

# Dialog Designs: Knowing Which Type of Speech Recognition to Use Can Make a Big Difference

BY PETER F. THEIS



**T**he idea of a company handling its customer service requests with automation has been around for years, but coupling the automation with speech recognition systems is still a relatively new process.

Automated call centers can handle more customers in less time than live agents, and at less cost, making them quite attractive to corporations who deal with ever-growing customer lists. But the

decision is no longer just a matter of choosing between live agents and automation — there's a matter of choosing what type of automation agent best serves a particular calling clientele.

The idea of a computer or automated system that can talk interactively with people involves a number of technologies, including artificial intelligence. The key for distributors of call center applications is technology that goes

beyond artificial to natural in the caller's perception. Many people have been looking for the ability to perform natural speech for a long time and claimed to be close to having natural speech. At many conferences, speakers say natural speech is just around the corner (with more advances being made in word recognition). Everybody is looking for that as the holy grail of speech technologies.

The real bottom line is not ➤

how well a system works the person operating the equipment, but how well it works with callers. That is measured by the yield.

If 100 people call, how many of those people complete the call? If you compare the natural speech yield against the yield of a conventional interactive voice response system's "push this and push this and say that" system, you get from 50 percent to more than 300 percent more callers satisfactorily responding.

The ability of a machine to answer a telephone used to be the sole criteria. How fast does the machine answer? How many calls can you answer at a time? The fact is, people who were answered by a machine and said, "Oh, not one of those!" and hung up were never factored into the value equation.

The user of the machine rationalized, "Well, hanging up was their choice, and that's the caller's problem, not mine." The evaluation of technology became based on largely irrelevant factors.

The trend for future designs puts the user in control throughout the process. For example, when voice mail first began to be used, it went in that direction coupled with the use of touch tone - "thank you and goodbye." Then the IVR came out, and that premise was that users could do everything by just punching in responses on the touchpad.

After IVR came speech recognition. Now there are systems that actually incorporate intelligent branching, which means the program can change depending upon what the caller says using natural speech. A more technical term for the intelligent branching would be called gisting, recognizing the gist of what a caller is saying.

Of course, at some call centers

customer service agents are timed to determine efficiency. In a perfect world the time spent with a customer shouldn't be an element, but in the real world where time is money it is very important.

The worst problem for a caller is the reasonable presumption that the operator knows something about a product or service. That caller has visions that the telephone agent is somewhat knowledgeable about the problem. When the operator can't answer a basic question or doesn't know anything about the product or service, the caller can get very upset.



As an example, I enjoy canoeing and recently called the State of Indiana Department of Tourism for information on the Wabash River. The call was handled by what must have been an out-of-state phone service. The operator said, "I think the name has been changed, do you know what the new name is?" and I replied, "I don't think the name has been changed. It's the main river that flows through Indiana."

The bottom line is that I thought I was speaking to someone from the Department of Tourism who would be knowledgeable about Indiana, but she had never heard of the Wabash River. The Indiana state song is "On the Banks of the Wabash."

The other issue is accuracy,

which can be challenging for most live operators. Getting a name with the correct spelling, the correct street address and correct information requires accuracy. With IVR, of course, those capabilities are very limited because first the information has to be entered manually and then, if voice responses are recorded in what is called voice capture, the accuracy goes way, way down because there is no intuitive response, just a recorded response.

Voice capture is an IVR term for when the IVR machine asks for voice information, which is then subsequently transcribed, generally off cassettes. In response to a question such as "What's your name and address?" the response from a caller answering, "Joe Sixpack," would be recorded, and that recording would be voice capture. It's the answering machine side of the IVR system.

The two bottom line criteria for success are accuracy and yield. Another analysis is that if results are compared from a natural speech system with results using live operators, the natural speech system will always equal or generally exceed live performance as measured by either yield or accuracy. For example, if people call a live operator and 85 of them leave the complete information sought, a properly programmed natural speech system will do 85 percent to 90 percent (according to in-house studies at ConServIT).

The problem with voice capture is that studies have shown it is inaccurate and a lot of callers do not tolerate it. It's a problem that was experienced when the business started in the '60s and '70s. Callers gave up or gave inaccurate and incomplete information, such as neglecting to mention the ZIP



code in an address.

The hidden cost of using automated machines, which is only now being widely recognized, is that if half the people are hanging up, call centers are really not performing a service. They're doing nothing more than just getting people to hang up. Some firms now recognize that servicing that caller is critical. The arrogant concept of the '90s of "if callers want service, they have to do it the way we tell them to" is passe.

Another fundamental issue is that marketers need to know is how the various technologies differ.

Dialog design companies are often asked which word recognition or speech recognition package is being used. Clients will also ask how the package recognizes speech.

It is necessary in working with natural speech to recognize that the fundamental building block is not speech recognition as the term is used today. Natural speech and word recognition are fundamentally different. Whereas dialog designers should be experts in natural speech, they do not have to be the ultimate experts in word recognition. The word recognition experts say they can recognize words within a spoken statement. The use of discrete speech recognition enables a system to recognize "Would you send me a book" by focusing on two or three key words.

Unfortunately for the designer, that is not the way people speak. Callers may say something that sounds like, "Wuhjasemya book?" And in that expression, not one of those words recognized by the discrete engine - "would - you - send - me - a" — appears. It's a single word, pronounced "wuhjasemya." It has nothing to do with an accent or voice pitch, it has nothing to do with the speed of the

voice, and someone else might pronounce it completely different.

If the designer attempts to decipher this using word recognition alone, mistakes are made. Designers try to point the caller toward saying specific words, such as, "Give me the city," placing the caller's choice in a limited range of alternative words that can be recognized.

Natural speech recognition pushes design into another arena. When the machine says, "Can I help you?" to a caller, there is no way of knowing what the caller is going to say, what words will be used, what expressions (wuhjasemya?) would be used. If the caller places an order, the word "order" or "buy" might never be used. A caller might simply say, "I'd like a book."

A caller could use a range of words to express what he wants without using one of a system's keywords. That's where natural language recognition comes into play.

The whole process has to be customized to fit the needs of the customer. It is specifically dependent upon the application and the caller.

The application must be examined first. If the system is designed to convey to the caller flight times of particular flights, natural speech is not necessary. Prompts must be designed to direct the caller to give specific information being requested. For example, prompts such as, "Please say the number of your flight," and, "Please say the name of your airline," etc., lead the caller into providing specific information, which can be easily recognized by the system. The system, after gathering the requested information with the proper prompts, simply answers. The machine may have

access to information about 10 million flights and based on the information provided, it can pick out a particular response.

On the other hand, if the system includes as one of its prompt questions, "May I have your name and address?" the parameters are completely different. In this country, the machine would be trying to pick out one of more than 280 million people in this country, with various pronunciations, accents, style of responses, etc. In this case, natural language understanding is needed to determine the answers to the questions.

There are still situations where a live agent is required to handle a customer's request. If the caller wants to complain about a bill, or get service for a computer, there is a necessary give-and-take between the caller and the agent.

The most important thing to remember is companies should never lock the caller into an imperfect structure. Customer service is the primary key, and call centers, by the varied nature of the duties they perform, do not fit into a one-style-fits-all template. No matter what system is used to handle calls, allowances must be made for all types of callers and requests.

Sometimes it can done with one technology, but in today's world it often can be accomplished only by combining technologies. ■

---

*Peter F. Theis is the founder and president of Gurnee, IL-based Conversational Voice Technologies Corp. (ConServIT), an automated voice inbound telemarketing service, and president of Theis Research & Engineering LLC. He can be reached at (800) 343-2882 or at [conservit@conservit.com](mailto:conservit@conservit.com).*