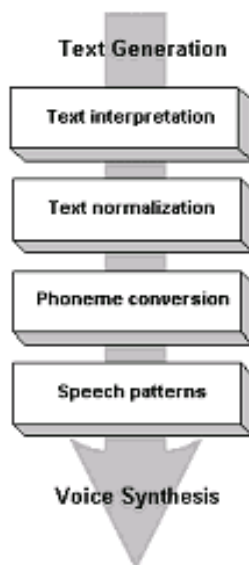


## Text to speech (TTS): a general description

**A system with text-to-speech (TTS) capabilities is able to interpret electronic text and generate audible speech from it.**

**Text-to-speech technology can be used in many different ways including aiding the sight impaired by reading email and other text-based information aloud.**

**TTS technology is also used in business applications such as interactive voice response (IVR) systems. IVRs provide information such as credit card balances, movie listings, etc., verbally to customers who access the system by phone.**



The process of converting text-to-speech involves breaking electronic text into its basic form and then converting that into audible speech, as shown in the figure below.

### Text Generation

Electronic text can come from many sources, including email, Web pages, and electronic documents like Word

and portable document format (PDF) files, but it can also come from printed material such as faxes, books, and magazines through a process called optical character recognition (OCR). An OCR application analyzes an image of a scanned page, recognizes the letters and words, and then creates an electronic text document. Once the electronic text is generated, it can be inputted to a TTS system and read aloud.

### Text interpretation

The first step a TTS system performs is the initial formatting and analysis of the text to be read. The interpreter function of the TTS application removes extra items like graphics and then formats the text according to its sentence structure.

### Text normalization

There is a significant difference between how humans write and how they speak. After the text has been formatted into its sentence structures, it must be restructured from the written word into a format that is appropriate for the spoken word. This process is called normalization. Normalization makes sure that symbols like "\$" are spoken as "dollar" and 10/1/02 is spoken "October first two thousand two," among other formatting rules.

### Phoneme conversion

Once the inputted text has been properly formatted and normalized it is broken down into phonemes. A phoneme is a basic linguistic unit for the spoken form of a language. A TTS system uses a series of rules and a pronunciation dictionary to identify and correctly break down a word into its correct phonemic pronunciation.

### Speech patterns

Correctly pronouncing a word is only a small part of replicating how humans talk. Other variables such as rhythm, speed, and the stress placed on words are underlying qualities of natural speech. After the phonemes are identified, sentence structure, grammar, punctuation, and other qualities are used to create natural speech patterns from the written word.

### Voice synthesis

The final step is to combine all the phoneme and speech pattern information with a system that generates sound to produce audible speech, allowing the text to be read aloud to the listener.