Translate speech with the Azure Speech service

1. Introduction

https://learn.microsoft.com/en-us/training/modules/translate-speech-speech-service/1-introduction

Introduction

Completed

- 2 minutes

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.

In this module, you'll learn how to:

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

The units in the module include important conceptual information about Azure Speech and how to use its API through one of the supported software development kits (SDKs), after which you're able to try Azure Speech for yourself in a hands-on exercise. To complete the hands-on exercise, you'll need a Microsoft Azure subscription. If you don't already have one, you can sign up for a free trial at https://azure.com/free

2. Provision an Azure resource for speech translation

https://learn.microsoft.com/en-us/training/modules/translate-speech-speech-service/2-speech-service

Provision an Azure resource for speech translation

Completed

- 3 minutes

The Azure Speech service provides robust, machine learning and artificial intelligence-based speech translation services, enabling developers to add end-to-end, real-time, speech translations to their applications or services. You can use either a dedicated Azure Speech resource or a multi-service Foundry Tools resource.

Before you can use the service, you need to create an Azure Speech resource in your Azure subscription.

After creating your Azure resource, you'll need the following information to use it from a client application through one of the supported SDKs:

- The **location** in which the resource is deployed (for example, eastus)
- One of the **keys** assigned to your resource.

You can view of these values on the **Keys and Endpoint** page for your resource in the Azure portal.

3. Translate speech to text

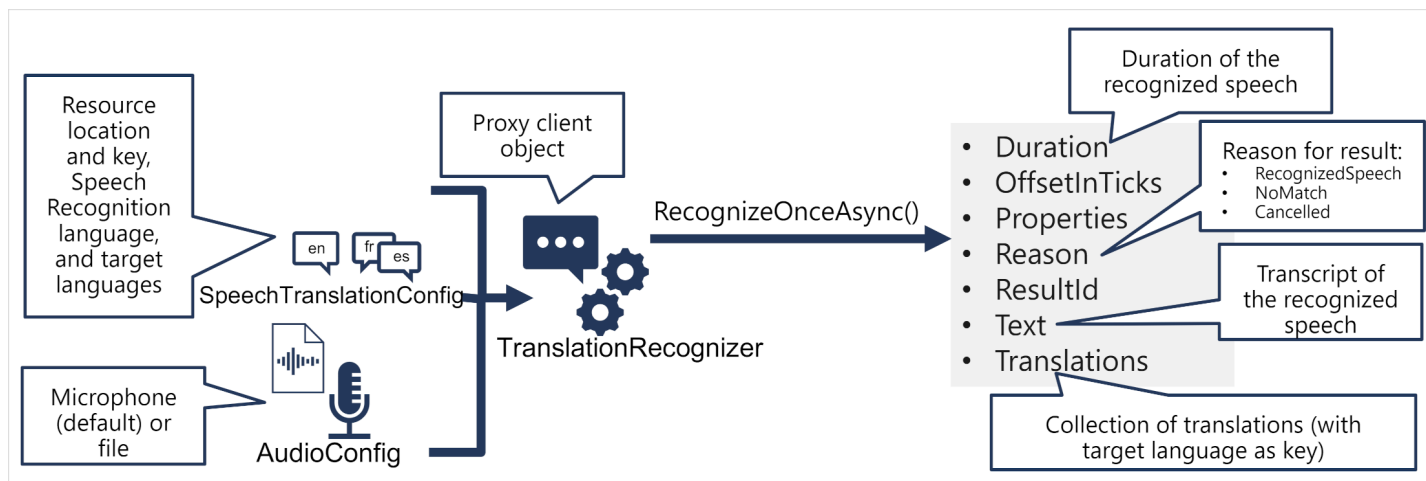https://learn.microsoft.com/en-us/training/modules/translate-speech-speech-service/3-translate-speech-text

Translate speech to text

Completed

- 3 minutes

The pattern for speech translation using the Azure Speech SDK is similar to speech recognition, with the addition of information about the source and target languages for translation:

1. Use a **SpeechTranslationConfig** object to encapsulate the information required to connect to your Azure Speech resource. Specifically, its location and key.

2. The **SpeechTranslationConfig** object is also used to specify the speech recognition language (the language in which the input speech is spoken) and the target languages into which it should be translated.

3. Optionally, use an **AudioConfig** to define the input source for the audio to be transcribed. By default, this is the default system microphone, but you can also specify an audio file.

4. Use the **SpeechTranslationConfig**, and **AudioConfig** to create a **TranslationRecognizer** object. This object is a proxy client for the Azure Speech translation API.

5. Use the methods of the **TranslationRecognizer** object to call the underlying API functions. For example, the **RecognizeOnceAsync**() method uses the Azure Speech service to asynchronously translate a single spoken utterance.

6. Process the response from Azure Speech. In the case of the **RecognizeOnceAsync**() method, the result is a **SpeechRecognitionResult** object that includes the following properties:

   - Duration
   - OffsetInTicks
   - Properties
   - Reason
   - ResultId
   - Text
   - Translations

If the operation was successful, the **Reason** property has the enumerated value **RecognizedSpeech**, the **Text** property contains the transcription in the original language. You can also access a **Translations** property which contains a dictionary of the translations (using the two-character ISO language code, such as "en" for English, as a key).

Synthesize translations

Completed

- 3 minutes

The **TranslationRecognizer** returns translated transcriptions of spoken input - essentially translating audible speech to text.

You can also synthesize the translation as speech to create speech-to-speech translation solutions. There are two ways you can accomplish this.

Event-based synthesis

When you want to perform 1:1 translation (translating from one source language into a single target language), you can use event-based synthesis to capture the translation as an audio stream. To do this, you need to:

Specify the desired voice for the translated speech in the **TranslationConfig**. Create an event handler for the **TranslationRecognizer** object's **Synthesizing** event. In the event handler, use the **GetAudio**() method of the **Result** parameter to retrieve the byte stream of translated audio. The specific code used to implement an event handler varies depending on the programming language you're using. See the [C#](#) and [Python](#) examples in the Speech SDK documentation.

Manual synthesis

Manual synthesis is an alternative approach to event-based synthesis that doesn't require you to implement an event handler. You can use manual synthesis to generate audio translations for one or more target languages.

Manual synthesis of translations is essentially just the combination of two separate operations in which you:

1. Use a **TranslationRecognizer** to translate spoken input into text transcriptions in one or more target languages.
2. Iterate through the **Translations** dictionary in the result of the translation operation, using a **SpeechSynthesizer** to synthesize an audio stream for each language.

5. Exercise - Translate speech

https://learn.microsoft.com/en-us/training/modules/translate-speech-speech-service/5-exercise-translate-speech

Exercise - Translate speech

Completed

- 30 minutes

In this exercise, build an app that recognizes and translates speech into a specific language.

> **Note**
>
> To complete this lab, you need an **Azure subscription**.

Launch the exercise and follow the instructions.

Launch Exercise

> **Tip**
>
> After completing the exercise, if you've finished exploring Foundry Tools, delete the Azure resources that you created during the exercise.

6. Module assessment

Module assessment

Completed

- 3 minutes

7. Summary

Summary

Completed

- 3 minutes

Now that you've completed this module, you learned how to:

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

For more information about speech translation, refer to the [Azure Speech documentation](#).