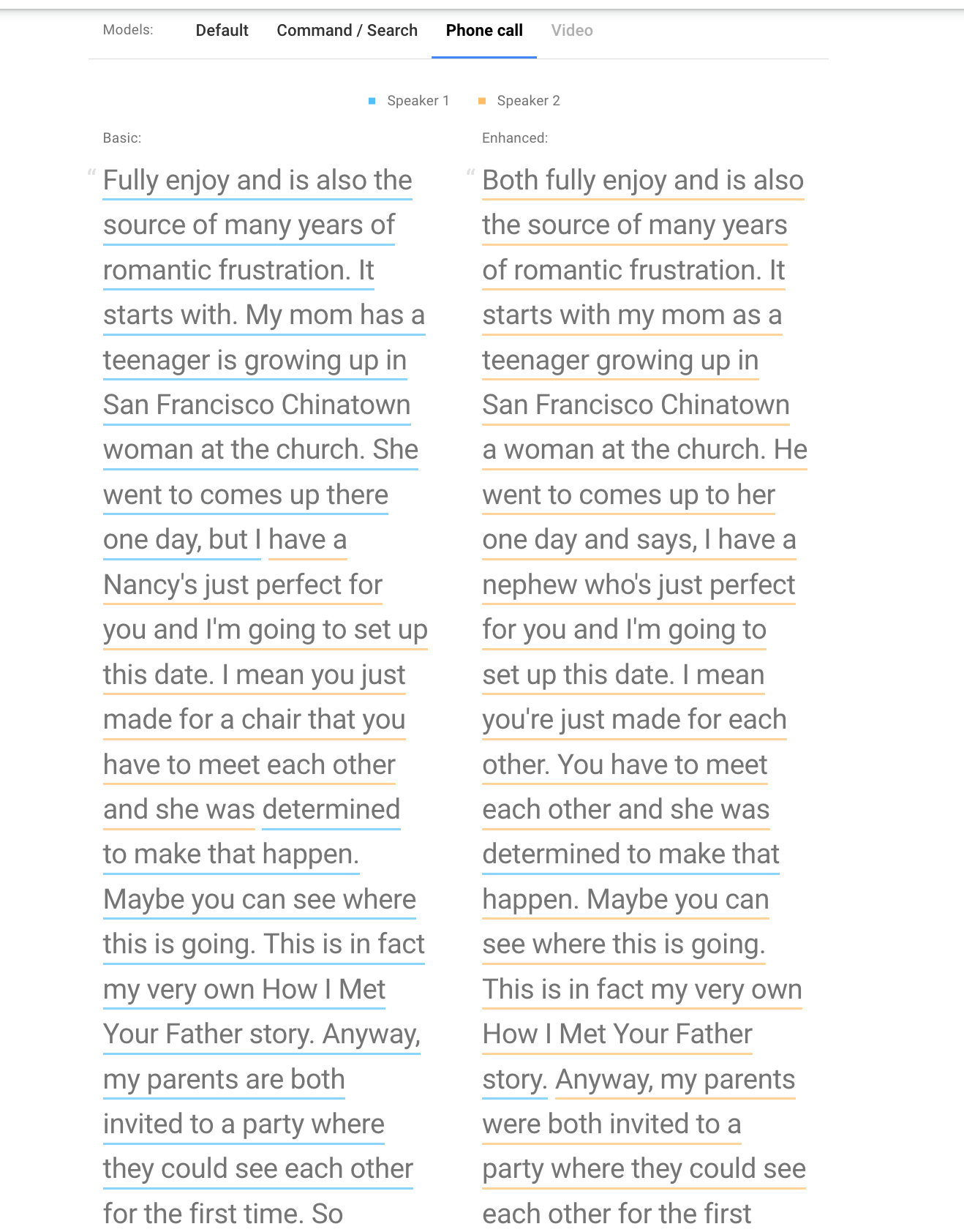
### Speech Separation and Transcription

#### Auto recognition

<https://cloud.google.com/speech-to-text>

audio tested for google service.mp3

$0.003 / minute = $0.18 / hour = $1.44 / day (8 hours)



It’s not accurate, but I guess this happens because this is a podcast so it lacks features of different toning, soundness, pauses, filters, etc.

After the poc is finished, use

<https://github.com/openai/whisper>

<https://elevenlabs.io/>

#### Record user’s voiceprint first, then use the voiceprint to separate

Theoretically it should have higher accuracy.