Desktop Dictation: Then and Now

By Robin Springer

Desktop dictation has changed in the 10 years I have been in the field. From available features to distribution channels, let's take a look at where we were, where we are, how we got here, and where we might expect to go.

Where We Were:

In 1994 speech recognition was more difficult to use than it is today. Installations could take upwards of four hours. The software required a dedicated sound card and did not work on notebook computers. While there was more competition among manufacturers, there were fewer programs on the market. DragonDictate, IBM VoiceType, and Kurzweil Voice for Windows were the three programs going. Kolvox Communications made a software overlay that was used with speech engines, making the interface smoother and more intuitive.

Speech recognition was used almost exclusively by people with disabilities. The legal market was minimally interested in the technology but most corporate demos were primarily comprised of people who had heard about "some computer you could talk to" and wanted to see it for themselves.

Significantly, the software was sold exclusively through the VAR channel. VARs added value to the product by demonstrating it to prospective users, corporate and end-user alike, becoming a strategic partner with the client, allowing users to construct realistic expectations as to what the technology could and could not do for them.

Was the technology perfect? No. You had to pause between words. You were tethered to the computer by a corded microphone. But people who purchased the technology had relatively high satisfaction levels. Why? Because the personal interaction allowed customers to discuss their needs with someone who was trained to understand the benefits and limitations of the software. Users were more likely to purchase training services to ensure the software was being used to the best of the abilities of the user and the capabilities of the software.

Where We Are:

According to Matt Revis, Senior Product Marketing Manager for Dictation Products at ScanSoft, 30 percent of Dragon NaturallySpeaking users are disabled, implying that the use of the software has grown in other verticals. Revis estimates approximately $16 billion are spent worldwide annually on manual transcription so speech provides an attractive alternative for health care and legal markets.

The technology does pretty much the same thing it did 10 years ago; put words on paper. We are able to speak to the computer faster and control more computer applications, with better recognition, than in the past. But some VARs have noticed a decrease in customer satisfaction. Jerry Thompson of SRT, whose company distributes speech recognition via the VAR channel, notes that end-user complaints are up.

He and many others believe the frustration experienced by end-users is in direct correlation to the proliferation of speech recognition being sold at rock-bottom prices on Internet storefronts. Using the Sales Triangle theory that there are three aspects to a sale; service, quality, and price, and at least one aspect needs to be sacrificed, what are Internet buyers giving up? They are not sacrificing price or quality.

Internet customers are forfeiting service. The service provided by VARs, while not being used as often as in the past, is more important now because there are more configuration options from which to choose. "The product has become more capable, more complex, more thorough," says Thompson. "It is logical that users need some type of teaching, learning, understanding, to make use of those capabilities."

How We Got Here:

Dr. Martin Tuori, former vice president of Development at Kolvox Communications, gives a brief history of the evolution of desktop dictation from niche product to commoditization. Tuori describes Geoffrey Moore's "Bowling Alley" theory; niche products expand into vertical markets before going mainstream. Speech recognition in the mid-1990s was making its way down the bowling lane.

Growth would have emerged into new verticals but, "Kurzweil started a price war, decreasing the engine price from $1,000 to $200, prematurely pushing desktop dictation from the bowling alley into the mainstream," says Tuori.

Tuori and others in the industry believed that every computer user would talk to their computer but, "when Kurzweil initiated the change in the market before the market was ready to receive it, we found that there was not a corresponding increase in sales volumes. Most users did not have an intensive need for talking typewriters." Desktop dictation became commoditized.

Tuori compares desktop dictation to spreadsheet programs. If a user does not know how to create formulas in a spreadsheet, he can still use the spreadsheet to manually input numbers but he is not using the program to its potential. Tuori believes the commoditization of speech has done a disservice to the industry, "like improperly using Excel but more so."

Where We May Be Going:

People using speech now have, "more knowledge of what it is versus what it might be," says Dr. Janet Baker, co-founder of Dragon Systems, but levels of research and competition have decreased, resulting in a loss of momentum in progress.

Users are often impressed when they get 98 percent recognition out-of-the-box but the real value of desktop dictation is the recognition one achieves down the road. What accuracy rates are users achieving after three months, six months, or nine months of use? If recognition is still at percent, that's impressive. If customers are not achieving these results they may return to the VAR channel or they may become more hesitant to use the technology at all.

Robin Springer is president of Computer Talk (www.comptalk.com), a consulting firm specializing in the design and implementation of speech recognition and other hands-free technology services. She can be reached at (888) 999-9161 or by e-mail at info@comptalk.com.