How do I get there from here? Exploring speech in CARS

Christine A. Halverson, Luc Julia and Jehan Bing

SRI International
Computer Human Interaction Center—CHIC!
333 Ravenswood Av.
Menlo Park, CA 94025
Christine.Halverson, Luc.Julia, Jehan.Bing@ sri.com

ABSTRACT

This paper describes the results of a preliminary study of an in-car navigation aid that uses voice and gesture to allow the user to interact in a natural way to get directions from web-based services.

Keywords

Multimodal interaction, speech recognition, speech user interfaces, auditory I/O, mobile computing.

INTRODUCTION

With the explosion of the web, cellular phones, and our increasing mobility it seems inevitable that we want to accomplish more when we spend time in our cars. Our previous work with speech recognition in autos—CARS: cooperative agents and recognition systems—focused on controlling onboard systems like the CD player and getting access to personal information like email. This system used both audio input and output as well as visually displayed information on a screen [1]. Here, we extend that idea to getting access to information that you want while driving, but are not likely to have personally, such as directions to a new location.

Getting and giving directions in a moving car poses special problems. Audio input and output is the obvious choice but integrating that information with the real world in a way that doesn't impact the main task—driving—provides special challenges.

In this paper we present a synopsis of the various components necessary, how they are brought together, and present the preliminary study we performed to explore the system's usability.

User Interface

For a user to be able to get directions without looking down we needed to provide a way to continue to look out into the world and get information interleaved with current visual information. By using a heads up display (HUD) we were able to provide labels positioned in the appropriate direction based on the orientation of the car. Labels appear in response to the user's query and are selected using an ultrasonic pointer that registers when the user is pointing at the label. Once selected, the label changes color. Now the user can ask for more specific directions relating to the label—like directions—and the system will respond verbally.

Technology

To drive this interaction requires a laundry list of equipment. The setup consists of two computers, the HUD, the pointer, a microphone, speakers, a GPS receiver, a wireless modem, and a cell phone modem. A standard UPS (uninterruptible power supply) provided power for all equipment for around 30 minutes.

A critical part of the scenario presented is the integration of several technologies necessary to support the different modalities of interaction, as well as the access to different pieces of information. The integration is possible because of the Open Agent Architecture™ (OAA) developed at SRI [2,3]. OAA is a distributed infrastructure that provides the means for bringing together multiple component technologies in a flexible, plug-and-play manner. Components can be written in different programming languages[1] and be distributed over multiple computers.

Navigation, audio input and output, and GPS tracking is handled on a sub-notebook 300Mhz Pentium II with 128MB of RAM. This machine displays and runs the CARS application. Other agents include speech output provided by TTS from IBM (ViaVoice Outloud) and speech input using the Nuance speech recognizer. Navigation is provided through a Garmin GPS III receiver connected to the serial port. We use Compaq's WebL agent to extract information from hmtl pages, and GeoDatabase

[1] Most of the agents (CARS, WebL, GeoDatabase) are written in 100% Java (JDK 1.2.2). We used JNI to interface the agent layer (Java) with the Nuance recognizer, written in C. The VrmlAgent uses Java3D and the Vrml97 add-on. The GPS agent uses the Serial Port (COMM) extension from Sun. The TTS agent is written in C

which is a very simple database to keep some piece of information locally for a faster access (grocery stores).

The second machine, is a Dell 450Mhz Pentium II with 128MB of RAM. It runs the VrmlAgent which display the labels on the HUD. The pointing device, Mimio from Virtual Ink, is plugged in one of the serial ports. The two machines communicate via a very simple network consisting of them and a hub. The sub-notebook uses a Breezecom wireless connection.

For the Internet connection, we used a cell phone from Nokia (model 5190) connected into the second serial port of the Dell machine. This connection is actually needed by the sub-notebook, but it can not have both the wireless and cell-phone modems at the same time. On the Dell SyGate software provides a simple DHCP server, which allows the sub-notebook to be configured automatically to use the DELL as a gateway to the Net.

METHOD

SubjectsFor this preliminary study we recruited four volunteer participants. None of them were directly connected with the project, but were recruited based on their willingness to drive a car of the future. All participants were male between the ages of 30 and 35. Participants could not have had any extended prior experience with speech recognition systems (one has limited experience) and all claimed unfamiliarity with the Stanford University campus. All four participants are computer professionals, but none work with SRI or with the kind of technology used.

Procedure

We were primarily interested in two issues: 1) could a complete novice use this system, and 2) what would they find easy or difficult about using it. Participants met us at SRI and were driven to the Stanford Oval. Once there, they were given a scenario to frame their session experience while final adjustments were being made to the hardware and software. They were then set up in the driver's seat and fitted with a pair of IO Display Systems heads up display glasses, with a microphone adjusted around their neck.. The pointing device, an ultrasonic pen from the Mimeo system electronic white board, was positioned near to hand. Once settled they were asked to back out of the parking space and begin the session. They could drive anywhere on campus they wanted and the monitor would intervene only if they were about to drive off campus.

Once they were driving the scenario started. Users received an audible message that they needed to pick up orange juice on the way home. To make it immediate they were told that the store near home was closed. Subjects then need to talk to the system to find out where they could find a grocery store and get directions to it. The session stopped once they had received directions to the store.

Measures

During the exercise, participants were video and audio taped, their speech was recorded on the computer for future use with the recognition system, and their behavior was observed and noted. After the study, spontaneous comments were also noted. Participants also took a questionnaire about the experience of using the system. One of the participants never saw labels or received directions because of set up problems. All of the other participants received directions, but none were able to successfully act on them.

RESULTS AND CONCLUSIONS

This study provided insight into some basic usability challenges as well as unforeseen conceptual challenges. All participants commented on the appropriateness of voice interaction for an automotive system. They found the modality was useful and did not interfere with driving. However, they were frustrated by delays in system response, which they assumed was caused by speech recognition. In reality, web access time caused the delay; the system provided no feedback that their request was understood and being acted upon.

Most usability feedback was related to the visual display of the labels and the process of interacting with them. Subjects were evenly split on the comfort of the glasses, but complained that it was difficult in some circumstances to see the labels. All three subjects who saw the labels were able to select the label they wanted and get directions to the grocery store. However, this process of selection revealed a usability problem based on the process of relating the label to the real world. The heads-up display only presented information about destinations in the direction the car was pointing. Subjects often wanted information about destinations in a different direction if they were about to make a turn or were forced to turn away from a given destination. Some tried turning their heads, assuming that would identify destinations in the direction they were looking, but were frustrated when they lost a reference to a destination..

Results from this study are being used to inform further design of the system and we expect to be testing more users shortly. This study has provided us with some concrete examples of the workability of our basic premise and is helping us to further our design and expectations of its eventual use.

ACKNOWLEDGEMENTS

Thanks to Robert Bowdidge & Adam Cheyer for their help, and our subjects for driving a car covered with equipment.

REFERENCES

1. Julia, L. and Cheyer, A. (1998) Cooperative Agents and Recognition Systems (CARS) for drivers and passengers, *in Proceedings of OZCHI'98,* (Adelaide, South Australia, 29 November - 3 December 1998), IEEE Press.
2. Web site on the Open Agent Architecture: http://www.ai.sri.com/~oaa/applications.html
3. Martin, D, Cheyer, A. and Moran, D. .(1999) The Open Agent Architecture: A framework for building distributed software systems. Applied Artificial Intelligence: An International Journal: 13(1-2), pp 91-128.