

Further Advances in Real-Time Voice to Text with Steno Interpreters

JONATHAN E. OAKEY

Director Deaf and Hearing Impaired Program
RAPIDTEXT, Inc.
Newport Beach, CA 92660

Abstract. Microcomputer steno-interpreting technology allows the deaf to receive full access to the spoken word. The technology also benefits a variety of disabled people including the dyslexics, brain damaged people, and the deaf-blind. The Rapidtext software on a PC computer translates the stenowriter's output into English words and displays on a computer screen. The spoken English words are converted into text at up to 200 words per minute and displayed on computer screen. The technology also helps international audiences, people speaking English as a second language, and even illiterate people who speak English as their native language.

Needs Justification

Deaf students face communications barriers to completing their university studies and doctoral programs. The deaf cannot relax their eyes while the professors are speaking. Lipreaders cannot easily simultaneously maintain lipreading while looking at the chart, diagram, or poster being illustrated by the teacher. The same problem applies to viewing sign language interpreters. Neither can the deaf look down to write their notes with the same ease as hearing students. Neither can they relax their eyes as a prelude to meditating on important concepts being verbally illustrated by the teacher. If they blink, they may lose a crucial word, or an important phrase illustrating a concept. Lipreaders and signers cannot quickly scan the classroom for other details and still retain mastery of receiving the spoken word. Concentrated eye focus on a moving and floating two inch target (either mouth or hand) at a distance of ten to a hundred feet can be quite taxing to the brain. The level of eye focusing and mental effort is similiar to target competition. That effort is simultaneously superimposed on learning sometimes difficult new information and concepts. The eyes become physically tired after concentrated use as unblinking receptors of information. Hearing students, on the other hand, rely on both their eyes and ears to simultaneously receive doubled sensory inputs of the same information. Deaf people are a visual people. A solution must be found to allow them easy and relaxed access to information. The senses of smell, touch, taste are not suitable for rapid processing of abstract concepts for the deaf. Visual presentation of information is still necessary. Reading is easier than reading lips or signing (especially on abstract or complex subject matter), and happily, reading rates are higher than speaking rates. The solution is found by stretching the spoken information into the past, and

simultaneously presenting past and present information together within one format. A way has been found to convert spoken information into written information as the words are being said. The solution is analogous to reading a book as it is being typed. The deaf student receives every word accurately and does not miss any words. The deaf student can go back a minute in time to mull over concepts, and relink them to concepts being presented. The student is looking at a full page of the spoken information. He is also free to wander his eyes over to the charts and diagrams, and observe the teacher's expressions and emphasis on key words, and then go back to reread words he may have missed. He is also freed to write notes. He is freed to think. If he is allowed to have control over the keyboard, he can make use of the scrolling function of a word processing program. He can scroll back into the past 10 or 60 minutes, or more.

The dyslexics can be helped by doubling/repeating the same information input through more than one sense. The dyslexics are both hearing and seeing the written information. They may be able to gain increased understanding of the material being presented. A sensory brain retraining effect may be in play to lessen their dyslexity. The deaf-blind is also helped by using this technology combined with screen output braille reader. Normally, the reader machine is plugged into a pc computer. A blind person 'reads' the 'contents' on the monitor screen through his fingers resting on the tactile reader machine. As the steno-interpreter translates the speech into text, the braille reader outputs the text into braille form for the deaf-blind to read. However, the maximum braille output on some tactile readers may be 60 wpm, even though the average speed of both the speaking and steno input is 160 wpm. When the class session is over, the deaf-blind can continue 'reading the class' at leisure, or save it and attend the next class session. Even this slow braille output is better than relying on memory alone. Relying on a tape recorder for recording notes is useless for a deaf-blind. The brain-injured may have short term memory loss. These brain-injured are helped by allowing them to go back in time to refresh their memory and to reconnect their lost thoughts while the speaker is speaking. Depending on the speaker's speed, and the formatting of the text, the visual display allows people to go back up to two minutes into the past without scrolling. If text scrolling is controlled by the reader, more spoken minutes deeper in the past is accessible to the brain damaged listener. This micro-computer based steno-interpreting technology may also open up a new communications channel to bypass other types of brain deficits (defective single or multiple sensory input/information processing) communications channel.

1. Background Review

The average speaking rate of speakers is 160 words per minute. Some speakers speak up to 300 words per minute. Typists, at 60 words per minute, are clearly inadequate to transcribe words being spoken in real time. Stenography offers the solution through phonetic representation. Stenographers use the stenewriter, a phonetic based machine. Fewer strokes are required to represent a word or phrase. The concept is similar to using macros on the computer to shorten the number of strokes on the keyboard to achieve the same results. The stenomachine's keyboard have fewer number of keys than the typewriter, and the keys are larger. However, the problem is

that the output is in steno, which is readable only to trained stenographers. Stenographers laboriously retype into English on an ordinary computer. One hour of live proceedings requires a minimum of an additional five to six hours of typing off-site to make a transcript.

2. Review Solution

The stenomachine keys were wired to send steno signals to a personal computer. Software was written to convert the steno into English words. This is analogous to a computer recognizing the keystrokes on a computer keyboard as representations of letters. This allows for realtime translation of the spoken word into written English words. The entire written text is free of any spelling mistakes, because only complete words appear on the computer screen. However, a stenographer may steno stroke the phonetic equivalent of an English word not yet stored in computer's memory. These visible steno code are called "untranslates". In other words, the untranslate appears whenever the stenographer has not yet entered that word into memory. This memory of stored English words "translates" is called "dictionary", and is a personal dictionary created on disk by the stenographer.

An experienced steno interpreter (my coined terminology to refer to those stenographers trained and equipped with realtime steno translation software and works in a similiar capacity as sign language interpreters for deaf people) has at least a 50,000 general word dictionary stored in memory.

3. Technical Information

There is a less than a quarter second electromechanical lag from pressing a steno stroke to English translation on the screen on AT 12 mhz speed computers. It executes faster on higher speed computers. The desirable requirements of a computer are a 486 processor with 40 MB hard drive, DOS 5.0, and a wordprocessing software. A laptop or notebook size computer is suggested for ease of transportation from location to location. For international conventions, the conduct of international scientific/technical/business meetings increases participation and better insightful understanding when the common spoken language among international participants, English, is simultaneously viewed in written form keeping pace along with the speaker.

4. Further Advances

Rapidtext's Rapid Caption software allows the steno-interpreter to send text to a variety of devices, including closed caption encoders, character generators, and the InfoSign, described below (Section 5). Caption encoders and character generators are used to caption a variety of video output devices such those found in television broadcast stations, televisions, videocassette recorders, and RGB video projectors.

The program allows steno-interpreters to control the devices and the format of output directly from the stenewriter. Multiple on-line and off-line scripts can be loaded into memory and accessed instantly as needed. Scripts can be saved as text or printed either in line-by-line mode, or in paragraph form. The program can interface with alternate captioning device, such as a teleprompter, allowing the operator to switch between its output and stenographic input. A teleprompter is usually a transparent (to the television camera) screen of prepared text read live by a television anchorman as he looks straight into the camera. A reporter would write the text of his speech into his computer. The text is then loaded into the teleprompter in advance of the reporter appearing live in front of television audiences. Most teleprompters in the USA have now been converted to feed into a caption encoder essentially instantaneously to be transmitted with the rest of the live video-signal being transmitted from the television station. In this way, the news are captioned simply by feeding the reporter's written teleprompter text into the caption encoder. Teleprompter captioning has now become a standard practice in the USA. The problem is that live breaking news, such as a disaster are ad-libbed. The remote reporters reporting into the station live through a satellite feed are also ad-libbing. Live interviews of subjects, or their comments are not pre-written into the script. Jokes and comments between anchorpeople are also ad-libbed and thus goes uncaptioned. So the deaf people are getting essentially the day's old news captioned, but immediate news are uncaptioned. Rapidtext's software has been made to address this problem by allowing teleprompter and live captioning to blend together seamlessly.

Macro recording language allows users to record often used key sequences and play them back with a keypress from the keyboard or steno machine. The program also takes advantage of additional memory found in 386, 486, and Pentium machines. The program can also control two captioning devices simultaneously, and one of these devices can attach to the computer's parallel port (optional converter box required). This is a useful feature, if you wish to have an audience to view the InfoSign, while another audience is viewing a projected video picture with live captions being fed into the video picture.

Both of Rapidtext software now uses windowed interface with on-line help. Dictionaries can be edited on-line. Steno-interpreter can change translation dictionaries in use instantly. This is useful in the case of a speaker speaking about physics using physics terminology, and switching to medical jargon in mid-speech. The steno interpreter can switch from her physics dictionary to her medical dictionary in mid-stream. Access has been improved to decoder functions under the new FCC specification. Macro recording language allows users to record often used key sequences and play them back with a keypress from the keyboard or steno machine. This is useful in the case of a television series where a phrase or a the theme song is repeated on every broadcast.

Developments currently under way include full compatibility with memory intensive, DOS graphical interface applications like WordPerfect 6.0. This reduces need for screen-enhancement software. A feature that allows the steno-interpreter to transcribe proceedings and the user (usually a student) to annotate simultaneously on the same computer without the overhead of multitasking software.

5. LED Display

There has been problems with displaying the text appropriately. The full potential of steno-interpreting technology is not reached with only a computer screen placed on the student's desk. Focusing the eyes from close-up to a distance repeatedly cuts down on the student's understanding and comprehension. The author invented a solution. The seed of the idea was found in a LED display commonly seen in bank lobbies. The words are typically a moving display from right to left. The display is usually only one line. Such display is unsuitable for absorbing complex concepts, and is used only to present simple messages. The author's idea was to convert the LED display into 2 or more lines of text, and made it scroll upwards, for easy reading. The software that is normally used for captioning broadcast, professional, and amateur videotapes was easily converted to be used with LED display.

Captioning creates opportunities that once were almost impossible for the deaf and hearing impaired. You may talk to the deaf author by voice at RapidText™, Inc., where he is the director of the Deaf and Hearing Impaired Program, 230 Newport Center Drive, Suite 250, Newport Beach, California 92660, USA. (714) 644-6500 or by fax (714) 644-5706.

BIOGRAPHY

The author became profoundly and prelingually deaf in infancy from spinal meningitis. He received special training until 4th grade, when he was mainstreamed into regular school. Graduated from regular high school as the top outstanding versatile senior of all area high schools. Majored in Physics at the University of California, and graduated with a degree in Aerodynamics in 1984. He became involved with pc's in 1981. He worked as a director creating a new division based on the author's vision for Climax Computers, a mini-computer manufacturer with PICK operating system, and laser disk photographic storage and retrieval applications. In 1991, he became director and creator of the deaf and hearing impaired program at RapidText™, a stenographic based computer software and peripheral firm.