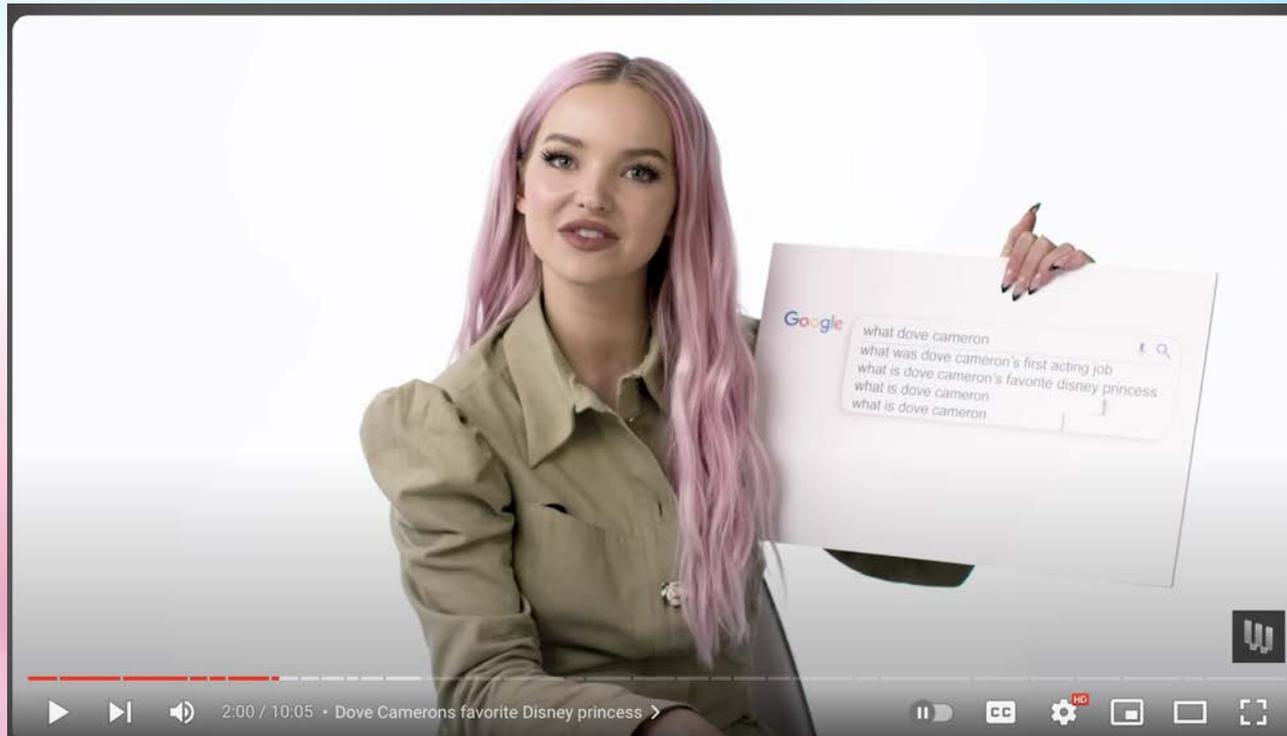# Voice Recognition

Sara Greenberg

Gathering Samples

- Gathered audio recordings from celebrity interviews on Youtube
- Tried to collect audio of just one person speaking without background noise
- Gathered 2 samples each for about 40 people

- Used computer to gather samples
- Converted audio files to proper format

```
!ffmpeg -i /content/TonyHawk1.m4a /content/TonyHawk1.wav
!ffmpeg -i /content/TonyHawk2.m4a /content/TonyHawk2.wav
!ffmpeg -i /content/RickyGervais1.m4a /content/RickyGervais1.wav
!ffmpeg -i /content/RickyGervais2.m4a /content/RickyGervais2.wav
!ffmpeg  -i /content/PatrickStewart1.m4a /content/PatrickStewart1.wav
!ffmpeg  -i /content/PatrickStewart2.m4a /content/PatrickStewart2.wav
!ffmpeg -i /content/NiallHoran1.m4a /content/NiallHoran1.wav
!ffmpeg -i /content/NiallHoran2.m4a /content/NiallHoran2.wav
!ffmpeg -i /content/MichaelKnowles1.m4a /content/MichaelKnowles1.wav
!ffmpeg -i /content/MichaelKnowles2.m4a /content/MichaelKnowles2.wav
!ffmpeg -i /content/MattWalsh1.m4a /content/MattWalsh1.wav
!ffmpeg -i /content/MattWalsh2.m4a /content/MattWalsh2.wav
!ffmpeg -i /content/Markiplier1.m4a /content/Markiplier1.wav
!ffmpeg -i /content/Markiplier2.m4a /content/Markiplier2.wav
!ffmpeg -i /content/MachineGunKelly2.m4a /content/MachineGunKelly2.wav
!ffmpeg -i /content/LucyHale1.m4a /content/LucyHale1.wav
!ffmpeg -i /content/LucyHale2.m4a /content/LucyHale2.wav
!ffmpeg -i /content/LillySingh1.m4a /content/LillySingh1.wav
!ffmpeg -i /content/LillySingh2.m4a /content/LillySingh2.wav
!ffmpeg -i /content/KumailNanjani1.m4a /content/KumailNanjani1.wav
```

Enrolling Users

- Enrolled user with first audio sample

```
[ ]  !python /content/Voice-Authentication-CNN/voice_auth.py -t enroll -n "Olivia Rodrigo" -f /content/OliviaRodrigo1.wav

/content/Voice-Authentication-CNN/voice_auth.py:103: SyntaxWarning: "is" with a literal. Did you mean "=="?
  if len(embeds) is 0:
2023-11-19 21:51:01.381469: E tensorflow/compiler/xla/stream_executor/cuda/cuda_dnn.cc:9342] Unable to register cuDNN factory
2023-11-19 21:51:01.381535: E tensorflow/compiler/xla/stream_executor/cuda/cuda_fft.cc:609] Unable to register cuFFT factory:
2023-11-19 21:51:01.381572: E tensorflow/compiler/xla/stream_executor/cuda/cuda_blas.cc:1518] Unable to register cuBLAS facto
2023-11-19 21:51:01.390281: I tensorflow/core/platform/cpu_feature_guard.cc:182] This TensorFlow binary is optimized to use a
To enable the following instructions: AVX2 FMA, in other operations, rebuild TensorFlow with the appropriate compiler flags.
2023-11-19 21:51:02.655531: W tensorflow/compiler/tf2tensorrt/utils/py_utils.cc:38] TF-TRT Warning: Could not find TensorRT
Loading model weights from [voice_auth_model_cnn]....
Processing enroll sample....
1/1 [==============================] - 12s 12s/step
Succesfully enrolled the user
```

Testing Samples

- Tested each user with second audio sample
- System outputs distance score for each user, and returns the name of the person with the lowest distance score

```
!python /content/Voice-Authentication-CNN/voice_auth.py -t recognize -f /content/AndersonCooper2.wav

/content/Voice-Authentication-CNN/voice_auth.py:103: SyntaxWarning: "is" with a literal. Did you mean "==
  if len(embeds) is 0:
2023-11-30 20:17:11.412949: E tensorflow/compiler/xla/stream_executor/cuda/cuda_dnn.cc:9342] Unable to r
2023-11-30 20:17:11.413016: E tensorflow/compiler/xla/stream_executor/cuda/cuda_fft.cc:609] Unable to re
2023-11-30 20:17:11.413054: E tensorflow/compiler/xla/stream_executor/cuda/cuda_blas.cc:1518] Unable to
2023-11-30 20:17:11.421898: I tensorflow/core/platform/cpu_feature_guard.cc:182] This TensorFlow binary
To enable the following instructions: AVX2 FMA, in other operations, rebuild TensorFlow with the appropr
2023-11-30 20:17:12.690185: W tensorflow/compiler/tf2tensorrt/utils/py_utils.cc:38] TF-TRT Warning: Coul
Loading model weights from [voice_auth_model_cnn]....
Processing test sample....
Comparing test sample against enroll samples....
1/1 [==============================] - 17s 17s/step
{'Ethan': 0.03927075231814088, 'Anderson Cooper': 0.012569070951842693, 'Naveed': 0.05184147468244125],
Recognized:   Anderson Cooper
```

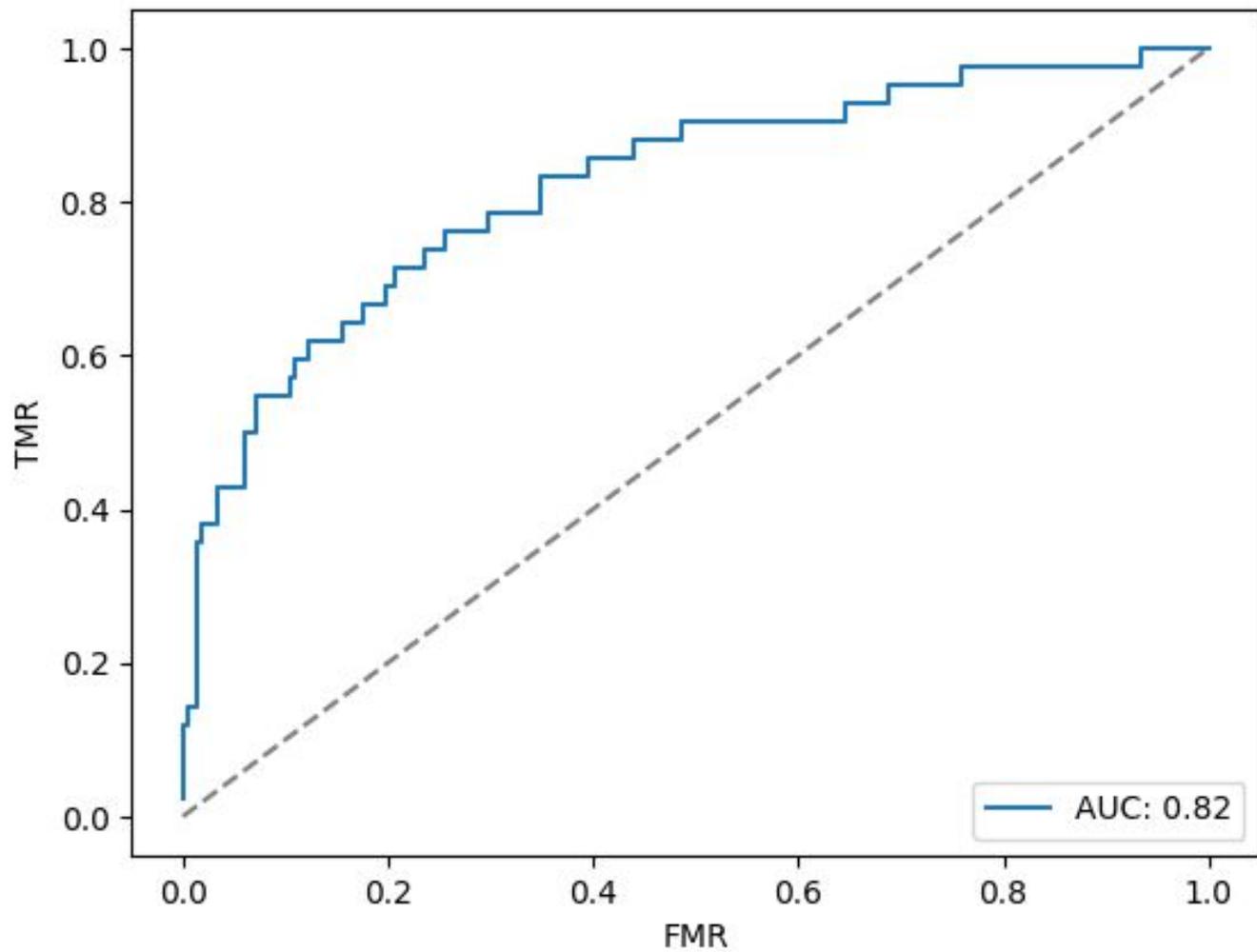# Performance

Score distribution, d'=1.17
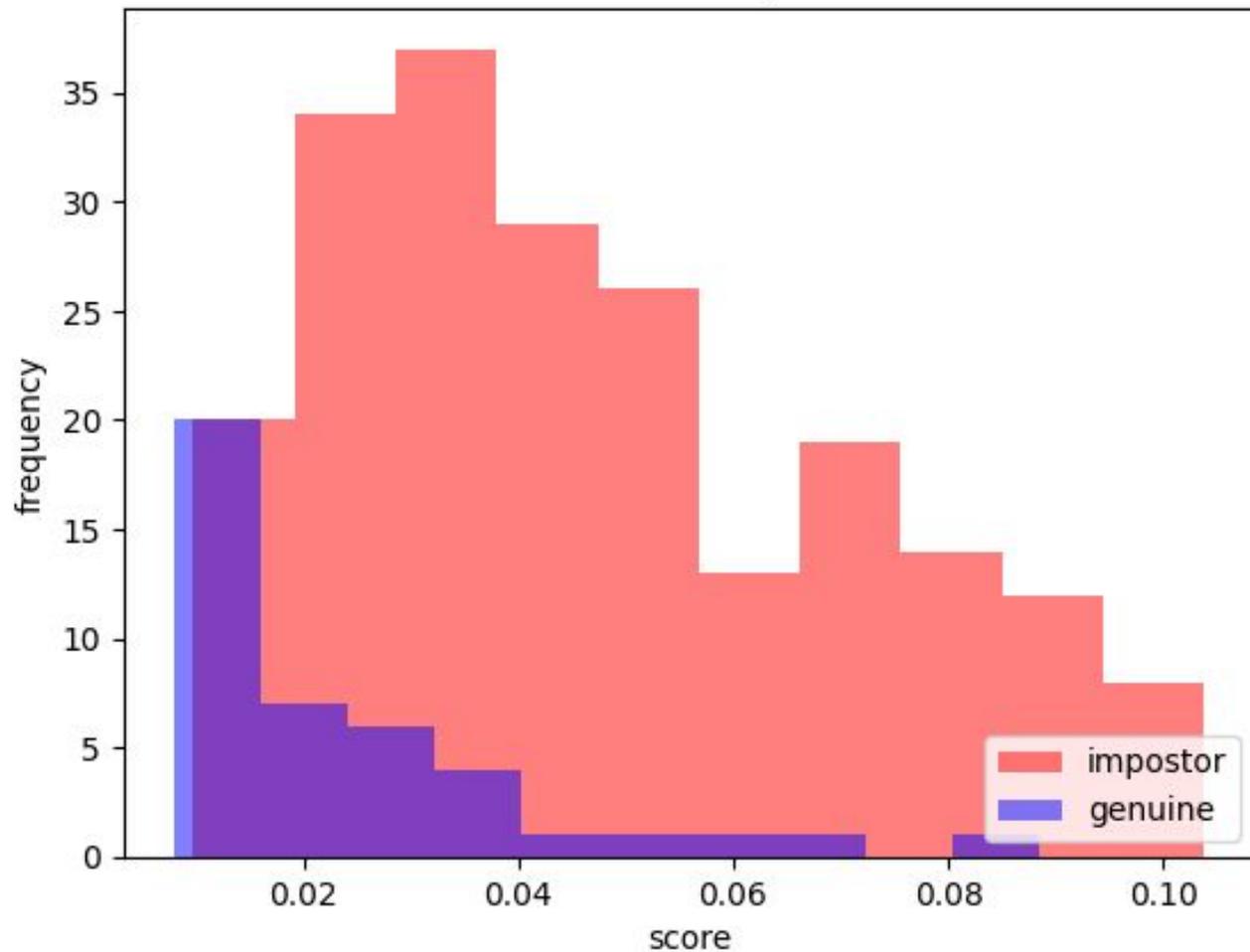
Finding the proper threshold value

- Gathered all values for each genuine pair and then gathered values for the same number of impostor pairs
- Combined all values into one file
- Ran file through program to find EER value

```
v1 = load_data('/content/VoiceScores.txt')
print(compute_fmr_fnmr_eer(v1,is_similarity=False))
```

```
(0.2619047619047619, 0.25471698113207547, 0.028436916023620556)
```
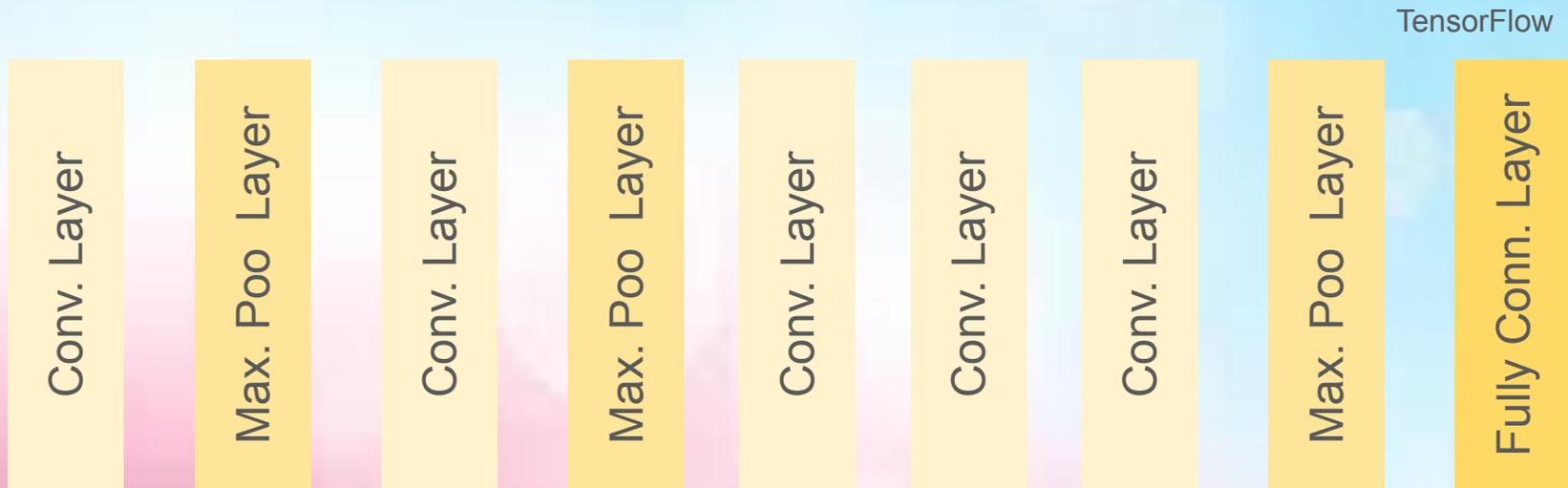
ROC curve

Explanation

Audio enhancement

Audio Signal → **Noise Removal** | **Frame Sampling** | **Fourier Transform** | **Quantization** → Audio Frames

Feature Extraction

Input: audio frames

TensorFlow

Conv. Layer

Max. Poo Layer

Conv. Layer

Max. Poo Layer

Conv. Layer

Conv. Layer

Conv. Layer

Max. Poo Layer

Fully Conn. Layer

Constant utterances are extracted from the frames

Output: single 1024 feature vector

Links

- Google Colab notebook - https://colab.research.google.com/drive/1-I2BzsOLo7asLfuqzIhbFB7VwwYcpbHu#scrollTo=ee_QtVTYfnY3
- Solution GitHub - https://github.com/NaveedShahid/Voice-Authentication-CNN/blob/main/voice_auth.py
- Method Source - https://arxiv.org/abs/1705.09422