



## The Cell Phone Does More Than Just Make Calls

By Stephanie Staton - Jul 9, 2007

Interactive Data Corp. predicts that there will be 850 million remote workers across the globe by 2009. To help those workers remain productive while away from the office, many firms and employees have turned to PDAs. But while these devices increase mobility, they also make for cumbersome menus and repeated clicking and typing.

"What we saw with our customers, and especially physicians who were starting to get e-prescriptions on a PDA, was that as neat as it was to do, it really was a pain in the neck," explains Bill Montgomery, national director of healthcare sales at Sprint. It could take doctors 35 to 70 seconds to create and file e-prescriptions using their PDAs. "That length of time spent clicking a wheel or pen or typing was annoying for most physicians. Only the die-hard techies didn't mind, so the adoption rate for e-prescribing was pretty horrific. What became clear to us was that this was too slow and too hard."

"To type through a menu is very painful. It takes about 40 clicks to find a Rolling Stones song by typing with your thumbs," adds Michael Thompson, vice president and general manager of search and communications at Nuance Communications.

Developers have toyed with ways to make mobile devices like these more user-friendly, but what if it were as simple as using what 3 billion people around the world already have in their pockets—their mobile phones?

According to Sprint, it is that easy. Working with companies like Nuance and Vocera, Sprint is rolling out speech-enabled solutions for its mobile phones across all of its enterprise and individual customer bases. Rather than going to a Web site and clicking through pages and pages of data or calling into an interactive voice response (IVR) system, Sprint makes it possible for airline travelers to push a button on their cell phones and say *status, Delta Flight 312* to have the answer delivered immediately to their phones. Getting that answer would take about two minutes through a company's IVR or Web site, but using speech access via the mobile phone cuts that to anywhere between two and four seconds.

"The speech interface can help to provide a much better experience for interacting on a device like the cell phone, especially in a situation where you can't use your fingers or you can't pay too much attention using your eyes on the display," explains Thilo Koslowski, vice president of the Automotive Manufacturing Industry Advisory Service at Gartner.

One of those situations is in the car. "Mobile phone usage-related traffic accidents are on the rise. In the United States, New York, California, Connecticut, and Washington, D.C., have all passed hands-free mobile driving laws. Similar regulations have been enforced in countries such as Australia, Austria, France, Germany, Japan, and the United Kingdom, to name a few," explains Daniel Hong, lead analyst at Datamonitor. "This opens up opportunities for the voice user interface in mobile devices and automotive navigation systems if vendors can render them truly dependable. The ability to operate mobile phones and automotive navigation systems with hands-free navigation is becoming more crucial and vendors must examine the opportunities for speech recognition."

That's exactly what the industry is doing. "Mobile applications are once again looking to speech as a possible solution. We have all been through a couple of iterations and false starts with speech technology. The positive side is that people are re-engaging in the possibilities of speech. In terms of timing in an industry, there are increasing legislation and social pressures around the interface while you are in the car," declares Victor Melfi, chief strategy officer and senior vice president of voice services at VoiceBox Technologies. "The time has come for speech."

"Realistically, if you look worldwide there are about 3 billion cell phones out there, and that is still growing very rapidly. The shift that is going on is that these cellular phones—or whatever device this is going to morph into, is going to be the primary device that people use from now on," says Betsy Wood, a multimedia applications evangelist at Nortel Networks.

What's more, Interactive Data Corp. predicts that shipments of converged mobile devices will grow from 80.9 million in 2006 to 304.4 million in 2011. The demand for access to mobile content using speech is becoming more than a matter of convenience and coolness; it is a matter of efficiency and productivity. "If you are going to provide service for those people and do it really well, speech is going to have to play a role in it," Wood says.

Granting users access to speech on mobile devices opens up many opportunities for individuals and enterprises alike. The convenience of this technology for the individual actually complements and creates value for the enterprise. The access to information and tasks that would generally be limited to the desk or an environment where hands and eyes must be used is expanded by freeing the users' hands and eyes for other tasks, such as driving. In addition, this technology enables users to multitask, which in turn makes them more productive and efficient throughout their day. This increased efficiency and productivity has obvious benefits to the enterprise; it also creates more free time for the user.

"Enterprises benefit if their employees are more productive. More productive means their employees can communicate and do their jobs on their mobile phones more effectively. What is happening around this market in the enterprise is that speech-enabled capabilities really unleash productivity for the enterprise customer," Thompson says.

### **Embedded Speech**

Traditionally in the mobile environment, speech has been used as a control and command application that is embedded into the device. Embedded speech consists of a speech recognition engine that is built into the device itself. When the device ships, the recognizer ships with the device and doesn't connect to anything. It can be used for MP3 players, games, toys, voiceactivated dialing, and more. There are also text-to-speech read-back capabilities.

"The embedded capabilities that are very important are things like voice-activated dialing and playing songs stored on your phone. There is local content sitting on your device. You can launch that and you are good to go," Thompson explains.

The embedded model has benefits that can outweigh those of networked applications in some cases. "The key benefit is that you get a more reliable experience. You don't have to rely on the availability of the network. You can have the device working in an environment where you have no network connectivity and that is important to users," Gartner's Koslowski states.

"The tradeoff is if you designed an app that has the horsepower and memory to work on the device and has all the information you need, then, of course, you would stay on the device because you don't inherit all the latencies involved in going up the mobile network. But people really want access to fresh information when they are mobile," VoiceBox's Melfi says.

However, embedding applications into the device can be complex and limited in its reach. "The footprint of the voice recognition software can be a significant obstacle for portable devices in particular, which don't have a lot of available memory.

The other problem is regarding the processor speed. "To get a reliable experience and have the recognizer understand your command from speech dialogue requires that the processor is used heavily. The device manufacturers have to consider exactly how much else the device can do while it is processing voice inputs from the user. In most cases that means you can't do too much in addition to the voice processing," Koslowski concedes.

"The first limitation on embedded devices is the amount of hardware horsepower required. The other limitation is that even if you solved that problem, you are still stuck with the data being

static," Melfi states. "With embedded you are stuck with the information on the hardware and there is not enough space to keep the information fresh and meaningful."

### **Network Links**

Not completely abandoning the embedded market, many speech vendors in the mobile arena are expanding their reach and capabilities through a network-based architecture. The network-based services work off a server pulling live information from the Internet, giving users real-time information about traffic, weather, directions, and anything else that can be searched online. Most of the benefits remain the same as using embedded: quicker, easier access to information, freeing up hands and eyes for other activities.

The only advantage the technology offers that embedded does not is access to live data, which is, of course, limited by the network itself. Naturally, if you are in an area that doesn't have service, you can't access the information. "Using your voice to navigate the device itself is an appropriate use of voice [technology]. But as soon as you deal with the issue of information access and contact, frankly I can think of no good embedded applications. You have to get off that device and get to that server in the clouds to get meaningful information," Melfi asserts.

"These little mobile phones are powerful little computers that can connect to data just like a PC can and you can get results via the browser just like a PC can," Nuance's Thompson explains. "It seems like a natural thing to push a button and talk into a phone because we have been doing it for over a hundred years around the world. It is just easier to speak than it is to type."

"Network-based voice recognition access sounds good, but it is difficult to realize, because, first of all, we have to have that network in place," Koslowski says.

Additionally, whether you are using an embedded application, a network application, or a combination of the two, there are required components and power levels that must be met. On the hardware side for embedded applications, you need a processor with a minimum of 225 MHz of processing power for the speech recognition technology, a minimum of about 20 MBs of storage for the application and data, and a microphone. For network access to speech, the mini client is just less than 14 KBs with the ability to record the voice and send it up the mobile network; there is no processing requirement on the device. The device just needs a microphone.

The most common tools available to developers of these applications come in software development kits (SDKs) provided by the speech technology vendors. Some vendors, such as VoiceBox, prefer to manage the development on behalf of the customer and include that in their services package.

When developing and designing speech applications, there seems to be one universal best practice to always keep in mind. Datamonitor's Hong describes this approach as a user-centric application where "more research on consumer behavior on the front end before application design and usability studies is required."

That is something that cellular equipment manufacturer Motorola has actively been pursuing. Through the company's Human Interaction Research Labs, it has tried to build an understanding of user interface architectures, development tools, prototyping, experience design, input interpretation, and output generation. Specific areas of focus include image understanding, speech recognition and synthesis, tactile generation, contextual reasoning, workload management, goal determination, and user interaction and preferences, says Tom McDonald, senior manager of technology marketing at Motorola.

### **Making It Work**

"In order for seamless mobility to become a reality, devices and networks must enable users to achieve their goals while having complete freedom as they move between various devices and environments," McDonald explains. "This requires a higher level of intelligent interaction between the user and the device and applications." There are also guidelines to keep in mind for developing and designing applications specifically for mobile devices.

"Looking at the resource requirements for the device is extremely important to do early on, as is playing out a couple of user scenarios where a user is using specific applications on the device

together in a voice-based format. You need to understand the processing requirements to see how smooth the applications are working, how positive the voice experience would be for that user—it is not enough to just put your speech engine into the device and hope that everything will just work out. Using actual usability studies is extremely important," Koslowski states.

According to Nuance's Thompson, it is best to look for large, telco-grade, high-end speech recognition software solutions that are scalable because "anything mobile needs to be big over time." He also recommends investing significantly in the user experience when designing the system. "This is not for engineering experimentation on how people behave. Consumers and sales reps that use mobile phones for business behave in very unique ways. Getting someone with experience on that behavior is critical. Especially in the search space, it is important to have someone who understands how to search very large grammars using voice. To search 1.2 million songs with your voice requires some pretty robust capabilities — large grammar and dictation experience and depth," he adds.

Beyond the productivity gains, "companies of all types have started to realize that up until this point [the mobile market] has been a missed or untapped opportunity for customer service, so what they are looking to do is come back and improve what they are doing to take advantage of this lost opportunity," Wood says.

"Companies are exploring new innovation as it relates to voice recognition and text-to-speech to make that interaction more human-like for the user. Because there are so many devices on the market today and some have limited speech recognition capabilities, consumers may shy away from using voice recognition initially due to bad experiences in the past. Awareness building and actual demonstration would help consumers significantly to get more comfortable with this technology," Koslowski adds.

Making the speech-enabled experience easier and quicker will motivate the user to use the voice interface. "The accuracy of typing on a mobile phone is less than 70 percent; with a regular keyboard it is in the 90s, so speed and simplicity are the most important reasons why speech will be a very, very powerful interface for the mobile phone," Thompson says.

"Going forward—because you would be able to use more intelligent application software with higher accuracy due to the more contextual type of market behind it—the footprint for embedded will be smaller. You will eventually see the network-based piece of this, but not in the short term or mid term. You still have to have the network in place that can actually reliably communicate the data back and forth. You will see more intelligence being offered that require a smaller embedded memory footprint using up less of the internal power of the device," Koslowski maintains.

"The world has entered an era of ubiquitous computing where one user interfaces with multiple computing devices, such as mobile phones, Blackberries, MP3 players, and gaming devices, on a daily basis. Speech is becoming more widely accepted as an interface between user and computing devices, and has the potential to grow in tandem with rising expectations about access to information while on the move. GPS-enabled navigational systems in automobiles and portable devices, voice search, and voice command-and-control interfaces are all potential areas for growth," Hong says. "If vendors can provide a consistently reliable solution, this could be a key moment for the expansion of speech into mobility."

As the mobile devices industry flourishes, the role of speech remains to be seen. However, the goal of the technology, as well as those who use it, is clear. "At the end of the day, what it is really about is productivity, how to get things done quicker, faster, simpler. I think it is all about, at a mobile front end, how to make it simple. We believe that the phones are way too complex right now for users. The more and more applications we put on, the more complex it becomes so adding a voice interface to that is what's going to really simplify it," Sprint's Montgomery says.

"Users don't care whether it is embedded on the device or networked; they just want a seamless, easy experience," Thompson concludes.