Multimodal Conversational Systems for Automobiles

Currently available in-vehicle speech recognition systems are designed around a single-utterance-command paradigm [2], with as many as 200 commands[1] that must be learned or referenced in a manual—an unpractical option while driving. The combination of a flexible dialogue-based speech system with a visual and haptic touch screen, while still an area of active research [1, 3], provides the opportunity for an intuitive and effective multimodal interface for vehicles.

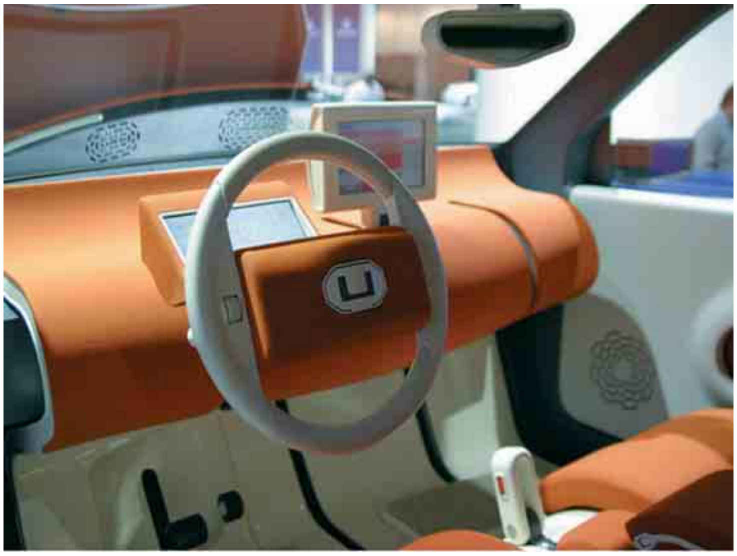
SpeechWorks[2] and Ford designed and realized a prototype targeted at relaxing the limitations of the current systems by adopting a conversational speech interface coupled with a touch-screen display. The system, which controls vehicle functions such as climate, telephone, navigation, MP3 player, and personalization, was installed in the Ford Model U concept vehicle and first shown at the 2003 North American International Auto Show in Detroit. Figure 1 shows an image of the actual car interior where the touch-screen interface is visible on the dashboard. Figure 2 shows a depiction of the GUI in one of its various configurations.

The multimodality of the system allows users to adapt to their environment, for example, interacting through the GUI when the car is stopped at a light versus when the car is moving. In addition, we designed the two modalities to complement

[1] Jaguar Cars Ltd. Publication Part No. JJM 18 09 24/22, Coventry, UK, (Oct. 2001).

[2] SpeechWorks is now part of Scansoft.

Figure 1. The interior and haptic interface on the dashboard of the Ford Model U concept car.



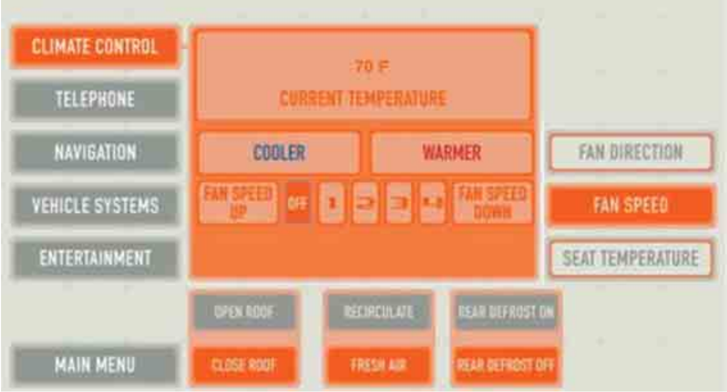


Figure 2. Example of GUI controls used in the Ford concept car multimodal application.

each other; the graphic interface controls providing hints of the corresponding voice commands. New users interact by speaking commands shown on the display, while the system engages in a directed dialogue and prompts for missing information. Experienced users can then adopt more effective commands and decide, at each turn, whether to interact by using speech or touch controls. The following is an example of the interaction:

User speaks: Climate Control.
System speaks: Climate Control. Warmer or cooler?
(The system displays changes to climate control display showing a list of options, including seat temperature, fan speed/direction, and so on).
User: Seat temperature.
System: Seat temperature. Please say driver, passenger, or both.
User: Driver.
System: Warmer or cooler?
User touches the "Warmer" button.
System: Increasing the driver seat temperature by two degrees.

This long interaction helps the driver learning the options and the words for completing the task. Experienced users learn to achieve the same goals with single sentences, such as "climate control driver's seat temperature down," or "turn the driver's seat temperature to 60°." Drivers are allowed to issue any command at any point, via speech or touch screen, for instance placing a telephone call while engaged in a navigation dialogue.

The speech recognition engine (SpeechWorks' Speech2Go) makes use of dynamic semantic models that keep track of the current and past contextual information and dynamically modify the language model in order to increase the accuracy of the speech recognizer. A conditional confirmation strategy, also based on contextual information, is used for improving the dialogue flow and reducing the time to completion.

The interaction is controlled by a dialogue manager that responds to input signals with output actions. In automobile applications, the input signals may come from the user as well as from the vehicle, and the actions may be directed to the user or the vehicle. All input signals can cause a change in the course of the interaction. For example, a low-fuel condition signal can cause the navigation system to engage in a dialogue for rerouting the driver to the nearest gas station. We developed a general multimodal dialogue-manager architecture that allows for a complete separation between the interaction logic and

THE *goal of our multimodal interface is to provide an intuitive and flexible means for controlling vehicle systems while providing a user with the option to operate the system with speech, touch, or any combination of the two.*

the input signals. The interaction logic is independent from the source of the signals (speech, GUI, vehicle signals) and is represented by a recursive transition network as described in [4, 5].

The goal of our multimodal interface is to provide an intuitive and flexible means for controlling vehicle systems while providing a user with the option to operate the system with speech, touch, or any combination of the two. Providing this flexibility, while maintaining UI capabilities, required careful design at the UI and architectural levels. The prototype described here is one of the first attempts to move this challenge from the research community to commercialization.

References

1. Bernsen, N.O., Dybkjaer, L. A multimodal virtual co-driver's problem with the driver. In *Proceedings of IDS02.* (Kloster Irsee, Germany, June 2002).
2. Heisterkamp, P. Linguatronic—Product level speech system for Mercedes-Benz Cars. In *Proceedings of HLT 2001.* Kaufmann, San Francisco, CA, 2001.
3. Minker, W., Haiber, U., Heisterkamp. Intelligent dialog strategy for accessing infotainment applications in mobile environments. In *Proceedings of IDS02.* (Kloster Irsee, Germany, June 2002).
4. Pieraccini, R., Carpenter, B., Woudenberg, E., Caskey, S., Springer, S., Bloom, J., Phillips, M. Multi-modal spoken dialog with wireless devices. In *Proceedings of ISCA Tutorial and Research Workshop—Multi-modal Dialog in Mobile Environments.* (Kloster Irsee, Germany, June 2002)
5. Pieraccini, R., Caskey, S., Dayanidhi, K., Carpenter, B., Phillips, M. ETUDE, a recursive dialog manager with embedded user interface patterns. In *Proceedings of ASRU01-IEEE Workshop* (Italy, Dec. 2001).
6. Pieraccini, R., Dayanidhi K., Bloom, J., Dahan, J.G., Phillips, M., Goodman B.R., Prasad, K.V. A multimodal conversational interface for a concept vehicle. *Eurospeech 2003* (Geneva, Switzerland, Sept. 2003).

Roberto Pieraccini (rpieracc@us.ibm.com) of IBM T.J. Watson Research Center; Krishna Dayanidhi (kdayanid@speechworks.com); Jonathan Bloom (jbloom@speechworks.com); Jean-Gui Dahan (dahan@speechworks.com); and Michael Phillips (mike@speechworks.com) of SpeechWorks International, New York, NY. Bryan R. Goodman (bgoodma2@ford.com) and K. Venkatesh Prasad (kprasad@ford.com) of Ford Motor Co., Dearborn, MI.