Voice Activation: The Merger of Linguistics and Technology

Michelle A. Hudgins

Writer's Comment: *When my Computer Science 15 class was assigned a paper dealing with computers and any field of interest to us, I was initially at a loss. My interests lie in the use of words, of language as a means of communication—both grammatically and aesthetically; how could this relate to something so scientific and unartistic as computers? After much thought, I decided to put my linguistics minor to good use, writing about the efforts of linguists to improve the accuracy and efficiency of voice-activated computers. While researching this paper, I discovered the wide range of uses for which companies employ voice activation, and also the potential threats these systems pose for less-skilled workers in the future. Before this class, I had only thought of computers as something to type papers into, or to communicate with friends long distance; now I understand what great potential they have to change life in very fundamental ways—both for better and for worse.*

I would like to thank Professor Walters for his encouragement, without which I never would have entered this paper into competition. I would also like to thank Professor Ward for her advice that I consider a linguistics minor, which has proven to be of great benefit to me. And finally, I'd like to thank my roommates for their tolerance while I inundated them with information about this topic while composing my paper; your suffering helped this paper become what it now is.

—Michelle A. Hudgins

Instructor's comment: *Those of us who offer General Education courses are faced with the challenge of helping students become more proficient in writing effective term papers. The degree of the challenge varies with students, but happily, in the case of Michelle Hudgins, it was no challenge at all. Her paper followed all of the guidelines provided to the students, but more than that, she turned what could have been a relatively dry paper into one that is a delight to read as one learns from her research. I will use this paper as a model for future students in this class.*

—Richard F. Walters, Department of Computer Science

1. Introduction

The possibility of communicating directly with computers via normal human speech has been of great interest ever since their creation. Science fiction especially was intrigued by the potential therein: the *USS Enterprise* on the *Star Trek: The Next*

Generation television show, for example, has been controlled by voice since the beginning of the series. Science has not quite caught up to science fiction, but strong advances have been made in this field. Software packages with the ability to process and understand speech have become widespread on the computer market. Both large corporations and individual users have jumped on this "linguistic bandwagon" and are enjoying the benefits it provides over the more traditional keyboard and mouse.

My own interest in this field lies not in its present technological aspects, but in its origins. As a linguistics minor, I find it amazing that a computer has the ability to process and act upon human speech. To think that someday people will be able to address and control their computers as easily as the actors on a television show can is an incredible prospect, and well within the reach of future computer programmers and linguistic scientists.

2. The History of Voice Activation

Voice activation research began in the 1970s with the Defense Department's Advanced Research Projects Agency (DARPA) underwriting studies at Carnegie Mellon and MIT. Initially, these projects centered on word matching: "using computers to compare the energies and frequencies of speech sounds with the stored acoustic profiles of words" [BYLI93, p. 91]. This use of spectrograms, the graphic portrayal of physiological aspects of spoken language, was scrapped because pronunciation of words varies too widely from speaker to speaker. No standard pronunciation guide for the computer was possible. Next, investigators tried using phones, the basic units of human speech (for example, the word "beet" has three phones: b, i, and t), instead of entire words. This process was somewhat simpler for a computer to manage, but the seemingly infinite combinations of sounds finally became unwieldy [BYLI93, p. 91].

In 1971, however, a breakthrough occurred. James Baker, a mathematician, employed a method called "Hidden Markov Modeling," named after the Russian mathematician Andrei A. Markov. Markov had "invented a method of statistically predicting the sequence of letters as they appeared" [BYLI93, p. 91]; Baker applied this process to the problem of voice activation. Computers using this system have the ability, once having determined the first sound in a sequence, to statistically predict the sound most likely to follow. In 1986, DARPA issued grants to discover a link between voice activation and natural-language studies. The latter field centers on the understanding of grammars in human language. With this combination of phonetic and grammatical studies, voice activation finally became a viable field, with the ability to recognize even tricky statements such as: "Our last two presenters were one hour too long," with two sets of homophones [BYLI93, p. 91]. Computers can now predict not only the most likely sound to follow, but also the most likely word.

The manner in which one primes a voice-activated system depends on the software involved. One program, Personal Dictation System, requires the user to read from Mark Twain's "A Ghost Story" for 90 minutes [PICK94, p. 51]. In most systems,

however, the user must simply repeat a specific set of words three times into a microphone [ENGL93, p. 70]. For best results, the user must speak slowly, leaving a slight pause between each word, allowing the computer to process each word separately. Background noise must be kept to a minimum, lest the computer accept the noise as part of the user's voice and not work without consistent background noise. The manner in which the system registers each command is as follows:

> Most systems record your voice commands and compare them to a database of trained commands using a pattern-matching algorithm. The program calculates a score that represents how close your voice command is to each trained command and chooses the trained command with the closest score. If the score is within acceptable limits, the program initiates the macro that's associated with the trained command. If the score is outside the acceptable limits, the program alerts the user that it didn't understand the command. [ENGL93, p. 70]

For example, with SayIt, a $295 speech-recognition program, if the computer doesn't understand a command, a character named Simon appears on the screen and scratches his head in confusion until the user repeats the command more clearly [BYLI93, p. 88]. Most systems have a similar program, allowing the user to correct or repeat an unrecognized command. With programs such as these, and with the combination of Hidden Markov Modeling and natural-language systems, voice-activation software has achieved an accuracy rate of between 80 and 95 percent. Mistakes still happen, but they occur rarely enough that the system is definitely a viable enterprise.

3. Business and Personal Uses of Voice Activation

Voice activation has several advantages over the more "traditional" keyboard and mouse. In the fast-paced world of business, a user may have to do several different actions at once. Voice activation frees the user's hands for other actions, including using the keyboard and/or mouse. For example, a program called VoiceEM allows doctors to use voice-activated computers to record diagnoses, without having to physically type the words in [BYLI93, p. 88]. It also has a "built-in 'knowledge base' of medical data that prompts the doctors to check for symptoms they may have overlooked, thus improving the accuracy of diagnoses and reducing the threat of malpractice suits" [BYLI93, p. 88].

Another use for this kind of system would be for handicapped individuals without the ability to use their hands for typing on a keyboard [ENGL93, p. 70]. Voice activation need not be directly connected to a PC for it to be a productive system. Audi and BMW are working on voice-activated cellular phones, which would allow drivers to use the phone without having to take their eyes off the road and their hands off the wheel [JORG91, p. 14]. The ability to use only one's voice to activate and work with a computer could be a great boon to these segments of society.

Other businesses are using voice-activation software to their advantage. Phone companies, for instance, use these systems to interact with their customers, rather than

taking up operator time with routine questions [BYLI93, p. 89]. Also, "in the New York City suburb of Bay Shore, Long Island, latchkey kids in 400 households can pick up the phone and simply say 'Mom' to reach their mothers at work." In Boise, Idaho, customers can dial "#44," say "Messages," and retrieve their voicemail. By saying "return call," they can get the phone to dial the number of the person who had called most recently [BYLI93, p. 89]. These voice-activated systems are very user friendly, allowing people who would normally never use computers the ability to use them with enhanced ease.

4. Advantages/Disadvantages of Voice Activation

One benefit of voice activation is its ability to understand the voice patterns of a user. One need not have the best grasp of English grammar in order to use these programs; one need not even have a decent understanding of the language's inflections and pronunciations. As long as these grammatical and pronunciation mistakes are consistent, the program will work well [PICK94, p. 52]. One program, Personal Dictation System, has the ability to put in its best guess for a word, and, if it feels the word is inaccurate based on later word choice, can backtrack and substitute a more appropriate word [PICK94, p. 51]. Many programs have the ability to store the voice patterns of several users in their memories. For example, VoiceAssist, which works within Windows, allows several users—each with a distinct set of commands—more than 29,000 commands per user and makes 1,024 commands available at once [KEIZ93, p. 48]. With this new technology, even the "computer illiterate" may be able to use their PCs more productively.

There are disadvantages to voice activation, however. The systems take up a good deal of memory; for example, Personal Dictation System takes up 8 megabytes of random access memory and 32 megabytes of disk space. Also, continuous speech is generally not allowable under these systems. The user must pause briefly between each word to allow the computer to catch up and to process each command. Personal Dictation System officially accepts only 70 words per minute, although it may allow up to 110 words per minute before it lags behind the user [PICK94, 51]. Still, this is not quite up to the rate of normal human speech. One other problem may be the number of users the program supports. Some allow only one user, making them inconvenient for personal and corporate use (if more than one member of a household / company will use the computer).

Under less than optimal conditions, the program may fail to function properly. If the user has a cold or laryngitis, or if there is an excess of background noise, the system may fail to recognize the voice of the user [AYCO89, p. 8]. Also, because users tend to speak slowly and loudly to their computers, there is a danger that the workplace will drown in a cacophony of voices, making it difficult for workers to work and think productively during normal office hours [BYLI93, p. 91]. In this case at least, the keyboard and the mouse have definite advantages, being relatively silent modes of communication with the computer.

5. The Future of Voice Activation

The next step for voice-activated computers is to learn by doing, and not simply by being programmed by scientists. For example, the new wave of computers should be able to formulate their own version of human grammar simply by listening to humans talk. They should be able to add new words to their vocabularies, add new definitions to existing words, and choose the most appropriate meaning for the same phrase. Natural-language processing becomes a factor in this system; charting sentences to break them into their component parts allows the computer to store each part and to carry over that meaning from one sentence to another. For example, if one charted the sentence "The boy has left," the computer could retain the knowledge that "the boy" is a noun phrase and carry that information over to the next sentence, starting with "he," connecting the two sentences in meaning. Another concept programmers are working on is proper punctuation, based on pitch, loudness, and timing [McWI93, p. 101]. With these advances, the accuracy rate of voice-activated systems should improve dramatically. Perhaps it might be possible to have true continuous speech in the near future.

Voice activation also may have more serious consequences. Many people fear that voice-activated computers will replace human workers for jobs that could be easily automated. For example, in Japan, Toshiba is working on a machine for fast-food restaurants which will, in response to a spoken command, provide a customer with a soft drink and a hamburger [BYLI93, p. 90]. In other areas of the business world, these new computers could buy and sell stock with a simple voice command, which might also result in the loss of jobs by clerks and other customer-service workers [McWI93, p. 100]. As with any new technology, voice activation has the potential to take away less-skilled jobs from human workers.

6. Personal Assessment

The field of voice activation is just coming into its adolescence. At the present time, it is being used in a very few areas, mostly as a curiosity rather than as a fully mature area of study. It has the potential to change even the basic parts of our lives—both at work and at home. Voice-activated VCRs are on the market now, leading the way to a whole slew of voice-activated appliances for the home or office. Voice activation can also free up the user's hands for using the mouse and the keyboard—creating the equivalent of having three hands—which would greatly increase productivity. Not everything is possible with these systems, however. They have the same limitations on commands that a keyboard would. It is unfair, for example, to expect them to perform tasks that no computer could do, with or without voice activation.

With these new systems, however, learning to use a computer would become a less daunting task for the "computer illiterate." Instead of having to memorize commands designed by "computer nerds," users could simply tell the computer what to do, using commands of their own choosing. This technology also has the capacity to greatly improve the lives of the physically disabled. If one cannot use one's hands, voice

activation can allow one a great deal more freedom of career and of movement—voice-activated computers for the office, voice-activated doors for the home, etc.

Not everything associated with voice activation is positive. True, voice activation can allow careers for a segment of society that would have been nigh impossible in the past, but the prospect of putting a large portion of society out of work is a highly unpleasant one. There are also potential abuses involved in using this system. One possibility is that a disgruntled ex-employee, whose voice is still registered in the computer, could yell out, "Format . . . hard . . . drive!" thus wreaking havoc upon his/her former employer. A problem may also arise in allowing workers privacy. Speaking to a computer in a communal working space is not only annoying to co-workers, but may also be troublesome to the person speaking, who may feel that his/her every action is being overheard.

7. Conclusion

Voice activation as a field is not yet being used to its full potential. It fails still to recognize true continuous speech, making working with it seem like working with a rather slow and dull-witted child. It has an accuracy rate of between 80 and 95 percent, which is good, but needs to improve, especially if the technology is going to be used in the medical field, where a miscommunication between doctor and computer could be deadly for the patient. All in all, however, this field has the potential to change the world as we now know it. Whether it will live up to that potential is up to the skill of scientists and the desires of the consumer.

8. References

AYCO89 Aycock, Heidi. "Computer: Take a Letter." *Compute!* 11, (May 1989): 8.

BYLI93 Bylinsky, Gene. "At Last! Computers You Can Talk to." *Fortune* 127 (May 3, 1993): 88.

ENGL93 English, David. "Computer, Status Report." *Compute!* 15 (Sept 1993): 70.

JORG91 Jorgensen, Barbara. "Voice Activation Goes Mobile." *Electronic Business* 17, 1 (Jan 7, 1991): 14.

KEIZ93 Keizer, Gregg. "$99 VoiceAssist Listens to You." *PC Magazine* 12 (Nov 23, 1993): 48.

McWI93 McWilliams, Gary. "Computers Are Finally Learning to Listen." *Business Week* 3343 (Nov 1, 1993): 100.

PICK94 Pickering, Wendy. "Computer: Take a Memo." *Datamation* 40 (Jan 7, 1994): 51.