Google Speech to Text

Demira Dimitrova - Iohan Sardinha - Jordi Bosh - Leonardo Menti

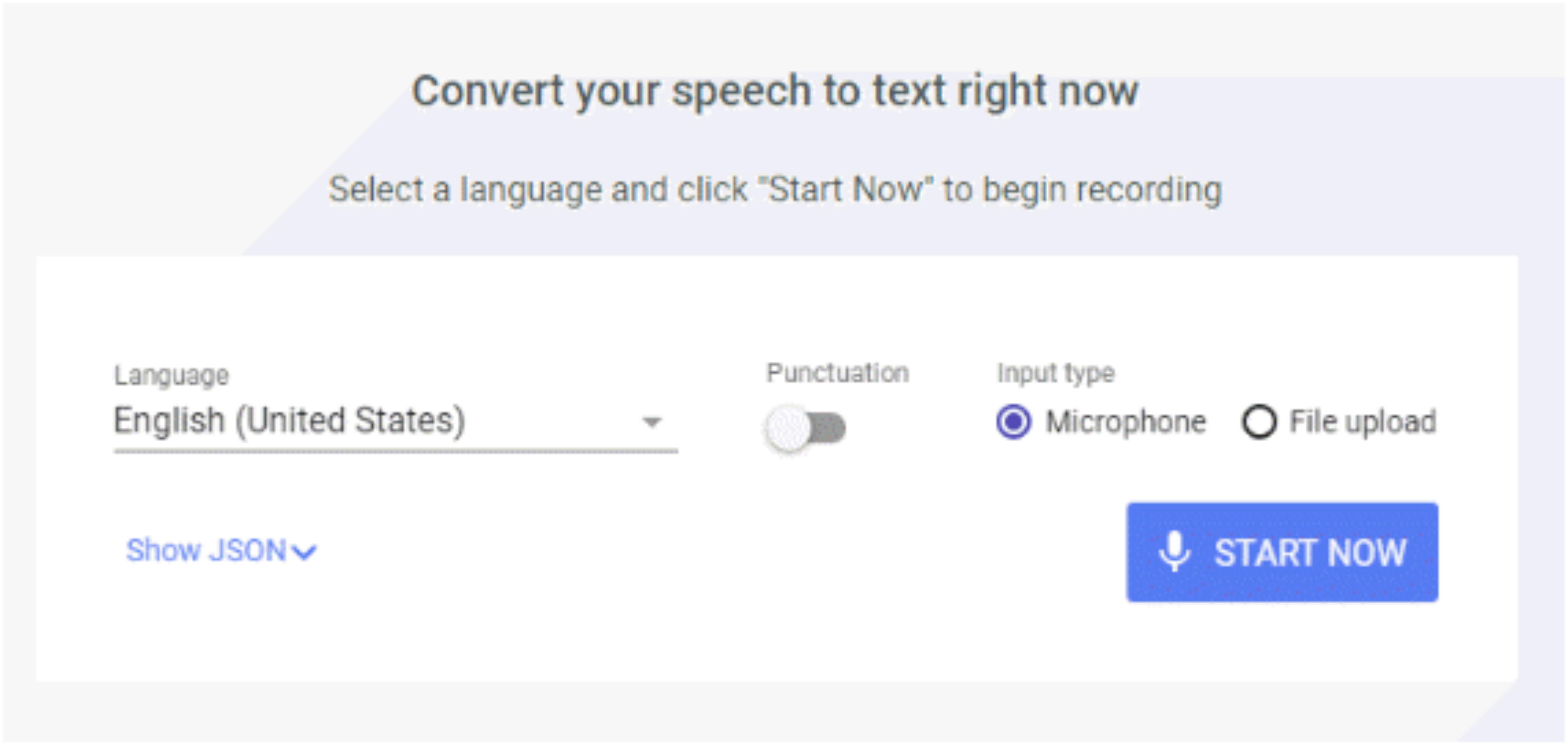
17/05/2022

What is Google Speech to Text

Convert speech into text with an API powered by the best of Google’s AI research and technology

Google Cloud

Speech to Text


Convert your speech to text right now

Select a language and click "Start Now" to begin recording

Language: English (United States)

Punctuation

Input type: ● Microphone ○ File upload

Show JSON

START NOW

What is Google Speech to Text

Speech to Text API has 3 main methods:

- Synchronous Recognition (REST and gRPC) - performs recognition on that data, and returns results after all audio has been processed
- Asynchronous Recognition (REST and gRPC) - initiates a Long Running Operation. Using this operation, you can periodically poll for recognition results
- Streaming Recognition (gRPC only) - on audio data provided within a gRPC bi-directional stream

Use Cases

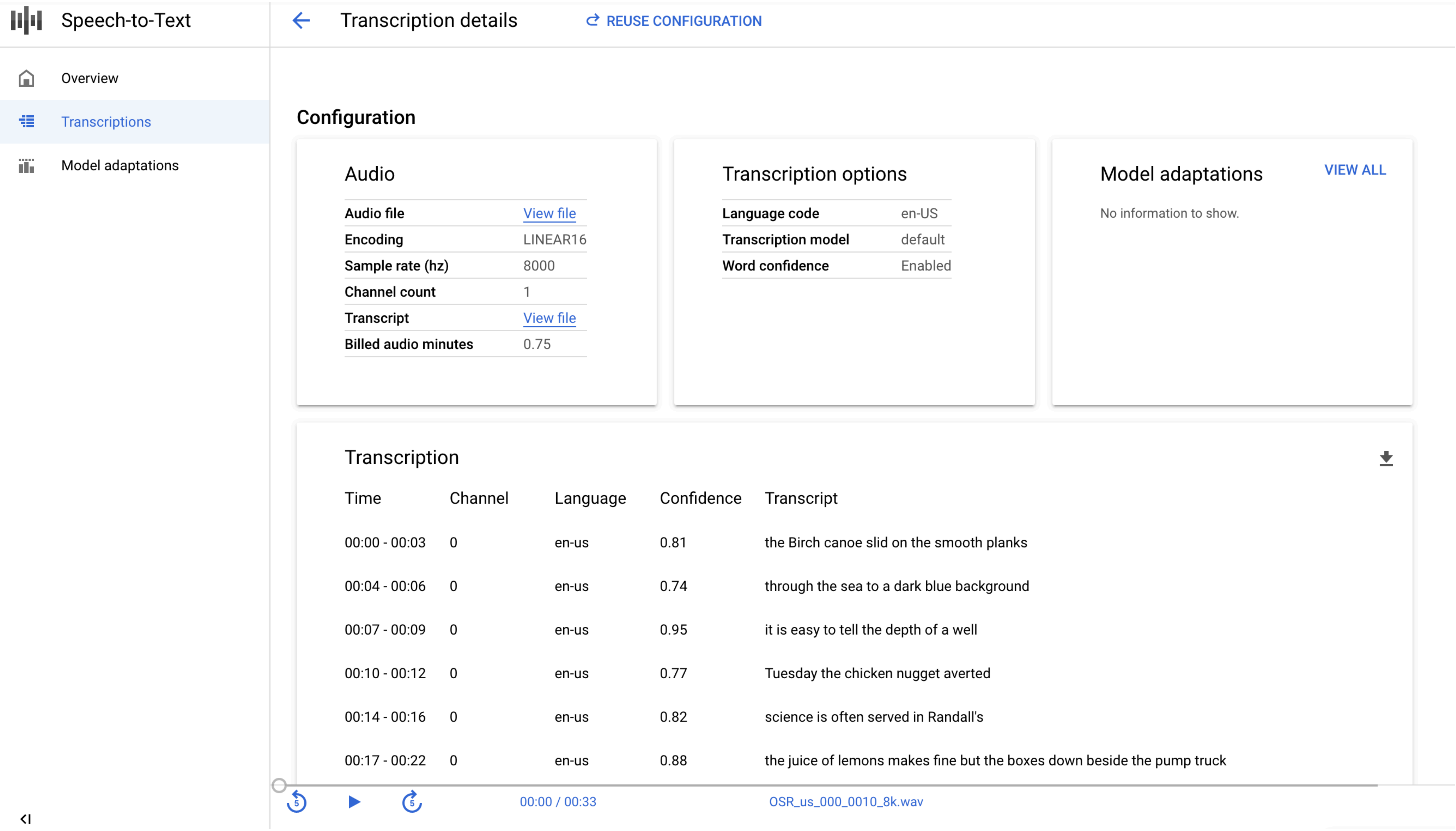
- Transcribe content with accurate captions - could be used to compute subtitles on recorded or live online meetings. Such example features could be seen on Zoom recording or live on Microsoft Teams

- Enable the power of voice to create better user experiences - Voice commands for personal robots or used for software capabilities helping people with different disability needs

- Speech-to-Text can use one of several machine learning models to transcribe your audio file

Tasks

- understand how the service works
- upload an audio file and analyse the transcription



Transcription details

Configuration

Audio

Audio file	View file
Encoding	LINEAR16
Sample rate (hz)	8000
Channel count	1
Transcript	View file
Billed audio minutes	0.75

Transcription options

Language code	en-US
Transcription model	default
Word confidence	Enabled

Model adaptations

No information to show.

Transcription

Time	Channel	Language	Confidence	Transcript
00:00 - 00:03	0	en-us	0.81	the Birch canoe slid on the smooth planks
00:04 - 00:06	0	en-us	0.74	through the sea to a dark blue background
00:07 - 00:09	0	en-us	0.95	it is easy to tell the depth of a well
00:10 - 00:12	0	en-us	0.77	Tuesday the chicken nugget averted
00:14 - 00:16	0	en-us	0.82	science is often served in Randall's
00:17 - 00:22	0	en-us	0.88	the juice of lemons makes fine but the boxes down beside the pump truck

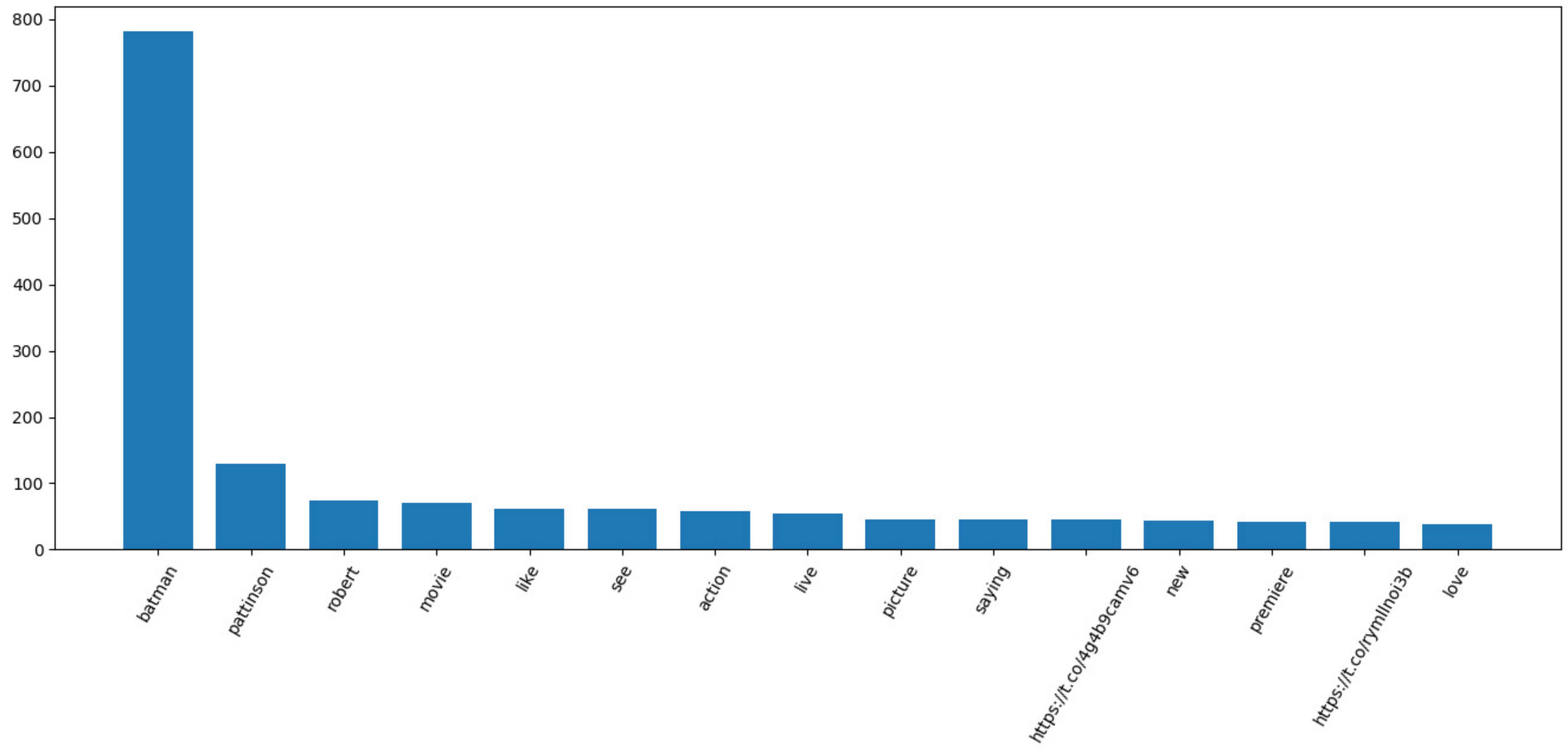
00:00 / 00:33 OSR_us_000_0010_8k.wav

Tasks

- learn how to work with the API by running local code
- edit access control to files on the cloud and export the access key json
- analyse obtained text from audio files located in the cloud bucket
- task 1 and 2 work with short audio files of less than 60 seconds

Google Cloud Storage

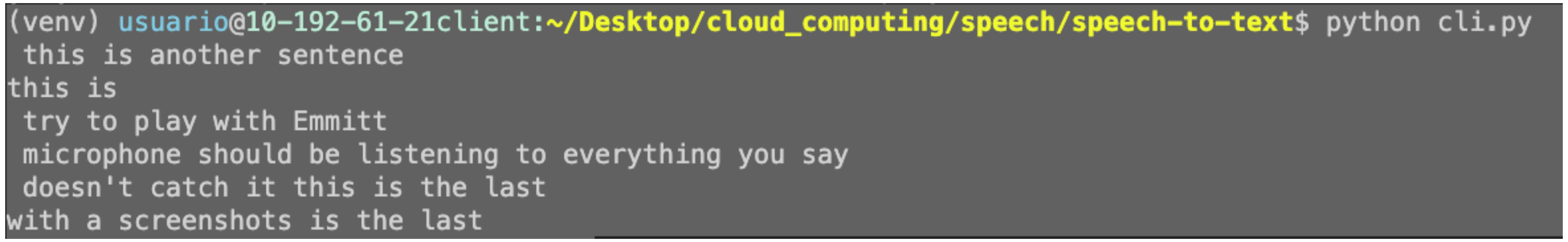
Tasks



- Long speech
- Upload file from your computer
- Analyse the common words in the speech
- Your turn!
  - Find files to analyse from any source that you want

Tasks

- In this task we are going to try to understand audio from microphone and execute corresponding CLI commands
- portaudio and pyaudio
- Interact real time with speech API
- the user is going to “create” the commands



```
(venv) usuario@10-192-61-21client:~/Desktop/cloud_computing/speech/speech-to-text$ python cli.py
 this is another sentence
this is
 try to play with Emmitt
 microphone should be listening to everything you say
 doesn't catch it this is the last
with a screenshots is the last
```

Conclusion

- power of cloud computing to the end user, use of already developed models for audio recognition on several languages and accents
- diminishing the boundary between human and computer
- possibility to analyse big chunks of available data in audio format that were before that unaccessible without human transcribing it

Thank you!