SPEECH RECOGNITION...THE NEW FRONTIER FOR CONVERSANT® VIS

Last month BCSystems announced significant speech recognition enhancements to the CONVERSANT Voice Information System (VIS), which allow callers to conduct various types of transactions more easily and more naturally than ever before.

The virtue is that it lets callers speak to the system, offering an alternative way to enter information into an interactive voice response system. What's more, this technology extends the use of these systems—once limited to callers with touch-tone service—to rotary telephone callers.

The new system was developed by AT&T Bell Laboratories' Voice Transaction System department in Columbus, Ohio—using Hidden Markov Modeling and the latest in algorithm technology developed at Bell Labs in Murray Hill, N.J.—and is produced at AT&T's own manufacturing facilities in Denver.

AT&T's speech recognition feature adapts to human speech. Human conversation is usually sprinkled with a variety of sounds that are not actually words, such as "um" and "uh." During conversation, people recognize these sounds as unnecessary, automatically weed them out, and focus on the vital words that convey a concept.

According to Bob Perdue, supervisor of the Algorithm Development group in Columbus, the CONVERSANT VIS speech recognition option uses intelligence similar to humans'. A caller may say, "Uh, yes, my account number is...let me see...one-six-four...No, it is two-five-one-three-eight." A capability called "word spotting," lets the system block out the extraneous sounds.

"Other systems may become 'confused' by additional sounds that they aren't trained to recognize," explains Perdue, "giving the caller another prompt, or asking the caller to speak only the digits, which slows the completion of the transaction, and makes it seem less natural to the caller."

Another capability of the new speech recognition feature, "prompt interrupt," allows callers to speed their transactions. The caller can verbally interrupt system prompts to proceed to the next activity.

"These capabilities are unique to AT&T's speech recognition product," notes Perdue.

Taking time out from their digit-recognition algorithm work (left to right): Bell Labs' Bob Perdue, Roy Grubbe, Prasad Mikkilineni,and Sedat Gokcen (bottom center).

The AT&T speech recognition feature of CONVERSANT VIS is currently supplied with a small vocabulary of 13 words for the general population. The technology can support more than a hundred key words on a customized basis. The feature is speaker-independent, meaning that it recognizes and responds to any caller's spoken words, and requires no advance preparation or instruction for users. The system processes the spoken commands using 12 to 24 AT&T microchips operating at as many as 600 million instructions per second, "for superior parallel processing power," according to Larry Rabiner, director of the Information Principles Research Lab at Murray Hill.

The system's accuracy rate over the telephone network is between 95 percent and 99 percent, depending on the application and the length of the string of digits that the system must recognize. "This is the highest accuracy rate for speaker-independent speech recognition systems," says Jay Wilpon, a Member of Technical Staff in the Speech Research department at Murray Hill. "Typical customers should experience well over 95 percent accuracy for most applications."

"AT&T is the only company that enables its customers to easily update their speech recognition investment as technology evolves," says Curtis Weeks, BCSystems' Product Management vice president.

CONVERSANT VIS V2.1 will begin controlled introduction this quarter. The speech recognition option is scheduled for general availability in June.

SPEECH RECOGNITION USES LIMITED ONLY BY CUSTOMER NEEDS...AND IMAGINATION

The new speech recognition capability presents a rainbow of opportunities, with a string of admirers already on board to employ it for some unique—and well-known—uses, both here and abroad.

The word spotting feature is scheduled for use in conjunction with the 1992 summer Olympic games in Barcelona, enabling Telephonica, Spain's telephone company, to offer toll-free information services. A system developed by the Advanced Services Technology Department at Indian Hill Park is being trained on Spanish digits using 8,000 speakers, and is providing essentially error-free performance.

And this is only the beginning. BCSystems and Bell Labs already have more than a dozen beta site tests with the CONVERSANT VIS speech recognition capability. These customers represent a variety of industries and disciplines—including transportation, law enforcement, franchise food service, retail sales, banking and brokerage.

Because the system recognizes digits, both individually and in strings—such as telephone numbers, social security numbers, catalog numbers, credit card numbers and warranty numbers—the speech recognition option can be used for bank account inquiry, credit card services, job postings, information dissemination and announcements, employee benefits inquiries, cargo status and location, order and inventory status.

For example, an airline could customize the capability to include the words "arrive," "depart," " flight," "from," "to," in the system's verbal repertoire, enabling travelers to access flight schedules simply by asking for it verbally.

For retail sales, a mail order company could offer its customers the ability to inquire verbally on the status of orders and deliveries.

For educational institutions, student registration would be fast, easy, and accessible to any student with a telephone.

The applications are only limited by customers' needs—and imagination.