Speech-To-Text Services
January 2013

Audience:
Administrators (AEA/LEA);
IEP Teams, including Parents

For situations where a sign language interpreter is not available, speech-to-text services may be a viable option. In order for this to be a successful option, the student "should have reading skills at least at a 4th grade level. Students in 5th or 6th grade and above typically have the note-use skills, or are developing these skills, to make full use of both aspects of the system - the communication access during class, and the notes after class. Students younger than this, with the appropriate reading level, typically benefit from the communication access more than from the notes." (Retrieved from http://typewell.com/adminfaq.html). The student's IEP team would be responsible for determining the appropriateness of this technology.

What it is:

Speech-to-text technology delivers spoken information, such as lectures and classes, as text on a computer screen in real time or as printed documents.

How it works:

A speech-to-text provider listens to a speaker and using an input device produces text on a computer screen for the student to read. The text appears simultaneously (in real time) on the computer screen used by the student who is deaf or hard-of-hearing. The text may also be projected onto a screen in a classroom to be accessible for all students in a particular educational environment.

Students Using Speech-to-text Services:

This service should be recommended by an IEP team as an appropriate accommodation to meet the individual needs of an individual student. Typical profiles of students who may successfully use speech-to-text include:

> ➢ students who read fluently and comprehend text with a minimum of a 4th grade reading level,
> ➢ those who lack sufficient sign language skills to benefit from sign language interpreting,
> ➢ students who have limited auditory access to classroom instruction and class discussions, and
> ➢ students who are hard of hearing or late-deafened.

Choice of Services:

a) ***Steno Based Communication Access Realtime Translation (CART)***

CART provides a verbatim output, much like court reporting. It requires specialized stenography equipment and at least two years of training. A steno machine and specialized software are needed to create a real-time text display on a laptop

computer or other display monitor. A print or electronic transcript of the text can be provided to the student after class. For more articles regarding CART: http://www.ncra.org/Membership/content.cfm?ItemNumber=9079&navItemNumber=11438 .

Remote CART transcription services may be needed if a provider isn't available. The audio feed of the teacher's instruction would be transmitted by use of a microphone used by the teacher which would be transmitted through a phone line to the provider. The instantaneous translation of what the CART reporter hears would then be transmitted to a student's laptop.

This can be provided on-site or from a remote site. Search for "remote speech-to-text services."

b) *Laptop-to-Laptop (Text Interpreting)*
This involves using a laptop computer and specialized software to create a real-time display on the same laptop or a second laptop connected by a cable or a wireless card. This process is referred to as text interpreting as it provides a meaning-for-meaning text rather than word for word. There are two types of laptop-to-laptop software programs currently being used in educational environments. Both produce similar outputs. The input differs and is either phonetic-based or spelling-based. These include:

1. **C-Print:**
The provider uses a standard keyboard and phonetic-based abbreviation system to deliver a meaning-for-meaning text display. There is an online training in the abbreviation system and text-condensing strategies (See http://www.ntid.rit.edu/cprint/ and Additional Information, below).

2. **TypeWell:**
The provider uses a standard keyboard and spelling- based system to deliver a meaning-for-meaning text display. Online training is available (See http://typewell.com/ and Additional Information, below).

Text interpreting can also be provided remotely, for example, in case of interpreter absence. Search for "remote speech-to-text services."

c) *Automatic Speech Recognition (Voice to Text)*
A speech-to-text service provider creates a real time text display using speech recognition software. The provider speaks into a microphone and the software translates the speech into text. For example, Dragon Naturally Speaking is the name of one such software program.

Additional Information:

- *Different speech-to-text services* What is speech- to- text? What is the difference between CART, C-Print, and TypeWell? http://www.pepnet.org/resources/faq04
- *Speech to Text Services: An Overview of Real-Time Captioning* http://www.pepnet.org/resources/speech-to-text

- *C-Print - Online Speech-to-Text Training*
  http://www.ntid.rit.edu/CPrint/captionist_online_training.php
  This distance education program is designed to give individuals the core preparation for providing speech-to-text services. The training incorporates a variety of topics that are essential for promoting success, not only for the captionist, but also for the client receiving services. The skill-building portion of the program includes training in a newly modified C-Print abbreviation system, condensing strategies, preparing real-time text and notes, and in using voice with automatic speech recognition to input text.

  Recommended skills for successful trainees include:
  - Typing speed of at least 60 words per minute
  - Awareness of phonetics
  - Excellent listening and English skills
  - Ability to process spoken information
  - Experience with computers and word processing applications
  - A genuine desire to learn

  To purchase the online training, see the website.

- *National Court Reporters Association*: http://www.ncra.org/

- *PEPNet*, *A Guide to Speech-to-Text Services in the Postsecondary Environment*, July 2005: http://www.pepnet.org/resources/guidespeechtextservices

- *TypeWell -* Online Speech-to-Text Training http://typewell.com
  This online training is provided in a distance learning course. The course teaches the trainee how to listen to lectures and discussions and rapidly type a meaning-for-meaning transcript using speed-typing techniques with the TypeWell software.

  Transcribers should gain at least 50 hours of solid experience providing communication access services before providing advanced services like transcribing math, science, or remote lectures.

  The 29 lessons in the training course take most trainees between 35 and 60 hours to complete. Most lessons take between 1 to 3 hours, depending on the material covered and the individual trainee's learning rate. The course can be completed in as few as 18 days of nearly full-time work, once the course start date arrives. It is more usually completed in 10 - 12 weeks of part-time work. To be accepted for training, a trainee must:
  - Type at least 55 words per minute, with no errors, as measured by our typing test (follow the *Registering* link above for details).
  - Have good listening and English skills (able to understand and restate complex English quickly and clearly).
  - Have basic computer knowledge.
  - Have no history of pain in the arms or wrist that might suggest a tendency toward repetitive motion disorders. This is not a requirement but is a strong recommendation.

In addition to the *training* qualifications listed above, a *transcriber* should have the characteristics listed below. Look for them when you check references and interview potential trainees.

- o A quick mind and good problem-solving skills.
- o Comfort working in a school setting.
- o Demonstrated professionalism and flexibility.
- o Ability to work as part of an educational team.
- o Experience at the school level in which he or she will be working (e.g., some personal college experience is desirable for transcribers who will work at the college level).

- *Comparison Chart of Speech-to-Text Systems (from the Fall 2008 PEPNet Newsletter, no longer available)*

	Description	Equipment	Training	Verbosity (pages/hour)
CART	Provider types nearly every spoken word. This steno-based approach is described as verbatim.	Eight-key steno machine is used to record court proceedings and specialized software is used to create a real-time text display.	Requires two to four years. Associate or Baccalaureate degrees offered. List of approved reporting programs at the National Court Reporters Association website.	15 - 20
Note: In Meaning- for- Meaning Systems (e.g., C-Print® and TypeWell®) the service provider uses a laptop computer and proprietary software to create a real-time text display on the same or a second laptop. The laptops are connected via cable or wirelessly. The service provider types only the meaning of the spoken words; repetitions, interjections and other extraneous material is ignored.				
C-Print®	Provider interprets (i.e. listens for content) what the speaker says and using a standard keyboard and phonetic-based abbreviation system delivers a meaning-for-meaning text display.	C-Print® software is used to produce text on a laptop computer or other display monitor.	Requires an initial (approximately) 60 hours of training in the abbreviation system and text-condensing strategies. Additional training is recommended for the provider to increase real-time skills before entering the classroom.	6 - 10
TypeWell®	Provider interprets (i.e. listens for content) what the instructor says and using a standard keyboard and spelling-based abbreviation system delivers a meaning-for-meaning text display.	Laptop computer is used with TypeWell® software	Requires an initial (approximately) 60 hours of training in the abbreviation system and text-condensing strategies. Additional training is recommended for the provider to increase real-time skills before entering the classroom.	6 - 10

Speech Recognition	A service option that enables a speech-to-text service provider to create a real-time text display using speech recognition software. The process often is described as voice-to-text.	Provider speaks into a microphone and the software translates the speech into text. A computer keyboard may be utilized during this process to change a word or phrase not recognized by the software or recognized incorrectly.	Requires extended voice training to achieve a high rate of accuracy. While the initial voice training may take anywhere from five to 30 minutes, it takes significantly more time to ensure quality output.	15 - 20

Same Speech, Three Different Transcripts

(Excerpt taken from PEPNet 2008 Conference presentation by Cindy Camp, Judy Colwell and Pam Francis titled, "Meaning-for-Meaning in Speech-to-Text Services: A Better Understanding.")

Verbatim
Here is a cross section of the Great Pyramid. One of the things they did, is you have all of this work going on here. Here is a chamber cut into the bedrock. All the way down here. What people can't figure out is: How did they get the fresh air -- this is like 300 feet or more down this shaft into this big chamber that was cut into the bed-rock. How did they get fresh air down in there? Because you are going to have people down there working, breathing, and exhaling. After a while, they would use up all the oxygen. So, how do they get fresh oxygen down there for those people. That is one problem.

Meaning-for-Meaning
Here is a cross section of the Great Pyramid. One of the things they did, you have all this work going on, a chamber built into the bedrock. How did they get the fresh air 300 feet or more down this shaft into this chamber? Because you will have people down there working, breathing and exhaling, and after a while use up all the oxygen.

Meaning-for-Meaning
Here is a cross section of the Great Pyramid. There is all this work going on. There is a chamber cut into the bedrock. How did they get fresh air in there? People are breathing and exhaling and they would use all the oxygen.