



# Realtime Speech Recognition Captioning

Computer Prompting & Captioning Co.

GSA #GS-35F-4672G

**A Certified 8(a) Company**

## CPC-500 VoiceCaptioner™

### Features

- Converts speech to captions in real time
- Helps comply w/ FCC, ADA & Section 508 requirements
- English and Spanish versions available
- Caption encoding is fully automated using time code
- (CPC-700, CPC-700NL, and CPC-700NLE)

The VoiceCaptioner is a cost-effective means of live captioning when good, but not great accuracy is required. It uses speech recognition software to take the place of a highly skilled and highly paid realtime stenographer.

### Typical Uses

- Television sports and weather
- Classrooms
- Meetings

### Frequently Asked Questions

#### **How accurate is it?**

Approximately 90% to 95% accurate after training.

#### **Is this level of accuracy high enough to satisfy FCC, ADA and Section 508 regulations?**

We believe the answer is yes. Please check with your legal department for a definitive answer to this question.

#### **Can the software recognize the speech of a number of people, such as the rapid banter back and forth between newscasters?**

No. It can only recognize the speech of the person whose speech model is currently loaded into the software. It takes about 30 seconds to load a new person's speech model into the software, and a large number of speech models can reside in the computer's memory.

#### **How can the software be used to caption a program in which there are many people speaking?**

The answer is through an off-camera person who previously trained the speech recognition software to their speech, and who repeats what the on-camera people are saying. This person is known in the industry as a "shadow speaker."

It takes a while to develop the ability to act as a shadow speaker, but it can be learned. For example, at the United Nations, simultaneous translators not only repeat what diplomats say, but also simultaneously translate their speech into another language. The shadow speaker only repeats what the talent is saying, and does not have to simultaneously translate it into another language.

### Ideal Realtime Package

The CPC-500 VoiceCaptioner combined with ViaVoice and the Link PCE-845D encoder is our most popular combination for realtime speech recognition based captioning.

The PCE-845D is a very reliable (it has a 10 year warranty) and cost-effective closed caption encoder.

### System Requirements

A notebook or desktop Windows computer running at 1 GHz with at least 512 MB of memory.

### Prices

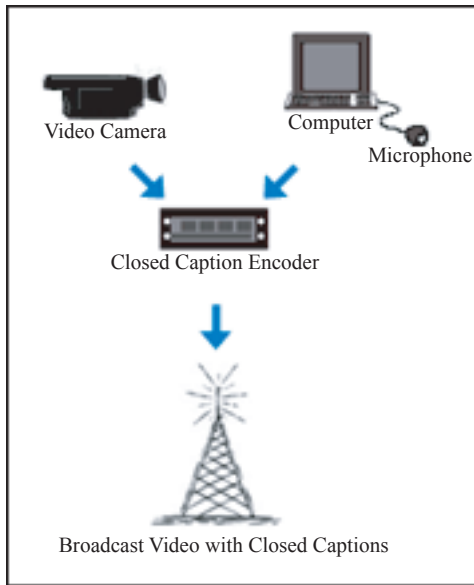
CPC-500 VoiceCaptioner Realtime captioning software	\$1,995
ViaVoice® Speech recognition software	\$250
Link PCE-845D Closed caption encoder	\$1,200
<i>Optional</i>	
Mask microphone	\$275
Headset microphone	\$95
USB microphone interface	\$59

### Complete Package: \$4,995

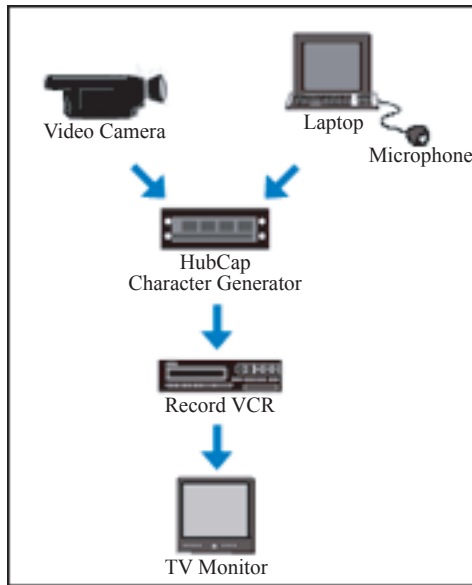
Includes:

- CPC-500 VoiceCaptioner
- ViaVoice Professional ver. 10
- Link PCE-845D CC encoder
- Dell Inspiron 600M notebook
- (1.6 GHz, 512MB ram, Serial/USB)
- Special speech recognition
- microphone with USB interface

## System Diagrams



*Television Use*



*Classroom Use*

## High Accuracy

A new generation of speech recognition software, such as version 10 of IBM's ViaVoice®, provides accuracy rates in the 90-95% range. The CPC-500 software interfaces to this off-the-shelf speech recognition software to produce realtime captions.

## Multiple Speakers

For multiple speakers a shadow speaker is used. A shadow speaker is a person who first trains the speech recognition software to recognize their speech. The shadow speaker then repeats what multiple speakers say into the system.

## How The System Works

A person speaks into a microphone attached to a computer. Their speech is transformed into a text file by speech recognition software, such as ViaVoice®. The text file is in turn formatted into a closed caption file by the CPC-500 software running on the same computer. This closed caption file is output to a standard closed caption encoder via a serial cable.

Video of the program is simultaneously input into the closed caption encoder through a standard video cable. The output of the encoder, which is standard video with embedded closed captions, is then broadcast out and/or recorded for later use.

There is typically a delay of about 3 seconds from the time a person says a word, until that word is output as a closed caption. This is very close to the delay time when a court reporter is used.

How long does it take to train the speech recognition software to attain 90% to 95% accuracy? The answer varies from person to person. A related, and more easily answered question is: what type of training will achieve this accuracy level. The answer is two types of training.

The initial training consists of a few hours of reading a pre-written text file.

The follow up training, which will be used to increase the accuracy from its initial level, consists of re-training the software to recognize the words, which it incorrectly recognized after the initial training. This is a key requirement of the training, and you probably will not achieve 90% to 95% accuracy if you do not do this.

Through progressive re-training sessions, the software's accuracy will slowly increase.

After a while, you will reach a level where re-training is rarely needed. At this point, you will only have to train the software to learn to spell names, cities and other terms not used previously.

Since the CPC-500 software can interface to most Windows based speech recognition software, if you or CPC find speech recognition software that is superior to ViaVoice®, there is a good chance that CPC will build an interface to allow you to use it.

© 2005 CPC. All trademarks and copyrights are the property of their respective owners. Prices subject to change without notice.

**A Certified 8(a) Company**

**GSA Schedule #GS-35F-4672G**



**Computer Prompting & Captioning Co.**  
1010 Rockville Pike, Suite 306, Rockville, MD 20852

**800-977-6678**  
**301-738-8487**

**info@cpcweb.com**  
**www.cpcweb.com**