speechTECHNOLOGY

Miami Children's Hospital

New Operations in Speech

Page **28**

New Operations in Speech



Photos by: Jon Kral, Edgar Estrada and Steven Llanes

By Stephanie Owens

Miami Children's Hospital, a world leader in pediatric health care, has a medical staff of more than 650 physicians and over 2,000 employees. The hospital specializes in all aspects of pediatric medical care from birth through adolescence. Although the hospital draws children with very specific needs from all over the world, it is also the only licensed specialty hospital exclusively for children in South Florida.

Originally opened in 1950 as Variety Children's Hospital, its name was changed to Miami Children's Hospital in 1986. MCH now treats more than 185,000 patients each year. It is a non-profit, free-standing hospital. Today, MCH is in the midst of a renovation project that will enlarge the hospital by 78,000 square feet, making it one of the most attractive health care facilities in South Florida.

With all that is going on at the hospital, MCH is also in the process of incorpo-

rating speech recognition technologies into their existing i-Rounds clinical documentation application from Teges Corporation. Teges Corporation visited IBM in November of 2004 and saw a demonstration of the X+V programming language. The demonstration used IBM's WebSphere
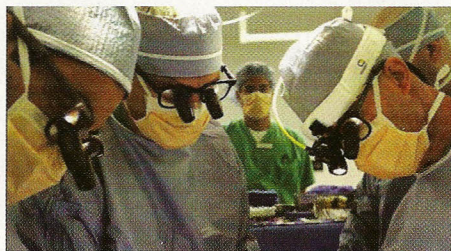
speech technology to speech-enable Web pages. Teges realized that IBM's speech technology could be used in their i-Rounds application. In January of 2005, Teges teamed with IBM to create a speech-enabled prototype of the i-Rounds application. The entire i-Rounds application was speech-enabled for navigation and other functional components, such as an

Operating Room (OR) timer and a Cardiac Intensive Care Unit (CICU) voice recording application, within 30 days. The resulting application can be used with traditional keyboard, mouse and display as well as with speech input and output.

It was MCH's need for hands-free and eyes-free access to computing that would help clinicians access or record critical information without having direct contact with equipment that led to their implementation of the multimodal i-Rounds. There are many areas within the hospital where clinicians do not have access to computers or hardware because the environment is sterile or the clinicians' hands are busy with surgery or other tasks. These are the areas that MCH wanted to focus on with speech technologies. Their goals with this project included hands-free access to all critical patient information from anywhere with any device, increased efficiency and speed with which clinicians access and create patient information (including voice recordings) within the hospital, an interac-

tive environment for users to communicate with the computer, text-to-speech (TTS) capability of the voice application to warn users of dangerous trends/events and TTS capability to remind users not to forget key elements of the medical treatment.

The speech application's ability to meet these goals would depend heavily on its ability to adapt to dynamic and noisy environments. MCH found ways to minimize ambient noise and increase recognition accuracy by finding, positioning and outfitting MCH staff and rooms with the appropriate microphone technology. The application had to be easy to use so clinical staff would quickly adopt the new technology. In the end, IBM's WebSphere Everyplace Multimodal Environment allowed the Hospital to design new features as well as enhance the existing i-Rounds application to become multimodal. The underlying technology for doing this was a Web-based programming language called XHTML+Voice (X+V). Based on Web standards, X+V sped the addition of speech to i-Rounds and it used application development skills to which Teges developers were accustomed. Also, since the speech technology is user independent, the medical staff was not required to do any speaker training to use the application.

The speech-enabled application underwent a litany of tests both for endurance and accuracy before it was used in a production environment. This included IBM's testing procedures such as usability, automation, lifecycle testing, and audio analysis. The browser with which the voice application is delivered has a series of debugging tools such as logging, tracing and audio capturing, that the IBM team used to reproduce the exact environment and interaction the user had with the computer. The OR, though an uncommon environment, shares many similar characteristics of other environments that IBM has speech-enabled, such as automobiles. For example, heart monitors provide constant noise to the "always listening" speech system. This situ-

ation is similar to constant noises in a car. Adjusting variables in the multimodal system such as microphone type, placement and speech engine settings enabled the system to be tuned to the operating room environment. The debugging functionality available in the system proved critical to the success of the project.

The speech features of i-Rounds have been enabled in the Cardiac Intensive Care Unit (CICU) and in the OR for use during pediatric cardiac surgery. They provide a hands-free mechanism for physicians to enter information, retrieve information and record their voice directly into the patient's medical record making the doctor's assessment/diagnosis available immediately without waiting for transcription services. In the OR, the computer speaks to the surgeon through four speakers that are embedded in the ceiling. The surgeon interacts with the computer using a cardioid wireless microphone, which activates the speech system when the surgeon utters the keyword, "Computer." In the CICU, clinicians access the system by using a wireless tablet PC. They use the tablet's built-in microphone and activate the speech system using a push-to-talk mechanism (depressing a button on the tablet).

When MCH started testing the speech recognition in different areas of the facility, they encountered three major hurdles. First, ambient noises, such as beeping oxygen saturation monitors, drills, and saws in the OR and bedside monitors, alarms, families, patients and staff making noise in the CICU, created a challenge for the speech recognition system. MCH researched microphones, both wireless and hard wired, to improve recognition. Once they determined the position and type of microphones to use, they were able to eliminate much of the ambient noise and significantly increase the speech recognition capability.

Second, MCH had to overcome resistance in some environments to pushing a button to indicate to the computer that they were going to talk. The users wanted to talk to the computer spontaneously and have it respond to their commands. In order to fulfill this request, IBM made changes in the WebSphere Everyplace Multimodal Environment that extended existing features built into speech technology giving the system the ability to first listen for a keyword (e.g "Computer") and then listen until the user stopped speaking. Certain design changes were made to the i-Rounds application that resulted in users having more flexibility with the system; for

instance, during surgery, the surgeon tells the system when they pass key phases of the operation such as going "on bypass" or "cross clamp off." The surgeon needs to discuss passing these phases with the team and saying the phrase "on bypass" could accidentally trigger the computer to process that event. With the keyword activation feature, the surgeon says "Computer, on-bypass," which solved the problem in a way that was natural and intuitive for the clinicians.

The third and final significant challenge was developing a common dialogue or nomenclature for navigating and delivering speech commands within the application. Early users were, literally, getting lost in the system. More specifically, some users of the speech application never look at a computer monitor when using the system and are only talking to the voice-activated room. To handle this, navigating to all the major sections of the application were made common by always saying "Go to . . . ." Then on each page, a pop-up window shows the speech commands that are active for that particular part of the application. Having a consistent way to navigate around the i-Rounds application and then having the application show which commands were valid at the time relieved users of the burden of remembering commands. This made the system easy as well as efficient for users who knew the commands.

Once all the challenges were overcome and the tuning of the application was complete, the system was implemented in the OR and CICU. In the OR, the technology serves as a safety net. It provides hands-free verification of the patient, diagnosis and planned procedure. It also provides timed reminders for the delivery of medication and other procedures that must be administered at specific intervals. The safety net of the speech application reduces the occurrence of human error by offering an alternative reference for the administering clinician to use before performing any life-threatening tasks on the patient. This may eventually work to improve the hospital's mortality rate, which is already among the best in the world at 98 percent.

Also, in the CCU, it works as a new way to record transcription by recording the user's voice directly into the Web browser which then becomes part of the patient's electronic medical record. Others can use the browser to access the information from anywhere and at any time. By capturing the doctor's audio transcription into the patient's database record, this information becomes immediately available to other doctors and medical staff. Speech-to-text occurs subsequently offline and the

results are updated in the patient's database record. Health care practitioners then have real-time access to the doctor's notes, which previously was not the case. Also, by having the transcribed text in the same electronic document, the data can be mined in the future. This system is an improvement over handheld recorders because the data is in the record under application control/protection and not the doctor's pocket or courier bag traveling to a transcription service.

MCH has only just started to use the multimodal i-Rounds solution and has put it to use in two departments. At this time, it's difficult to predict exactly what impact the system will have on further reducing mortality rates or preventable medical errors throughout the hospital. Research is available suggesting the relationship between improved outcomes and increased availability of patient information. The multimodal i-Rounds application provides speech input and output as yet another vehicle for medical staff to access patient information faster and more efficiently than ever before.

Another potential benefit is that speech-enabled systems use built-in libraries of words and phrases. This has the positive side effect of enforcing common spelling and phraseology with the staff. It improves data accuracy and avoids handwriting and typing errors. Another potential, but as yet unproven benefit, is more efficient collection of patient data from caregivers because entry is quick, easy and can be done on the spot.

Making a huge paradigm shift, MCH's users had to adapt from both a clinical documentation and computer interaction stand point. Therefore, MCH has received mixed reactions from these users. One of the main concerns expressed by physicians is that they don't have the time to learn new technology at the expense of losing time with their patients. Since IBM's speech engine is user independent, it does not require voice training, which solves one of MCH's major obstacles for implementing speech technology.

In an effort to continue to tear down barriers to implementing speech recognition hospital-wide, MCH is updating its IT Vision and Strategy and aligning it with the clinical focus. Once complete, the multimodal capability will be an option that can be rolled out to physicians wishing to use it for their documentation needs, and potentially to other caregivers including nurses, technicians, etc. Also,

because the multimodal i-Rounds application is a fully working traditional application for keyboard, display and mouse, users can experiment with using speech at their own pace. That is, they can keep using the system the way they're used to doing it and use speech when they want to.
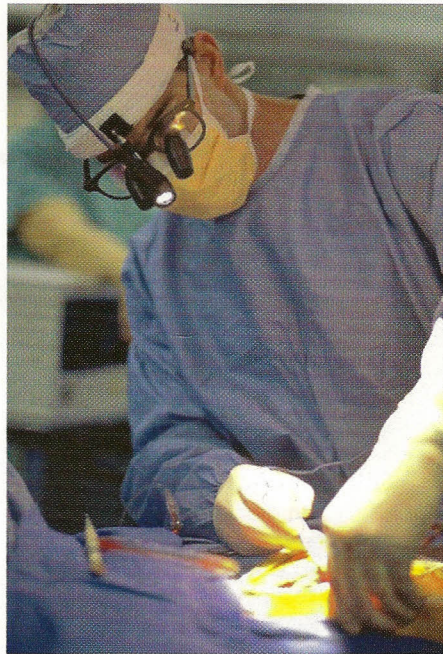
MCH will continue with its testing to determine if this technology will be widely accepted by the clinical staff. Physicians are often conservative and MCH anticipates some resistance, only because it is new technology and a new process compared to the way care is historically delivered. Once practicing physicians see the benefits gained through the system's efficiency and accuracy, MCH predicts that their resistance will be replaced with excitement and adoption of the speech system.

Special thanks to Jeffrey A. White, head of research and information systems for the Cardiac Intensive Care Unit and the Division of Cardiovascular Surgery at Miami Children's Hospital (MCH), and David Jaramillo, senior software engineer, embedded voice solutions development, IBM Software Group, for all of their help in coordinating this article.

Stephanie Owens is the associate editor for Speech Technology Magazine. She can be reached at stephanie@amcommpublications.com.

....

As devices become smaller, modes of interaction other than keyboard and stylus are a necessity. In particular, small handheld devices like cell phones and PDAs serve many functions and contain sufficient processing power to handle a variety of tasks.

Present and future devices will greatly benefit from the use of multimodal access methods.

Multichannel access is the ability to access enterprise data and applications from multiple methods or channels such as a phone, laptop or PDA. For example, a user may access his or her bank account balances on the Web using MicrosoftR Internet Explorer when in the office or at home and may access the same information over a regular phone using voice recognition and text-to-speech when on the road.

By contrast, multimodal access is the ability to combine multiple modes or channels in the same interaction or session. The methods of input include speech recognition, keyboard, touch screen, and stylus. Depending on the situation and the device, a combination of input modes will make using a small device easier. For example, in a Web browser on a PDA, you can select items by tapping or by providing spoken input. Similarly, you can use voice or stylus to enter information into a field. With multimodal technology, information on the device can be both displayed and spoken.

Motorola, Opera Software ASA, and IBM submitted to the W3CR a proposal for a multimodal markup language standard called XHTML+Voice (X+V for short) that provides a way to create multimodal Web applications (i.e., Web applications that offer both a voice and visual interface).

Multimodal applications using X+V offer a natural migration path from today's VoiceXML-based voice applications and XHTML-based visual applications to a single application that can serve both of these environments as well as multimodal ones.

For addtional information:

1. Please contact your IBM Representative
2. Please view the IBM Mutimodal web site @ http://www-306.ibm.com/software/pervasive/multimodal

IBM®