

SpeechToText API comparison and choice

We have tried to transcribe a WAV audio file using multiple methods:

1. **Google Cloud Speech-to-Text API**
2. **speech_recognition library with Python library API**
3. **OpenAI's Whisper API (assuming from the function name and comments)**

The test examples include male voice, female voice, and background noise.

Male voice example

```
C:\Users\User\AppData\Local\Microsoft\WindowsApps\python3.11.exe C:\Users\User\Documents\GitHub\AI-RedBack\src\SpeechToText\Comparison.py
Google Speech API Transcript: but what if somebody decides to break it be careful that you keep adequate coverage but look for places to save money baby it's taking longer to get things squared away then the bankers expected
Speech Recognition Transcript: but what if somebody decides to break it be careful that you keep adequate coverage but look for places to save money maybe it's taking longer to get things squared away then The Bankers expect
OpenAI Transcript: but what if somebody decides to break it? Be careful that you keep adequate coverage, but look for places to save money. Maybe it's taking longer to get things squared away than the bankers expected. Hiring
```

Female voice example

```
C:\Users\User\AppData\Local\Microsoft\WindowsApps\python3.11.exe C:\Users\User\Documents\GitHub\AI-RedBack\src\SpeechToText\Comparison.py
Google Speech API Transcript: conveyor the closest distance but this one no state yet I think it's at over thing but it's everything in the middle keeps up there so we'll know staying here no messy if you feeling if you ple
Speech Recognition Transcript: embarrassed because of the stains but this one no state I think it's at over saying that keeps everything in the middle and you feel clean all white
OpenAI Transcript: I usually get so embarrassed because of the stains, but this one, no stains. Yeah, I think it's that oval thing that keeps everything in the middle. Yeah, it keeps it there, so no stains. Yeah, no messy,
```

Human voice with background noise

```
C:\Users\User\AppData\Local\Microsoft\WindowsApps\python3.11.exe C:\Users\User\Documents\GitHub\AI-RedBack\src\SpeechToText\Comparison.py
Google Speech API Transcript: sons of Welland away for races two miles of the Moon in 1992 by solar sales
Google Speech API Transcript: director malls is to commemorate Columbus is Johnny to the new world 500 years ago on the one to the moon it's to promote the use of Solace sales in space exploration
Speech Recognition Transcript: space exploration
OpenAI Transcript: Plans are well underway for races to Mars and the Moon in 1992, by solar sail. The race to Mars is to commemorate Columbus' journey to the New World 500 years ago, and the one to the Moon is to promote the
```

Based on the above test results.

Google Speech API:

This transcript has several inaccuracies and seems to misinterpret many words.

Speech Recognition (using Python library free API):

This transcript is more accurate than the Google Speech API's result but still has a few errors. The length of the converted speech does not meet the requirements.

OpenAI's Whisper API:

This transcript seems to be the most accurate and coherent of the three. It captures the context and provides a clear message.

Conclusion:

Based on this example, OpenAI's Whisper API appears to be the most accurate and provides the best transcription.