

Goodbye, Operator

Speech recognition is revolutionizing customer support—in multiple languages.



by Steve Adams

Airline Representative: “Hello, Mr. Zhu. Will you be scheduling your monthly trip to Paris today?”

Mr. Zhu: “Yes, I’d like to leave after 8:00 PM on Thursday.”

The conversation continues, and Mr. Zhu hangs up, satisfied with his flight times and holding his confirmation number.

Things aren’t always as they seem. Mr. Zhu’s “airline representative” was actually a computer programmed with the latest speech-recognition technology and linked with the airline’s customer database. The computer identified Mr. Zhu’s voice, looked up his flight records, and began a personalized exchange that reflected Mr. Zhu’s previous use of the system. Had Mr. Zhu spoken in another language, the computer would have understood his request and responded in his native tongue.

This article will examine how global companies are starting to give their customers improved service with ASR (automated speech recognition) technology—regardless of what languages they speak. As companies expand, merge, and otherwise “go global,” they must have increased options for supporting all their customers in a variety of languages—without having to invest millions in hiring native-language speakers.

Goodbye Touch-Tone

Traditional touch-tone technology provided a way for companies in many countries to conduct very simple requests for their customers, such as “For your account balance, press one.” Such touch-tone interfaces fall short for anything more complicated, such as transferring 250,000 yen from a savings to a checking account tomorrow. And even simple applications often leave callers frustrated or stuck in seemingly endless loops.

Further, touch-tone severely limits the type of transactions that can be accomplished, since the basic vocabulary is 0-9, *, and # on a keypad. Think of how difficult it would be to buy and sell stocks, or any other typical phone transaction, using only numeric codes.

Touch-tone also poses a variety of additional problems for global companies. Some countries such as those in South America cannot support touch-tone interfaces because the “tones” are not recognizable in adjoining countries. In India and Malaysia only rotary-dial phones exist, making it impossible to attain information with touch-tone.

There are also safety issues to consider when using touch-tone applications on cell phones. With the recent surge of cell-phone use in many countries, particularly in Asia, entering long strings of numbers into the phone can be dangerous indeed.

Imagine the difficulty of entering a long list of numbers while driving a car.

Hello ASR

Speech-recognition technology attempts to convert human speech into a string of text that accurately represents what the person on the phone is saying. This complex task requires a thorough understanding of such disciplines as digital-signal processing, electrical engineering, statistics, and perhaps most importantly, global linguistics.

Converting speech into meaningful data requires these steps:

- 1: The system captures the caller’s spoken words and utterances in the form of an analog acoustic signal.
- 2: The system converts the analog acoustic signal into digital components it can understand.
- 3: The computer splits the digital version into distinct sounds or segments. Each of these segments corresponds to a specific sound, such as consonant or vowel sounds like “s,” “p,” “i,” etc.
- 4: The speech-recognition software attempts to classify each sound by determining possible matches between the sound segments and their phonetic representations (e.g., trying to match an “s” sound to the phoneme “s”).

5: The speech-recognition engine searches for the word or phrase closest to what the caller said. This involves mapping the sets of possible phonemes to the words or phrases that form the vocabulary for the particular application. A travel service, for example, is listening for things like city names, dates, and times. If you start asking about 401(k) plans, it probably won't be able to process the transaction.

ASR at Work in Mandarin

Recently, the Singapore Stock Exchange launched a Mandarin ASR application to allow the public to attain live stock quotes using natural language. As Mandarin is one of the world's most popular languages, and many adult Singaporeans own stock, this application allows a very sizable population group to call a hotline and inquire about stock prices in their native tongue. For example, the computer can engage in the following dialog:

Computer: "For the current stock price, please say the company's name."

Customer: "Singapore Airlines."

The computer will respond with the latest stock price by linking into the Singapore Stock Exchange database, and then ask the caller if other transactions are needed such as, "Are there other stocks you are interested in?"

This ASR application even understands such sentence "fillers" as "lah," the common utterance Singaporeans attach to the ends of sentences.

With current speech-recognition capabilities, applications not only provide information, but also allow the caller to transact business over the phone such as making airline reservations, trading stocks, transferring funds, finding restaurants, and much more. These speech systems aren't replacing Web or call-center activity, they're complementing it: a brokerage client, for example, can place a trade on the Web at night, then check the order status via phone on her drive into work the next morning without ever talking to a live broker.

Is ASR Right for My Global Company?

Once a company recognizes the need to interact with its customers and prospects in a variety of languages, it may consider ASR technology. These are some important questions to think about when investigating ASR in a global application:

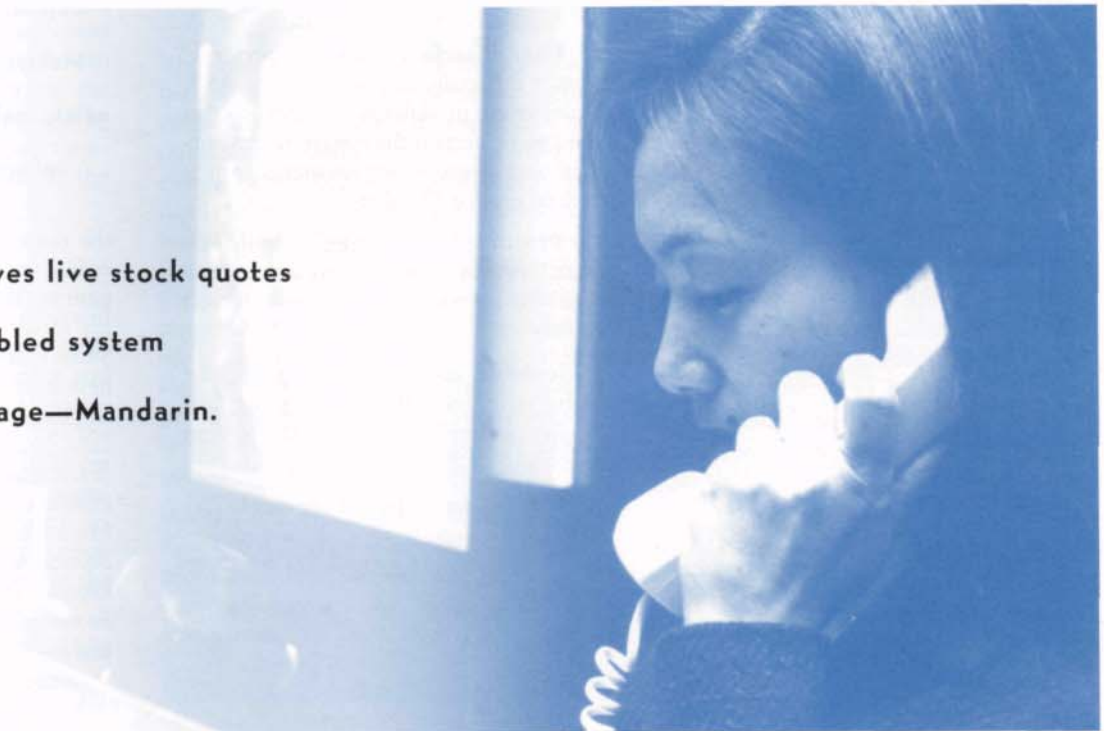
- What are the most important languages to support?
- What are the different dialects or accents in my target market?
- Where is my company expanding?
- What are my customer-service goals?

Benefits of ASR in a global company

Speech-recognition applications in a global call center allow companies to:

- Communicate and help customers in their native languages
- Extend the hours when customers can get information or conduct transactions
- Increase sales revenue by providing customers an easy way to get information or conduct transactions
- Reduce hold times by gathering caller information up front, and provide customer-service representatives (CSR) with information about who is calling and why
- Reduce costs by increasing the number of calls processed by a call center
- Improve service by handling routine inquiries and freeing up CSRs to focus on complex questions

A Singaporean retrieves live stock quotes through a speech-enabled system that speaks her language—Mandarin.



How global companies can customize ASR

Establish a distinctive voice and style. Consider the level of formality and speech style appropriate to the culture in question. For example, a public-transportation system in Hong Kong may want to project an image of speed and accuracy.

Their prompts may start with: "Tell me what time of day you are travelling." A cruise line in Norway on the other hand, may choose to stress value or pleasure. Their initial prompt may be, "Are you celebrating a special event or anniversary on the cruise?"

Offer multiple navigation options. Be sure to accommodate both experienced and novice users. While the first-time user may need to listen to every prompt and follow step-by-step instructions ("directed dialog,") the experienced user may wish to interrupt prompts ("barge in") and speak in complete sentences ("natural language").

It's helpful to guide your callers in the use of natural-language short-cuts by prompting with hints such as, "Next time, you can just say, 'I want to fly to Tokyo this Thursday in the morning.'"

Be friendly. Use polite, conversational phrases such as "I'm sorry, I didn't understand that," as compared to the more technical, "That was an invalid entry."

In short, don't make your callers feel uncomfortable because they're unsure of what they are expected to say.

Be culturally aware. Be aware of the phone culture of the people using the system. For example, some people may be more sensitive to providing personal information over the phone than others.

Others may not be used to using technology to conduct business transactions. The prompts should reflect the culture of the users, and be sensitive to their needs.

How to evaluate an ASR system

Companies interested in implementing an ASR system should consider the following:

1. Number of languages and dialects supported. Can the system understand Mandarin? French? Spanish? What about the 12 different Australian dialects?

2. State-of-the-art recognition accuracy. How accurate are the systems? No system is 100-percent accurate, however a 95-percent call completion rate is considered good.

3. Natural-language processing. Will the system be able to understand callers that speak in complete, natural sentences? What data are available to substantiate such claims?

4. Barge-in deployed with customers. Many global users will want to bypass the prompts and get the information they need quickly.

5. Advanced development tools. Some systems offer development tools that get systems up and running quickly. They do this by bundling complex speech science concepts and user interface design expertise into pre-tested, application "building blocks" which can be easily arranged and configured to create a variety of telephony applications.

6. User-interface design process. This is especially important when ASR is deployed in different countries. You want to be certain the system meets cultural and language requirements so it is easy to use for all callers.

7. Prototyping center. Companies must "try out" their system in a prototype before deploying it in all call centers.

8. Scalable architecture. Critical for global companies that are growing fast, flexible architecture enables the system to expand as the company expands.

9. Operational support services. The best technology is of little use unless robust systems and procedures are there to maintain it.

ASR is ideal if many of the following statements are true:

- A significant percentage of users are rotary or mobile callers
- Callers speak in a variety of languages
- Customers pay for access to information
- An existing touch-tone system contains complex menus and/or long lists of caller choices
- Menu items include alphanumeric strings (such as tracking number "3Z78UN8T")
- The cost per call has been identified and is perceived as high
- The target application accesses the same information database as an existing Internet-based system
- There is a large percentage of high-volume, repetitive information requests that do not require human assistance (such as account balances, price quotes, basic travel information, etc.)

ASR can offer many companies a moderately-priced way to show the world that they're serious about "going global," while making it easier for their customers to do business with them. For example, now that

Touch-tone poses a variety of problems for global companies.

In certain parts of South America, the "tones" are not recognizable in adjoining countries. In India and Malaysia only rotary-dial phones exist, making it impossible to obtain information with touch-tone.

the euro is in place, privatization is in full swing, and Europe is more globally competitive than ever before, German banks are looking to attract French, Italian, and Swiss customers. One way to entice these new customers is to start "speaking their languages" with an ASR system.

Steve Adams is the vice president international of SpeechWorks International, Inc., a company dedicated to enhancing the development of speech-activated services for information delivery and e-commerce through its patent-pending DialogModules and state-of-the-art recognition engine.
www.speechworks.com