelines on general speech recognition interface design and physical and cognitive performances during driving and using speech recognition. However, the current research results do not conclude on how to design a speech recognition interface for in-vehicle systems, thus for the third section, an actual case-study from our organization was evaluated to identify usability issues. It describes how to apply general speech recognition guidelines into an in-vehicle speech recognition interface and introduces new solutions to solve the found usability issues.

Keywords

Speech recognition, Design guidelines, In-vehicle multimedia systems

1. INTRODUCTION

Speech recognition converts spoken words to machine-readable input, this input allows the machine to identify words the person is speaking and subsequently process the command [17]. It lets users manipulate the machine verbally without having to manually control it. This has the benefit of helping users to complete their work more efficiently while doing multiple tasks simultaneously.

Compared with the traditional control interfaces, speech recognition interfaces reduce the amount of attention the user has to spend on the mechanics of recording information of selecting functions and instead allows users to concentrate on their primary task. The advantages of using speech recognition include reduced user training time, increased worker productivity, and reduced secondary key input, and improved timeliness and accuracy of information made available via voice. [16] It has been widely applied in various domains. For example, in health care, it facilitates disabled to access computers; in military, the speech recognition application could reduce pilots' workload.

Speech recognition technology has also been applied within in-vehicle systems. Some Bluetooth phones and multimedia systems in cars can be connected, letting users dial or answer their mobile phones through speech. Other multi-media systems even have built in speech recognition interfaces that allow drivers to manipulate audio in the car while driving. However, there is a lot of debate surrounding using speech while driving. A lot of research have been done in order to prove and disprove the hypothesis that using speech recognition while driving can cause safety issues [4-6, 8, 13, 14]. These research have typically focused on one or more measurement of objective driving performance; e.g. task performance and subjective workload. Adriana B. and Paul G [1] summarized fifteen articles on the use of speech recognition while driving, but because of a lack of common definitions of dependent measures, unique test methodology and insufficient statistical data, in the end, there were few firm conclusions that could be drawn.

Regardless, it is clear that more automotive companies – both auto manufacturers as well as Original Equipment Manufacturers (OEM) - are implementing speech recognition technology into their systems to provide high-end products to drivers. However, there are no systematic guidelines still on how to design a user friendly speech recognition interface in order to reduce negative impacts to drivers.

In order to provide guidelines for designing an in-vehicle speech recognition system interface, the following was done:

1. Review existing research on general speech recognition interface design and identify areas applicable to in-vehicle design.
2. Review a driver's performance and cognitive workload during driving and using speech recognition
3. Evaluate a case-study from our organization to identify usability issues and describe how general speech recognition guidelines can be applied to in-vehicle speech recognition interfaces.
4. Develop speech recognition interface guidelines for driver needs.

2. LITERATURE REVIEW

2.1 Speech recognition Technology

Speech recognition technology has been applied in many domains. However, it should not be used indiscriminately and

AutomotiveUI'10, November 11-12, 2010, Pittsburgh, Pennsylvania
ACM 978-1-4503-0437-5

careful attention to their design and the complexity of the underlying system is critical [6].

Compared with traditional screen-based interfaces, they have a few differences. Firstly, while screen-based interfaces display parallel visual information output to users; speech recognition provides only serial auditory information. Secondly, screen-based interfaces use both hard keys and soft keys to control the device while speech recognition interfaces depend on voice recognition control. Lastly, visual information being on screen facilitates short-term memory visually while speech recognition is highly dependent on the user’s short term memory.

There has been substantial research and guidelines published on how the disadvantages of speech recognition may be reduced, with most of them focused on providing visual display information and shortcuts for speech recognition interface. For example, Min Yin and Shumin Zhai [10, 11] presented a series of experiments examining the benefits of augmenting telephone voice menus with coordinated visual displays and keyword search.

The first experiment qualitatively studied a callers’ experience of having a visual menu on screen in sync with the telephone voice menu tree. The second experiment quantitatively measured callers’ performance with and without visual display augmentation. Finally, the third experiment tested keyword search in comparison to visual browsing of telephone voice menu trees. The experiment approved that on average, with visual hints; callers could navigate phone trees 36% faster with 75% fewer errors, and made choices ahead of the voice menu over 60% of the time.

Nicole Yankelovich [18] gave recommendations on how to facilitate users to know what to say while using speech recognition interfaces. These recommendations included suggestions based on visual feedback prompts which validate the importance of visual display for speech recognition interface.

Other than displaying visual information, shortcuts are also useful. . Saverio Perugini, etc. [15] introduced the "out-of-turn" technique which empowers the user (who is unable to respond to the current prompt) to take the conversational initiative by supplying information that is currently unsolicited, but expected later in the dialog. The technique permits the user to circumvent any flows of navigation hardwired into the design and navigate the menus in a manner which reflects their model of the task. Experiments showed that out-of-turn interaction significantly reduced task completion time and improved usability.

Another shortcut speech interface called "Flexible Shortcuts" was introduced by Teppei Nakano [12]. This allowed users to select any command by using “continuous keyword input” which is a voice input method using a series of keywords related to the command. Experimental results show that "Flexible Shortcuts" is superior to the conventional approach from both objective and subjective points of view.

2.2 Using Speech Recognition While Driving

While applying speech recognition in vehicles, researchers need to consider that drivers will be interrupted by the system while driving and the effects that present.

Tijerina et al. [7] state that older drivers perform as well as the younger drivers when responding verbally to an experimental task. Brouwer et al. [2] reports that speech recognition technologies may aid older drivers in their performance of concurrent tasks while driving.

Adriana B. and Paul G. [1] summarized fifteen research papers which studied safety and usability issues of in-vehicle speech recognition systems for driving performance, driver behavior, task performance and subjective workload. They concluded that speech interfaces typically led to better driving performance and often resulted in better task performance with exceptions.

Most experiments used NASA – TLX ratings to determine the subjective workload and showed that speech interfaces led to less workload than manual interfaces [1]. John Lee. etc [6] compared two email systems – simple speech vs. non speech based. The research found that speech-based interaction introduced a significant cognitive load. However, it could be possible that it was the complexity of the email interface that caused the cognitive workload. Also, they did not compare the manual interface to the speech interface to validate if the speech recognition did increase driver's workload.

There has been little research that has considered the effects of the complexity of speech recognition interface and how it could possibly be a significant factor in affecting driver performance and cognitive workload.

2.3 Summary

Although there has been much research done on speech recognition, most of them have focused on driver behavior, task performance and cognitive workload. There are still arguments as to whether speech recognition applications negatively affect drivers and decrease driving safety. There are also usability experiments conducted using different interfaces, e.g. cell phone, email, and navigation. Unfortunately, these did not consider the effect of the complexity of the application interface affecting the users' behavior in testing. There have been few suggestions provided on how to design a user friendly speech recognition interface for drivers in order to improve driving performance and reduce workload. Even though Andrew William Gellatly [3] provided a recommendation table for designing an in-vehicle speech recognition system, his guidelines more focused on speech recognition system design rather than interface design.

3. CASE STUDY

A case study was conducted on an existing speech recognition product developed within the organization. The objective was to improve the voice menu structure so that the driver would not get 'lost' and also provide a better rate of recall.

3.1 The Product

This product – the Wireless Media Gateway Generation 1 (WMG 1) – connects to the driver's mobile phone using Bluetooth. Once connected, it then allows the driver to use speech to make phone calls, search his phone address book, play streaming audio and even have SMS'es read out to him. As a first generation product, it relied purely on voice to inform the driver, there was no visual information, and the driver had to rely on the various speech menu's available to know where he was in the system. There are 3 buttons on the device – Send, End and Bluetooth. (See Fig 1) The Send and End keys are used to receive and hang up phone calls respectively while the Bluetooth key makes a Bluetooth connection between the WMG1 and a mobile phone. It also turns on and off the speech recognition feature.

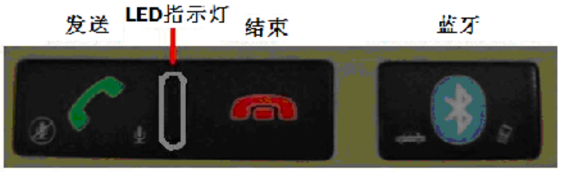

Figure 1. Insert caption to place caption below figure.

There are three different states the system can take: Bluetooth connected, not connected and call in progress. In WMG1, the speech recognition commands are organized in a hierarchical tree structure. For each system state (connected, not connected, call in progress), there is a separate voice command branch (See Table 1). When the driver gives a voice command and the system recognizes it to be valid, it will then feedback via system speech a confirm message. If it does not recognize the command, the system will then provide a short list of available options from which the user can make a selection.

Table 1. Voice command branches for different system states

System state = Bluetooth unconnected	System state = Bluetooth connected but phone is idle	System state = Bluetooth is connected and call is in progress
Setup Menu	Phone Menu	Call Menu
Pair a new phone	Dial number	Mute
Remove device	Contact list	Hand Free
Music	Music	
- USB Music	- USB Music	
- BT Music	- BT Music	
- Play	- Play	
- Pause	- Pause	

In this study, a heuristic evaluation was used to identify usability issues on of the WMG1. The following driver issues were identified:

3.2 Menu options in various states

Menu controls are different for different device statuses. For example, when Bluetooth is not connected, the available menu options allow the user to pair a new phone and remove a device. However upon connection, the options change to dial number, access the contact list. Once call in progress, menu list changes to private mode and enter extension number.

There are two usability issues with this menu structure; (1) The number of different options are too many for a driver to remember. Under phone menu, there are over than 10 direct commands and over 30 sublevel commands user can use. However, option items more than seven are unlikely to remember [9]. (2) Users do now know what are the available commands are in each status, which has been addressed in Nicole Yankelovich's report [18]. She recommended providing visual prompts to facilitate users.

3.3 Tedious and inflexible voice path

In order to give a command, the user has to go through a lengthy and fixed path of the voice menu. For example, to play music from his phone

```
        User:    Music;
        System:  Music, Please give a
command,
        User:    Bluetooth Music;
        System:  Bluetooth Music, Please
give a command;
        User:    Play Music
```

One way to improve this would be to use shortcuts. The system could implement either out-of-turn [15] or "Flexible shortcuts" [12] to reduce the path driver has to go through. The other method is to simplify the menu tree, reduce hierarchy level and remove an unnecessary category like "Music". By associating the keyword 'Play' to music, the user is able to give a direct voice command to the system to start playing music.

3.4 Unnatural voice commands

The current commands are also designed to sound very different from each other. This was done to increase the recognition success rate. However, for commands that are linked functionally, this sounds unnatural to the user and also requires the user to remember extra commands. E.g. For Bluetooth pairing, the logical voice commands would be 'Pair device' vs. 'Remove device', however, to increase differentiation, the system only recognizes 'Pair a new phone' vs. 'Remove device'. This, coupled with fixed commands described above, imposes a high cognitive load on the user.

By allowing the user to trigger the same function using various commands, e.g. 'Phone book', 'Address book' and 'Contacts' launch the contacts list function, the user is not forced to remember only one trigger command. This increases the probability of users being able remember the correct command. Providing a visual list of available options will also reduce the need for command recall from memory.

3.5 Uninterrupted path to complete a command

The system requires the driver to finish the voice commands once a dialogue is initiated. This is difficult because when driving, the changing road conditions may cause interruptions and a command delay. This would result in the dialogue being lost. The system can solve this problem by keeping the previous feedback in memory and having an extended 'no command' window. A command repeat button to allow the user to go back to the previous command will undoubtedly be very useful as well.

3.6 Unsuitable tasks for voice recognition

Basic tasks such as music play/pause are also controlled by voice. Compared to pressing a key on the dashboard, this method is slower and more tedious as well. There are thus situations where voice should not be used as the primary method of control. For example, a user can use the steering wheel control to control next song and previous song. This allows direct access for the driver without taking his hands and eyes off the road ahead. However, when searching for a song from a list, using voice to do so would clearly be the faster method.

Also, when a song is playing, interrupting the audio with a voice command could possibly lead to reduced recognition rate due to increased background noise.

4. GUIDELINES

From the conclusion of the literature review and the findings on the case study, we summarized guidelines to solve the issues above and to facilitate future designing of speech recognition interface for in-vehicle system.

4.1 Use a broad and shallow hierarchy structure

Navigating in speech recognition interface is more difficult than navigating in visual interface as there is no persistent information. A broad menu tree allows users to access available menu options at top hierarchy level without unnecessarily going too deep into any particular category. Shorter menu paths will also reduce task completion times efficiently. The ideal number of levels in a menu should not be more than three for a user to complete a final command.

After the improvement, WMG1 phone menu options reduced from ten to seven. Two sublevels were reduced for each option menu.

4.2 Provide visual feedback and memory aids

Visual feedback is critical for drivers while using speech recognition interface. Visual feedback corresponding to the voice menu will facilitate users in command recall. It will also enable them to know what commands had been given, and if the system is expecting a response from them.

A repeat button that triggers the last system prompt will let a driver return to where the dialogue left off before the driving interruption occurred.

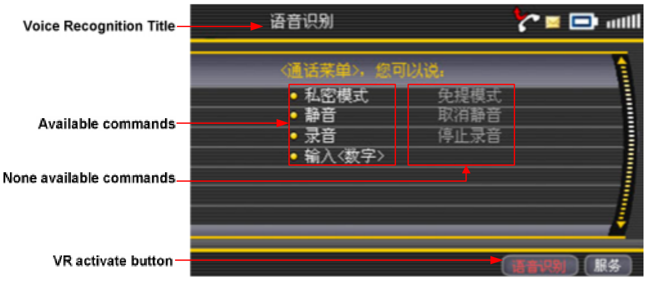

Figure 2. Recommended visual feedback for SR interface

4.3 Provide vocal shortcuts

To provide a quick access to a final speech recognition command, vocal shortcuts are necessary. It is especially useful for frequently used commands or when users forget where they are in the system. In previous literature, two approaches, "Flexible Shortcuts" [12] and "Our-of-turn" [15] solved this problem and statistically proven to effectively and efficiently reduce task completion time.

For example, music play menu path changed to as follow:

```
User:    Play Music;
System:  <starts playing 1st song
from phone>
```

To enhance this even further, other keywords like 'Artist', 'Song' could be associated by the system to give even more direct and immediate searches.

```
User:    Play Artist Avril
Lavigne;
System:  Avril Lavigne;
```

System: <starts playing all songs by Avril Lavigne from phone >

4.4 Implement hard keys for often used tasks

Functions that the user performs in the car should be analysed before porting all their trigger mechanisms to voice. Those that the user needs immediate and quick access to would be best activated by hard keys or steering wheel controls. Also, since speech recognition interfaces use serial auditory input to control devices and provides serial auditory output information, designers should avoid features which input conflicts with output, and use hard keys to facility controls. For example, when music is playing, commands like next song, previous song will conflict with music output. Using hard keys on steering wheel will solve this problem.

Multi-step functions like scrolling through a list, navigating within a menu or inputting text can be made easier by being voice activated.

5. Conclusion and acknowledgement

In this paper, we summarized literature on speech recognition applications and driving performance using speech recognition; evaluated an in-vehicle speech recognition device; and provided qualitative guidelines for designing an in-vehicle speech recognition interface. Due to the limitations of timing and funding, no quantitative data was collected to prove the efficacies of the design recommendations. However, further studies will be done focusing on improving the WMG2 interface based on the above guidelines. Experimental studies will also be conducted.

6. ACKNOWLEDGMENTS

We acknowledge the contributions of Continental Automotive Shanghai; Ming Jing, who gave us the opportunity to work on this project. Wu Ming and Sam Ma, who collaborated on the speech recognition work and provided their expertise

7. REFERENCES

[1] Barón, A. and Green, P. Safety and Usability of Speech Interfaces for In-Vehicle Tasks while Driving: A Brief Literature Review, 2006.

[2] Brouwer, W.H., Waterink, W., Van Wolffelaar, P.C. and Rothengatter, T. Divided attention in experienced young and older drivers: Lane tracking and visual analysis in a dynamic driving simulator. Human Factors, 33 (5). 573 - 582.

[3] Gellatly, A.W. The use of speech recognition technology in automotive applications Industrial and Systems Engineering, Virginia Polytechnic Institute and State University, 1997.

[4] Green, P., Potential Safety Impacts of Automotive Navigation Systems. in Automotive Land Navigation Conference, (1997).

[5] Joanne L. Harbluk and Lalande, S., Perofrming E-Mail Tasks While Driving: The Impact of Speech-Based Tasks on Visual Detection. in 3rd International Driving Symposium on Human Factors in Driver Assessment, (Rockport, Maine, 2005).

[6] John D. Lee, Brent Caven, Steven Haake and Brown, T.L. Speech-based Interaction with In-vehicle Computers: The Effect of Speech-based E-mail on Drivers' Attention to the Roadway. Human Factors, 43. 631 - 640.

[7] Louis Tijerina, S. Johnston, E. Parmer and Winterbottom, M.D. Driver Distraction with Route Guidance Systems National Highway Traffic Safety Administration, Washington, DC, 2000, DOT HS 809-069.

[8] Marvin C. Mccallum, Arvin C. Mccallum, John L. Campbell, Joel B. Richman and Brown, J.L. Speech Recognition and In-Vehicle Telematics Devices: Potential Reductions in Driver Distraction. International Journal of Speech Technology, 4. 25 - 33.

[9] Miller, G.A. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. The Psychological Review, 63. 81-97.

[10] Min Yin and Zhai, s., The benefits of augmenting telephone voice menu navigation with visual browsing and search in CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems (2006).

[11] Min Yin and Zhai, S. Dial and See — Tackling the Voice Menu Navigation Problem with Cross-Device User Experience Integration. UIST '05: Proceedings of the 18th annual ACM symposium on User interface software and technology

[12] Nakano, T., Flexible Shortcuts: Designing a New Speech User Interface for Command Execution. in CHI (Florence, Italy, 2008).

[13] Omer Tsimhoni, Daniel Smith and Green, P. Destination Entry while Driving-- Speech Recognition versus a Touch-Screen Keyboard, Transportation Research Institute, University of Michigan, Ann Arbor, Michigan,, 2001.

[14] Salvucci, D.D. Predicting the effects of in-car interface use on driver performance: an integrated model approach. Human-Computer Studies, 55. 85 - 107.

[15] Saverio Perugini, Taylor J. Anderson and Moroney, W.F., A Study of Out-of-turn Interaction in Menu-based, IVR, Voicemail Systems. in CHI Conference, (San Jose, CA, USA, 2007).

[16] Strass, A.R., Robillard, M., Schedler, S., Peterson, M. and Rabin, R. Speech recognition as a computer graphics input technique (Panel Session) ACM SIGGRAPH Computer Graphics, 16 (3).

[17] Wikipedia. Speech recognition, Wikipedia, 2009.

[18] Yankelovich, N. How do Users Know What to Say? ACM Interactions, 3 (6). 32 - 43.