

# Speech Recognition

Grettel Juárez



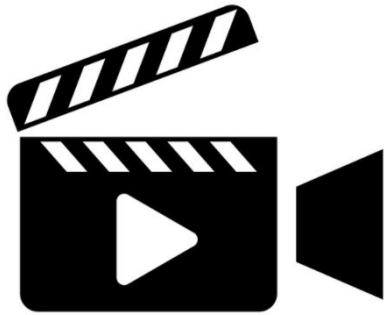
# Use Cases



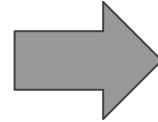
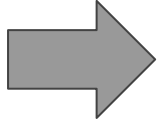
- Video subtitles
- Meeting transcription
- Voice assistants
- Dictation

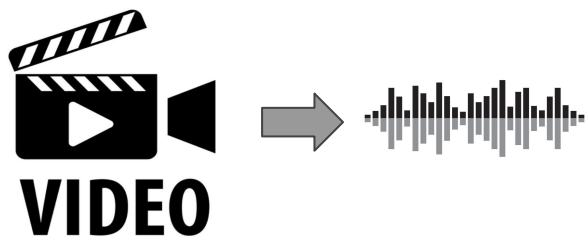
---

# Document Conversion



**VIDEO**



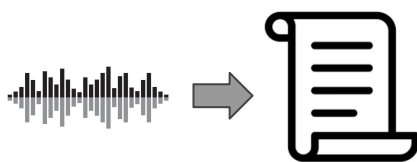


```
#pip install SpeechRecognition moviepy
import moviepy.editor as mp

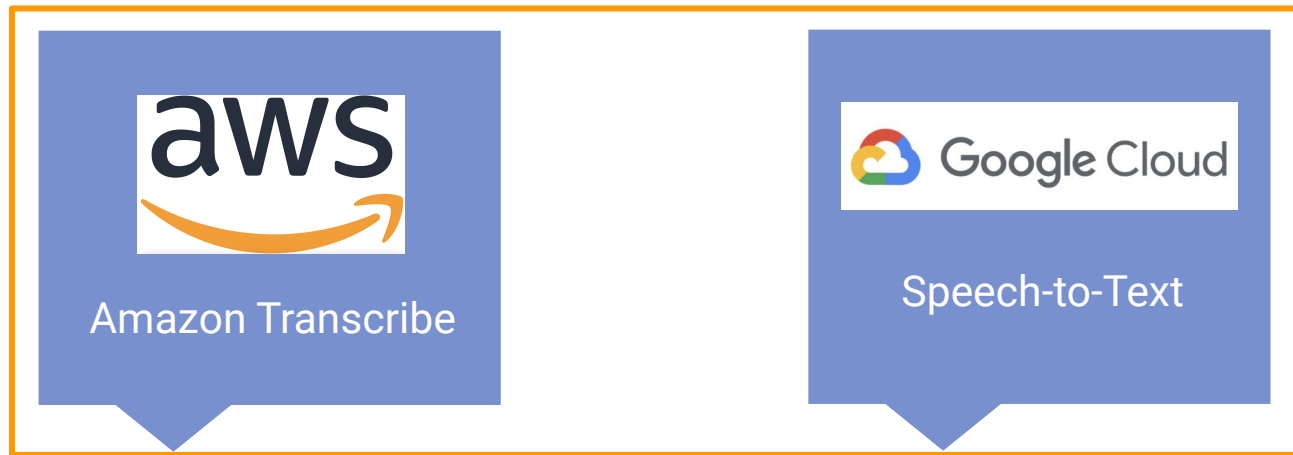
# Load video file
clip = mp.VideoFileClip(r"fold_in_the_cheese.mp4")

# Create .wav file
clip.audio.write_audiofile(r"fold_in_the_cheese.wav")
```





# Several options



Best for **medical**  
transcription



# Google Speech-to-text request types



Request	Description	File length limit
Synchronous Recognition	API must return response before processing next request	~ 1 minute
Asynchronous Recognition	Request sent and API polled for result	Max 8 hours
Streaming Recognition	Real-time transcription	N/A

Speech-to-Text typically processes audio faster than realtime, processing 30 seconds of audio in 15 seconds on average.

# Google Speech-to-text Setup



**Google Cloud  
Speech API**

Download/set JSON path  
(synch and asynch recognition)



**Google Cloud Storage**

Create storage bucket  
(asynchronous recognition)

# Speech Recognition Process

## Set up Account

- Create Storage
- Create Speech-to-text service

## Prep File

- Convert file to meet requirements
- Upload file
- Set configs

## Transcribe

- Call API
- Collect result
- Write to file
- Delete file from storage



# Google Speech-to-text Requirements and Configurations

## Requirements

- Sample rate must be 16K Hz or greater
- .wav file must be single channel (mono)

## Configurations

- Punctuation
- Speaker count if more than 1
- Specify model (ex. video, phone call)
- Word timestamps
- Alternatives
- More... in documentation

# Fold in the Cheese



**speaker 1:**

**speaker 2:** This is not your mother's recipe.

Yes, and now I'm passing it on to you. So try to keep

**speaker 1:** up. Oh next step is to fold in the

**speaker 2:** cheese. What does that mean? What does fold in the cheese mean?

He holds it in I understand that but how do you fold it?

Do you fold it in half like a piece of paper and drop it in the pot or what do you do and I cannot show you everything.

Okay? Well, can you show me one thing

**speaker 1:** you just what you do? You just fold it in.

**speaker 2:** Okay. I don't know how to fold broken cheese like that.

I don't know how to be any clearer.

You take that thing that's in your head, huh?

And you if you say fooled in one more time.

I'm that's follows it in.

This is your recipe you fooled in the cheese, then don't you dare you fold it in David?



**speaker 1:**

**speaker 2:** This is not your mother's recipe.

Yes, and now I'm passing it on to you. So try to keep

**speaker 1:** up. Oh next step is to fold in the

**speaker 2:** cheese. What does that mean? What does fold in the cheese mean?

He holds it in I understand that but how do you fold it?

Do you fold it in half like a piece of paper and drop it in the pot or what do you do and I cannot show you everything.

Okay? Well, can you show me one thing

**speaker 1:** you just what you do? You just fold it in.

**speaker 2:** Okay. I don't know how to fold broken cheese like that.

I don't know how to be any clearer.

You take that thing that's in your head, huh?

And you if you say fooled in one more time.

I'm that's follows it in.

This is your recipe you fooled in the cheese, then don't you dare you fold it in David?



This is not your mother's recipe.

Yes. And now I'm passing it on to you, so try to keep up.

Um oh. Next step is to fold in the cheese.

What does that mean? What does fold in the cheese mean? He holds it in.

I understand that. But how? How do you fold it? Do you fold it in half like a piece of paper and drop it in the pot? Or what do you do, David?

I cannot show you everything.

Okay, Well, can you show me one thing?

You just guess what you do? You just fold it in.

I don't know how to fold broken cheese like that, and I don't know how to be any clearer.

You take that thing that's in your head, huh?

If you say fold in one more time has folded in.

This is your recipe. You fold in the cheese then, don't you? You fooled it in David.

# Simon Sinek



And the reason it's a wonderful hotel is not because of the fancy beds any hotel can go and buy a fancy bed. The reason it's a wonderful hotel is because of the people who work there if you walk past somebody at the Four Seasons in this and they say hello to you. You get the feeling that they actually wanted to say hello to you. It's not that somebody told them that you have to say hello to all the customers say hello to all the guests, right? You actually feel that they care now in their Lobby they have a coffee stand and I want afternoon I went.

# Cedar Toolbox



**speaker 1:** Hey

**speaker 2:** has Paul Bunyan finished his

**speaker 1:** box who? I'm going to need some things for the chest like the number for a carpenter like a workbench in miter saw to Bar clamps and some towels. Okay?

**speaker 2:** Do you know how to use a miter saw

**speaker 1:** um, no might as well back and he's asked me to get these things for him. We're building the chest together. So Wow,

**speaker 2:** this whole thing just got a lot weirder. There's a tool shed out back the other side of the

**speaker 1:** motel. Okay, will you be requiring a toolbox? Maybe let's go with yes, just to be safe.

Will you be needing your

**speaker 2:** basic toolbox or your Cedar Chest

**speaker 1:** toolbox? Obviously the Cedar Chest toolbox

**speaker 2:** that's in the shed in the show.

# How to decide



- Fast
- More languages



- Leader
- Known for accuracy
- Timestamps seem more granular

# Just try it out!



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

## Speech-to-Text

[Benefits](#)

[Demo](#)

[Key features](#)

[Customers](#)

[What's new](#)

## Documentation

## Use cases

[Improve customer service](#)

[Enable voice control](#)

[Transcribe multimedia content](#)

## All features

## Pricing

## Take the next step

DEMO

## Put Speech-to-Text into action

As in this demo, you can easily infuse speech transcription into your applications with the Speech-to-Text API.

Input type

☐ Microphone ☒ File upload

Language

English (United States) ▾

Speaker diarization **BETA**

Recognize multiple speakers in single channel ▾

Speakers

2 speakers ▾

Punctuation

☒

[Show JSON ▾](#)

CHOOSE FILE

Models:

[Default](#)

[Command / Search](#)

[Phone call](#)

[Video](#)

■ Speaker 1 ■ Speaker 2

# Just try it out!



<https://aws.amazon.com/getting-started/hands-on/create-audio-transcript-transcribe/>

A screenshot of the AWS Management Console interface. At the top is a dark navigation bar with the AWS logo, "Services" and "Resource Groups" dropdowns, a notification bell, and regional settings for "Ohio" and "Support". Below this is a breadcrumb trail: "Amazon Transcribe > Transcription jobs". The main content area is titled "Transcription jobs" and includes a "Status: All" dropdown, "Download" and "Create Job" buttons, and a search bar labeled "Search job names". A table below shows a single job. The job's name, "sample-transcription-job", and its status, "Complete" (indicated by a green checkmark), are highlighted with red rectangular boxes. The table has columns for Name, Language, Output location, Creation Time, and Status.

	Name	Language	Output location	Creation Time	Status
<input type="radio"/>	sample-transcription-job	English	Amazon Transcribe	2023-08-16 10:10:10 AM	✓ Complete





# Thanks!

**Any questions?**



TEST

