Speech-Enabled Memory Assistants
Patti Price
Principal, PPRICE Speech and Language Technology Consulting

On the lookout for speech applications beyond dictation and 'your call is important', I am interested in applications aimed at consumers or small businesses. Since I have often wanted to upgrade my own memory as easily as that of my devices, I looked at a few speech-enabled memory assistants. Businesses, like individuals, also need memory assistance if they are to "let the right hand to know what the left hand is doing".

Interface designers for such applications face the usual dilemma of providing powerful features while keeping the interface simple. One could separate interfaces into the 'Don Norman' vs. 'Doug Engelbart' approaches. In this simplification, the 'Norman approach' is characterized as "make it obvious" and the Engelbart approach as "make it powerful". In reality Norman has argued that the goal is not simplicity, but rather making our complexity understandable. Analogously, Engelbart would argue that no interface should be *needlessly* complex.   Both would agree that power is good and understandable is good. In the real world, differences in approaches will show up in the inevitable compromises needed.

Speech interfaces are particularly problematic in that, as individuals and as a species, we have vast experience and expectations about how spoken language interactions work. To quote Norman: "develop a system that recognizes words of speech and people assume that the system has full language understanding, which is not at all the same thing."
jnd.org/dn.mss/how_might_people_interact_with_agents.html

While these expectations are real, it is also true that people have vast experience accommodating to those who differ from us. For example in talking with children, foreigners, the hearing impaired, etc., we learn to accommodate to differences in linguistic, cognitive, social and other norms. Therefore, humans can likely adapt to a new, machine-based system and might be willing to adapt more if the potential reward is larger.

Several services now transcribe voice mail, with varying degrees, of success (e.g., YouMail, CallWave, SpinVox, PhoneTag, GotVoice, Google Voice). While such services assist memory in making voice messages as searchable as text, I'm looking for more. Many of these seem to be a hybrid: automatic transcriptions reviewed by humans. This is a clever approach that bootstraps current technology while providing data for potential next generation technology. The use of keywords can trigger a reminder later and is a clever trade off of speed for accuracy. Although the pioneering "I want Sandy" closed operations December 2008, I found a few speech-enabled memory assistants currently available.

me2me.com, covered in this issue, does not transcribe the message itself, but recognizes separate voice tags used as categories.

JustKnow.com does not do any recognition, but leverages text to speech so that web information (directions, phone numbers, etc.) can be sent now or later as text or voice to devices or calendars. The service is sold to the website not the consumer. While this could help jog memory, it depends on information push from the company rather than information pull from the person who needs it.

Jott.com offers several information pull products. **Voicemail** ($9.95/month for 40 messages) transcribes speech via software with human review. The resulting text can be shared with others, searched and organized like regular email. Memory assistance can come through call-back reminders or by sharing messages. Jott's **Assistant** (under $4/month or pay per minute) focuses on memory allowing you to: get voice notes transcribed, integrate them with Outlook Tasks and other applications, send text messages by voice, set up reminders and calendar events that integrate with other calendars, connect with about 30 web applications by voice (e.g., Twitter, Facebook, Amazon), and get RSS feeds read to you. This definitely should assist personal and corporate memory and information access! Particularly interesting for businesses is Jott for **Salesforce** ($25/month including Assistant). This plan allows you, in addition, to add voice updates (with confirmation) to Salesforce accounts and opportunities, add tasks and notes, schedule appointments and set reminders. The interface is simple: "Who do you want to Jott?" You must know, of course, that the possibilities include 'notes', 'reminder', etc. Depending on the answer to the initial question, you may be prompted for the message, or for time and date and then the message.

ReQall.com is intriguing: Don Norman is "Chief Mentor" and Sunil Vemuri, the technical co-founder, did his PhD work at the MIT Media Lab on "Memory Prostheses", web.media.mit.edu/~vemuri/wwit/wwit-overview.html The free version enables adding items by voice or text, unlimited voice transcriptions, organization of items based on keywords (dates, times, buy, note, meet), categorization by time, things, people, and sharing of reminders with others. Free version users get a free month of 'Pro', which (for $25/year) adds: access to the keywords 'at home', 'at work', setting up places as keywords, reminders based on your calendar, organization based on categories you define, location awareness (if your device has GPS). Imagine you're near a hardware store you put on your list of places. You suddenly get a message with your list of hardware store items.

On the interface, Norman's website jnd.com says, "We worked hard to make it really simple, to eliminate all the features that came to mind. No features, therefore no fuss. Simple and powerful." This recalls the Onion News Network spoof of Apple allegedly releasing a laptop without a keyboard-- "just one big button --- what could be simpler?" Despite this comment, as a fan of Norman's work and Vemuri's idea of memory prostheses, I tried it out. The interface at the top level is a step more structured than Jott's, asking if you'd like to 'add', 'reQall' or 'share'. The website gives helpful examples, but it is hard at first to remember all the keywords and when things must be in a certain order. Although humans don't require that of you, once you know the rules, it feels like a reasonable accommodation. Keywords include: dates (put item on calendar), 'buy' (create list), and 'remind/tell/ask' (share with others). In my trials, about a half hour later the transcription appeared (Jott says theirs take 10-20 minutes).

I added one contact ("Gene") and recorded a few items. His name was transcribed correctly once, but more often it was not: twice 'Jean', and once each 'Jim' and 'Keane'. Although proper names are tough for humans or machines, with 'remind' a key word and 'Gene' the only contact, 'Jean' could have been automatically changed to 'Gene', and a request for verification might have been generated for 'Jim' and 'Keane'. The recognition was otherwise reasonable with some annoying exceptions, e.g., "can" misrecognized as "can't" (these are so frequently confused by humans that Noam Chomsky, the linguist, often said "can yes" for "can"). Verification, though tricky with a delay of 30 minutes or more, probably needs to be addressed. I also observed some errors hard to imagine coming from either humans or a speech system with a reasonable trigram language model making, e.g., "I need a ground cloth to sheets, a blankets, …"

In sum, Jott and ReQall and others will likely find users. Jott seems poised to address a largely untapped market in organizational tools and information access for small businesses. Their integration with other applications and with Salesforce in particular, should position them well for this. ReQall, on the other hand, aimed at consumers, could particularly appeal to aging baby boomers as memories fade. Both companies, as well as the intriguing Wolfram Alpha (blog.wolfram.com), combine automated with human effort. This is a great idea, but observed error rates can vary with the transcriber you happen to get, which complicates assessment. However, Google adds value and has been successful despite the notorious difficulty of determining search accuracy; a pragmatic solution should be possible here, too. I am hopeful that these companies can help us achieve Engelbart's vision of using technology to bootstrap our collective intelligence.