Commute UX: Voice Enabled In-car Infotainment System

Ivan Tashev, Michael Seltzer, Yun-Cheng Ju, Ye-Yi Wang, Alex Acero

Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA

{ivantash, mseltzer, yuncj, yeyiwang, alexac}@microsoft.com

ABSTRACT

Voice enabled dialog systems are well suited for in-car applications. Driving is an eyes-busy and hands-busy task and the only wideband communication channel left is speech. Such systems are in the midst of a transformation from a cool gadget to an integral part of the modern automobile. In this paper we highlight the major requirements for an in-car dialog system including usability during conditions of cognitive load, efficiency though multimodal user interface, dealing with locations, handling the noise in the car with better sound capture and robust speech recognition. We then present Commute UX, our prototype multi-modal dialog system for in-car infotainment system.

Categories and Subject Descriptors

H.5.1. [**Multimedia Information Systems**]: Audio input/output

H.5.2. [**User Interfaces**]: Natural language, Voice I/O, Interaction styles, Prototyping.

General Terms

Algorithms, Design, Human Factors.

Keywords

Speech interfaces, in-car infotainment, multimodal UI.

1. INTRODUCTION

Driving is an eyes-busy and hands-busy activity with substantial cognitive load. According to a 2001 national travel survey in the United States, 77% of adults commute to work alone using personal vehicles with an average commute time of 25.5 minutes one way [1]. The number of extreme car commuters who travel more than 90 minutes to work, one way, is estimated at 3.4 million, and is the fastest growing commuter group. Drivers use this time to listen to music, or news, and to make phone calls. In general they want to be informed and entertained in order to better utilize the commute time. With widespread use of mobile phones, portable and navigation devices, there is a temptation to operate them while driving. This is a serious safety issue as substantial number of the car accidents in the United States are result of a distraction. Voice enabled dialog systems are one of the most attractive features of the in-car infotainment system, as the speech is the only wideband communication channel not significantly engaged during the driving process.

Voice command systems first appeared in luxury cars as an addition to the integrated hands-free telephone system. They allowed users to call names in the address book of the Bluetooth enabled phone, paired with the system. Later their functionality was extended to cover the control of most of the non-driving related car features such as climate control, radio. Such systems had large number of voice commands (up to 400 in some models!) the driver had to memorize. Such systems have more recently started to move to mass market cars. Currently they are in transition from a expensive add-on feature to an integral part of the modern automobile.

One of the first speech enabled infotainment systems in mass production cars was Blue&Me deployed initially on selected FIAT models [2]. The system is based on Microsoft Auto platform and besides handling the Bluetooth phone allows music selection and radio control. Next followed SYNC [3] as an option on most Ford, Lincoln, Mercury, and Flex models. SYNC is also based on Microsoft Auto platform, but provides richer set of features. Besides initiating and answering phone calls the system reads received text messages and allows users to reply with one of 20 preset messages. SYNC handles all major portable media players, such as iPod, Zune, and even MP3 files stored in USB memory stick. More recent software versions provide car health information, retrieved from the vehicle's CAN network, and can make an emergency phone call in case of accident or air bag deployment.

In parallel were designed and deployed specialized or professional speech enabled dialog systems for in-car information. We should mention here Project 54 from the University of New Hampshire [13] designed for police cars. In this group is Ford Works as well, a version for professional drivers with features related to manage and control a car fleet.

Deploying SYNC in many car models sparked interest and is a serious selling factor. The spoken dialog system is one level above a simple voice commands and works well when the driver speaks the correct commands and queries. The frustration some users had trying to use the system was mostly related to improper operation/commands, stuck in the wrong branch of the menu, not remembering the exact name of the song title or artist. The overall impression is that while the system is technically well designed there is room for improvement in its usability and handling, especially in making the system robust to the most non-predictable errors – those caused by the driver.

In this paper we describe our observations and experience gained while building Commute UX – a prototype of a speech enabled in-car infotainment system. Our interest was mostly concentrated on the spoken dialog aspects of the system, working with the other parts of such system was inevitable. Section 2 describes the over-

all architecture of these systems and in Section 3 we provide our design tenets. Sections 4 and 5 are dedicated to the core technologies and the features of our system, in 6 we provide information about our evaluation methodology. We conclude and provide ideas for future work in Section 7.

2. IN-CAR INFOTAINMENT SYSTEMS

2.1 Architecture

Typical architecture of an on-board dialog system is shown in Figure 1. The input/output devices for communication with the driver are:

- Microphone or microphone array, installed on the dashboard, in the rearview mirror, or in the headliner.

- Display, typically mounted on the dashboard, integrated with the entertainment unit. More frequently the display is a touch screen, which gives ability to select items directly from the screen.

- Buttons, usually two groups:

  o placed around the display in the entertainment unit;

  o a cluster of buttons placed on the steering wheel.

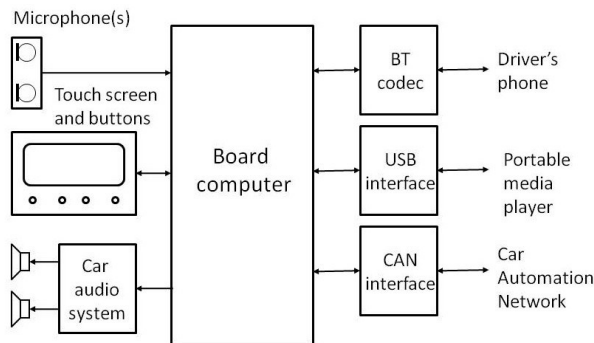- The car audio system (amplifier, loudspeakers).

Figure 1. Block diagram of in-car infotainment system.

Practically all in-car infotainment systems have a Bluetooth interface for pairing with the driver's mobile phone and many systems have a USB interface for connecting to the portable media player.

The latest systems connect to the automobile's CAN network, which allows access to information about the internal status of the car, including the speed, engine temperature, and airbag deployment status.

The central computer of the initial SYNC systems consisted of 200 MHz ARM9 processor and 64 Mbytes of memory, but newer models use more powerful CPU and have more memory. Latest systems already run x86 CPUs and have up to 2 Gbytes of memory.

2.2 Pluggable devices

The in-car infotainment system work with external devices, connected via various interfaces. Historically the first device is the Bluetooth enabled phone. After pairing with the device the on-board computer can retrieve the address book from the phone and create the proper grammar for the speech recognizer. This enables voice dialing by name. During driving the Bluetooth interface is used for detecting incoming or placing outgoing calls, and receiving and sending text messages. Today in rare cases the mobile phone is used for retrieving data from Internet if the phone has a data plan, but is believed that a data connection will be more common and the mobile phone will become the data portal for the in-car computer.

Portable media players usually are connected through the USB port.

Bluetooth as wireless interface is used by some other devices of interest. Bluetooth GPS sensors can be purchased and used in the navigation system as far as they support some of the standard protocols, such as NMEA.

2.3 Expected features and functionality

Implementations of in-car infotainment systems vary from manufacturer to manufacturer, and even from model to model from the same manufacturer. Still there is a set of features and functionality which practically all voice enabled in-car systems have today.

2.3.1 Handling the phone

The historically first function of the in-car system is handling the driver's Bluetooth enabled mobile phone. This is due to the requirement in some states to use a hands free device to talk on the mobile phone and, of course, that it is more convenient and less distracting for the driver to use the mobile phone. The in-car system retrieves the address book from the phone; builds the proper grammars for the speech recognition engine on the fly, which allows voice dialing by name or by number. The incoming conversations are announced usually with the caller ID (by name or number) and it is up to the driver to accept the call.

Once the conversation started the playback audio system of the car plus the microphone is converted into a big, high quality speakerphone.

Most of the systems deal in one or another way with the text messages. Usually the system can detect the incoming message and read it to the driver when accepted. The functionality to reply typically limited to small set of preset messages.

2.3.2 Media players and radio

The in-car system retrieves the metadata from the tracks on the connected portable media player and builds the necessary grammars for the speech recognizer. They typically include recognition by title, artist, album, and genre. Some systems support pre-prepared playlists.

Voice controlled radio is usually limited to selecting one the preset stations, but there are systems which allow tuning by saying the frequency or the station name (if the radio can read the metadata).

Handling the CD or cassette player is limited to trivial voice commands and buttons on the steering wheel.

2.3.3 Navigation system

Even for cars with integrated navigation system in most of the cases it is separated from the in-car infotainment system. This means separate microphone (if the navigation system uses speech recognition) and listening to the directions from the tiny loudspeaker the navigation system has. Operation of the entertainment

system and the phone is completely detached from the navigation system. While this modular structure makes it more convenient for the car manufacturer it increases the complexity of operation for the driver and reduces the safety.

2.3.4 Information

Currently it is almost entirely in the drivers hands. As Internet browsing on the mobile phone is extremely distracting and dangerous, to get information about traffic, weather, stocks, or anything else, the driver can listen to the news on the radio or dial a phone number. Such telephone services, speech recognition enabled, and providing specific for the drivers information, were prototyped by the research community several years ago. The system in [12] provides information about the current traffic, gas prices and station locations, weather, and stock quotes. Good step forward is the recently deployed by Tellme, a Microsoft subsidiary, telephone service for the next version of Ford SYNC models. The system is semi-integrated with the in-car infotainment; the driver can dial the service by voice and continue the dialog with the remote service, which provides traffic, directions, and other information. It doesn't require data plan for the driver's mobile phone as the on board computer can send and receive some data (mostly position information from the integrated GPS sensor) trough the voice channel.

2.3.5 Business information

Business drivers have their specific information needs. They want not only to informed and entertained, but the in-car infotainment system should have specific to their needs features. They are mostly related to handling the calendar and the meetings schedule. Integration of these features with the in-car computer has a lot of advantages: one style voice dialog system, easy entering the destination point to the navigation from the meeting place (retrieved from the calendar), reminders, default dialing the administrative assistant or the meeting room if you are late, automatic entering the corrected schedule, etc. Usually in these cars wideband Internet access is presented, the problems are with integration of several software systems.

2.3.6 Safety information

Currently most of the safety related information exchange is out of the scope of the in-car infotainment system. Inside the car it is going through shielded part of the CAN network – car manufacturers open for the infotainment system just a limited set of messages. Some cars send emergency message to the ONSTAR system in the case of a car accident or air bag deployment. This functionality in the latest SYNC models is replaced by placing an emergency call trough the driver's mobile phone.

2.3.7 Common usability requirements

The in-car systems can vary, but all of them should be designed in a way to put the driver's safety first. All dialog turns and interaction with the on board system should be designed in the least distractive way.

3. Design tenets

Based on SYNC users' interviews, usability videos, and discussions for the in-car infotainment system requirements we established the following design tenets.

3.1 Speech enabled

In most of the cases speech is the least distracting and safest way for communication between the driver and the in-car infotainment system. With eyes on the road and hands on the wheel the driver can hear and speak, the last two senses are not heavily engaged in operating of the car.

3.2 Minimizing the cognitive load

Unlike most of the voice dialog systems for in-car scenarios we can't have 100% of the user's attention. On the contrary, the user is busy with a primary task, driving. The portion of his or her attention the driver can safely devote to the dialog system varies based on the driving conditions. Reducing the distraction and the cognitive load can be achieved by minimizing the number of keyword, making the system simple to use, successfully handling human errors.

3.3 Single layer menu

One of the most common problems with the existing systems is a deeply branched menu system. Users frequently forget in which menu branch they are and which are the valid commands for this branch. In addition going through all of these steps is quite annoying for expert users. The more usable menu structure is a single layer menu, say anything anytime. Each speech query starts with a command word and parameters. The speech recognizer can reliably recognize the minimized number of command words and focus on the following parameters. If some or all of them are missing follows a clarification dialog. This improves the experience of first time users as well.

3.4 Multimodal user interface

Speech queries are a very good way to search large number of items, such as songs, albums, artists, not only in the car. Finding one of the more than ten thousand songs in the portable media player using buttons or touch screen only can be quite painful. On the other hand a simple spoken query can bring the list of four or five most relevant songs. However selecting one of them with speech only interface means listening to the full names of the song titles and saying "one", "two", etc., which can be quite annoying. Glancing at the screen and pointing to the right song is less distractive and way more efficient. In short, multimodal user interfaces are a necessity for designing user friendly human-computer interface in the car.

3.5 Multithreaded dialog system

In a telephone based voice dialog system user enters the dialog with the machine with one specific goal (obtaining the flight schedule, booking a room in the hotel). The system obtains the data (dates, itinerary) and provides the necessary information. In the car many things happen in parallel – in the middle of selecting song the navigation system may want to send message about approaching turn or an incoming phone call can arrive. The user does not have one single task to achieve. This leads to additional design requirements such as handling the interruptions and preserving the dialog context. From usability standpoint are the issues for announcing the interruption, how deep is the human "context stack", and how long after the interruption we can return to the previous context. Most probably the driver will be willing to continue the song selection after a 30 seconds phone call, but

the song selection context should probably be deleted from the stack after a ten minute telephone conversation.

3.6 Situation aware

Numerous user studies showed that a telephone conversation, even using a hands-free system, is more distracting than a conver-

sation with the passenger in the car. The reason for this is that the passenger is aware of the current driving situation, will not talk during lane changing or hard breaking, and will not expect an immediate response during a complex traffic situation. Making the in-car dialog system situation aware from a safety perspective is practically the same. The dialog system should detect the driving situation (lane change, hard braking, heavy traffic, even more than usual nervous driving) using the information from the CAN network, and properly modify its behavior.

3.7 Personalization

Personalization is using prior information about the user, such as point of interest names (Julie's school), favorite songs/albums/artists, most frequently called people (the first four most frequently called numbers are actually 80% of the total mobile phone calls). It is a powerful way to increase the recognition results (by adding weights in the grammar), to increase the usability and user satisfaction [12]. In the car this is much easier to achieve due to the limited number of users. Recognizing the user shouldn't be a problem; this can be done by the phone or by the car key. Many car models already recognize the driver by the key and adjust the rearview mirrors and driver's seat.

3.8 Integration in-car and in-cloud

In general the speech recognizer and the dialog system should be where the data is. This allows building and swapping the grammars for increased recognition results. This means that the song/title recognition should happen in-car, while the dialog system about the weather or traffic should be in the cloud. From driver's standpoint, however, this is the same infotainment system he is talking with, which requires seamless integration of the two parts.

4. Commute UX core technologies

To successfully design and implement in-car infotainment system with voice enabled dialog we have developed several core technologies.

4.1 Sound capture system

While almost every car has sound playback system – the car radio plus the loudspeakers, the sound capture is a relatively new element.

4.1.1 Dual purpose

The sound capture system works together with the car audio system in two main modes: communication and speech recognition. In communication mode the car is just a big speakerphone. From the Bluetooth enabled mobile phone the in-car system is just a headset. In speech recognition mode it is controlled by the driver's push-to-talk button on the steering wheel and the output goes to the speech recognizer.

4.1.2 Microphone(s) and their position

From mechanical and electrical design standpoint the best position of the microphone(s) is in the dashboard, integrated with the rest of the electronics. However in the car is noisy and the microphone should be placed as close as possible to the driver's mouth. The closest place for microphone placement in this case is the headliner, right in front of the driver's head. This position, however, is optimal for the driver's voice only and microphones there will not capture well the voice of the passenger or people on the back seat. Integration of the microphones in the rearview mirror is one of the most common solutions today – the rearview mirror of today's

cars is packed with electronics and a cable form is going there anyway. This solution doesn't increase manufacturing costs and captures the voices of the driver and the passenger well. Table 1 shows the average SNR for a mix of driving conditions for single microphone, placed in these three positions. As expected the headliner is best, followed by the rearview mirror and dashboard, each around 5 dBC apart.

Table 1. Average SNR for microphone positions

Microphone position	Average SNR, dBC
Dashboard	4.16
Rearview mirror	10.11
Headliner	15.41

In some high end cars we can see microphone arrays with two or four microphones, all of the existing models are in the rearview mirrors. The microphone array allows localization of the direction the sound comes from and performing spatial filtering, i.e. suppressing the sounds coming from unwanted directions. The microphone array acts as highly directional microphone, forming a listening beam towards the sound source we want to capture. The listening beam can be electronically steered for capturing the driver and passengers.

4.1.3 Sound processing

The overall diagram of the sound processing chain is shown in Figure 2. We use a two element microphone array placed in the rearview mirror. In communication mode each of the microphone channels is processed by separate acoustic echo cancellers (AEC) to remove the unwanted sounds from the loudspeakers. Then the two channels are processed by the beamformer (BF) [11] and the spatial filter (SF) – a non-linear processor for additional removal of noise [9]. Following the SF is a noise suppressor (NS) [10] that has been optimized for in-car use and automatic gain control
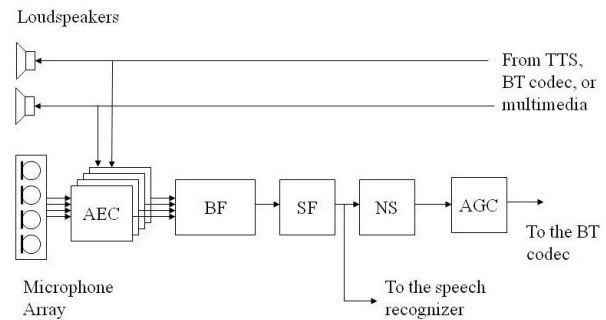

Figure 2. Sound processing chain diagram.

(AGC) block. Then the processed sound is sent to the mobile phone by the Bluetooth codec. Note that even in communication mode the sound processing stack should not only improve the perceptual sound quality for humans, but it should be speech recognition friendly –this is the communication channel with the in-cloud part of the dialog system.

Overall the SNR in the car varies from -10 dB to +30 dB. To have acceptable sound quality we need at least 20 dB SNR. This means that the sound capture system should provide the substantial noise suppression of 30 dB in the worst case. Such aggressive suppression tends to introduce distortions and artifacts in high SNR con-

ditions, i.e. the wide range of SNRs poses additional challenges. Our current system achieves this goal and not only substantially improves the perceptual sound quality, but reduces almost by 50% the word error rate of the speech recognizer.

4.2 Speech recognizer

Training the speech recognizer with in-domain speech corpus contaminated with automotive noise is essential for achieving good speech recognition results. In our case we use noise robust front end as described in [14] which explains why the signal for the speech recognizer is taken before the stationary noise suppressor.

4.3 Location handling

The car is a moving object and in many cases the driver needs to enter addresses, e.g. for driving directions, to get the closest gas station, or to get the traffic information. Understanding and conveying an address is a common task even when the car is equipped with a GPS system. In our studies we found that the short term human memory has difficulties to remember a spoken long address and using intersections is more convenient for the driver for both conveying and remembering the address [12]. In addition we take measures to handle cases when numbers sound similar (thirty and thirteen, for example) by checking parallel paths with the address database, see [7].

4.4 Voice search and information retrieval

The speech recognizer output contains a spoken language query, distorted by the speech recognition and human errors. The next task is to find the most relevant match (or rank the top four matches) based on this query and some prior information. This is the common task for the in-car infotainment system, regardless of the searched data: address book, song/album metadata, other. While each of these tasks has its own specifics, the underlying set of technologies is quite similar and commonly denoted as voice search. It is a spoken language understanding technology for processing user's queries in spoken language and searching for most relevant match.

5. Commute UX features

In our prototyping we focused on the spoken dialog aspects of different applications. This minimalistic design allowed gathering data with seemingly functional prototype, which we exercised for the desired scenarios only. Here we discuss just three of the prototype features: the music player, replaying to SMS by voice, and owner's manual search by voice. The system employs the design tenets from previous section. A GUI screen shot is shown in Figure 3.

5.1 Music player

The existing speech enabled in-car infotainment systems require a structured spoken language query such as "Play track <song name>", where the song name should be exactly as it is in the metadata. A modern portable music player can contain 10-20,000 songs. For humans is not possible to remember the exact name of all songs. Many songs are widely known with different than the "official" name and that second name is usually placed in the metadata in parenthesis. Our user studies showed that in more than 60% we have non-exact spoken query. Humans realize that the sound title is not correct and subconsciously adds more information – the artist or album, for example (not always correct as well). More than one quarter of the spoken queries contain information from multiple fields ("Play track The Yellow Submarine

Figure 3. Commute UX GUI screenshot.

by The Beatles"), which renders clarification track/album/artist meaningless. Not to mention the fact that in some cases it is just wrong – people mistake the song title with the artist or album.

In our system we experimented with several approaches and built a system which is robust to mismatches to the metadata and supports multiple fields. More details can be found in our papers [8] and [5].

5.2 SMS reply by voice

SMS reply by voice systems are limited to replying to urgent messages, that can't wait till the end of the trip. The SMS reply system in Ford SYNC can detect and read incoming text messages. It successfully handles language and abbreviations specific for these messages. The system provides a set of 20 pre-prepared messages to choose from for the reply. Texting is definitely way more distracting than acceptable during driving, but people are tempted to reply when they can't find a proper response. Drivers' need an SMS reply system with richer functionality and non-distracting speech input. The classic dictation technologies with ability to correct the speech recognition errors are too distractive, so we decided to improve the reply with pre-prepared messages. The

1. System: "*Message from Iris. ETA? Say 'Reply', 'Delete', 'Call back' or 'Skip'*"
2. User: "*Reply. Bad traffic in 20 minutes.*"
3. System [earcon]: "*'20 minutes bad traffic' say 'Yes' or a number on the list.*"
4. User: "*Yes*"
5. System: "*Got it. Message sent.*"

Figure 4. A sample SMS reply dialog and screenshot.

main idea is to have large number of SMS replies and to select the most relevant based on the speech input [4]. It may not be the same phrase, but should carry the same informational content. Our

system contains more than 2000 potential replies and in most of the cases finds satisfactory reply in the top four ranked search results. A sample dialog is presented in Figure 4.

5.3 Owner's manual search

The frustration when you have to find something in the owner's manual is quite common, especially with rented cars. Usually this happens in urgent situations, such as flat tire, need to install a child seat, etc. The index at the end of the manual is usually very sparse and only contains a handful of entries. As we already mentioned voice search technology is strong in searching in large amount of data. Considering every page of the owner's manual as a separate document and properly indexing it an TF/IDF based technology allows finding the proper page in the car manual with simple voice query, such as "How do I change a flat tire?". A screenshot with the system response on this query is shown on Figure 5. While this is activity which is not performed during driving and the output is just showing the proper pages and list ranked by relevancy, it is a quite useful feature without any additional expenses – the microphone and the touch screen are already part of the in-car infotainment system.
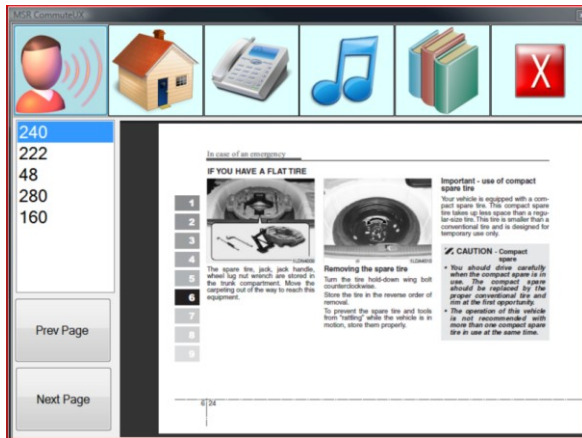

Figure 5. Owner's manual voice search.

6. Evaluation and testing

While each component of the in-car infotainment system can be evaluated with the domain specific measures (SNR for the sound capture system, WER for the speech recognizer, etc.) the evaluation of the end-to-end system goes beyond that. We believe that one of the most important usability parameters is the task completion rate. We define it as the percentage of dialogs when the driver achieved the initial intention. This was the main evaluation criterion for our telephone based information system for drivers. Higher task competition ratio leads to higher perceived accuracy of the dialog system.

Measuring how distracting a design approach may be is more complex and doing it in real vehicles can be dangerous. For evaluation of our system we have a driving simulator, shown in Figure 6. It provides 180° angle of view for the driver and has 7.1 surround sound system for realistic audio picture of the passing cars. The system is fully programmable for the road, driving environment (city, suburbs, mountains, etc.) and the traffic (statistical generation of give intensity of semi-smart and full controlled cars). More importantly it can log more than 40 driving parameters, which allows evaluation of the driver's behavior, correctness

Figure 7. Driving simulator.

of the driving, log accidents and crashes. One of the most useful parameter we found to be the deviation of the distance between the center line of the car and the road lane.

7. Conclusion and future work

The designed prototype of the in-car infotainment system was tested initially as a standalone application and later deployed in the driving simulator. While the initial results allowed us substantially to improve the usability and make it less distractive the main work in this direction is still ahead. We want to find the fine balance between the speech/touch/GUI in the multimodal HMI for the car which is both useful and safe.

Another big area for work and improvement is the situation aware in-car infotainment system. It starts with proper recognition of various driving situations based on the data from CAN, GPS, all other sensors. Then comes clarifying the proper system design under any of these conditions, even mimicking the passenger's behavior.

We just scratched the surface of the seamless integration of the in-car and in-cloud components of the in-car infotainments. This is a big area which is going to develop with propagation of the wide-band 3G and 4G telephones.

8. ACKNOWLEDGMENTS

9. REFERENCES

[1] Bureau of Transportation Statistics. 2001-2002 National Household Travel Survey http://www.bts.gov/programs/national_household_travel_survey/. Commute UX project in Microsoft Research http://research.microsoft.com/en-us/projects/CommuteUX/.

[2] FIAT Blue&Me http://www.microsoft.com/auto/fiat.mspx.

[3] Ford SYNC http://www.syncmyride.com.

[4] Ju, Yun-Cheng; Paek, Tim. 2009. A Voice Search Approach to Replying to SMS Messages in Automobiles. Proceedings of Interspeech, Brighton, UK (submitted).

[5] Ju, Yun-Cheng; Seltzer, Michael; Tashev, Ivan. 2009. Improving Perceived Accuracy for In-Car Media Search. Proceedings of Interspeech, Brighton, UK (submitted).

[6] Popeik, D.; Gray, G. 2003. High-tech distraction? Cell phones and driving: how risky? Consumer Research, v. 86, pp. 14-16.

[7] Seltzer, Michael; Ju, Yun-Cheng; Tashev, Ivan; Acero, Alex. 2007. Robust Location Understanding in Spoken Dialog Systems Using Intersections, in Proceedings of Interspeech 2007, Antwerp, Belgium.

[8] Song, Young-In; Wang, Ye-Yi; Ju, Yun-Cheng; Seltzer, Mike; Tashev, Ivan; Acero, Alex, 2009. Voice Search of Structured Media Data, in proceedings of International Conference on Acoustics, Speech and Signal Processing, Taipei, Taiwan.

[9] Tashev, Ivan; Acero, Alex. 2006. Microphone Array Post-Processor Using Instantaneous Direction of Arrival, in Proceedings of International Workshop on Acoustic, Echo and Noise Control IWAENC 2006, Paris, France.

[10] Tashev, Ivan; Lovitt, Andrew; Acero, Alex. 2009. Unified Framework for Single Channel Speech Enhancement. Proceedings of IEEE Pacific Rim Conference on Computers, Communications, and Signal Processing. Victoria, Canada (submitted).

[11] Tashev, Ivan; Malvar, Henrique. 2005. A new beamformer design algorithm for microphone arrays, in Proceedings of International Conference of Acoustic, Speech and Signal Processing ICASSP, Philadelphia, USA.

[12] Tashev, Ivan; Seltzer, Michael; Ju, Yun-Cheng; Yu, Dong; Acero, Acero. 2007. Commute UX: Telephone Dialog System for Location-based Services, in Proceedings of SIGdial Workshop on Disclosure and Dialogue 2007, Antwerp, Belgium.

[13] Turner, L.; Kun, A. 2005. Evaluating the Project54 speech user interface. In proceedings of Third International Conference on Pervasive Computing, Munich, Germany.

[14] Yu, Dong; Deng, Li; Droppo, Jasha; Wu, Jian; Gong, Yifan; Acero, Alex. 2008. Robust speech recognition using cepstral minimum-mean-square-error noise suppressor, IEEE Trans. Audio, Speech, and Language Processing, vol. 16, no. 5.