Azure Cognitive Speech Services

Azure Speech to Text can be used for processing audio files or real-time audio streams. The service supports multiple languages. The full list is available here <https://learn.microsoft.com/en-us/azure/cognitive-services/speech-service/language-support?tabs=language-identification#supported-languages>

The Speech to Text service can be used for a few different use cases. The work at Humana will be batch processing. You will most likely store your unpressed audio files on Azure cloud storage which is encrypted at rest. The Batch capability can be performed one of two ways.

* REST API
  + <https://learn.microsoft.com/en-us/azure/cognitive-services/speech-service/language-support?tabs=language-identification#supported-languages>
* Azure CLI
  + <https://learn.microsoft.com/en-us/azure/cognitive-services/speech-service/spx-basics>

You would provide multiple files per request or point to an Azure Blob Storage container with the audio files to transcribe. The batch transcription service can handle a large number of submitted transcriptions. The service transcribes the files concurrently, which reduces the turnaround time.

Batch transcription jobs are scheduled on a best-effort basis. You can't estimate when a job will change into the running state, but it should happen within minutes under normal system load. When the job is in the running state, the transcription occurs faster than the audio runtime playback speed.

## Supported Audio Formats

Batch transcription supports the following audio files.

| **Format** | **Codec** | **Bits per sample** | **Sample rate** |
| --- | --- | --- | --- |
| WAV | PCM | 16-bit | 8 kHz or 16 kHz, mono or stereo |
| MP3 | PCM | 16-bit | 8 kHz or 16 kHz, mono or stereo |
| OGG | OPUS | 16-bit | 8 kHz or 16 kHz, mono or stereo |

For stereo audio streams, the left and right channels are split during the transcription. A JSON result file is created for each input audio file. To create an ordered final transcript, use the timestamps that are generated per utterance

## Python Speech SDK Requirements

The Speech SDK for Python is available as a Python Package Index (PyPI) module. The Speech SDK for Python is compatible with Windows, Linux, and macOS.

You must install the [Microsoft Visual C++ Redistributable](https://learn.microsoft.com/en-us/cpp/windows/latest-supported-vc-redist?view=msvc-170&preserve-view=true) for Visual Studio 2015, 2017, 2019, and 2022 for your platform. Installing this package for the first time might require a restart.

On Linux, you must use the x64 target architecture.

Install a version of Python from 3.7 to 3.10. First check the [SDK installation guide](https://learn.microsoft.com/en-us/azure/cognitive-services/speech-service/quickstarts/setup-platform?pivots=programming-language-python&tabs=windows%2Cubuntu%2Cdotnet%2Cjre%2Cmaven%2Cnodejs%2Cmac%2Cpypi) for any more requirements.

## Python Speech SDK Sample code

Main.py will process audio files that are in an azure cloud storage container. This code just prints out the response but you could write to file or database.

There is a notebook file you can use to create azure storage container and another notebook you can use to test the speech sdk with your microphone.

You will need to update with your subscription keys