1. Sign Up for a Free Tier Account

Google Cloud offers a Free Tier plan, which will be used in this tutorial. An account is required to get an API key.

2. Generate an API Key

Follow these steps to generate an API key:

1. Sign-in to Google Cloud Console
2. Click “API Manager”
3. Click “Credentials”
4. Click “Create Credentials”
5. Select “Service Account Key”
6. Under “Service Account” select “New service account”
7. Name service (whatever you’d like)
8. Select Role: “Project” -> “Owner”
9. Leave “JSON” option selected
10. Click “Create”
11. Save generated API key file
12. Rename file to `api-key.json`

Make sure to move the key into speech-to-text cloned repo, if you plan to test this code.

3. Convert Audio File to Wav format

There are a lot of tools you may use to convert audio files.

4. Break up audio file into smaller parts

Google Cloud Speech API only accepts files no longer than 60 seconds. To be on the safe side, I broke my files in 30-second chunks. To do that I used an open source command line library called ffmpeg. I ran it on Windows, you can install ffmpeg using instruction from this site and then run in you command line (cmd.exe) below instruction:

Clean out old parts if needed via rm -rf parts/*

ffmpeg -i source/genevieve.wav -f segment -segment_time 30 -c copy parts/out%09d.wav

5. Install required Python modules

Install:

- google-api-python-client
- httplib2
- oauth2client
- pyasn1
- pyasn1-modules
- rsa
- six
- SpeechRecognition
- tqdm
- uritemplate

6. Run the Code

1. Loads API key from step 2 in memory
2. Gets a list of files (chunks)
3. For every file, calls speech to text API endpoint
4. Adds results to a list
5. Combines all results and adds a timestamp (every 30 seconds)
6. Saves results to `transcript.txt`