Using the Speech to Text SDKs

Jared Rhodes
INDEPENDENT CONSULTANT

@qimata www.jaredrhodes.com

Overview

SDK w/ C# Examples

SDK Demo

REST API Overview

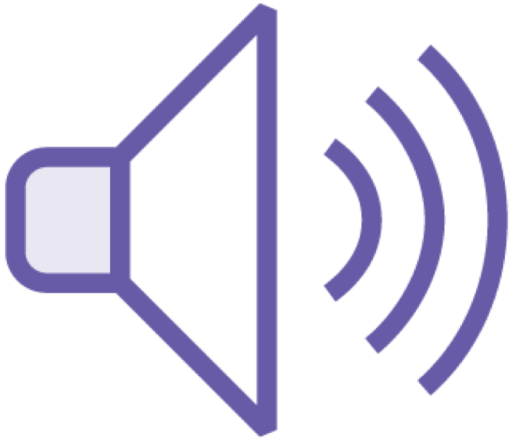
REST Demo w/ Postman

Shared Capabilities

Transcribe a short utterance

Understand speaker intents via LUIS

SDK Capabilities



Transcribe a short utterance

Transcribe a long utterance

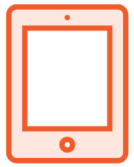
Transcribe streaming audio

Understand speaker intents

C# Platform Support

.NET Framework Windows

.NET Core



Universal Windows Platform

Unity

Speech Recognizer

```
var config = SpeechConfig.FromSubscription(
                        "YourSubscriptionKey",
                        "YourServiceRegion");
using (var recognizer = new SpeechRecognizer(config))
{
    ...
}
```

Audio File Transcription

```
using (var audioInput =
    AudioConfig.FromWavFileInput(“pluralsight.wav"))

using (var recognizer =

            new SpeechRecognizer(config, audioInput))

{

    ...

}
```

Microphone Input

```
using (var audioInput =
    AudioConfig.FromMicrophoneInput("<device id>"))

using (var recognizer =

            new SpeechRecognizer(config, audioInput))

{

    ...

}
```

```
using Windows.Devices.Enumeration;
var devices = await DeviceInformation
        .FindAllAsync(DeviceClass.AudioCapture);

devices.Select(device => device.Id).First();
```

Selecting Audio Devices: UWP

\\\\?\\SWD#MMDEVAPI#{0.0.1.00000000}.{5f23ab69-6181-4f4a-81a4-45414013aac8}#{2eef81be-33fa-4800-9670-1cd474972c3f}

```
using NAudio.CoreAudioApi;
var enumerator = new MMDeviceEnumerator();
var devices = enumerator
            .EnumerateAudioEndPoints
            (DataFlow.Capture, DeviceState.Active);
devices.Select(device => device.Id).First();
```

Selecting Audio Devices: Windows

{0.0.1.00000000}.{5f23ab69-6181-4f4a-81a4-45414013aac8}

```
arecord -L
```

Selecting Audio Devices: Linux

hw:1,0

hw:CARD=CC,DEV=0

```
[[AVAudioSession sharedInstance]
setCategory:AVAudioSessionCategoryRecord
withOptions:AVAudioSessionCategoryOptionAllowBluetooth
error:NULL];
```

Selecting Audio Devices: iOS

Streaming Audio

```
using (var pushStream =
    AudioInputStream.CreatePushStream())

using (var audioInput =

    AudioConfig.FromStreamInput(pushStream))

using (var recognizer =

            new SpeechRecognizer(config, audioInput))

{

    ...

}
```

Custom Audio Input Stream

Identify the Format

```
byte channels = 1;

byte bitsPerSample = 16;

int samplesPerSecond = 16000;

var audioFormat =
    AudioStreamFormat

    .GetWaveFormatPCM(

        samplesPerSecond,
        bitsPerSample,

         channels);
```

- PCM format
- 1 Channel
- 16 Bits per sample
- 16000 samples per second
- 32000 bytes per second
- 2 block align
- 16 bits per sample

Custom Audio Input Stream

Identify the Format

Meet Format Specifications

Create Custom Class

Create Custom Class

```
public class ContosoAudioStream :
    PullAudioInputStreamCallback {

    public int Read(byte[] buffer, uint size) {

        // returns audio data to the caller.

    }

     public void Close() {

        // close and cleanup resources.

     }

};
```

Custom Audio Input Stream

Identify the Format

Meet Format Specifications

Create Custom Class

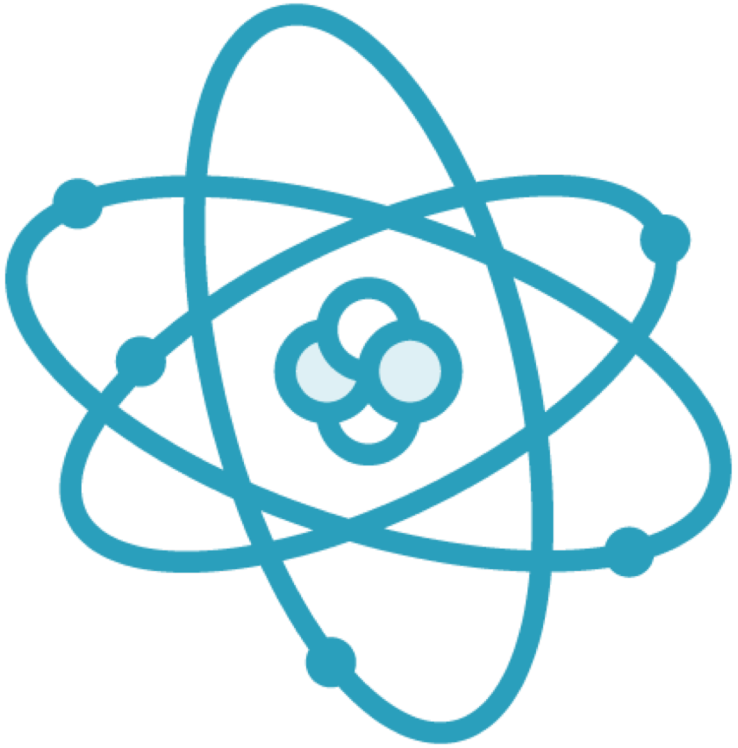
Create Audio Configuration

```
var audioConfig = AudioConfig.FromStreamInput(new
ContosoAudioStream(config), audioFormat);


var speechConfig = SpeechConfig.FromSubscription(...);

var recognizer = new SpeechRecognizer(speechConfig,
audioConfig);
```

Create Audio Configuration

Language Customization

```
var config = SpeechConfig.FromSubscription(
                        "YourSubscriptionKey",
                        "YourServiceRegion");


config.SpeechRecognitionLanguage = "de-DE";
config.OutputFormat = OutputFormat.Detailed;
```



- Authorization Token
- EndpointId
- OutputFormat
- Region
- SpeechRecognitionLanguage
- SubscriptionKey

Transcribe a Short Utterance

```
var result = await recognizer.RecognizeOnceAsync();
```

Handle the Result

```
result.Reason == ResultReason.RecognizedSpeech

result.Reason == ResultReason.NoMatch

result.Reason == ResultReason.Canceled

var text = result.Text;

var cancellation = CancellationDetails.FromResult(result);

cancellation.Reason == CancellationReason.Error
```

Continuous Results

```
await recognizer.StartContinuousRecognitionAsync();

...

await recognizer.StopContinuousRecognitionAsync();
```

Continuous Results

```
recognizer.Recognizing += (s, e) =>

recognizer.Recognized += (s, e) =>

recognizer.Canceled += (s, e) =>

recognizer.SessionStarted += (s, e) =>

recognizer.SessionStopped += (s, e) =>
```

Language Understanding

```
using (var recognizer = new IntentRecognizer(config))
{
    ...
}
```

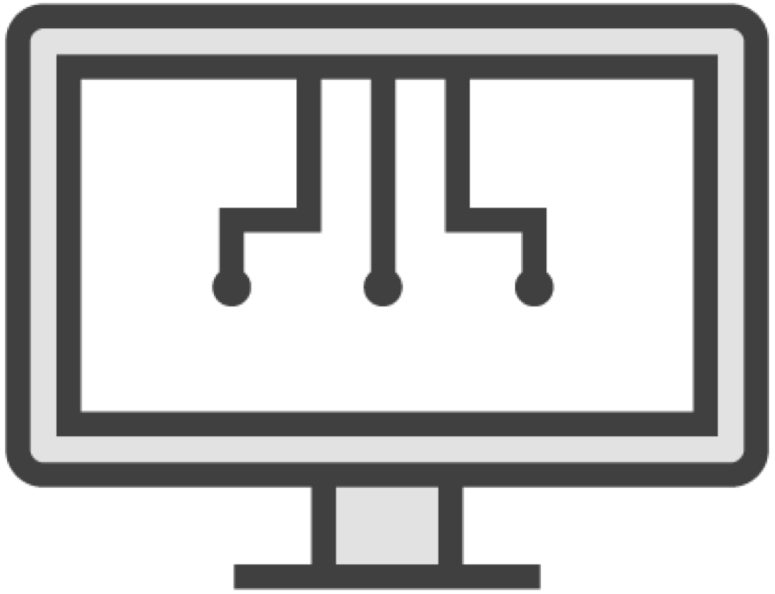
Language Understanding

```
var model =
    LanguageUnderstandingModel.FromAppId(“LUISAppId");

recognizer.AddIntent(model,“LUISIntentName1","id1");

recognizer.AddIntent(model,"LUISIntentName2","id2");

recognizer.AddIntent(model,"LUISIntentName3",“id3");
```

Language Understanding

```
result.Reason == ResultReason.RecognizedIntent
var intentId = result.IntentId;
var luisJson = result.Properties
    .GetProperty(PropertyId
        .LanguageUnderstandingServiceResponse_JsonResult);
```

Other SDKs



C++: Windows / Linux

Java: Android / Windows / Linux

Node.js: Windows / Linux / macOS

JavaScript: Browser

Objective-C: iOS

Python: Windows / Linux / macOS

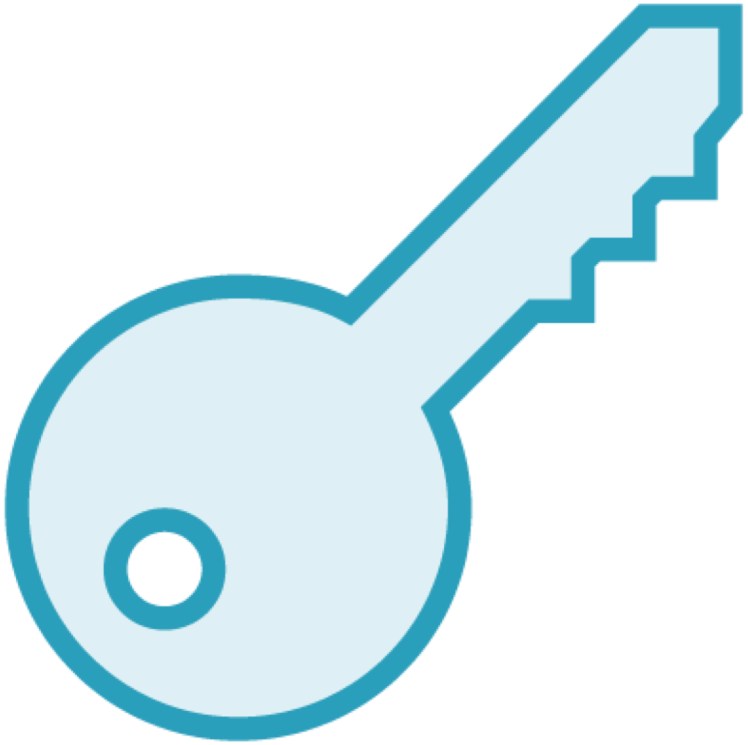
Demo

Create an Account

Transcribe a short utterance

Speech to Text REST API

Authentication



Each request to the REST API requires an authorization header

Ocp-Apim-Subscription-Key header is supported

Bearer Tokens are only valid for 10 minutes

```
POST /sts/v1.0/issueToken HTTP/1.1

Ocp-Apim-Subscription-Key: YOUR_SUBSCRIPTION_KEY

Host: westus.api.cognitive.microsoft.com

Content-type: application/x-www-form-urlencoded

Content-Length: 0
```

Create Your Access Token

Replace YOUR_SUBSCRIPTION_KEY with your Speech Service subscription key

If your subscription isn't in the West US region, replace the Host header with your region's host name

Using the Access Token

```
POST /cognitiveservices/v1 HTTP/1.1

Authorization: Bearer YOUR_ACCESS_TOKEN

Host: westus.tts.speech.microsoft.com

Content-type: audio/wav; codecs=audio/pcm; samplerate=16000

{Body}
```

Query Parameters

Language

Format

Profanity

Headers

Ocp-Apim-Subscription-Key

Authorization

Content-Type

Transfer-Encoding

Expect

Accept

Audio Formats

WAV

OGG

Sample Request URL

```
speech/recognition/conversation/cognitiveservices/v1?language=en
-US&format=detailed
```

Sample Request

```
POST speech/recognition/conversation/cognitiveservices/v1?language=en-US&format=detailed HTTP/1.1
Accept: application/json;text/xml
Content-Type: audio/wav; codecs=audio/pcm; samplerate=16000
Ocp-Apim-Subscription-Key: YOUR_SUBSCRIPTION_KEY
Host: westus.tts.speech.microsoft.com
Transfer-Encoding: chunked
Expect: 100-continue
```

Response Status Codes

100

Continue

200

Ok

400

Bad Request

401

Unauthorized

403

Forbidden

Simple Response Parameters

Duration

DisplayText

Offset

Recognition Status

Recognition Status

- Success
- NoMatch
- InitialSilenceTimeout
- BabbleTimeout
- Error

Simple Response


```
{
  "RecognitionStatus": "Success",
  "DisplayText": "Remind me to follow the author.",
  "Offset": "1236645672289",
  "Duration": "1236645672289"
}
```


NBest Fields

Confidence

Lexical

ITN

MaskedITN

Display

Detailed Response


```
{
  "RecognitionStatus": "Success",
  "DisplayText": "Remind me to buy 5 pencils.",
  "Offset": "1236645672289",
  "Duration": "1236645672289“,
  “NBest": [ ... ]
}
```

Detailed Response


```
[
    {
      "Confidence" : "0.87",
      "Lexical" : "remind me to buy five pencils",
      "ITN" : "remind me to buy 5 pencils",
      "MaskedITN" : "remind me to buy 5 pencils",
      "Display" : "Remind me to buy 5 pencils.",
    },
    ...
]
```

Detailed Response


```
[
        ...,
        {
          "Confidence" : "0.54",
          "Lexical" : "rewind me to buy five pencils",
          "ITN" : "rewind me to buy 5 pencils",
          "MaskedITN" : "rewind me to buy 5 pencils",
          "Display" : "Rewind me to buy 5 pencils.",
        }
]
```

REST API Capabilities

- Transcribe a short utterance
- Create Accuracy Tests
- Batch transcription
- Upload datasets for model adaptation
- Create & manage speech models
- Create & manage model deployments
- Manage subscriptions

Demo

Transcribe a short utterance

Change the detail level

Review

SDK w/ C# Examples

SDK Demo

REST API Overview

REST Demo w/ Postman