**Implement speech-to-text**

Azure AI Speech service offers advanced speech to text capabilities. This feature supports both real-time and batch transcription, providing versatile solutions for converting audio streams into text.

**Core Features**

The speech to text service offers the following core features:

* [Real-time](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/speech-to-text#real-time-speech-to-text) transcription: Instant transcription with intermediate results for live audio inputs.
* [Fast transcription](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/speech-to-text#fast-transcription): Fastest synchronous output for situations with predictable latency.
* [Batch transcription](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/speech-to-text#batch-transcription-api): Efficient processing for large volumes of prerecorded audio.
* [Custom speech](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/speech-to-text#custom-speech): Models with enhanced accuracy for specific domains and conditions.

**Real-time speech to text**

Real-time speech to text transcribes audio as it's recognized from a microphone or file. It's ideal for applications requiring immediate transcription, such as:

* **Transcriptions, captions, or subtitles for live meetings**: Real-time audio transcription for accessibility and record-keeping.
* **Diarization**: Identifying and distinguishing between different speakers in the audio.
* **Pronunciation assessment**: Evaluating and providing feedback on pronunciation accuracy.
* **Call center agents assist**: Providing real-time transcription to assist customer service representatives.
* **Dictation**: Transcribing spoken words into written text for documentation purposes.
* **Voice agents**: Enabling interactive voice response systems to transcribe user queries and commands.

Real-time speech to text can be accessed via the Speech SDK, Speech CLI, and REST API, allowing integration into various applications and workflows. Real-time speech to text is available via the [Speech SDK](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/speech-sdk), the [Speech CLI](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/spx-overview), and [Speech to text REST API for short audio](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/rest-speech-to-text-short).

**Fast transcription**

Fast transcription API is used to transcribe audio files with returning results synchronously and faster than real-time audio. Use fast transcription in the scenarios that you need the transcript of an audio recording as quickly as possible with predictable latency, such as:

* **Quick audio or video transcription and subtitles**: Quickly get a transcription of an entire video or audio file in one go.
* **Video translation**: Immediately get new subtitles for a video if you have audio in different languages.

**Batch transcription API**

[Batch transcription](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/batch-transcription) is designed for transcribing large amounts of audio stored in files. This method processes audio asynchronously and is suited for:

* **Transcriptions, captions, or subtitles for prerecorded audio**: Converting stored audio content into text.
* **Contact center post-call analytics**: Analyzing recorded calls to extract valuable insights.
* **Diarization**: Differentiating between speakers in recorded audio.

Batch transcription is available via:

* [Speech to text REST API](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/rest-speech-to-text): Facilitates batch processing with the flexibility of RESTful calls. To get started, see [How to use batch transcription](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/batch-transcription) and [Batch transcription samples](https://github.com/Azure-Samples/cognitive-services-speech-sdk/tree/master/samples/batch).
* [Speech CLI](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/spx-overview): Supports both real-time and batch transcription, making it easy to manage transcription tasks. For Speech CLI help with batch transcriptions, run the following command:



**Custom speech**

With [custom speech](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/custom-speech-overview), 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](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/speech-to-text), [speech translation](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/speech-translation), and [batch transcription](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/batch-transcription).

Out of the box, speech recognition utilizes a Universal Language Model as a base model that is trained with Microsoft-owned data and reflects commonly used spoken language. The base model is pretrained with dialects and phonetics representing various common domains. When you make a speech recognition request, the most recent base model for each [supported language](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/language-support?tabs=stt) is used by default. The base model works well in most speech recognition scenarios.

Custom speech allows you to tailor the speech recognition model to better suit your application's specific needs. This can be particularly useful for:

* **Improving recognition of domain-specific vocabulary**: Train the model with text data relevant to your field.
* **Enhancing accuracy for specific audio conditions**: Use audio data with reference transcriptions to refine the model.

**Usage Examples**

A close-up of a document

Description automatically generated

**using the client library and REST API**

**Create an Azure resource**

To use the code sample below, you'll need to deploy an Azure resource. This resource will contain a key and endpoint you'll use to authenticate the API calls you send to the Speech service.

1. Use the following link to [Create a Speech resource](https://portal.azure.com/#create/Microsoft.CognitiveServicesSpeechServices) using the Azure portal. You will need to sign in using your Azure subscription.
2. In the **Create Speech Services** screen, provide the following information:

A computer screen with a blue and white screen

Description automatically generated

A close up of a white card

Description automatically generated

1. Select **Review + Create** at the bottom of the page.
2. In the screen that appears, make sure the validation has passed, and that you entered your information correctly. Then select **Create**.

**Get your key and endpoint**

* Get the Speech resource key and region. After your Speech resource is deployed, select **Go to resource** to view and manage keys.

**Create a new .NET Core application**

Using the Visual Studio IDE, create a new .NET Core console app.

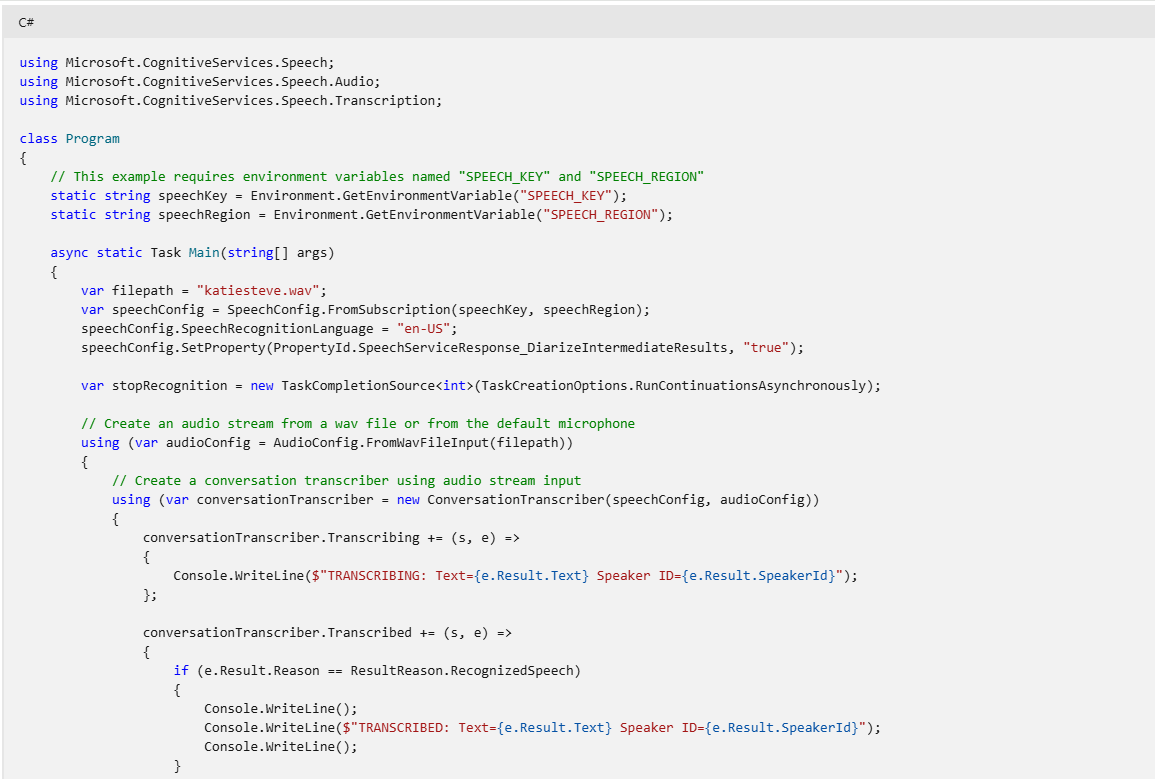
Install the client library by right-clicking on the solution in the Solution Explorer and selecting Manage NuGet Packages. In the package manager that opens select Browse and search for   
**Microsoft.CognitiveServices.Speech**.

**Code example**

We recommend Microsoft Entra ID authentication with [**managed identities for Azure resources**](https://learn.microsoft.com/en-us/azure/active-directory/managed-identities-azure-resources/overview) to avoid storing credentials with your applications that run in the cloud.

If you use an API key, store it securely somewhere else, such as in [**Azure Key Vault**](https://learn.microsoft.com/en-us/azure/key-vault/general/overview). Don't include the API key directly in your code, and never post it publicly.

* To set the **SPEECH\_KEY** environment variable, replace *your-key* with one of the keys for your resource.
* To set the **SPEECH\_REGION** environment variable, replace *your-region* with one of the regions for your resource.

****

**A screenshot of a computer program

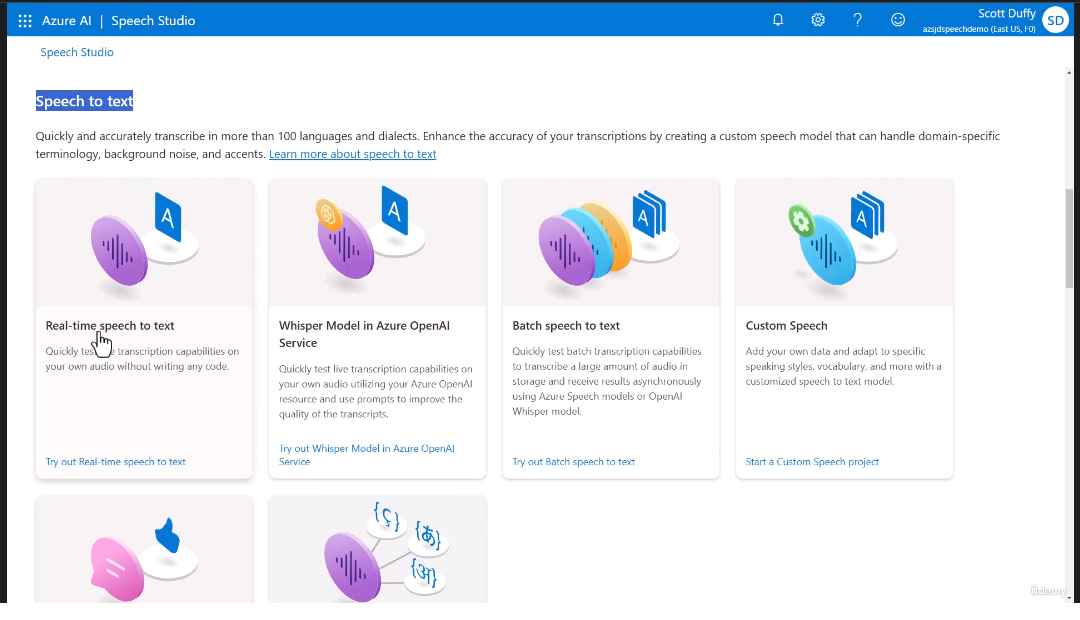
Description automatically generated**

**Output**

A screenshot of a computer

Description automatically generated

**Use Speech Studio**



A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated