

Computer Speech Recognition as an Assistive Device for Deaf and Hard of Hearing People

Joseph Robison

Carl Jensema

Institute for Disabilities Research and Training, Inc.
Silver Spring, Maryland

Introduction

Technology challenges us to keep up with it, adapt to it, and grow with it. This may seem to be an overwhelming challenge, but the benefits are too far-reaching to ignore. One technology that is growing at an extremely rapid pace is computer speech recognition. Developed as a dictation tool for business applications, computer speech recognition will eventually have many applications for deaf and hard of hearing people, but most of these applications are still years away. One area where it has several immediate applications is interpreting. The technology has advanced such that it can be used by sign language interpreters where the usual interpreting process encounters problems. As the speed and accuracy of computer speech recognition improves, it is likely to become a standard interpreting tool.

The Institute for Disabilities Research and Training, Inc. (IDRT) is currently involved in a three-year U.S. Department of Education grant to study how speech recognition can be used by sign language interpreters as an assistive tool to provide more complete interpreting for deaf and hard of hearing (D/HH) students in mainstream classes. Sign language is, and always will be, an effective means of communicating information to D/HH students. However, computer speech recognition can provide a useful communication tool in certain circumstances.

When the Interpreting Process Breaks Down

Sign language is an extremely effective means of communication for D/HH people in most cases. A good interpreter can keep pace with normal rates of speech accurately while providing the D/HH person with critical facial and body gestures needed to convey the speaker's emotion. However, American Sign Language (ASL) contains roughly 5,000 signs while a typical abridged English language dictionary contains about 80,000 words and speech recognition dictionaries contain up to 160,000 words. There are an estimated 500,000 words in the English language. The lack of an extensive sign vocabulary does not normally present a problem in daily conversation because people commonly use only a few thousand words of their vocabulary. For example, Jensema and McCann analyzed the captioned text of 183 television programs. Because of the volume of words collected in this study, it is an accurate representation of language spoken on a daily basis. Of well over 800,000 words in the captions, only 16,000 unique words were used. Of these, just 250 words accounted for two-thirds of all 800,000 in the text (Jensema & McCann, in press).

Extensive vocabulary is therefore not a problem in signing most communications. The estimated 5,000 signs, supplemented with fingerspelling, is adequate for most situations. The problem comes when technical or complicated vocabulary is used. Many high school and college-level courses contain complicated vocabulary for which there are no signs. Interpreters may fingerspell many of these words, but how many interpreters can correctly fingerspell the names of countries and places like Czechoslovakia, Uzbekistan, Kfar Ezion and Kealakekua Bay, or names of world leaders such as Gamal Abdel-Naser and Binyamin Netanyahu? The following words are taken from a partial list of specialized vocabulary used in a 46-minute high school anatomy and physiology class. These are words that would have to be fingerspelled or represented with made-up signs:

astrocytes	epidermal cells	ventricle
microglia	meiboid	psuedopod
dendrites	ligodendroglia	paraplegic
quadriplegic	myalin sheath	oligodendroglia cell
mitochondria	adipose	node of Ranvier

Fingerspelling words such as these slows down the interpreting process while potentially creating confusion if the interpreter or student is not familiar with the correct spelling. Made-up signs can be used to communicate this vocabulary, but the D/HH student can encounter difficulties when different interpreters are used.

Foreign languages also present a problem for a sign language interpreter and student. Even if an interpreter is fluent in the language being studied, translating the target language to signed English will not benefit the student. The only alternatives are fingerspelling everything, writing all instructions and exercises on the chalkboard, learning to lip-read the foreign language, or tutoring the student on an individual basis. However, none of these alternatives is practical. Fingerspelling everything in a foreign language is too slow and difficult. Writing everything on the chalkboard is too time-consuming for the teacher and class. Lip-reading is difficult and cannot be perfectly mastered. Individual tutoring is possible, but this defeats the goal of providing equal access to the classroom for D/HH students.

Many foreign-language classrooms now focus on using a conversational mode from the very beginning. It is critical for all students to interact in the target language to develop their language skills. Many times oral exercises are not written on the chalkboard because of time limits. Fifty minutes a day does not give an instructor much time to review old lessons, teach a new lesson, and focus on conversational, listening, and writing skills. A D/HH student who does not have access to oral classroom activities is not only denied important developmental exercises but also their right to equal access to education. It is also quickly becoming the norm for colleges and universities to require students to have completed two or more years of high school foreign language courses. This increasing emphasis on foreign languages makes equal classroom accessibility more important than ever for D/HH students.

Speech Recognition Development

Research and development for speech recognition has been going on for four decades, since Davis, Biddulph, and Balashek at AT&T's Bell Laboratories began doing research on a machine capable of understanding isolated spoken digits in 1952 (Davis, Biddulph, & Balashek 1952). From this early work, speech recognition research has expanded and is now a worldwide effort being pursued on many different fronts.

Although computers are not yet capable of understanding speech in the manner of HAL in the movie *2001 - A Space Odyssey*, or the robots in the *Star Wars* series, much has been accomplished in the area of speech recognition. There are a number of systems on the market which can be trained to understand the speech of a specific user with better than 90 percent accuracy at speeds of about 65 words per minute.

The "Holy Grail" of computer speech recognition is a system which will understand continuous speech spoken by anyone at normal conversational speeds of 120 to 140 words per minute. Although this has not yet been realized, advancing technology and an increasingly competitive speech recognition market are moving this goal toward reality. Currently, computers must be trained to understand the speech of each individual user, and users are required use "discrete speech" when dictating into the computer. Discrete speech means that the speaker must pause briefly (approximately one-tenth of a second) between each word. The computer needs this pause for two reasons: to have time to analyze the input, and to prevent the acoustic patterns of each word from overlapping and distorting the word boundaries (Markowitz, 1996). Until continuous speech is mastered, the computer must be able to identify the beginning and ending of each word to recognize it. Although this slows the user's rate of speech, dictation speeds of 65-70 words a minute with a 90-95% accuracy rate are possible and continue to improve as computers become faster and speech recognition programs improve.

Leading commercial speech recognition systems such as Dragon Dictate, Kurzweil, and IBM are "speaker dependent" systems. This means that every user must create a voice file which is based on his or her particular speech patterns. Users begin with voice files copied from a basic voice template. Voice files are modified as the system learns more about the user's unique voice characteristics. The more the voice files are used, the more likely they are to fit the particular user, and the more accurate the speech recognition process is likely to be. The process of building a voice file is essential in achieving high word recognition accuracy. Fortunately, the development of faster computers and newer versions of speech recognition systems is reducing the time needed to build accurate voice files.

The basic principles of a speech recognition system can be made to fit most any language. Several of the best-known speech recognition systems are available in a variety of languages. For example, Dragon Dictate is available in U.S. English, U.K. English, French, German, Italian, Spanish, Latin-American Spanish, and Swedish. The IBM system is available in U.S. English, U.K. English, Spanish, French, German, Italian, and Arabic.

Speech Recognition in the Classroom

For speech recognition to be used as an interpreting tool in the classroom, it must be unobtrusive and mobile. Because system operators sometimes need to speak while the teacher is speaking, they must dictate at low enough voice level to avoid distracting the teacher and students. Most speech recognition systems have various settings which allow the operator to adjust the microphone volume and sensitivity levels. It is equally essential that the background noise level of the classroom does not interrupt the word recognition process. Excessive background noise such as laughter or slamming books can cause the program to hear phantom words or distort words which are being dictated. Most speech recognition programs also provide settings to adjust to the amount of background noise. These options give the speech recognition operator flexibility to customize the system according to the class dynamics.

As an experiment, IDRT set up speech recognition in an advanced European history class in a local high school. The class included two deaf students who shared a sign language interpreter. The class lectures contained many long and complicated European names, many of which the interpreter did not know how to spell correctly. The speech recognition system could not keep pace with the class lecture, but it was able to retrieve the difficult names with little effort. Both the students and the interpreter began to use the speech recognition screen to see how to spell certain words. It was found that speech recognition was useful as a vocabulary reinforcer in this particular situation.

IDRT spent three months using speech recognition to help interpret for a deaf high school student taking second-year Spanish. He had earned a B in Spanish I, but was falling behind quickly in the second year. His interpreter signed for him when the teacher spoke English but could not help him when Spanish was spoken. IDRT put a personal computer with Spanish speech recognition on a small cart and set it up every day next to the student's desk where the student could see the screen easily during Spanish class. A Spanish-speaking person who had trained on the speech recognition system sat next to the student and took notes in Spanish with speech recognition. Reading the screen was made easier with boldfaced and enlarged fonts. At the end of class the generated file was saved and later printed out to create a hard copy. This procedure was very helpful to the deaf student and provided much information he would otherwise have missed.

Learning to Use Speech Recognition

Learning to operate speech recognition systems is fairly simple, and an extensive background in computer operation is not necessary. Dragon Dictate, Kurzweil, and IBM provide interactive training programs to help new users establish their voice file and learn how to use the system. Once the user has created a voice file, some voice training is required. Voice training is simply dictating into a word processing program, correcting all incorrectly recognized words, and storing the corrected data in the voice file. The computer's recognition improves as more data is entered, and the voice file becomes a more accurate representation of the user's voice.

To facilitate the process of building a voice file, IDRT has developed a workbook to train the computer to recognize the 3,000 most commonly used words in the English language. This list covers the

majority of words used in daily speech. After about 15 hours of voice training, most users can dictate 65-70 words per minute with 90-95% recognition accuracy.

The key to accurate speech recognition is consistency. Users must consistently correct mistakes and pronounce words exactly the same way. If mis-recognized words (i.e. user says "can" and the computer hears "and") are not corrected, the user's voice files will become corrupted and there will be a loss of recognition accuracy. Similarly, if consistent pronunciation is not maintained while dictating, speech recognition accuracy will decline. The computer does not care if a word is pronounced in an unusual way, as long as it is pronounced exactly that way every time.

The Future of Speech Recognition and Sign Language Interpreting

Speech recognition computer technology continues to improve at a rapid pace. Computer companies are extremely eager to develop a system which can be trained to understand continuous speech. They are close to achieving this goal. Such systems are likely to be on the market by the year 2000.

Speech recognition currently has many different applications, and more will be added as accuracy and speed improves. Current applications are perhaps broader than most people realize. Medical transcriptionists are reducing the common risk of repetitive motion injury by using speech recognition to enter medical records, while many business people are using it to increase productivity by dictating their own documents. People with physical disabilities who normally cannot operate a computer with their hands can now do so with speech recognition. Among other things, this opens up the increasingly resourceful world of the Internet to severely disabled people. Furthermore, in the field of deaf education, speech recognition currently provides a special interpreting tool to help D/HH students gain greater access to mainstream classrooms.

In the future, the role of computer speech recognition in interpreting will grow as the speed and accuracy of the systems improve. Far from making interpreters obsolete in the foreseeable future, computer speech recognition is more likely to expand and enhance the role played by interpreters in communication involving deaf and hard of hearing people. It represents a new tool to be mastered and applied, and those who do so will open new markets for their services, especially among the great mass of late-deafened people who never learned sign language.

References

Davis, L., Biddulph, R., & Balashek, S. (1952). Automatic recognition of spoken digits. The Journal of the Acoustic Society of America, 24(6), 637-642.

Jensema, C. J., & McCann, R. (in press). Presentation speed and vocabulary in closed captioned television. American Annals of the Deaf.

Markowitz, J. A. (1996). Using speech recognition. Upper Saddle River, NJ: Prentice Hall PTR