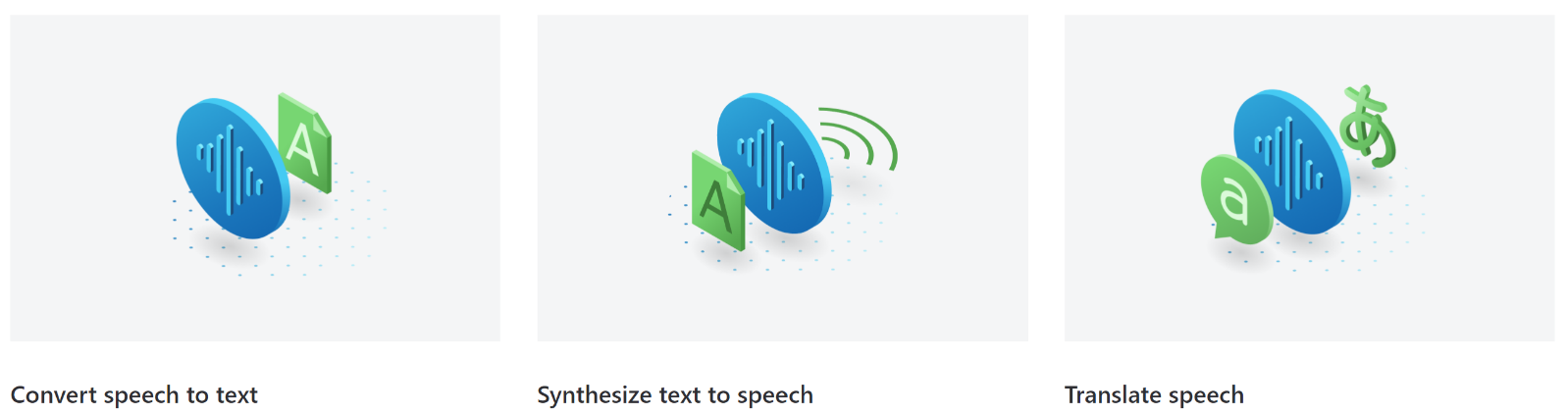
**Azure AI Speech Service**

The Speech service provides speech to text and text to speech capabilities with a [Speech resource](https://learn.microsoft.com/en-us/azure/ai-services/multi-service-resource?pivots=azportal). 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.

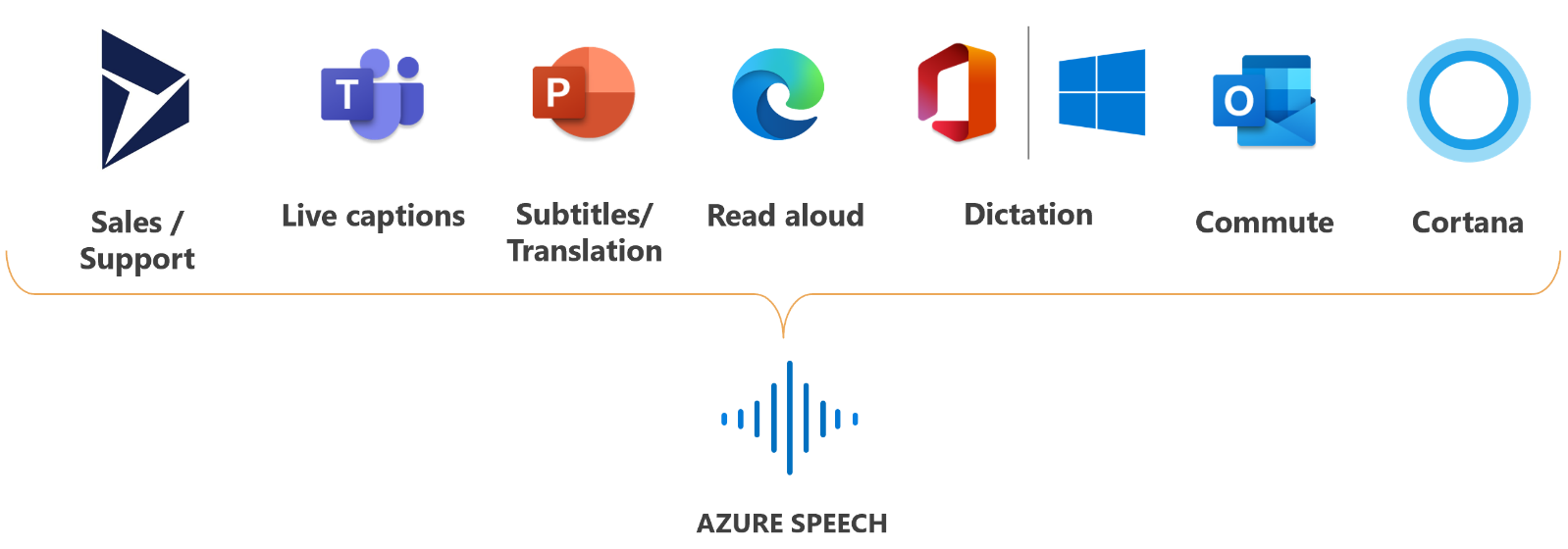
****

**Speech scenarios**

Common scenarios for speech include:

* [Captioning](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/captioning-concepts): 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](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/text-to-speech#more-about-neural-text-to-speech-features): 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](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/call-center-overview): 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](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/language-learning-overview): 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](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/voice-assistants): Create natural, human like conversational interfaces for their applications and experiences. The voice assistant feature provides fast, reliable interaction between a device and an assistant implementation.

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

****

**Speech capabilities**

**Speech to text**

Use [speech to text](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/speech-to-text) to transcribe audio into text, either in [real-time](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/overview#real-time-speech-to-text) or asynchronously with [batch transcription](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/overview#batch-transcription).

Convert audio to text from a range of sources, including **microphones, audio files, and blob storage**. Use speaker diarization to determine who said what and when. Get readable transcripts with automatic formatting and punctuation.

The base model might not be sufficient if the audio contains ambient noise or includes numerous industry and domain-specific jargon. In these cases, you can create and train [custom speech models](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/custom-speech-overview) with acoustic, language, and pronunciation data. Custom speech models are private and can offer a competitive advantage.

**Real-time speech to text**

With [real-time speech to text](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/get-started-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](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/get-started-stt-diarization)
* [Pronunciation assessment](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/how-to-pronunciation-assessment)
* Contact center agents assist
* Dictation
* Voice agents

**Fast transcription API**

Fast transcription API is used to transcribe audio files with returning results synchronously and much 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, subtitles, and edit.
* Video translation

To get started with fast transcription, see [use the fast transcription API](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/fast-transcription-create).

**Batch transcription**

[Batch transcription](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/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 prerecorded audio
* Contact center post-call analytics
* Diarization

**Text to speech**

With [text to speech](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/text-to-speech), you can convert input text into human like synthesized speech. Use neural voices, which are human like voices powered by deep neural networks. Use the [Speech Synthesis Markup Language (SSML)](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/speech-synthesis-markup) to fine-tune the pitch, pronunciation, speaking rate, volume, and more.

* Prebuilt neural voice: Highly natural out-of-the-box voices. Check the prebuilt neural voice samples the [Voice Gallery](https://speech.microsoft.com/portal/voicegallery) and determine the right voice for your business needs.
* Custom neural voice: Besides the prebuilt neural voices that come out of the box, you can also create a [custom neural voice](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/custom-neural-voice) that is recognizable and unique to your brand or product. Custom neural voices are private and can offer a competitive advantage. Check the custom neural voice samples [here](https://aka.ms/customvoice).

**Speech translation**

[Speech translation](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/speech-translation) enables real-time, multilingual translation of speech to your applications, tools, and devices. Use this feature for speech to speech and speech to text translation.

**Language identification**

[Language identification](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/language-identification) is used to identify languages spoken in audio when compared against a list of [supported languages](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/language-support). Use language identification by itself, with speech to text recognition, or with speech translation.

**Speaker recognition**

[Speaker recognition](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/speaker-recognition-overview) provides algorithms that verify and identify speakers by their unique voice characteristics. Speaker recognition is used to answer the question, "Who is speaking?".

**Pronunciation assessment**

[Pronunciation assessment](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/how-to-pronunciation-assessment) evaluates speech pronunciation and gives speakers feedback on the accuracy and fluency of spoken audio. With pronunciation assessment, language learners can practice, get instant feedback, and improve their pronunciation so that they can speak and present with confidence.

**Intent recognition**

[Intent recognition](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/intent-recognition): Use speech to text with conversational language understanding to derive user intents from transcribed speech and act on voice commands.

**Use Speech in your application**

The [Speech Studio](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/speech-studio-overview) is a set of UI-based tools for building and integrating features from Azure AI Speech service in your applications. You create projects in Speech Studio by using a no-code approach, and then reference those assets in your applications by using 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), or the REST APIs.

The [Speech CLI](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/spx-overview) is a command-line tool for using Speech service without having to write any code. Most features in the Speech SDK are available in the Speech CLI, and some advanced features and customizations are simplified in the Speech CLI.

The [Speech SDK](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/speech-sdk) exposes many of the Speech service capabilities you can use to develop speech-enabled applications. The Speech SDK is available in many programming languages and across all platforms.

In some cases, you can't or shouldn't use the [Speech SDK](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/speech-sdk). In those cases, you can use REST APIs to access the Speech service. For example, use REST APIs for [batch transcription](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/batch-transcription) and [speaker recognition](https://learn.microsoft.com/en-us/rest/api/speakerrecognition/) REST APIs.

**Speech Studio scenarios**

* [Captioning](https://aka.ms/speechstudio/captioning): Choose a sample video clip to see real-time or offline processed captioning results. Learn how to synchronize captions with your input audio, apply profanity filters, get partial results, apply customizations, and identify spoken languages for multilingual scenarios. For more information, see the [captioning quickstart](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/captioning-quickstart).
* [Call Center](https://aka.ms/speechstudio/callcenter): View a demonstration on how to use the Language and Speech services to analyze call center conversations. 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. For more information, see the [call center quickstart](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/call-center-quickstart).

**Speech Studio features**

In Speech Studio, the following Speech service features are available as project types:

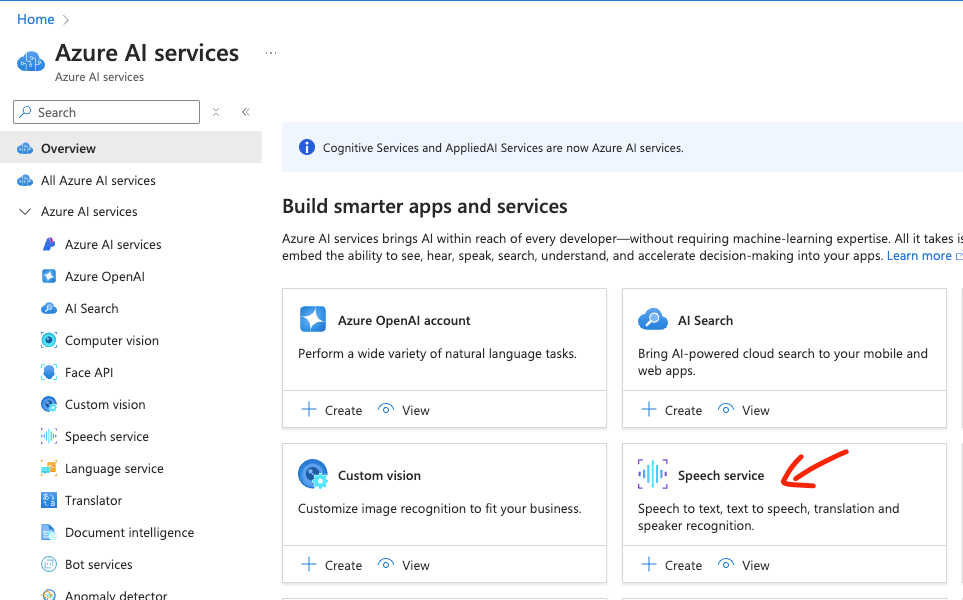
* [Real-time speech to text](https://aka.ms/speechstudio/speechtotexttool): Quickly test speech to text by dragging audio files here without having to use any code. Speech Studio has a demo tool for seeing how speech to text works on your audio samples. To explore the full functionality, see [What is speech to text](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/speech-to-text).
* [Batch speech to text](https://aka.ms/speechstudio/batchspeechtotext): Quickly test batch transcription capabilities to transcribe a large amount of audio in storage and receive results asynchronously, To learn more about Batch Speech-to-text, see [Batch speech to text overview](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/batch-transcription).
* [Custom speech](https://aka.ms/speechstudio/customspeech): Create speech recognition models that are tailored to specific vocabulary sets and styles of speaking. In contrast to the base speech recognition model, Custom speech models become part of your unique competitive advantage because they're not publicly accessible. To get started with uploading sample audio to create a custom speech model, see [Upload training and testing datasets](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/how-to-custom-speech-upload-data).
* [Pronunciation assessment](https://aka.ms/speechstudio/pronunciationassessment): Evaluate speech pronunciation and give speakers feedback on the accuracy and fluency of spoken audio. Speech Studio provides a sandbox for testing this feature quickly, without code. To use the feature with the Speech SDK in your applications, see the [Pronunciation assessment](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/how-to-pronunciation-assessment) article.
* [Speech Translation](https://aka.ms/speechstudio/speechtranslation): Quickly test and translate speech into other languages of your choice with low latency. To explore the full functionality, see [What is speech translation](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/speech-translation).
* [Voice Gallery](https://aka.ms/speechstudio/voicegallery): Build apps and services that speak naturally. Choose from a broad portfolio of [languages, voices, and variants](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/language-support?tabs=tts). Bring your scenarios to life with highly expressive and human-like neural voices.
* [Custom voice](https://aka.ms/speechstudio/customvoice): Create custom, one-of-a-kind voices for text to speech. You supply audio files and create matching transcriptions in Speech Studio, and then use the custom voices in your applications. To create and use custom voices via endpoints, see [Create and use your voice model](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/professional-voice-train-voice).
* [Audio Content Creation](https://aka.ms/speechstudio/audiocontentcreation): A no-code approach for text to speech synthesis. You can use the output audio as-is, or as a starting point for further customization. You can build highly natural audio content for various scenarios, such as audiobooks, news broadcasts, video narrations, and chat bots. For more information, see the [Audio Content Creation](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/how-to-audio-content-creation) documentation.
* [Custom Keyword](https://aka.ms/speechstudio/customkeyword): A custom keyword is a word or short phrase that you can use to voice-activate a product. You create a custom keyword in Speech Studio, and then generate a binary file to [use with the Speech SDK](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/custom-keyword-basics) in your applications.

**Transform Video into Different Language:**

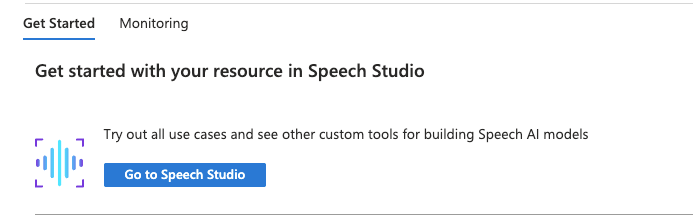
* GitHub Introduction Video : We will try to convert this in Finnish language

Let’s get started on Azure portal :

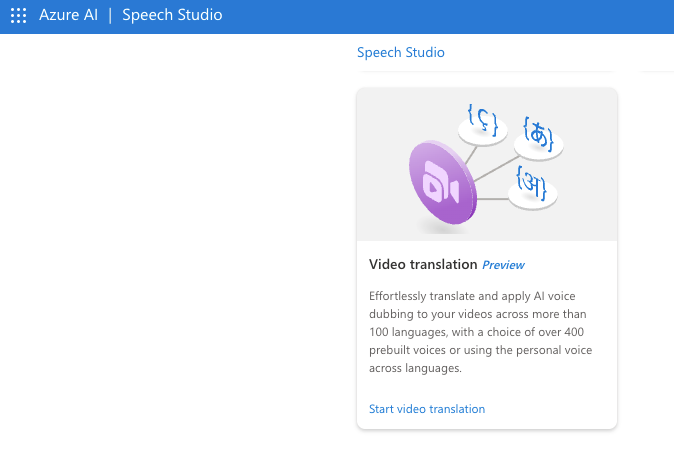
1. Goto Azure AI services and choose Speech Service from the catalogue. Create a new speech service resource.



2. After creating the service , we need Speech Studio to translate video. [Speech Studio](https://aka.ms/speechstudio/) is a set of UI-based tools for building and integrating features from Azure AI Speech service in your applications. You create projects in Speech Studio by using a no-code approach, and then reference those assets in your applications by using 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), or the REST APIs.



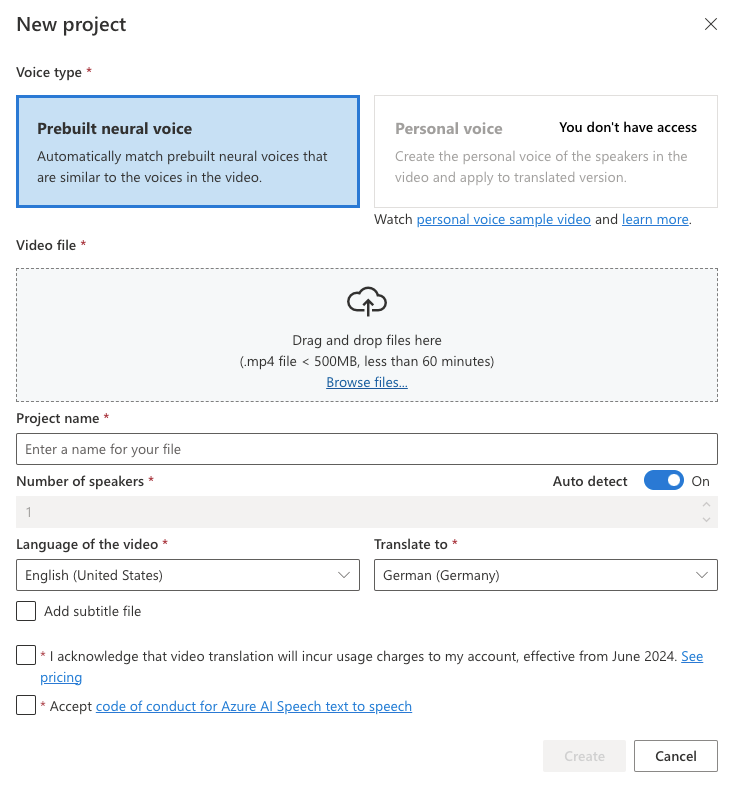
3. Select the **Video Translation** tile from the service catalogue.



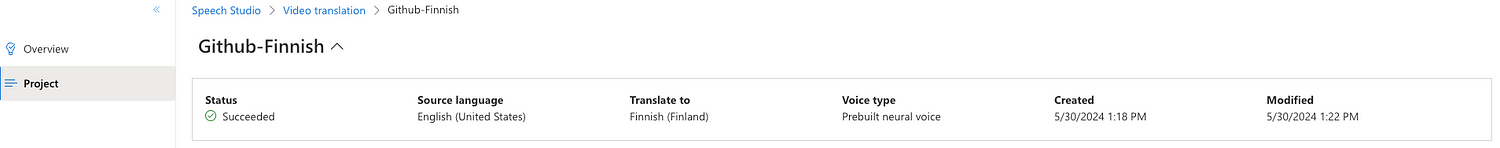
4. Create a new project and select a voice type. By default, you’ll get a *prebuilt neural voice*, but you also have the option to use a custom voice by providing a personal voice sample.

After choosing the voice type, upload the video and assign a unique project name. Next, select the original language of the video and the target language for translation.

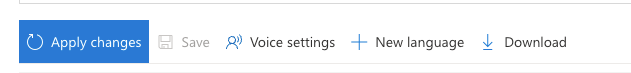
There’s also an option to add a subtitle file.



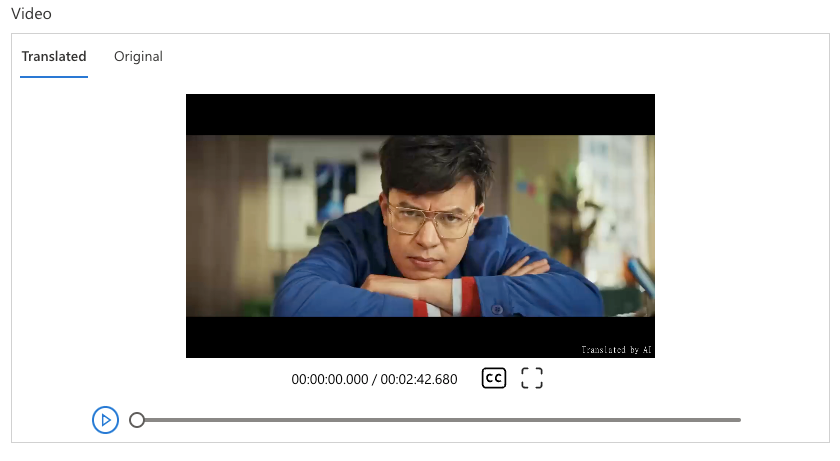
5. After creating the project, your video will be ready for processing. Once completed, you’ll see the translated video within your project. The top bar displays information such as the processing status, source and target languages, voice type, and the date and time of creation and last modification.



6. Just below the top panel , you have options to change settings like voice settings or new languages. Also you have a button to download the video.



7. In the main panel, On the left hand side you get the translated video that you can play to check out the results as well as you can see the original video.

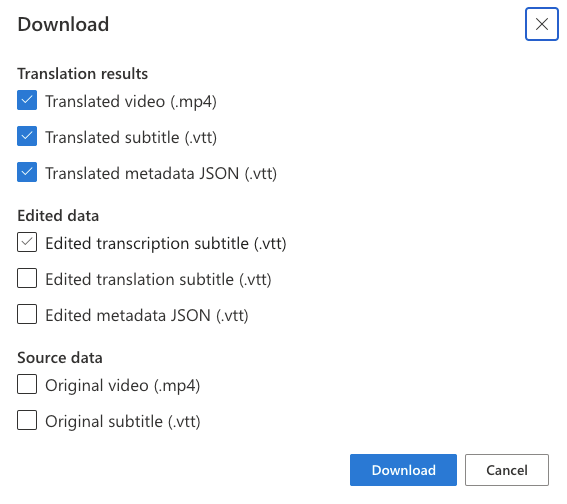


8. On the right hand side , you get the text from original video in the source language and the translated



9. This feature is especially helpful because you can review the translation and make any necessary corrections; your changes will be applied automatically.

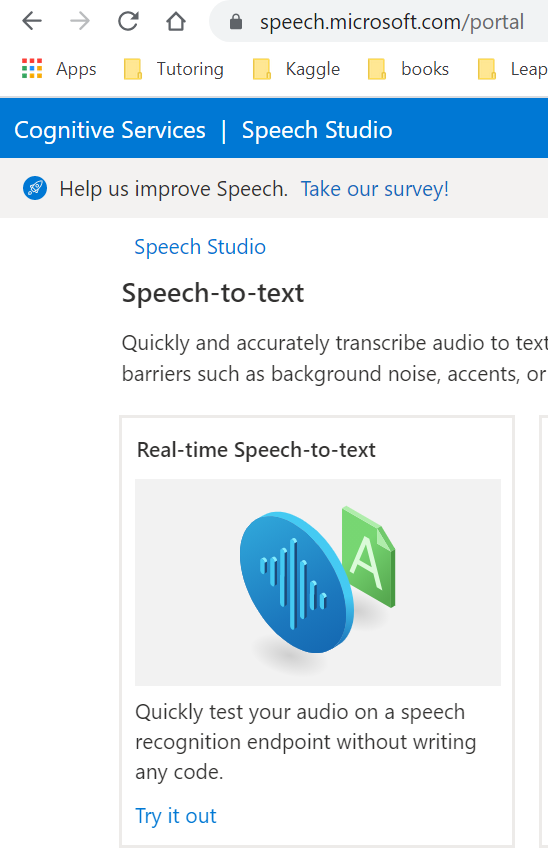
10. To download the results, use the **Download** button on the top panel. This will provide you with three files: the translated video in .mp4 format, the subtitle file, and metadata in a JSON file.



**Speech-to-text Service via Speech Studio**

The out of the box speech-to-text Service is available for quick real-time Speech-to-text service and transcription of WAV audio file(s) (16kHz or 8kHz, 16-bit, and mono PCM).

1. Sign in to [Speech Studio](https://speech.microsoft.com/) with your Azure account.
2. Select the speech service resource you need to get started.
3. Select Real-time Speech-to-text.



Once you upload the WAV file after creating the service. You can see the transcription below the audio player as follow.

