

# SIT789 – Robotics, Computer Vision and Speech Processing

## Pass Task 9.1: Speech recognition with IBM Watson

### Objectives

The objectives of this lab include:

- Getting hands-on experience with speech-to-text service developed by IBM Watson
- Building a speech recognition application using Watson's API

### Tasks

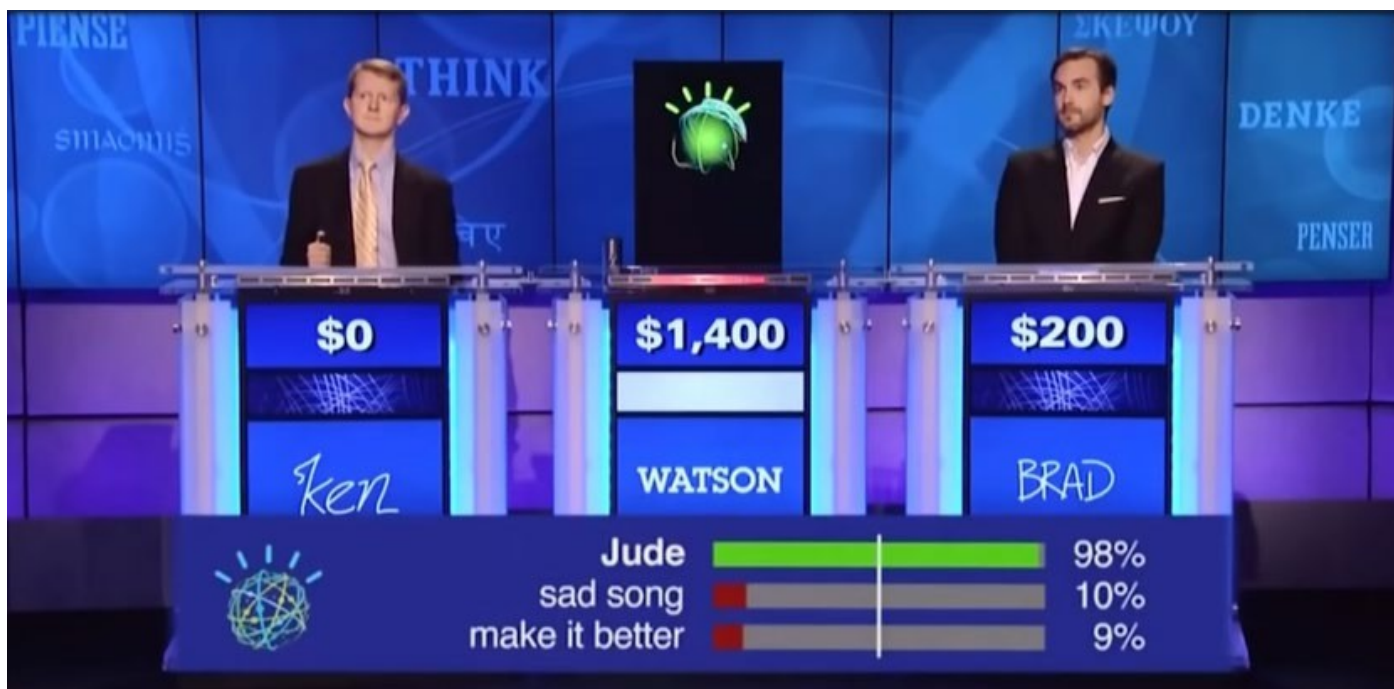
#### 1. Speech-to-text with IBM Watson

IBM cloud is a cloud computing platform that provides solutions for real-world problems and drives business value with applications, infrastructure and services across IBM Watson (AI), IoT, mobile, and more.

The screenshot shows the IBM Cloud Services catalog page. The browser address bar displays <https://cloud.ibm.com/catalog#services>. The page features a dark sidebar on the left with navigation links: Catalog, IBM Cloud catalog, Services (highlighted), Software, and Consulting. Below these is a 'Category' section with a list of checkboxes: Compute, Containers, Networking, Storage, AI / Machine Learning, Analytics, Blockchain, Databases, Developer Tools, Logging and Monitoring, and Integration. The main content area is titled 'Services' and includes a search bar. Below the search bar, it says 'Explore our broad portfolio of managed services for infrastructure, developer tools, and more to build your apps on the public cloud.' and 'All Categories 169 items'. A grid of service cards is displayed, each with an icon, name, provider, description, and status. The 'Actifio GO' card is highlighted with a blue border. A 'FEEDBACK' button is visible on the right side of the grid.

Service Name	Provider	Description	Status
Accern-API	Third party • Developer Tools	Get the most advanced breaking news technology for your investment strategies.	Free
AccountScore	Third party • Analytics	AccountScore Open Banking & transaction analytics	Free
Actifio GO	Third party • Storage	SaaS platform for backup/DR/cloning/migration of Enterprise workloads in IBM Cloud	IAM-enabled
Alloy	Third party • Developer Tools	API for identity (KYC, AML & fraud)	Free
Analytics Engine	IBM • Analytics	Flexible framework to deploy Hadoop and Spark analytics applications.	Lite • Free • IAM-enabled • Service Endpoint Supported
Annotator for Clinical Data	IBM • AI / Machine Learning	Analyze text to extract medical codes and concepts such as diseases, lab values, medications, procedures and...	Lite • Free • IAM-enabled • Service Endpoint Supported
API Connect	IBM • Integration	Create, manage, enforce, and run APIs.	Lite • Free • IAM-enabled
API Gateway	IBM • Developer Tools	A developer-focused solution for creating, securing, and sharing API proxies and custom APIs backed by IB...	Lite • Free • IAM-enabled

Watson was originally a question answering computer system built by IBM.



With Watson on IBM Cloud, you have access to the widest range of cognitive technologies available today to quickly and securely build smart applications. In this workshop, you will learn to use speech-to-text service by IBM Watson. First, you need to register an IBM Cloud account by following instructions given at <https://cloud.ibm.com/login>. You need to use your Deakin email for account registration.



Log in to IBM Cloud

Don't have an account? [Create an account](#)

Sign in with

IBMId

[Continue](#)

Forgot ID?

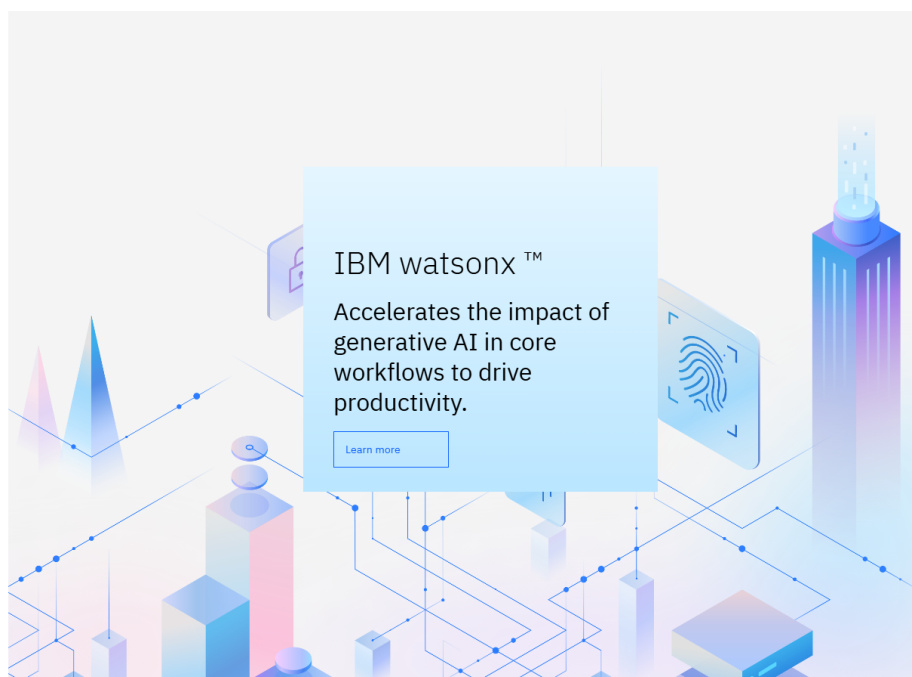
☐ Remember ID

Or

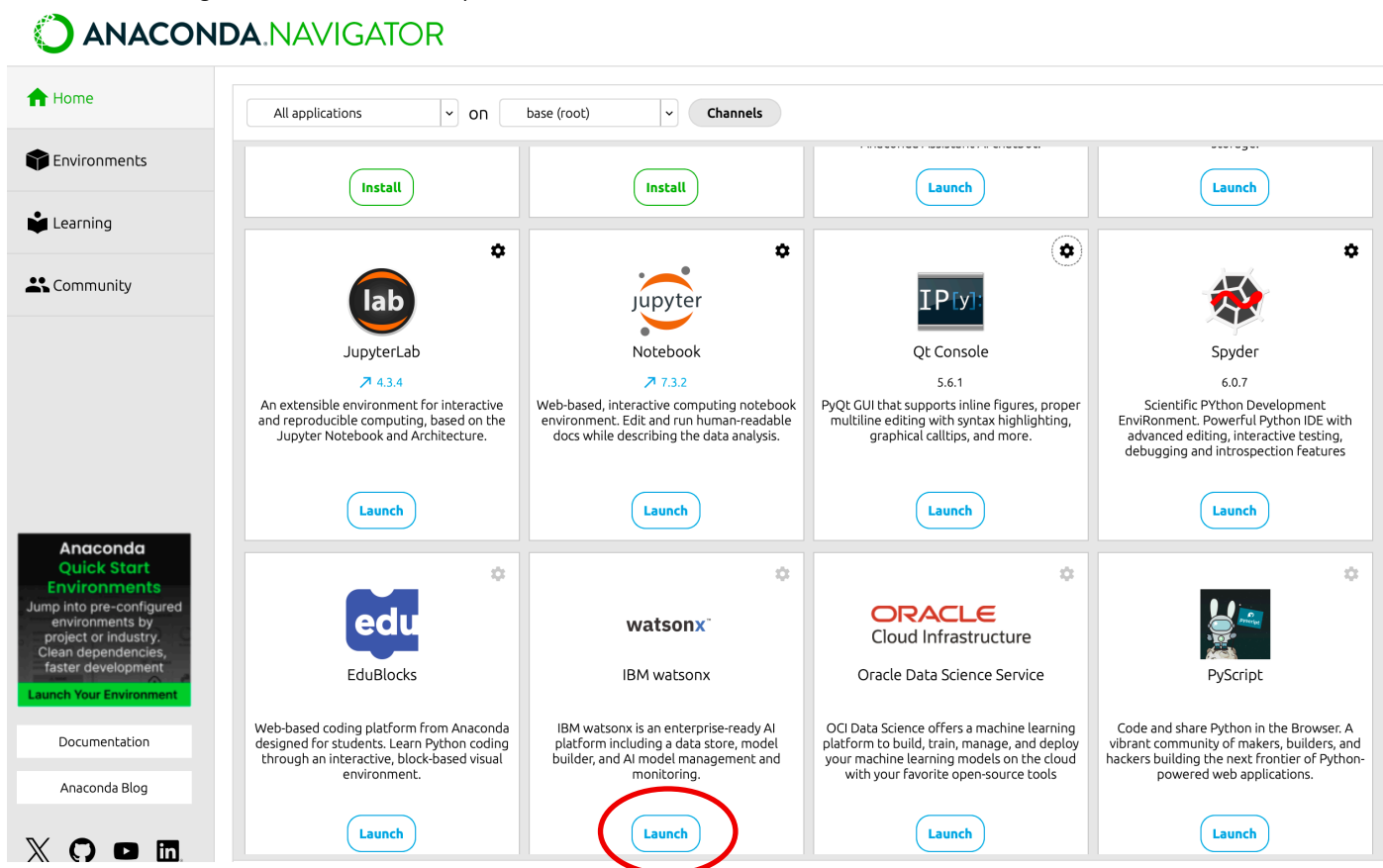
[Continue with Red Hat](#)

[Continue with Google](#)

© Copyright IBM Corp. 2014, 2025. All rights reserved. [Privacy](#)



You can also navigate to IBM Cloud directly from the anaconda navigator (which can be launched by the command: “anaconda-navigator” from a terminal)



While registering your account, you can choose “Personal” for the Account Type (see the figure below).

# Create an IBM Cloud account

Already have an IBM Cloud account? [Log in](#)

Account information

Email

→

Enter an email address.

Password

👁️

OR

Sign up with Google

Sign up with Red Hat

Next

↓

Verify email

Personal information

Account notice

⌚ Unlock the full IBM Cloud Catalog

Get started with everything you need to take your projects to the next level. Including always-free products like IBM Watson® APIs. They never expire and you can't be charged for them—ever.

⌚ 200 USD Cloud Credit for free

Try any IBM Cloud product with \$200 in credit to use over the next 30 days.

⌚ Get started for free

We ask for your credit card to verify your identity. You will not be charged for any usage within the free tier. Learn more: [www.ibm.com/cloud/free](https://www.ibm.com/cloud/free).

You then may be asked to provide credit card information. Note that we will use a free service for this workshop and thus your will not be charged even if you provide a credit card. However, in case that you do not want to provide a credit card, you can apply for a free feature code as follows.

- Under Account Information, select Registration with a code (see the highlighted ellipse in the figure above)
- Apply for a free feature code by following the instructions [here](#). Note that you need to use your Deakin email to get a free feature code.

Finish setting up your account

You're just one step away from creating your account.

Enter your billing information and payment method to confirm your identity and secure your account.

**Account information**

Register with a code ⓘ

Account type

☐ Company ☒ Personal

Next ↓

Billing information

Credit card information

Upgrade account ⓘ

After obtaining the feature code, come back to the Account settings page and enter the code in the provided text box as shown in the figure below.

Finish setting up your account

You're just one step away from creating your account.

Enter your billing information and payment method to confirm your identity and secure your account.

**Apply code**

Register without a code ⓘ

Enter the code that has been provided to you. You can apply the code to only one account, and it can't be removed.

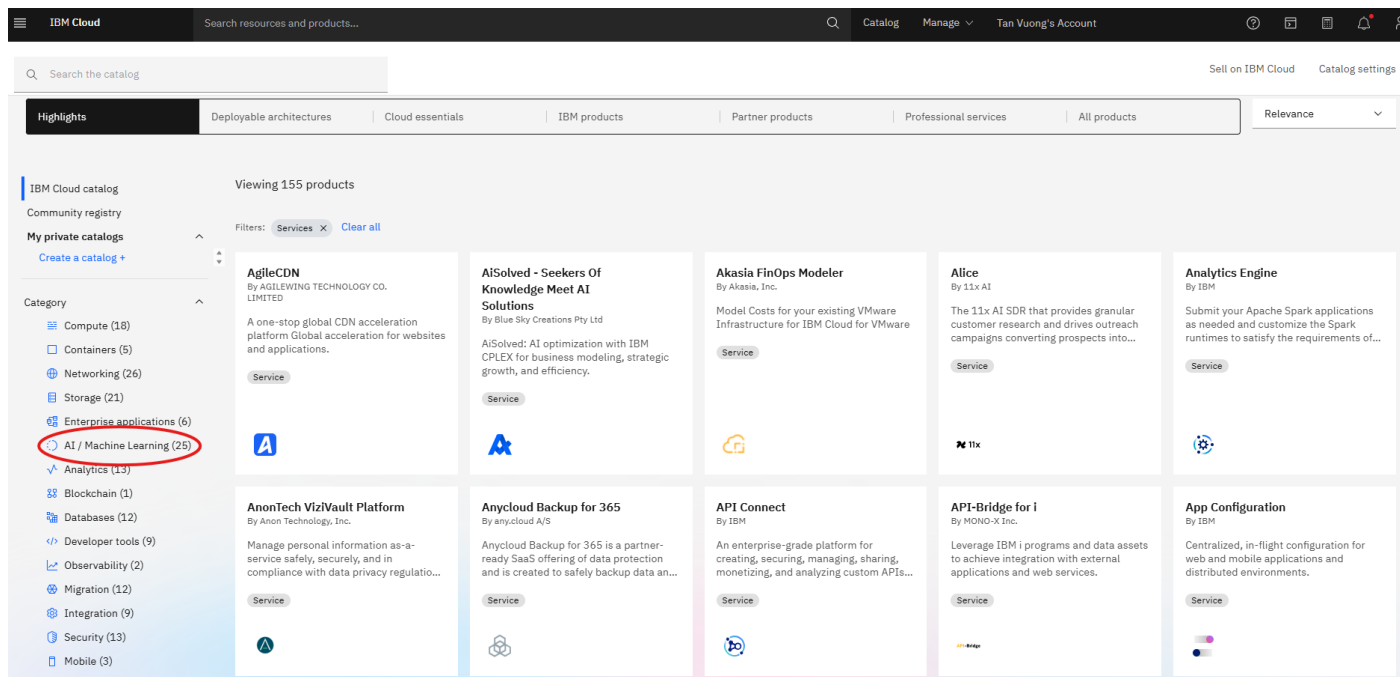
Enter code

Account ID: f9c08d5158f0402599414fa541652bda ⓘ

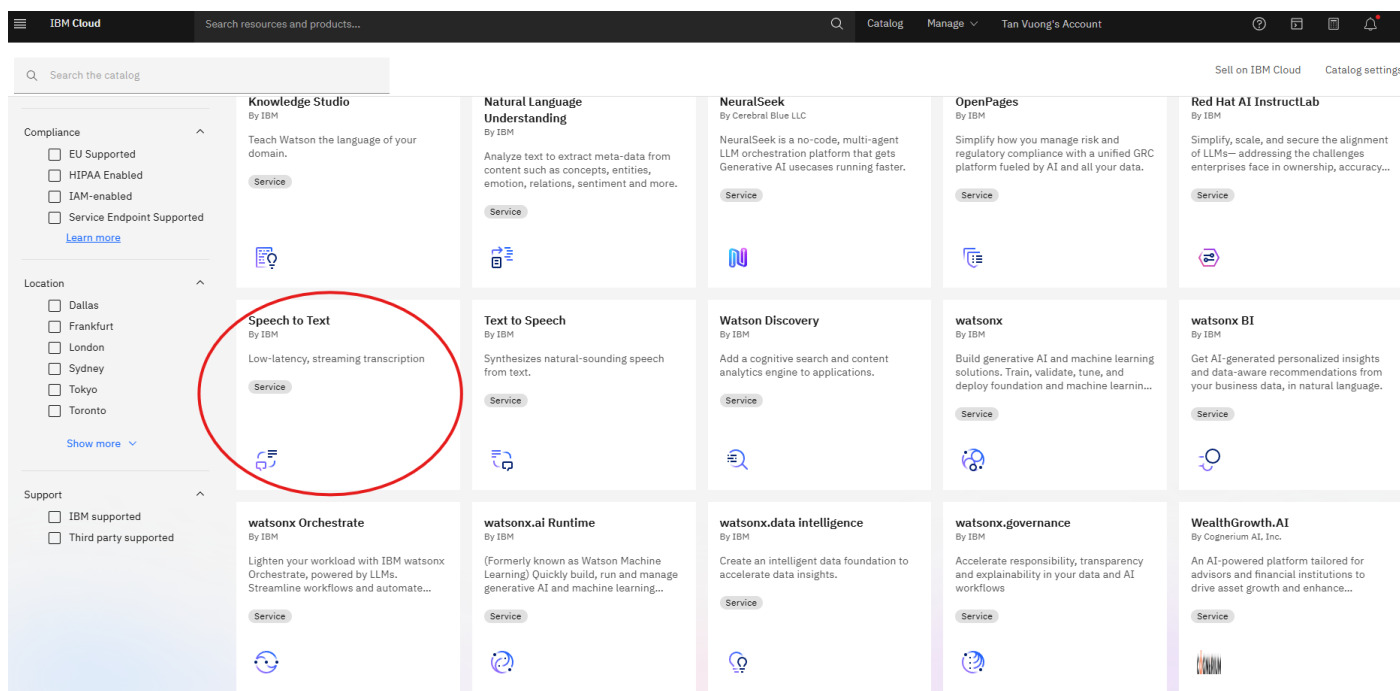
Upgrade account ⓘ

After registering your account, you can find services provided by IBM Cloud at <https://cloud.ibm.com/catalog#services>.

You then select **AI/Machine Learning** for **Category** as shown below,



and then **Speech to Text** as shown below,



Select **Lite** with free pricing for your **Plan** (unless you wish more advanced features) and then click **Log in** (if you are not logged-in) as shown below.

The screenshot shows the IBM Cloud catalog page for the 'Speech to Text' service. The page is titled 'Speech to Text' and includes a 'Create' button. The 'Pricing plans' section displays three plans: Lite, Standard, and Premium. The 'Lite' plan is highlighted with a red circle and a 'choose this' label. The 'Lite' plan details include '500 Minutes per Month' and 'Free' pricing. The 'Standard' plan details include 'Standard Minutes' and 'Multi-Tiered' pricing. The 'Premium' plan details include 'Everything in Standard plus...' and 'Usage and Training Data is Private + Stored in an Isolated Single Tenant Environment'. The right sidebar shows the 'Summary' section with 'Speech to Text' and 'Plan: Lite'. The 'Log in' button is circled in red.

Plan	Features	Pricing
Lite	500 Minutes per Month The Lite plan gets you started with 500 minutes per month at no cost. When you upgrade to a paid plan, you will get access to Customization capabilities. Lite plan services are deleted after 30 days of inactivity.	Free
Standard	Standard Minutes Multi-Tiered	Click to view tiers and pricing detail
Premium	Everything in Standard plus... Usage and Training Data is Private + Stored in an Isolated Single Tenant Environment High Availability and Service Level Uptime Guarantee HIPAA - Washington DC Only IBM Cloud Service Endpoints	

After being logged in successfully, you will be given a name for your service (as highlighted in the following figure). You then need to click **Create** to create a service,

The screenshot shows the 'Configure your resource' page for the 'Speech to Text' service. The 'Service name' field is highlighted with a red circle and contains the text 'Speech to Text-jc'. The 'Select a resource group' dropdown is set to 'Default'. The right sidebar shows the 'Summary' section with 'Speech to Text' and 'Plan: Lite'. The 'Service name: Speech to Text-jc' is circled in red. The 'Create' button is visible at the bottom of the sidebar.

Configure your resource

Service name:

Select a resource group:

Tags:

You are now linked to a tutorial page that shows you how to use the **Speech to Text** service.

The screenshot shows the IBM Cloud console interface. At the top, there's a navigation bar with 'IBM Cloud', a search bar, and user account information. Below the navigation bar, the main content area is titled 'Speech to Text-jc'. On the left, there's a sidebar with 'Manage' and 'Getting started' (selected). The main content area has a heading 'Getting started with Speech to Text' and a sub-heading 'Last Updated: 2025-02-21'. The text describes the IBM Watson® Speech to Text service and provides a note about using the 'curl' command-line utility. Below the text, there's a video player with a play button and the title 'Get started with Watson Speech to Text'. At the bottom, there's a section titled 'Before you begin'.

Following the instructions given in the tutorial, you need to test the service with the speech data in the supplied [SpeechtoTextData](#) folder. The supplied data is collected from various well-known datasets including: the [CMU ARCTIC](#), the [LibriSpeech](#), and the [Valentini](#) dataset. Speech clips also vary in their quality, e.g., clips from the Valentini dataset contains various levels of noise. Transcripts for the clips are also provided.

#### Note:

1. Depending on the format of clips, you need to choose the right content type. In particular, for .flac files, you need to use "Content-Type: audio/flac" for the [header](#) option. For .wav files, you need to set "Content-Type: audio/wav".
2. Command lines need to be executed one by one, i.e., one command is executed at a time. For example, after entering the API key, you will be asked to provide the content type (e.g., wav).
3. You could make your own data to test this service. However, make sure that your test clips are sampled at (or above) 16kHz.

Below is a screenshot of the speech recognition results of the file [arctic\\_a005.wav](#).

```
{
  "result_index": 0,
  "results": [
    {
      "final": true,
      "alternatives": [
        {
          "transcript": "will we ever forget it ",
          "confidence": 0.95
        }
      ]
    }
  ]
}
```



## 2. Speech recognition using Speech to Text API

In this section, we will learn how to use **Speech to Text** API for speech recognition. You can find the documentation of this API at <https://cloud.ibm.com/apidocs/speech-to-text?code=python>. The API is compatible with different programming languages such as Java, Python, Ruby, etc.

First, you need to install **watson\_developer\_cloud** and **ibm-watson** by using the following command lines.

```
pip install --upgrade watson_developer_cloud
```

```
pip install --upgrade "ibm-watson>=4.6.0"
```

You then need to authenticate your service using your API key and url as follows.

```
from ibm_watson import SpeechToTextV1
from ibm_cloud_sdk_core.authenticators import IAMAuthenticator

authenticator = IAMAuthenticator('{APIkey}') #replace {APIkey} by your API key
speech_to_text = SpeechToTextV1(authenticator=authenticator)
speech_to_text.set_service_url('{url}') #replace {url} by your URL
```

You can test the **Speech to Text** service with a speech clip, e.g., [arctic\\_a0005.wav](#), as follows.

```
import json
with open('SpeechtoTextData/arctic_a0005.wav', 'rb') as audio_file:
    speech_recognition_results = speech_to_text.recognize(
        audio = audio_file,
        content_type='audio/wav').get_result()
print(json.dumps(speech_recognition_results, indent = 2))
```

The speech recognition result can be saved to file (in json format) using the command:

```
with open('SpeechtoTextData/arctic_a0005.json', 'w') as outfile:
    json.dump(speech_recognition_results, outfile)
```

To load the result from file, you can use the following code

```
with open('SpeechtoTextData/arctic_a0005.json') as infile:
    data = json.load(infile) # load data from a json file
print(data)
```

Like Section 1, your task is to test the **Speech to Text** service using the provided API with the speech data supplied in the [SpeechtoTextData](#) folder and discuss your observations on recognition results.

## Submission instructions

1. Perform tasks required in Section 1 and 2.
2. Complete the supplied answer sheet and submit it (in .pdf format) to OnTrack.