TOWARDS ADVANCED SPEECH DRIVEN NAVIGATION SYSTEMS FOR CARS

Stefan W. Hamerich

University of Hamburg
Department of Computer Science
Natural Language Systems Division
Hamburg – Germany

Harman/Becker Automotive Systems GmbH
Speech Dialog Systems
Dialog Research and Tools
Ulm – Germany

shamerich@harmanbecker.com
Fax: +49/731/3994250

Keywords: Speech-control, speech system for cars, navigation system, proactive dialog, TMC

Abstract

In this paper ideas of the next generation of speech-controlled navigation systems are presented allowing proactive dialogs. These proactive dialogs are used in the context of TMC messages informing about real-time traffic situations. Additionally the combination of a calendar with the navigation device is described leading to early information about delayed arrivals at meetings.

1 Introduction

Speech-controlled systems for cars have been available for more than ten years now. The first device in a car controlled by speech was the telephone, which originally was built-in. Nowadays all infotainment devices in a car can be speech-controlled. Even destination input by voice is possible, although the size of the vocabulary makes embedded speech recognition hard.

As stated by Cameron in [2] people are using speech systems more likely when at least one of the following criteria is met:

- They are offered no choice,
- It corresponds to the privacy of their surroundings,
- Their hands or eyes are busy on another task,
- It is quicker than any alternative.

Following on these points speech systems for cars meet three of four criteria. When driving in a car, the privacy of the surroundings is met, since other persons in a car are usually known to the driver or in most cases the driver is the only person in a car. While driving, the eyes should be on the road ahead and the hands on the wheel. And in a typical driving situation it is much faster (and safer) to just speak a command than push several control buttons.

Therefore speech systems for cars have been a success since the first product system was presented in 1996 in the Mercedes S-Class [6]. This system already allowed the speaker independent control of the built-in phone with a vocabulary of 30 words with continuous recognition of e.g. digits.

In this paper the next generation of speech-controlled navigation systems is presented. First current speech control systems for the automotive market are described including their technology and functional range. Next ideas for advanced speech-controlled navigation systems are presented. The possibilities of this approach will allow for an advanced speech interface, which is discussed in more detail when illustrating the implemented prototype. Next the planned evaluation phase is mentioned. Finally the major findings are concluded and future work is presented.

2 Speech Control Systems for Cars

After Mercedes-Benz being the first car manufacturer offering speech control in cars, other manufacturers followed soon making such systems nowadays an important feature for middle and upper class cars. For luxury cars as the Maybach speech has even emerged as a standard component.

2.1 Technology

Speech control systems in cars are embedded solutions, provided either as a separate hardware box (as shown on Figure 1) or integrated into infotainment systems running under a variety of real-time operating systems.

Figure 1: Speech control unit for Audi A8

Due to different climatic conditions which cars are exposed to, to minimize the mean time between failures, and since identical electronic spare parts must be available for a certain period of a cars lifetime, electronic components in cars
are very different from consumer products. Therefore expensive flash memory is used for speech control systems leading to limited resources. This means for current systems they must manage with only 128 KB to 8 MB as totally available memory.

To meet the requirements of the automotive market special algorithms and tools are used to model speech dialogs. One important fact is the impossibility of using standard dialog modeling approaches as e.g. VoiceXML, alternatively special dialog descriptions for embedded devices are used, as described in [5]. The peculiarity of this approach is the connection of external devices, such as a tuner, phone, or navigation system to the speech dialog itself. Another advantage is the memory consumption, which has been minimized by compiling the complete dialog description. Finally a special speech recognizer is used, which was trained on speech-data collected in cars. For the data collection refer e.g. to [7].

2.2 Integration

Generally the speech control system is directly attached to the car entertainment bus, which connects all entertainment devices. That means the dialog system itself is sending the control commands to the devices, using direct communication between the speech system and controlled devices. This of course explains the need of a complete integration into the cars entertainment infrastructure. Additionally the speech control is using the loudspeakers of the car for system output and needs a push-to-talk (PTT) button or lever to start the system. The latter is needed since due to the surround noise an open microphone technology cannot be used. Another advantage of the PTT is the full check on the system it is giving to the driver, since he activates the system pushing PTT. As described, the integration of such systems into cars is complex and costly, see [4] for more details.

2.3 Functional Range

In the last ten years the functionality of speech control systems for cars grew. Today systems allow controlling the phone (connected to the speech system by a cradle or by Bluetooth), climate-control, tuner, CD-Player, CD-Changer, DVD-Player, MP3-Player, TV, and the navigation system.

As described before, a PTT button or lever is used to activate the speech recognizer and start the speech dialog. Nevertheless system initiated (i.e. proactive) dialogs are possible. E.g. in the Becker Traffic Pro and Cascade systems an incoming telephone call results in a proactive speech dialog, automatically asking the driver if he likes to accept the call.

The most current feature available on the market allows spoken destination input for navigation systems without the need for entering the letters of city and street names haptically. Due to the huge number of alternatives in the vocabulary and the limited resources available for speech recognition in a car, the first approach of implementing speech-controlled destination input was by spelling the first letters of a city name. This feature was first presented in 2002 by Temic SDS (now with Harman/Becker Automotive Systems) and implemented in Audi A8 and Mercedes E-Class cars. With hardware getting cheaper and more powerful today full-word destination input is available on the market e.g. in the Honda Acura RL and the new Mercedes C-Class. The major benefit of speech-controlled navigation input is its safety, since no buttons or haptic spellers have to be used any more. This allows entering a destination comfortably even while driving. To give a better impression on current systems an example dialog for a spoken destination input is shown below:

<table>
<thead>
<tr>
<th>User:</th>
<th>System:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enter destination</td>
<td>First you have to enter a state.</td>
</tr>
<tr>
<td>Your state please.</td>
<td></td>
</tr>
<tr>
<td>California</td>
<td></td>
</tr>
<tr>
<td>California has been selected.</td>
<td></td>
</tr>
<tr>
<td>The city name please.</td>
<td></td>
</tr>
<tr>
<td>San Francisco</td>
<td></td>
</tr>
<tr>
<td>San Francisco has been selected.</td>
<td></td>
</tr>
<tr>
<td>Do you want to enter a street?</td>
<td></td>
</tr>
</tbody>
</table>

Future systems will allow entering a complete address in one utterance, making destination input much more natural and easy. That applies in particular for addresses in Japan and Korea, which are composed of up to seven layers.

The next feature to come and currently under development is the direct selection of music from CDs or attached MP3 players by speaking the names of artists, albums or titles [11]. This approach will allow a direct and quick access to music by just saying e.g. "play album 'Bad' from Michael Jackson".

3 Advanced Navigation

To extend the possibilities of current speech-controlled navigation systems, the following ideas are currently tested.

3.1 Dynamic Speech Output

Within speech advisories for route guidance current systems in cars could not prompt dynamic vocabulary, since only pre-recorded prompts are used. These have been recorded by professional speakers in sound studios having an excellent speech quality. For dynamic output pre-recorded prompts fail, therefore speech synthesis is to be used.

Due to the mentioned memory requirements only very small speech synthesizers are to be embedded. These are developed especially for the automotive market providing a fair sound quality. However, future systems will offer more memory and computing power surely leading to improved text-to-speech (TTS) quality.
Having the possibilities of a speech synthesizer this allows for new functionalities. Currently e.g. the new Mercedes C-Class has a TTS integrated in its Linguatronic system allowing read out of address book entries, TMC messages (see section 3.2 for details on TMC) and incoming SMS messages. Additionally the availability of a TTS will also allow for enhanced navigation prompts in the upcoming next product generation. Instead of saying "keep right in 200 meters" future systems will prompt the driver with e.g. "keep right in 200 meters into fifth Avenue".\(^1\) This will give more information to the driver allowing immediate verification of navigation systems route calculation capabilities. This could strengthen user’s confidence in the navigation system itself.

### 3.2 Proactive Speech Interface to Dynamic Route Guidance

In most European countries and the United States TMC (Traffic Message Channel) \(^{10}\) is available. This is an application of the FM Radio Data System (RDS), which is used for broadcasting real-time traffic information. The data is received silently by the TMC decoder of the tuner and then sent to the navigation system. There the information is taken into account for the routing. If e.g. a motorway is closed due to an accident, a detour is calculated automatically. TMC is currently supported by many navigation systems. It allows for dynamic route guidance taking into account real-time traffic situations and is making phone-based traffic information systems (as e.g. [3]) superfluous. Currently a navigation system, when receiving a TMC message automatically calculates a new route if the current route is affected. The driver is only informed by a short prompt saying the current route is changed because of a new traffic situation. Advanced systems additionally read out the contents of the message.

Since future systems will offer more memory and computing power, the possibilities for speech interfaces will grow as well. One of the ideas for future systems is to inform the driver about the reason of the change and ask him for the actions to take. This makes sense, since unfortunately TMC messages can be old, incorrect or vague. A study has found out, that especially when driving fast (e.g. on a motorway) it could take between five and fifteen minutes until a message is received by the navigation device [1]. Since e.g. traffic jams can disappear quite fast, this can make calculation of a detour needless.

If on the other hand a valid message has been received reporting a current traffic situation several detour routes could exist. E.g. one of the routes is known to the driver, nice to drive, etc.

So why not ask the driver, if he wants to stay on his route or if a detour should be calculated? Then the system can ask, if the fastest way should be selected or better a scenic road. The system could even recommend important sights and inform the driver about them.

This scenario will then lead to proactive dialogs, which are not any more initiated by the driver; instead the initiative comes from the system itself. However, these proactive dialogs of course can be aborted by the driver.

### 3.3 Combination with Calendar

Today mobile phones often are equipped with calendar functionality and PDA devices offer telephony functionality. Both kinds of devices can easily be connected to the hands-free system of speech interfaces in cars via Bluetooth. Therefore it would be possible, to read out the calendar data and provide the driver additional functionalities. E.g. the next appointment can be read out by the system informing the driver on his way about participants etc. In the same manner, new meetings could be added just using speech. Now when combining the knowledge of the calendar and the telephone with the navigation system, several possibilities appear, which are described below:

- **Automatic destination adaption:** when entering the car and connecting the mobile phone to the infotainment system, it can use the knowledge of the phone’s directory, to get to know the place of the next meeting and transfer it automatically to the navigation system to set it as new destination.
- **Early information about late arrival at an appointment due to traffic conditions:** when driving the navigation system checks the internal calendar against the current TMC messages. Then calculates the time to destination and informs the driver, in case of delay. The advantage is the early information allowing to inform others and to save time. E.g. when coming later to an appointment following meetings can be postponed, if the driver tells the system to do so.
- **Notification of other meeting participants in case of a delay:** since the mobile device knows about appointments and other participants, the latter can be called upon request and the driver can inform about the situation or depending on the delay make a new appointment. In that case the infotainment system can assist the driver in telling which time slots in his calendar are free.
- **Providing information on nearby contacts when driving:** when e.g. a detour has been selected the navigation can inform the driver about contacts from the directory, which are living/working nearby. The system can even call the person to see, if he or she is there.

### 4 Prototypical Implementation

A prototype of such a speech-control system has been implemented running on a PC environment. The additional navigation functionality is simulated. The implementation of the prototype has been done using the dialog modeling approach, tools, and speech recognition of Harman/Becker as described in section 2.1. Though this approach originally has been developed to be used with simple command

---

\(^1\) The information is stored within the navigation database and could be used to precise route guidance prompts. This feature is already available in some mobile navigation devices.
& control dialogs, these more complex dialogs could be implemented without any need to extend the existing dialog description language or dialog manager.

Below an example dialog of a proactive dialog is shown, where the system received a TMC message and informs the driver about this. As described in section 3.2 the system asks about detour calculation and finally offers a road to sights or the fastest way.

System: New traffic information has been received.
User: Do you want to hear the message?
System: 'Motorway A7 Memmingen to Würzburg is closed due to an accident between exit Hittistetten and exit Nersingen. It is recommended to leave the motorway.' Shall a detour be calculated?
User: Yes please
System: Do you want the fastest way or a scenic road?
User: Scenic road
System: Ok, scenic road. Do you want information on the sights nearby?
User: Yes
System: You are next to Ulm, a nice city with the tallest church spire in the world. Do you want to go there?
User: Yes
System: Ok, calculating a route to Ulm.

When on the way to a scheduled appointment the system can inform the driver about a possible delay, showing the possible interaction of several applications within the speech-control system as described in section 3.3:

System: The current traffic situation will not allow arriving at your next meeting in time. The delay will be approximately 20 minutes. Do you want to call Mr. Smith?
User: Yes
System: Calling Mr. Smith
[telephone number is dialed]

5 Evaluation

The described prototype will be evaluated to find out, if the described features can be easily used while driving. Especially the usability of the dialog is to be evaluated with focus on its proactivity. The evaluation will be done using driving simulator software as the primary tool with additional tasks to solve with the speech system. The simulator software used is a standard tool and described in [8]. After the first evaluation round, the system will be updated based on the test results and a new evaluation round be started. This user-centered development for in-car speech-systems is described in more detail in [9].

6 Conclusion

In this paper current speech control systems for cars have been described, including their technology and functional range. Next new ideas for advanced navigation systems have been illustrated allowing complex and proactive dialogs to ensure safer driving and inform about sights or contacts nearby. The implemented prototype has been presented with example dialogs. Finally the planned evaluation has been covered.

7 Future Work

As already mentioned the prototype will be evaluated, to find out about its usability using driving simulator software. After these tests the prototype will be revised and then be tested in a real car. It will take some more iterations and capable hardware to make this functionality available to end customers in some years.

References