

# VOICE CLONING

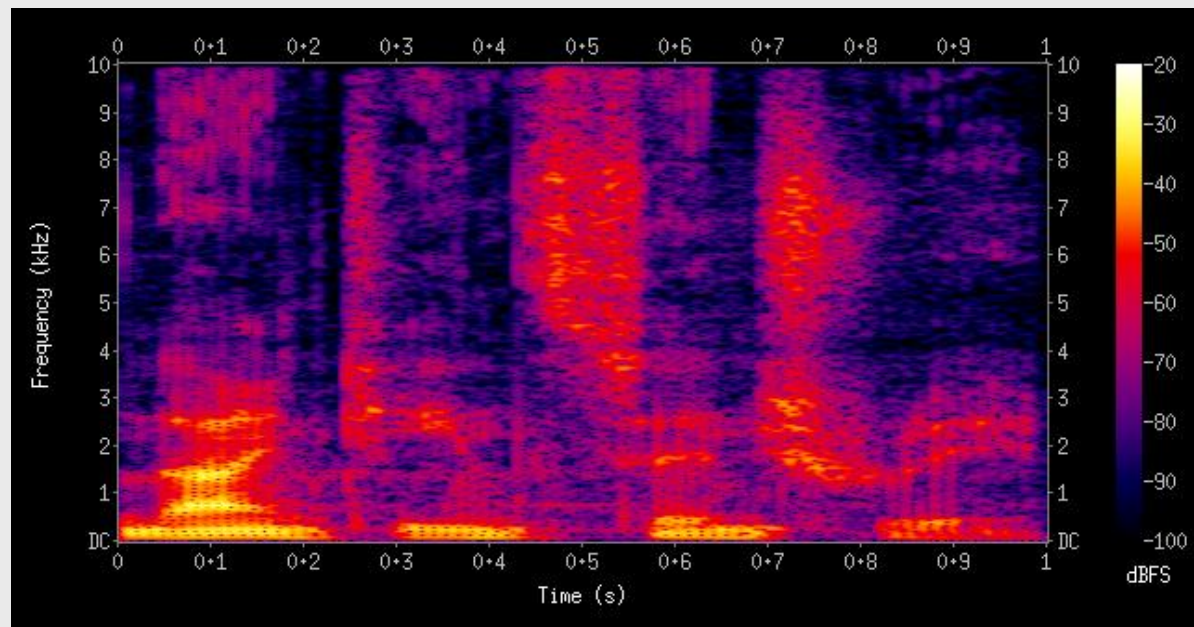


# **AI Voice Cloning**

- **AI can easily mimic someone's speaking style in text with transformers**
- **This is possible because the AI is able to represent text with a good embedding vector**
- **If the AI wanted to represent the sound of someone's voice, what would the embedding vector represent?**

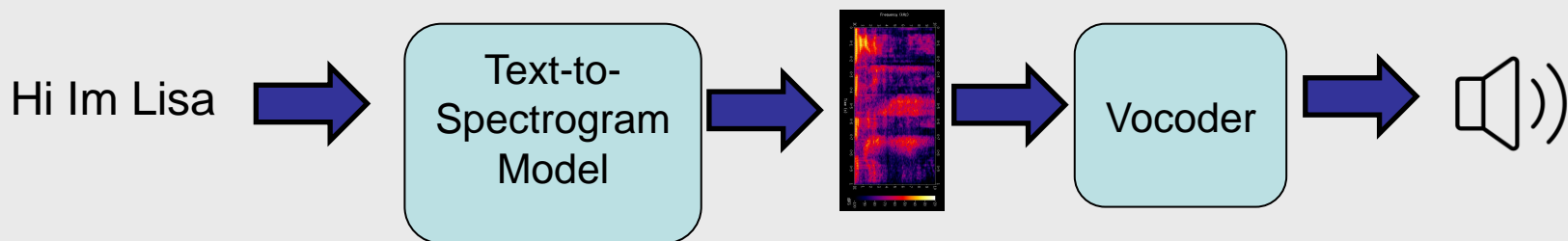
# Spectrograms: How Computers Represent Voices

- Spectrogram – represents the frequencies in an audio signal as it varies in time
  - Like an audio fingerprint



# Text-to-Speech (TTS) Models

- Text to speech (TTS) models are neural networks that convert text into spectrograms
- The spectrograms are then turned into voice audio using neural networks called vocoders



# First Good TTS Model: Tacotron

## TACOTRON: TOWARDS END-TO-END SPEECH SYNTHESIS

**Yuxuan Wang<sup>\*</sup>, RJ Skerry-Ryan<sup>\*</sup>, Daisy Stanton, Yonghui Wu, Ron J. Weiss<sup>†</sup>, Navdeep Jaitly,**

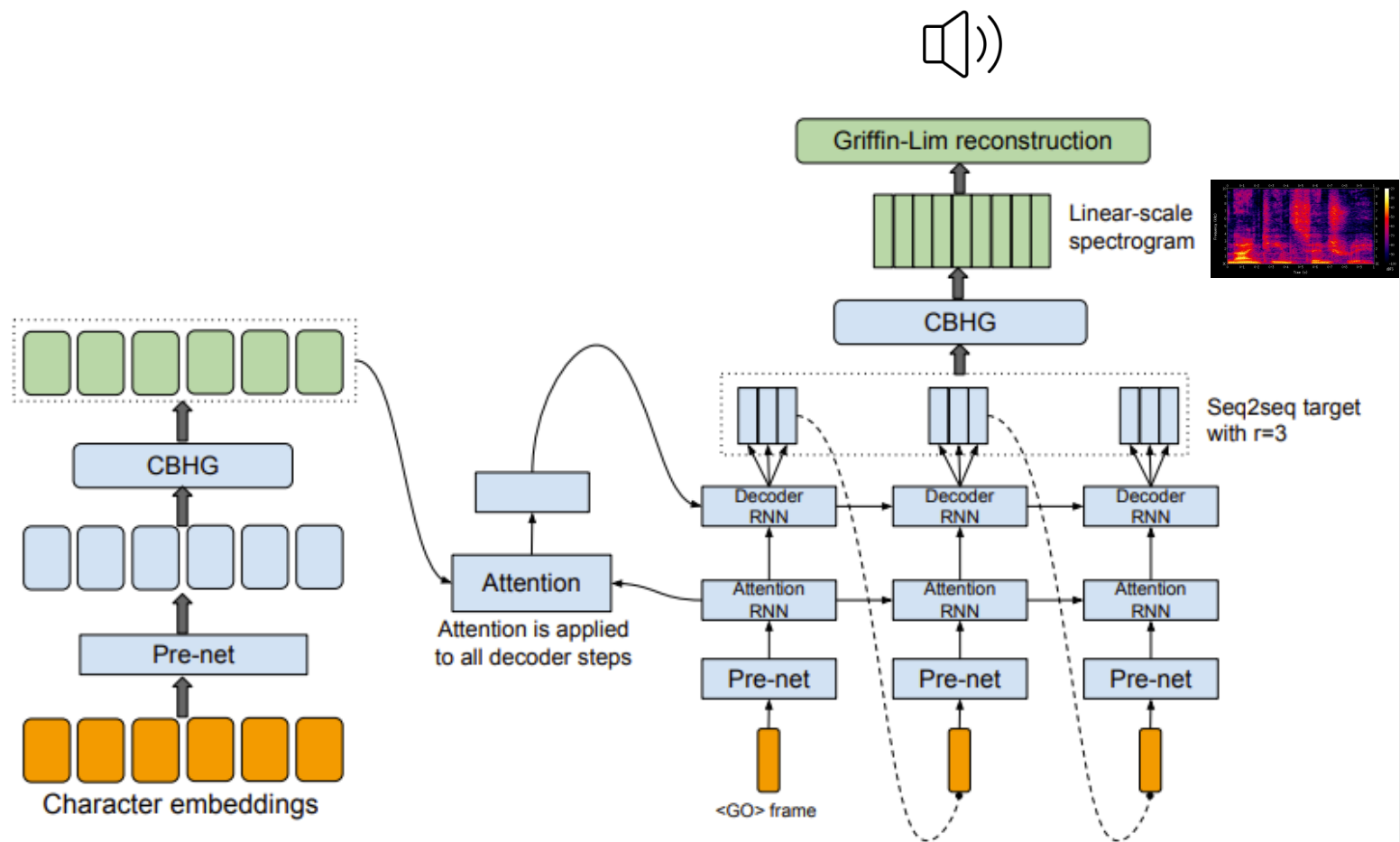
**Zongheng Yang, Ying Xiao<sup>\*</sup>, Zhifeng Chen, Samy Bengio<sup>†</sup>, Quoc Le, Yannis Agiomyrgiannakis,**

**Rob Clark, Rif A. Saurous<sup>\*</sup>**

Google, Inc.

`{yxwang, rjryan, rif}@google.com`

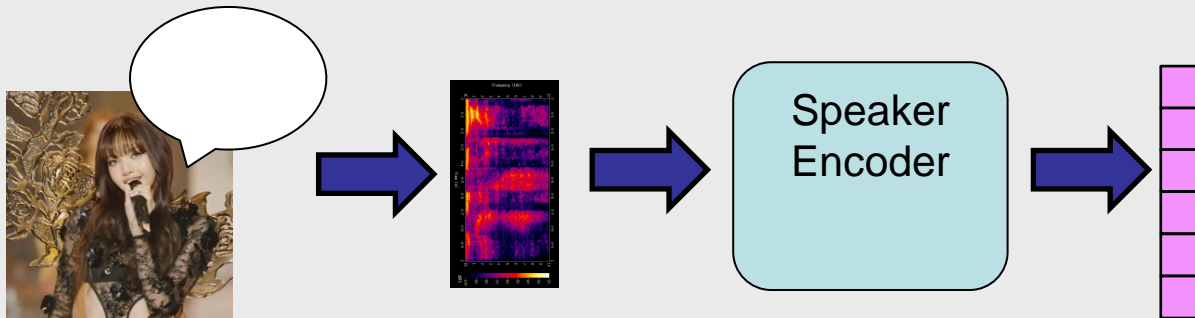
# Tacotron Architecture



Hi Im Lisa

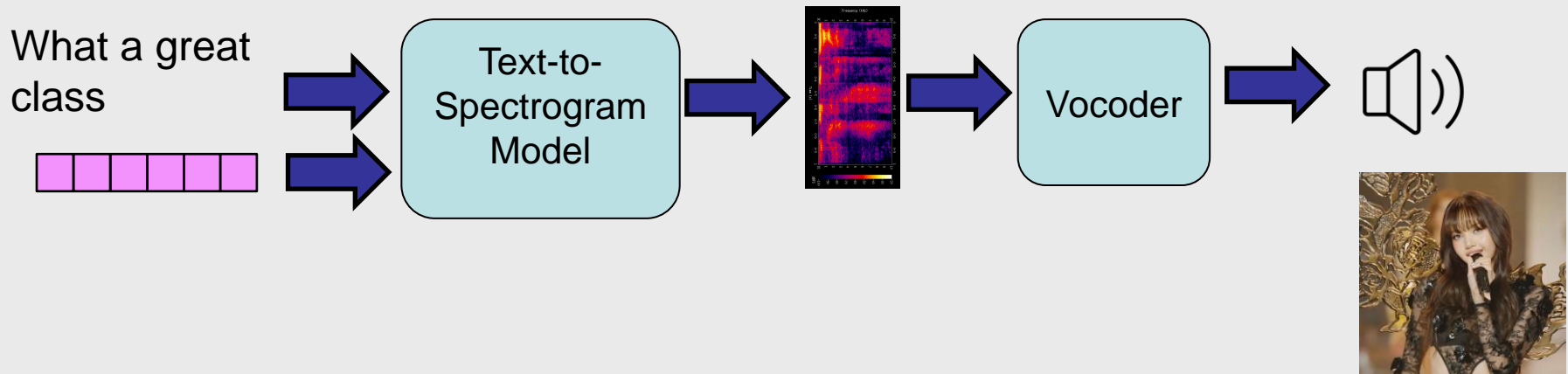
# Representing a Speaker

- Spectrograms are like voice fingerprints
- Voice cloning neural networks called “speaker encoders” take spectrograms and turn them into vectors
- Trained using contrastive learning



# Voice Cloning

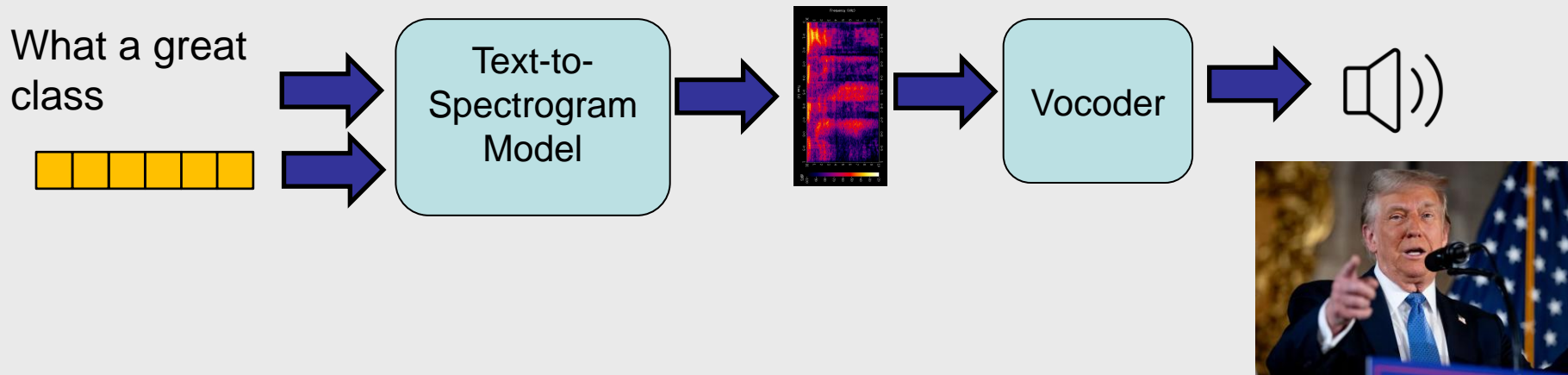
- Once we have the speaker vector, we can just add it to the TTS model





# Voice Cloning

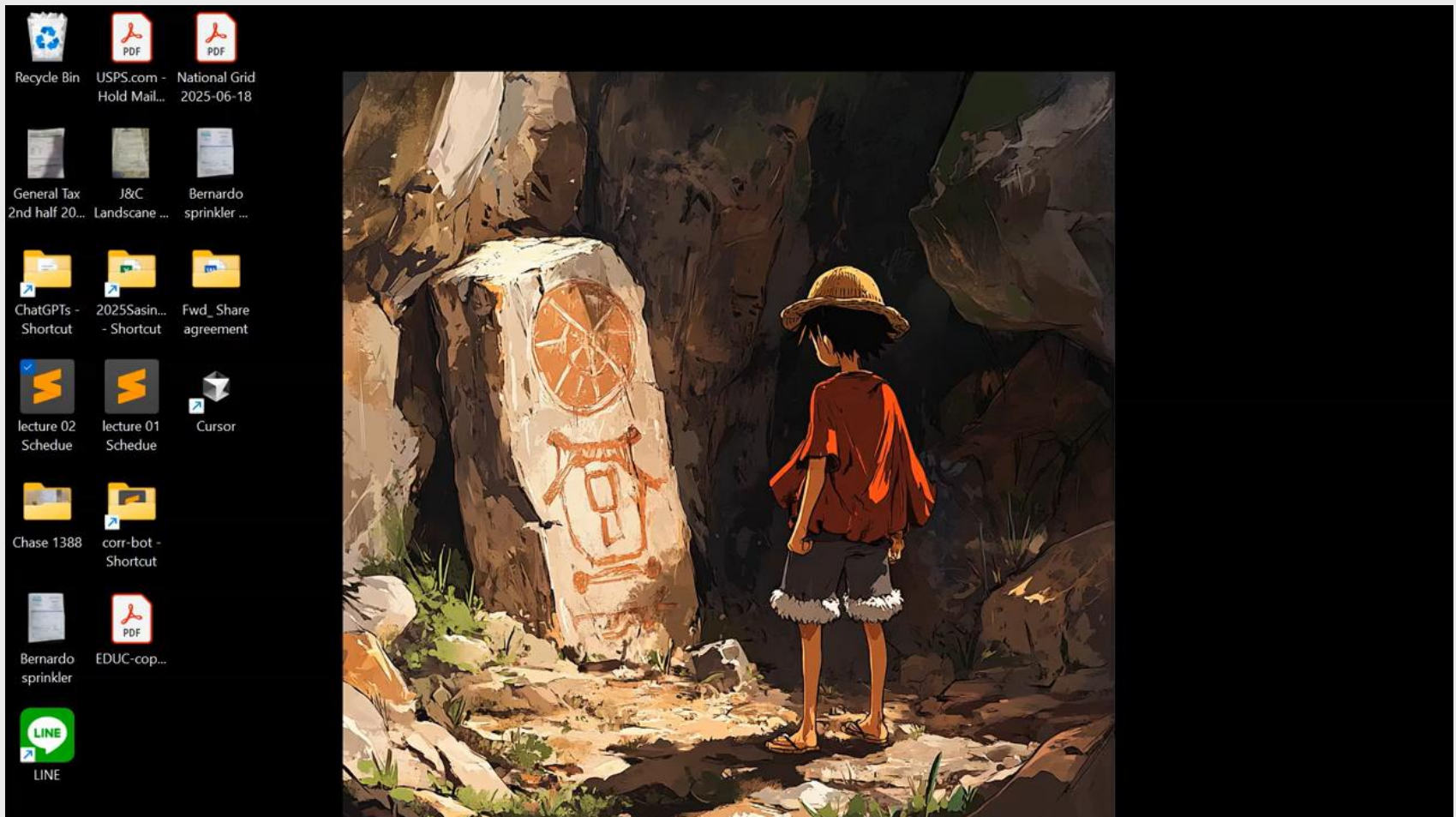
- Once we have the speaker vector, we can just add it to the TTS model



# ElevenLabs

- ElevenLabs is one of the most advanced voice cloning tools: <https://elevenlabs.io/>
- Many voice cloning features
  - Voice library (use voices others cloned)
  - Instant voice cloning (10 seconds of audio)
  - Professional voice cloning (30 minutes of audio)
  - Voice design (create voice by text prompt)
  - Voice changer (change the voice in an audio file)

# Voice Changer



# Permission Granted to Voice Clone

IIElevenLabs

## Iconic Voices



# Permission Granted to Voice Clone

## James Earl Jones Signed Over Rights For AI To Recreate Darth Vader's Voice

By [Tim Lammers](#), Contributor. © I cover Hollywood and entertainment.

[Follow Author](#)

Published Sep 09, 2024, 06:08pm EDT, Updated Sep 09, 2024, 11:55pm EDT





# Permission Not Granted to Voice Clone

## Robert Downey Jr. vows there will never be a digital AI replica of him on-screen

News

By Eric Hal Schwartz published October 31, 2024

Ultron isn't Iron Man's only AI foe



# **Voice Cloning Ethics**

- **If using a voice clone, be transparent**
- **Don't use anyone's voice without their permission**
- **Don't make the voice clone say offensive things**
- **Don't create deepfake audio that will cause panic...**

# Voice Cloning Ethics

- Or get someone fired

## Thailand's prime minister suspended over leaked phone call with former strongman



By Helen Regan and Kocho Olarn, CNN

3 minute read · Updated 3:44 AM EDT, Tue July 1, 2025





# Voice Cloning Ethics

- **Definitely don't do this:**



# How to Clone a Voice

- ElevenLabs allows us to clone voices with only 10 seconds of audio
- Can obtain audio files from YouTube
  - Choose a YouTube video of the subject speaking alone or mostly alone
  - Download audio of YouTube video (<https://tuberipper.com>)
  - Clip the audio to 10 to 30 second clips (<https://audiotrimmer.com/>)

# Video Narrations

- If we give the frames of a video to AI plus some instructions, it can give us a narration for the video
- We already have methods to give multiple images and context text to AI and get a text description
- We just need to extract the images from the video
  - We will usually sample around 10 images per video
  - This is near the limit of what OpenAI allows


# Old HW Problem

- **Make a narration of this video**
- **Bonus – clone a voice and have it narrate the video**



Richard Ogu  
Yale MAM  
Class of 2024

# Video Narration App




## VoxOver: AI Narration Studio

Create professional AI voiceovers for your videos with ease.

Step 1 of 7: Upload Video



### Step 1: Upload Your Video

Choose a video file



Drag and drop file here  
Limit 200MB per file • MP4, MOV, AVI, WMV, MPEG4

Browse files

 leomessi\_Argentina.mp4 13.6MB 

Video uploaded successfully: leomessi\_Argentina.mp4

### Step 2: Provide Voiceover Instructions

Describe the style and content you want for your voiceover:



Example: Create a professional, enthusiastic narration that explains the key points shown in the video. Use a conversational tone suitable for a marketing presentation.

Generate Voiceover Text

Share ☆ ✎

### Preview

Original Video Video with Voiceover



# Coding Session

- **We will clone a voice**
- **We will clone a GitHub repository with some useful code to get us started**
- **We will build an app that narrates a video using the cloned voice**
- **AI tools used**
  - **OpenAI**
  - **ElevenLabs**