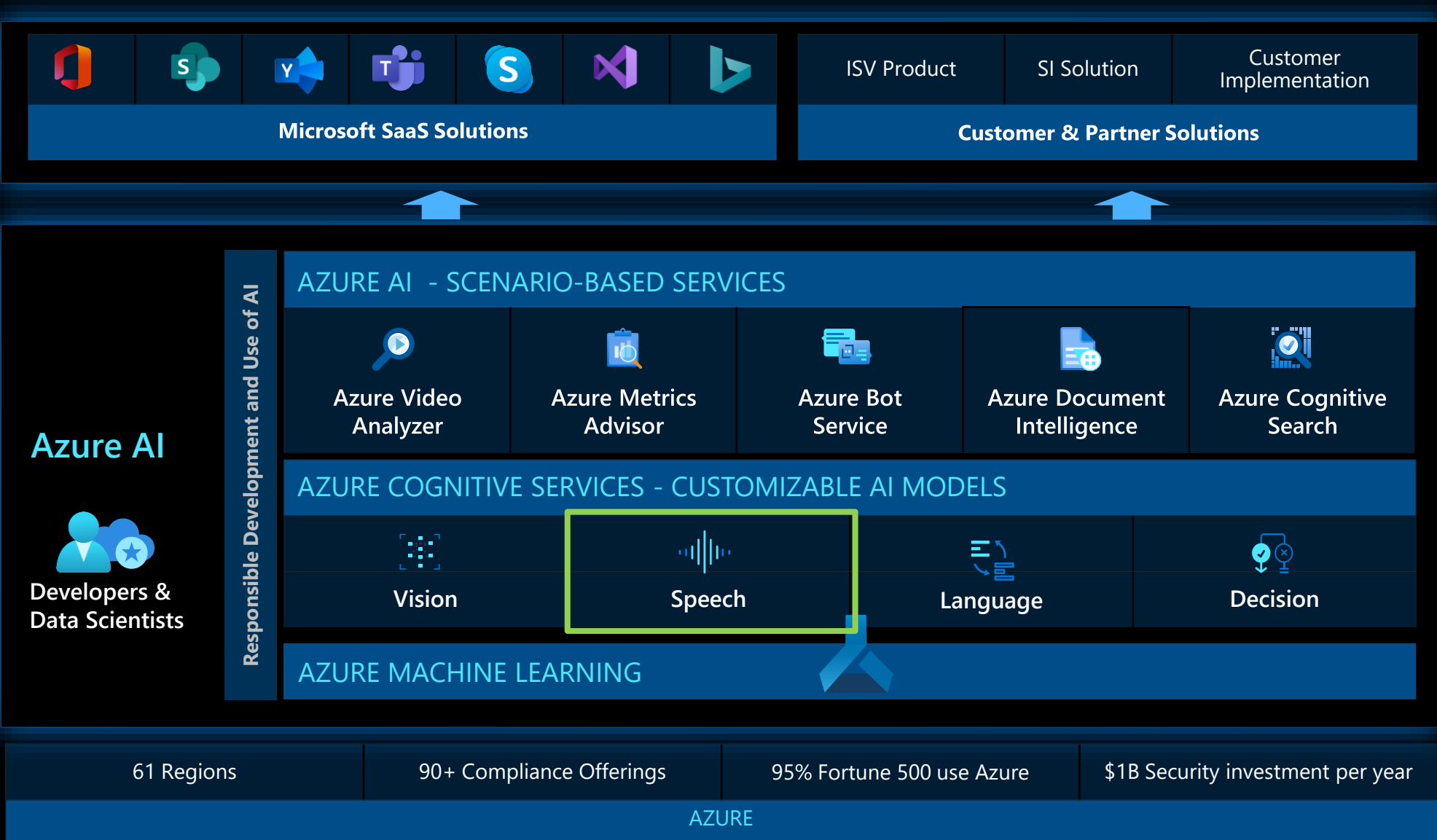




Azure AI: Speech Service Overview

Azure AI



Speech Service Overview

The Speech service provides speech to text and text to speech capabilities with a Speech resource. You can transcribe speech to text with high accuracy, produce natural-sounding text to speech voices, translate spoken audio, and use speaker recognition during conversations.

Speech capabilities

Speech to text

Text to speech

Speech Translation

Language identification

Speaker recognition

<https://learn.microsoft.com/en-us/azure/ai-services/speech-service/overview>

Speech Service Overview

Common scenarios for speech include:

- [Captioning](#): Learn how to synchronize captions with your input audio, apply profanity filters, get partial results, apply customizations, and identify spoken languages for multilingual scenarios.

- [Audio Content Creation](#): You can use neural voices to make interactions with chatbots and voice assistants more natural and engaging, convert digital texts such as e-books into audiobooks and enhance in-car navigation systems.

- [Call Center](#): Transcribe calls in real-time or process a batch of calls, redact personally identifying information, and extract insights such as sentiment to help with your call center use case.

- [Language learning](#): Provide pronunciation assessment feedback to language learners, support real-time transcription for remote learning conversations, and read aloud teaching materials with neural voices.

- [Voice assistants](#): Create natural, humanlike conversational interfaces for their applications and experiences. The voice assistant feature provides fast, reliable interaction between a device and an assistant implementation.

Speech Service Overview

Microsoft uses Speech for many scenarios, such as captioning in Teams, dictation in Office 365, and Read Aloud in the Edge browser.



**Sales /
Support**



Live captions



**Subtitles/
Translation**



Read aloud



Dictation



Commute



Cortana



AZURE SPEECH

Speech to text

Real-time speech to text

With real-time speech to text, the audio is transcribed as speech is recognized from a microphone or file. Use real-time speech to text for applications that need to transcribe audio in real-time such as:

- Transcriptions, captions, or subtitles for live meetings
- [Diarization](#)
- [Pronunciation assessment](#)
- Contact center agent assist
- Dictation
- Voice agents

Real-time speech to text is available via the [Speech SDK](#) and the [Speech CLI](#).

Batch transcription

Batch transcription is used to transcribe a large amount of audio in storage. You can point to audio files with a shared access signature (SAS) URI and asynchronously receive transcription results. Use batch transcription for applications that need to transcribe audio in bulk such as:

- Transcriptions, captions, or subtitles for pre-recorded audio
- Contact center post-call analytics
- Diarization

Custom Speech

With Custom Speech, you can evaluate and improve the accuracy of speech recognition for your applications and products. A custom speech model can be used for real-time speech to text, speech translation, and batch transcription.

Text to speech

The text to speech feature of the Speech service on Azure has been fully upgraded to the **neural text to speech engine**. This engine uses deep neural networks to make the voices of computers nearly indistinguishable from the recordings of people. With the clear articulation of words, neural text to speech significantly reduces listening fatigue when users interact with AI systems.

The patterns of stress and intonation in spoken language are called **prosody**. Traditional text to speech systems break down prosody into separate linguistic analysis and acoustic prediction steps that are governed by independent models. That can result in muffled, buzzy voice synthesis.

Text to speech features

- **Real-time speech synthesis:** Use the Speech SDK or REST API to convert text to speech by using prebuilt neural voices or custom neural voices.
- **Asynchronous synthesis of long audio:** Use the batch synthesis API (Preview) to asynchronously synthesize text to speech files longer than 10 minutes (for example, audio books or lectures).
- **Prebuilt neural voices:** Each prebuilt neural voice model is available at 24kHz and high-fidelity 48kHz.
- **Fine-tuning text to speech output with SSML:** Speech Synthesis Markup Language (SSML) is an XML-based markup language that's used to customize text to speech outputs. With SSML, you can adjust pitch, add pauses, improve pronunciation, change speaking rate, adjust volume, and attribute multiple voices to a single document.
- **Visemes:** Visemes are the key poses in observed speech, including the position of the lips, jaw, and tongue in producing a particular phoneme. Visemes have a strong correlation with voices and phonemes.

Speech Translation

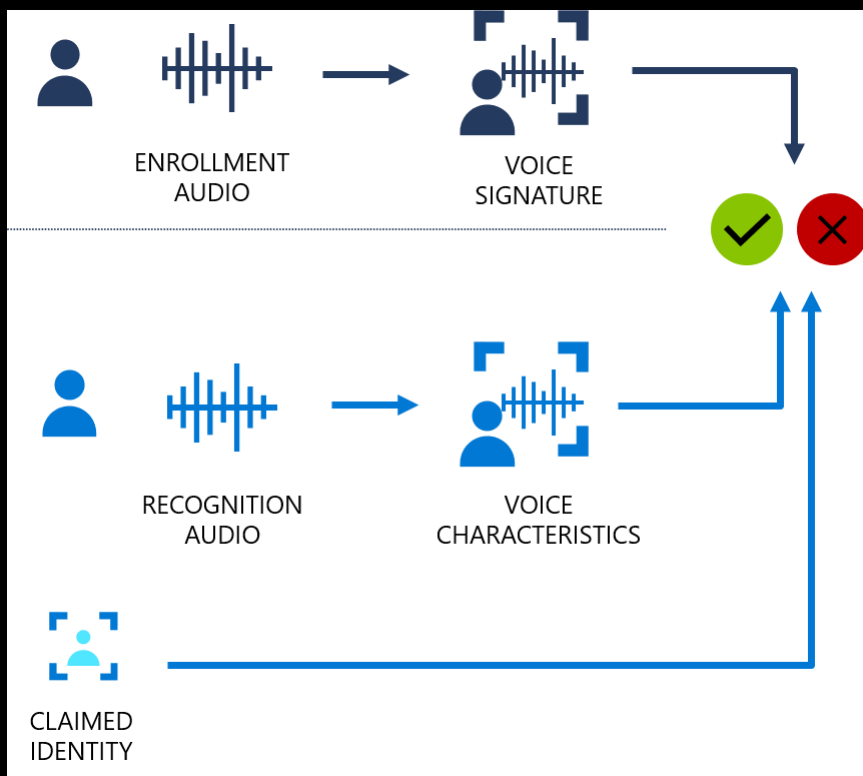
By using the Speech SDK or Speech CLI, you can give your applications, tools, and devices access to source transcriptions and translation outputs for the provided audio. Interim transcription and translation results are returned as speech is detected, and the final results can be converted into synthesized speech.

Speech Translation Core features

- Speech to text translation with recognition results.
- Speech-to-speech translation.
- Support for translation to multiple target languages.
- Interim recognition and translation results.

Speaker Recognition

Speaker recognition can help determine who is speaking in an audio clip. The service can verify and identify speakers by their unique voice characteristics, by using voice biometry.




Speaker verification can be either text-dependent or text-independent. Text-dependent verification means that speakers need to choose the same passphrase to use during both enrollment and verification phases. Text-independent verification means that speakers can speak in everyday language in the enrollment and verification phrases.

For text-dependent verification, the speaker's voice is enrolled by saying a passphrase from a set of predefined phrases. Voice features are extracted from the audio recording to form a unique voice signature, and the chosen passphrase is also recognized. Together, the voice signature and the passphrase are used to verify the speaker.

Text-independent verification has no restrictions on what the speaker says during enrollment, besides the initial activation phrase when active enrollment is enabled. It doesn't have any restrictions on the audio sample to be verified, because it only extracts voice features to score similarity.

Exercise 1



600 XP

Translate text and speech with Azure AI services

29 min • Module • 5 Units

★★★★★ 4.7 (5,215)

Beginner AI Engineer Data Scientist Developer Solution Architect Student Azure AI services

Automated translation capabilities in an AI solution enable closer collaboration by removing language barriers.

Learning objectives

After completing this module, you will understand how to perform text and speech translation using Azure AI Translator and Azure AI Speech.

Start > + Add

Prerequisites

Ability to navigate the Azure portal

This module is part of these learning paths

Microsoft Azure AI Fundamentals: Explore natural language processing

Introduction

3 min

Get started with translation in Azure

3 min

Exercise - Explore translation

20 min

Knowledge check


2 min

Summary

1 min

<https://learn.microsoft.com/en-us/training/modules/translate-text-with-translation-service/>

Exercise 2

800 XP

Translate speech with the Azure AI Speech service

47 min • Module • 7 Units

★★★★★ 4.7 (2,562)

Intermediate AI Engineer Developer Solution Architect Student Azure AI services

Translation of speech builds on speech recognition by recognizing and transcribing spoken input in a specified language, and returning translations of the transcription in one or more other languages.

Learning objectives

In this module, you will learn how to:

- Provision Azure resources for speech translation.
- Generate text translation from speech.
- Synthesize spoken translations.

[Start >](#) [+ Add](#)

Prerequisites

Before starting this module, you should:

- Be familiar with Azure services and the Azure portal.
- Have experience programming with C# or Python.
- Have experience of using the Azure AI Speech service to transcribe speech to text.


This module is part of these learning paths

Develop natural language processing solutions with Azure AI Services

Process and Translate Speech with Azure AI Speech Services

<https://learn.microsoft.com/en-us/training/modules/translate-speech-speech-service/>

Exercise 3



Create speech-enabled apps with Azure AI services

58 min • Module • 9 Units

★★★★★ 4.8 (2,371)

Intermediate AI Engineer Data Scientist Developer Solution Architect Student Azure AI services

The Azure AI Speech service enables you to build speech-enabled applications. This module focuses on using the speech-to-text and text to speech APIs, which enable you to create apps that are capable of speech recognition and speech synthesis.

Learning objectives

In this module, you'll learn how to:

- Provision an Azure resource for the Azure AI Speech service
- Use the Azure AI Speech to text API to implement speech recognition
- Use the Text to speech API to implement speech synthesis
- Configure audio format and voices
- Use Speech Synthesis Markup Language (SSML)

[Start >](#) [⊕ Add](#)

Prerequisites

Before starting this module, you should:


- Be familiar with Azure services and the Azure portal
- Have experience programming with C# or Python

This module is part of these learning paths

Process and Translate Speech with Azure AI Speech Services

<https://learn.microsoft.com/en-us/training/modules/transcribe-speech-input-text/>

Exercise 4



Recognize and synthesize speech with Azure AI Speech

29 min • Module • 5 Units

★★★★★ 4.8 (5,835)

Intermediate AI Engineer Developer Solution Architect Student Azure Azure SDKs

Learn how to recognize and synthesize speech by using Azure AI Speech.

Learning objectives

In this module you will:

- Learn about speech recognition and synthesis
- Learn how to use Azure AI Speech

[Start >](#) [⊕ Add](#)

Prerequisites

Ability to navigate the Azure portal

This module is part of these learning paths

Microsoft Azure AI Fundamentals; Explore natural language processing

600 XP

<https://learn.microsoft.com/en-us/training/modules/recognize-synthesize-speech/>



Invent with purpose.

Thank you