

Get started with speech on Azure

3 minutes

Microsoft Azure offers both speech recognition and speech synthesis capabilities through **Azure AI Speech** service, which includes the following application programming interfaces (APIs):

- The **Speech to text** API
- The **Text to speech** API

Azure resources for Azure AI Speech

To use Azure AI Speech in an application, you must create an appropriate resource in your Azure subscription. You can choose to create either of the following types of resource:

- A **Speech** resource - choose this resource type if you only plan to use Azure AI Speech, or if you want to manage access and billing for the resource separately from other services.
- An **Azure AI services** resource - choose this resource type if you plan to use Azure AI Speech in combination with other Azure AI services, and you want to manage access and billing for these services together.

The Speech to text API

You can use Azure AI Speech to text API to perform real-time or batch transcription of audio into a text format. The audio source for transcription can be a real-time audio stream from a microphone or an audio file.

The model that is used by the Speech to text API, is based on the Universal Language Model that was trained by Microsoft. The data for the model is Microsoft-owned and deployed to Microsoft Azure. The model is optimized for two scenarios, conversational and dictation. You can also create and train your own custom models including acoustics, language, and pronunciation if the pre-built models from Microsoft do not provide what you need.

Real-time transcription

Real-time speech to text allows you to transcribe text in audio streams. You can use real-time transcription for presentations, demos, or any other scenario where a person is speaking.

In order for real-time transcription to work, your application will need to be listening for incoming audio from a microphone, or other audio input source such as an audio file. Your application code streams the audio to the service, which returns the transcribed text.

Batch transcription

Not all speech to text scenarios are real time. You might have audio recordings stored on a file share, a remote server, or even on Azure storage. You can point to audio files with a shared access signature (SAS) URI and asynchronously receive transcription results.

Batch transcription should be run in an asynchronous manner because the batch jobs are scheduled on a *best-effort* basis. Normally a job will start executing within minutes of the request but there is no estimate for when a job changes into the running state.

The text to speech API

The text to speech API enables you to convert text input to audible speech, which can either be played directly through a computer speaker or written to an audio file.

Speech synthesis voices

When you use the text to speech API, you can specify the voice to be used to vocalize the text. This capability offers you the flexibility to personalize your speech synthesis solution and give it a specific character.

The service includes multiple pre-defined voices with support for multiple languages and regional pronunciation, including *neural* voices that leverage *neural networks* to overcome common limitations in speech synthesis with regard to intonation, resulting in a more natural sounding voice. You can also develop custom voices and use them with the text to speech API

Supported Languages

Both the speech to text and text to speech APIs support a variety of languages. Use the links below to find details about the supported languages:

- [Speech to text languages.](#)
 - [Text to speech languages.](#)
-

Next unit: Exercise - Explore Speech Studio

[Continue >](#)
