The mobile computing revolution has fundamentally altered how we interact with technology. As our electronic devices have increased in capability, so too has the complexity of their interfaces, presenting a forest of icons, menus, hierarchies and sub-hierarchies that even the most elegant multi-touch interfaces may struggle to tame. Device manufacturers are turning increasingly to voice interfaces to cut through the forest. Why bother to remember the exact app, screen and icon that will allow you to add a new appointment to your calendar, when you can just say “Schedule a haircut for next Tuesday at 4:00pm”? Google’s Voice initiative, Apple’s Siri, Microsoft’s Cortana, and other OEM solutions, all herald the importance – or perhaps even primacy – of voice interactions in the future.

Industry acceptance
Automotive manufacturers have been quick to develop in-vehicle systems that offer the driver access to a multitude of functions on the go, including live weather information, radio control, playlist management, navigation systems, integrated calling and emergency assistance. Some features require the input of strings of information, which can be challenging and possibly unsafe to do using traditional visual/manual means while driving. These driver-vehicle interactions have the most to gain from a well-designed voice interface. More than simply being a matter of convenience, voice driver-vehicle interfaces have been promoted as a way...
Volvo will offer Apple’s CarPlay in the new XC90. Drivers will be able to use voice and steering wheel controls to access Apple features and services.

VOICE INTERFACES

Rosie the Robot is alive and well

The idea of working with a synthetic assistant that makes the power and resources of a machine available via the intuitive ease of a conversation has been in the public imagination for decades — from The Jetsons’ Rosie the Robot through Knight Rider’s KITT to Iron Man’s JARVIS. We can’t seem to resist the urge to personify technology, to assign an emotion to an error message, or to treat a screen as if it were a face.

Throughout the studies described in this article, the researchers noticed a curious phenomenon: when issuing voice commands to the system, drivers would frequently turn toward the in-vehicle screen, addressing it as if the system were located inside the console (in fact, the system’s microphones were dispersed in the cabin and could detect the driver’s voice regardless of where he or she was looking). This behavior — called an orienting response — often took the form of subtle, seemingly unconscious shifts in posture as the driver spoke. The effect was more pronounced when drivers became frustrated with the system, as when repeated missteps caused them to restart a task multiple times. Older drivers, particularly women, were more likely to orient themselves to the system as they spoke. Younger drivers, who are perhaps more accustomed to mobile technology and voice interfaces (or possibly less socialized to in-person face-to-face communication), were much less likely to make orienting responses.

This personification of technology, as studied by the late Clifford Nass and others, is another example of the multimodal demands placed on drivers by voice interfaces. An engineer designing a voice interface must therefore consider not just the optimal series of commands needed to perform a task, but the desires, assumptions, and misperceptions that the average driver is likely to bring to the experience.

Unexpected results

Before the study began, the researchers expected that the voice system would draw a driver’s eyes away from the road less frequently than visual-manual controls would. The point of the research was to gain better
understanding as to whether, and to what extent, engagement with a voice-command system might introduce cognitive demands that affect drivers, possibly causing them to slow the vehicle, control it less smoothly, and elevate stress indicators such as heart rate. Instead, the results (obtained from 117 subjects) showed that they rated the voice tasks as being about as demanding as those that used buttons and knobs. These ratings were backed up by objective measures of vehicle control and physiological measures of arousal, which showed roughly equivalent changes regardless of which interface was being used. The data also showed that while glance time decreased with voice radio tuning compared with the traditional alternative, activities such as full alphanumeric destination entry drew visual attention away from the road for sizeable periods of time.

In essence, while a voice interface can offer advantages in some situations, interactions can still be associated with relatively extended periods of visual demand. In hindsight, this is easy to understand – the workflows presented by the voice interface often required visual confirmation, such as picking a number from a list of options presented on screen. Drivers also had a habit of looking to the screen to make sure that the system correctly recognized a
VOICE INTERFACES

Apple’s CarPlay gathers momentum

In March 2014, Apple announced the first wave of manufacturers – Ferrari, Mercedes-Benz and Volvo – to roll out CarPlay. It should be available in certain vehicles in 2014.

The system enables drivers to use Lightning-enabled iPhones for an array of functions through voice commands or manually through the car’s interface. For example, it can be used for making and receiving calls, listening to and dictating messages, listening to voicemail, and accessing the iPhone’s audio files.

One can easily connect this to experience with smartphone-based voice systems, where a common behavior is to speak and then inspect the input recognition for errors.

The destination entry task was the most time-consuming, requiring an average of 111 seconds to complete in the first two studies. Task completion time was not a matter of problematic speech recognition, as most of the time the system had little trouble interpreting drivers’ voices. Rather, it was a matter of interface design. Task workflows were characterized by a series of tedious confirmatory steps, and drivers had to use very specific command phrases, such as ‘destination street address’, rather than more natural phrases. While this structure generally did an effective job of promoting successful task completion, many drivers had trouble remembering some commands, despite undergoing extensive training beforehand.

A third study examined the voice interface’s Expert mode, which removed some of the audio prompts and most command confirmations. This reduced the average time taken to enter the destination to about 93 seconds. Interestingly, this reduction in task time did not translate into a statistically significant reduction in the overall time spent looking away from the road ahead, indicating that little of the visual engagement observed in the earlier studies could be attributed to simple confirmatory interactions.

Multimodal demands

Far from being a clear-cut case of ears on the interface, eyes on the road, the research demonstrates that voice-command interfaces may engage the user in multimodal resource demands – drivers will listen and talk, but they will also continue to check the system visually, regardless of whether their attention is drawn intentionally or not. If the screen presents something to look at, drivers will often look. In designing such interfaces, care should be taken to limit visual engagement purely to

Another feature is that CarPlay can be used with Maps for navigation. This consists of spoken turn-by-turn directions, ETA and information on traffic conditions, plus a map shown on the car’s display. One clever feature is that the system anticipates destinations based not only on recent trips, but also on information gleaned from emails and texts.

As well as giving access to a lot of iPhone functions, CarPlay supports some third-party audio apps, including Spotify and iHeartRadio.

“iPhone users always want their content at their fingertips and CarPlay lets drivers use their iPhone in the car with minimized distraction,” says Greg Joswiak, Apple’s vice president of iPhone and iOS product marketing.

Others working on integrating CarPlay include BMW Group, Ford, General Motors, Honda, Hyundai Motor Company, Jaguar Land Rover, Kia Motors, Mitsubishi Motors, Nissan Motor Company, PSA Peugeot Citroën, Subaru, Suzuki and Toyota Motor Corp.

Radio Preset (Button) Radio Tuning (Knob) Radio Preset (Voice) Radio Tuning (Voice) Destination Entry (Voice) Destination Cancel (Voice)

Study I Study II (Default Mode) Study III (Expert Mode)

Total Gaze Time (Seconds)

Task Completion Time (Seconds)

Study I Study II (Default Mode) Study III (Expert Mode)
VOICE INTERFACES

“DESIGNERS MUST REMAIN AWARE THAT, ALTHOUGH THEY MAY BE TARGETING THE EAR AND THE VOICE, WHAT THEY ARE REALLY DEALING WITH IS THE ENTIRE PERSON”

information that supports the driver in a truly useful manner.

The voice interface used in these studies had generally excellent speech recognition capabilities, but the task workflows occasionally relied on specific phrases that were, at least initially, hard to remember. In some instances, auditory or visual support prompts cued the driver on response options and seemed to keep cognitive demand relatively low. However, in one of the tasks studied – canceling a destination – many drivers were unclear how to proceed.

Whole picture
User interface engineers are only just beginning to understand the delicate balancing act that vocal interactions require. Some initiatives, such as Apple’s CarPlay, which deactivates the in-vehicle screen whenever possible, seem to be designed around an awareness of these multimodal distractions. It is, however, unclear to what extent removing visual support will affect cognitive demand. Having to listen to and make a selection from a remembered list has its own demands. Further work is needed to help provide designers with guidance on how to assess these trade-offs.

Voice interfaces are becoming part of our everyday interactions, and associated design considerations will affect how quickly and fully these technologies are embraced, especially in safety-critical situations such as driving. Designers must remain aware that, although they may be targeting the ear and the voice, what they are really dealing with is the entire person. New data suggests that the results of this research generalize across vehicles beyond those considered in these initial studies. The MIT AgeLab is investing in a number of activities, including the Advanced Human Factors Evaluator for Automotive Distraction (AHEAD) consortia, which aim to ascertain how best to quantify driver workload, optimize the driver-vehicle interface so that consumers can engage with information safely while driving, and enhance the driver’s focus.

Further information