

Understand speech recognition and synthesis

3 minutes

Speech recognition takes the spoken word and converts it into data that can be processed - often by transcribing it into text. The spoken words can be in the form of a recorded voice in an audio file, or live audio from a microphone. Speech patterns are analyzed in the audio to determine recognizable patterns that are mapped to words. To accomplish this, the software typically uses multiple models, including:

- An *acoustic* model that converts the audio signal into phonemes (representations of specific sounds).
- A *language* model that maps phonemes to words, usually using a statistical algorithm that predicts the most probable sequence of words based on the phonemes.

The recognized words are typically converted to text, which you can use for various purposes, such as:

- Providing closed captions for recorded or live videos
- Creating a transcript of a phone call or meeting
- Automated note dictation
- Determining intended user input for further processing

Speech synthesis is concerned with vocalizing data, usually by converting text to speech. A speech synthesis solution typically requires the following information:

- The text to be spoken
- The voice to be used to vocalize the speech

To synthesize speech, the system typically *tokenizes* the text to break it down into individual words, and assigns phonetic sounds to each word. It then breaks the phonetic transcription into *prosodic* units (such as phrases, clauses, or sentences) to create phonemes that will be converted to audio format. These phonemes are then synthesized as audio and can be assigned a particular voice, speaking rate, pitch, and volume.

You can use the output of speech synthesis for many purposes, including:

- Generating spoken responses to user input
 - Creating voice menus for telephone systems
 - Reading email or text messages aloud in hands-free scenarios
 - Broadcasting announcements in public locations, such as railway stations or airports
-

Next unit: Get started with speech on Azure

[Continue >](#)