SIT789 - Applications of Computer Vision and Speech Processing
Pass Task 9.1: Speech recognition with IBM Watson

```
:\Users\sagar\Documents\SpeechtoTextData>
:\Users\sagar\Documents\SpeechtoTextData> curl -X POST -u "apikey:AIdYnFXSjDKkwpGaobgf93L7_X
uQStfLlUnNZaZRTYGA" ^
More? --header "Content-Type: audio/flac" ^
More? --data-binary @367-130732-0000.flac ^
More? "https://api.au-syd.speech-to-text.watson.cloud.ibm.com/instances/8922baf7-1c62-4e42-93
14-46f4a8fde0ce/v1/recognize"
{
  "result_index": 0,
  "results": [
    {
      "final": true,
      "alternatives": [
        {
          "transcript": "lobsters and law officers ",
          "confidence": 0.74
        }
      ]
    }
  ]
}
:\Users\sagar\Documents\SpeechtoTextData>
:\Users\sagar\Documents\SpeechtoTextData> curl -X POST -u "apikey:AIdYnFXSjDKkwpGaobgf93L7_X
uQStfLlUnNZaZRTYGA" ^
More? --header "Content-Type: audio/flac" ^
More? --data-binary @367-130732-0001.flac ^
More? "https://api.au-syd.speech-to-text.watson.cloud.ibm.com/instances/8922baf7-1c62-4e42-93
14-46f4a8fde0ce/v1/recognize"
{
  "result_index": 0,
  "results": [
    {
      "final": true,
      "alternatives": [
        {
          "transcript": "when is a lobster not a lobster when it is a crayfish ",
          "confidence": 0.87
        }
      ]
    }
  ]
}
```

```
C:\Users\sagar\Documents\SpeechtoTextData> curl -X POST -u "apikey:AIdYnFXSjDKkwpGaobgf93L7_X
uQStfLlUnNZaZRTYGA" ^
More? --header "Content-Type: audio/flac" ^
More? --data-binary @367-130732-0004.flac ^
More? "https://api.au-syd.speech-to-text.watson.cloud.ibm.com/instances/8922baf7-1c62-4e42-93
14-46f4a8fde0ce/v1/recognize"
{
    "result_index": 0,
    "results": [
        {
            "final": true,
            "alternatives": [
                {
                    "transcript": "a book could be written about this restaurant and then all woul
d not be told for all its secrets can never be known ",
                    "confidence": 0.92
                }
            ]
        }
    ]
}
C:\Users\sagar\Documents\SpeechtoTextData> curl -X POST -u "apikey:AIdYnFXSjDKkwpGaobgf93L7_X
uQStfLlUnNZaZRTYGA" ^
More? --header "Content-Type: audio/wav" ^
More? --data-binary @arctic_a0001.wav ^
More? "https://api.au-syd.speech-to-text.watson.cloud.ibm.com/instances/8922baf7-1c62-4e42-93
14-46f4a8fde0ce/v1/recognize"
{
    "result_index": 0,
    "results": [
        {
            "final": true,
            "alternatives": [
                {
                    "transcript": "author of the danger trail Phillips steals etcetera ",
                    "confidence": 0.78
                }
            ]
        }
    ]
}
C:\Users\sagar\Documents\SpeechtoTextData> curl -X POST -u "apikey:AIdYnFXSjDKkwpGaobgf93L7_X
uQStfLlUnNZaZRTYGA" ^
More? --header "Content-Type: audio/wav" ^
More? --data-binary @arctic_a0003.wav ^
More? "https://api.au-syd.speech-to-text.watson.cloud.ibm.com/instances/8922baf7-1c62-4e42-93
14-46f4a8fde0ce/v1/recognize"
{
    "result_index": 0,
    "results": [
        {
            "final": true,
            "alternatives": [
                {
                    "transcript": "for the twentieth time that evening the two men shook hands ",
                    "confidence": 0.86
                }
            ]
        }
    ]
```

```
C:\Users\sagar\Documents\SpeechtoTextData> curl -X POST -u "apikey:AIdYnFXSjDKkwpGaobgf93L7_X
uQStfLlUnNZaZRTYGA" ^
More? --header "Content-Type: audio/wav" ^
More? --data-binary @arctic_a0005.wav ^
More? "https://api.au-syd.speech-to-text.watson.cloud.ibm.com/instances/8922baf7-1c62-4e42-93
14-46f4a8fde0ce/v1/recognize"
{
   "result_index": 0,
   "results": [
      {
         "final": true,
         "alternatives": [
            {
               "transcript": "will we ever forget it ",
               "confidence": 0.95
            }
         ]
      }
   ]
}
```

```
C:\Users\sagar\Documents\SpeechtoTextData> curl -X POST -u "apikey:AIdYnFXSjDKkwpGaobgf93L7_X
uQStfLlUnNZaZRTYGA" ^
More? --header "Content-Type: audio/wav" ^
More? --data-binary @p232_009.wav ^
More? "https://api.au-syd.speech-to-text.watson.cloud.ibm.com/instances/8922baf7-1c62-4e42-93
14-46f4a8fde0ce/v1/recognize"
{
   "result_index": 0,
   "results": [
      {
         "final": true,
         "alternatives": [
            {
               "transcript": "there is according to legend a boiling pot of gold at one end "
,
               "confidence": 0.98
            }
         ]
      }
   ]
}
```

```
C:\Users\sagar\Documents\SpeechtoTextData> curl -X POST -u "apikey:AIdYnFXSjDKkwpGaobgf93L7_X
uQStfLlUnNZaZRTYGA" ^
More? --header "Content-Type: audio/wav" ^
More? --data-binary @p232_010.wav ^
More? "https://api.au-syd.speech-to-text.watson.cloud.ibm.com/instances/8922baf7-1c62-4e42-93
14-46f4a8fde0ce/v1/recognize"
{
    "result_index": 0,
    "results": [
        {
            "final": true,
            "alternatives": [
                {
                    "transcript": "no one ever finds ",
                    "confidence": 0.19
                }
            ]
        }
    ]
}
```

```
C:\Users\sagar\Documents\SpeechtoTextData> curl -X POST -u "apikey:AIdYnFXSjDKkwpGaobgf93L7_X
uQStfLlUnNZaZRTYGA" ^
More? --header "Content-Type: audio/wav" ^
More? --data-binary @p232_014.wav ^
More? "https://api.au-syd.speech-to-text.watson.cloud.ibm.com/instances/8922baf7-1c62-4e42-93
14-46f4a8fde0ce/v1/recognize"
{
    "result_index": 0,
    "results": [
        {
            "final": true,
            "alternatives": [
                {
                    "transcript": "the Hebrews it was a token that there would be no more universa
l floods ",
                    "confidence": 0.92
                }
            ]
        }
    ]
}
```

```
C:\Users\sagar\Documents\SpeechtoTextData> curl -X POST -u "apikey:AIdYnFXSjDKkwpGaobgf93L7_X
uQStfLlUnNZaZRTYGA" ^
More? --header "Content-Type: audio/wav" ^
More? --data-binary @p232_030.wav ^
More? "https://api.au-syd.speech-to-text.watson.cloud.ibm.com/instances/8922baf7-1c62-4e42-93
14-46f4a8fde0ce/v1/recognize"
{
    "result_index": 0,
    "results": [
        {
            "final": true,
            "alternatives": [
                {
                    "transcript": "they were described as being in a serious but stable condition
yesterday ",
                    "confidence": 0.94
                }
            ]
        }
    ]
```

API:

```python
import json
with open('/content/drive/My Drive/SpeechtoTextData/arctic_a0003.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))
import json
with open('/content/drive/My Drive/SpeechtoTextData/arctic_a0001.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))
import json
with open('/content/drive/My Drive/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))
import json
with open('/content/drive/My Drive/SpeechtoTextData/p232_009.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))
import json
with open('/content/drive/My Drive/SpeechtoTextData/p232_010.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))
import json
with open('/content/drive/My Drive/SpeechtoTextData/p232_014.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))
import json
with open('/content/drive/My Drive/SpeechtoTextData/p232_030.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))
import json
with open('/content/drive/My Drive/SpeechtoTextData/367-130732-0000.flac', 'rb') as audio_file:
    speech_recognition_results = speech_to_text.recognize(
    audio = audio_file,
    content_type='audio/flac').get_result()
    print(json.dumps(speech_recognition_results, indent = 2))
with open('/content/drive/My Drive/SpeechtoTextData/367-130732-0001.flac', 'rb') as audio_file:
    speech_recognition_results = speech_to_text.recognize(
    audio = audio_file,
    content_type='audio/flac').get_result()
    print(json.dumps(speech_recognition_results, indent = 2))
with open('/content/drive/My Drive/SpeechtoTextData/367-130732-0004.flac', 'rb') as audio_file:
    speech_recognition_results = speech_to_text.recognize(
    audio = audio_file,
    content_type='audio/flac').get_result()
    print(json.dumps(speech_recognition_results, indent = 2))
```

```json
{
  "result_index": 0,
  "results": [
    {
      "final": true,
      "alternatives": [
        {
          "transcript": "for the twentieth time that evening the two men shook hands ",
          "confidence": 0.86
        }
      ]
    }
  ]
}
{
  "result_index": 0,
```

```json
    "results": [
      {
        "final": true,
        "alternatives": [
          {
            "transcript": "author of the danger trail Phillips steals
etcetera ",
            "confidence": 0.78
          }
        ]
      }
    ]
}
{
  "result_index": 0,
  "results": [
      {
        "final": true,
        "alternatives": [
          {
            "transcript": "will we ever forget it ",
            "confidence": 0.95
          }
        ]
      }
    ]
}
{
  "result_index": 0,
  "results": [
      {
        "final": true,
        "alternatives": [
          {
            "transcript": "there is according to legend a boiling pot of
gold at one end ",
            "confidence": 0.98
          }
        ]
      }
    ]
}
{
  "result_index": 0,
  "results": [
      {
        "final": true,
        "alternatives": [
          {
            "transcript": "no one ever finds ",
            "confidence": 0.19
          }
        ]
      }
```

```
    ]
  }
  {
    "result_index": 0,
    "results": [
      {
        "final": true,
        "alternatives": [
          {
            "transcript": "the Hebrews it was a token that there would be
no more universal floods ",
            "confidence": 0.92
          }
        ]
      }
    ]
  }
  {
    "result_index": 0,
    "results": [
      {
        "final": true,
        "alternatives": [
          {
            "transcript": "they were described as being in a serious but
stable condition yesterday ",
            "confidence": 0.94
          }
        ]
      }
    ]
  }
  {
    "result_index": 0,
    "results": [
      {
        "final": true,
        "alternatives": [
          {
            "transcript": "lobsters and law officers ",
            "confidence": 0.74
          }
        ]
      }
    ]
  }
  {
    "result_index": 0,
    "results": [
      {
        "final": true,
        "alternatives": [
          {
```

```
            "transcript": "when is a lobster not a lobster when it is a
crayfish ",
            "confidence": 0.87
          }
        ]
      }
    ]
  }
}
{
  "result_index": 0,
  "results": [
    {
      "final": true,
      "alternatives": [
        {
          "transcript": "a book could be written about this restaurant
and then all would not be told for all its secrets can never be known ",
          "confidence": 0.92
        }
      ]
    }
  ]
}
```

The speech recognition of IBM Watson was able to convert the speech to text of the given dataset very efficiently, except for one particular case, where the speech to text couldn't recognize the data completely due to the noise present in the data provided ' p232_010.wav'.