

Thursday, August 15, 2002



From the Desk of

David Pogue

Speech Recognition Follies

In this column a few weeks ago, I wrote about the futility of trying to predict the future of technology. I focused on the limits of miniaturization: After a certain point, computers will never be any smaller, because we still need a screen and some way to input information.

Many of you wrote to me suggesting that these are easily surmounted problems. We won't need screens on our computers, you said, because we'll all wear goggles that project an enormous virtual monitor before our eyes. We won't need keyboards anymore, either, because scientists will perfect speech-recognition software. We'll just dictate text into our computers, palmtops and watches.

Well, I have my doubts about the goggles thing. These virtual monitor glasses already exist, but you don't exactly see lines forming outside of Circuit City. I've tried these things out at trade shows, and found them to be pretty annoying compared with a nice big flat-panel screen. I'm not saying they'll never happen; I'm just saying it's not a sure thing by any means.

I will, however, bravely stick my neck out to say this: speech recognition will never replace the keyboard. Never — no matter how sophisticated software gets.

The problem isn't the accuracy of the transcription. Thanks to a nasty wrist ailment called tenosynovitis, I do most of my writing using ScanSoft's NaturallySpeaking 6, which I dearly love. I've been using this program since version 3, correcting each transcription error, thereby continually perfecting its understanding of my voice. After all these years, I get 99 percent accuracy. (I've dictated this entire column so far without a single error.)

But dictation software will never reach 100 percent, and therefore we'll always need a keyboard or stylus to correct typos (or "wordos"). Not because the software isn't good enough, but because in the English language, too many words sound alike.

I spent my first ten years out of college working as a Broadway conductor and arranger. The day I became sure that speech recognition would never replace the keyboard was during rehearsals for a show called "The Will Rogers Follies." The actress, trying out a new song, sang: "I'm filled with an aimless feeling."

I was sitting next to a stage manager who had the script open on her lap. Just for fun, I looked down to follow along with the singer — and realized that she hadn't sung "an aimless feeling" at all. What she actually sang was: "I'm filled with a nameless feeling."

When spoken at conversational speed (or sung), "an aimless" and "a nameless" sound identical, and no amount of body language or context would ever tell you which is the correct interpretation.

Nor is that the only example: over the years, I've kept a log of the "mistakes" made by my speech recognition software. These examples are hilarious, but they make a very serious point. In most of these cases, what I really said and what the computer typed out are sonically identical.

What I Said → What Was Transcribed:

- bookmark it → book market

- Motorola → motor roll a
- modem port → mode import
- a procedure → upper seizure
- and then stick it in the mail → and dense thicket in the mail
- movie clips → move eclipse
- I might add → I my dad
- inscrutable → in screw double
- hyphenate → -8
- suffocate → Suffolk 8
- a case we summarily dismissed → a case we so merrily dismissed
- or take a shower → Ortega shower
- the right or left → the writer left
- oxymoron → ax a moron
- ArialPhone guy → aerial fungi

Do you see the problem? Everybody says that speech recognition will eventually become perfect, as software becomes sophisticated enough to "understand" the context and inflection of our speech. But what about situations when even context and inflection are no help? Without reading my mind, how would even you ever determine whether I said "a case we summarily dismissed" or "a case we so merrily dismissed"? In short, how can we expect computers to understand us perfectly, when half the time we can't understand each other?

I'm guessing that keyboards will always be with us. Still, watch this space in 2030. If I'm proven wrong, I'll be the first to celebrate.

Visit David Pogue on the Web at DavidPogue.com.