

A series of thin, light-brown lines forming an abstract geometric pattern in the top-left corner of the slide. The lines intersect to create various triangular and polygonal shapes.

SAMPLING VOCALS USING DEEP LEARNING METHODS TO ISOLATE HUMAN VOICES

Darren Fok 301461164

Chris Sim 301448214

Tony Chen 301458138

Nicholas Lagasse 301451354

ENSC 429 - Group 3

METHODOLOGY AND ANALYSIS

ENGINEERING DESIGN / APPROACH

Milestones, project distribution, iterative design layout

IMPLEMENTATION

Wrote initial drafts of pre/post processing scripts, training notebook, and main.py file

ITERATION DESIGN

Tried different loss methods, input training file formats, number of epochs, training set size

FINAL VERSION

Usable Trained Model

ENGINEERING DESIGN

PROJECT ENVIRONMENT

Python, Github

MILESTONES

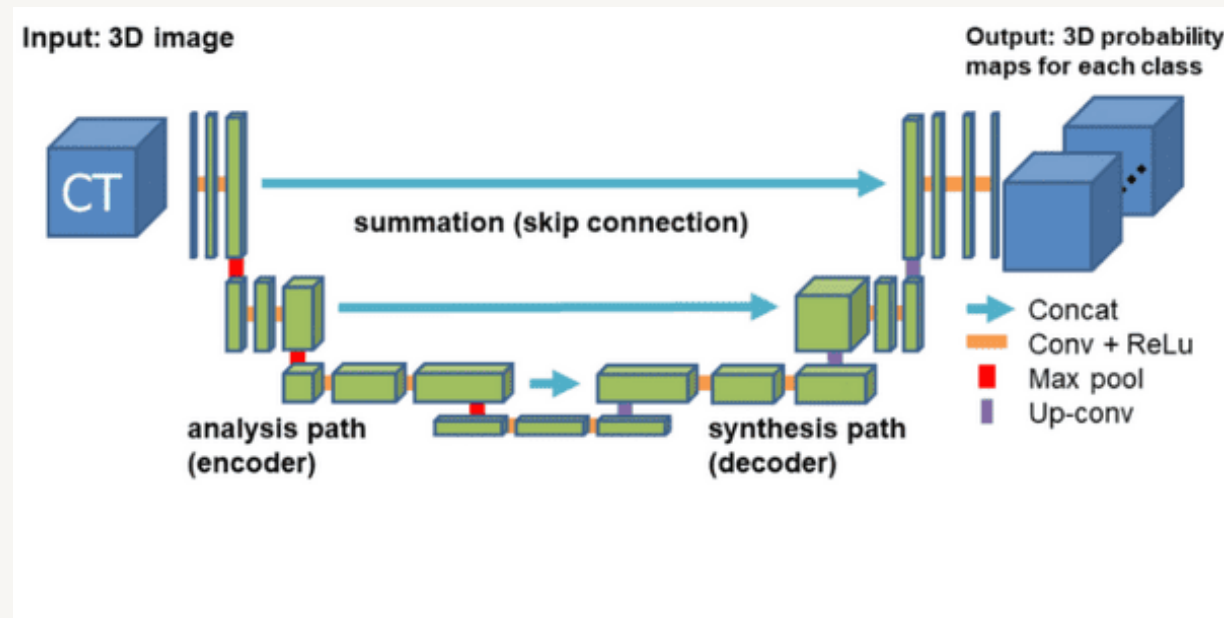
File Conversion, Model Creation

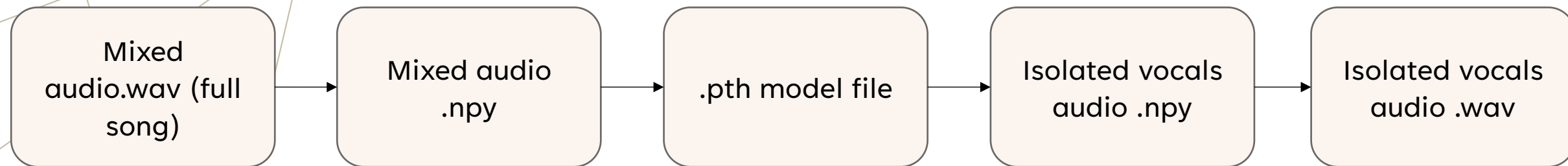
WORK DISTRIBUTION

Data Gathering, Coding, Testing

DATA CONVERSION

mp3, wav, xlsx, npy





IMPLEMENTATION

- **Short Time Fourier Transform (STFT)** performed on mixed source .wav file
 - **Logarithm taken of magnitude** of result
- Log-mag stored in **.npy** for speed, compact size and stability
- .npy file fed into **.pth model file**, which outputs .npy file with isolated vocals
- Output .npy converted back to .wav file with isolated vocals via **inverse STFT** and **Griffin-Lim algorithm**

DISCRETE TIME ANALYSIS

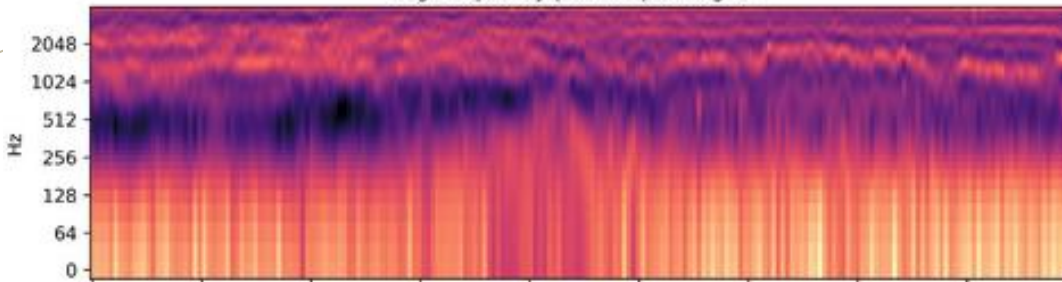
Short Time Fourier Transform (STFT)

Creation of Spectrogram Matrix Object via
Short Time Fourier Transform of Audio Signal

$$STFT\{x[n]\} = \sum_{n=0}^{N-1} x[n]w[n-m]e^{-i\omega n}$$

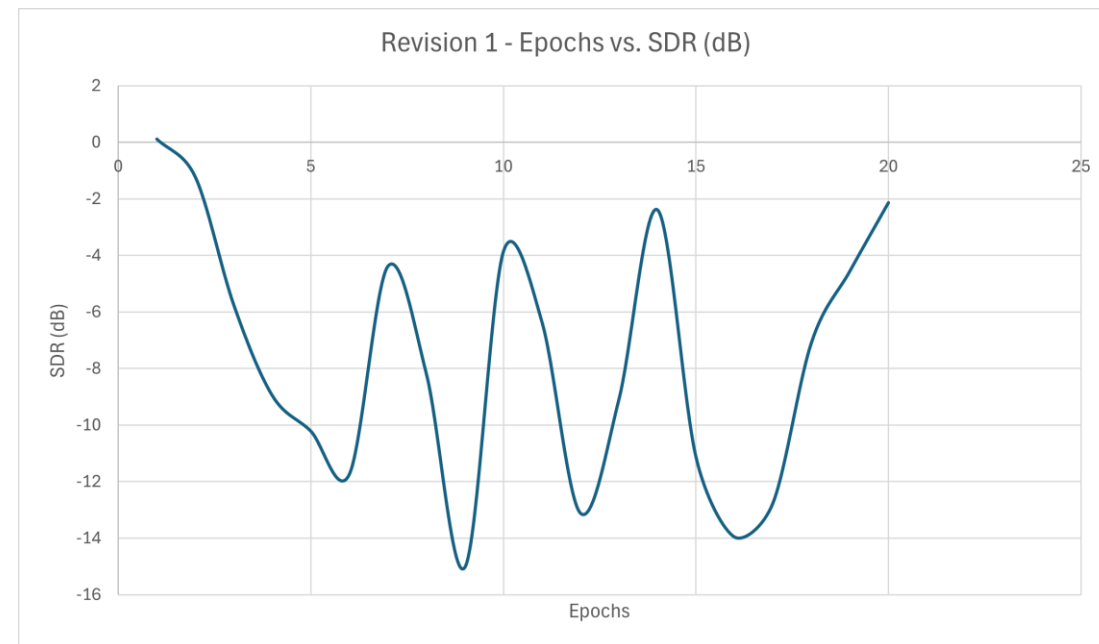
- Records the Fourier Transform of a signal as it changes over time
- Continuous time version of STFT uses windows to segment the signal
- Discrete time version splits up signal into chunks and then applies the Fourier transform to each chunk
- The Fourier transforms of each chunk are then stored into a matrix which can then be saved for U-Net Modelling

Log-frequency power spectrogram



ITERATION DESIGN

- Used U-Net training to attempt to create a model that can separate detected vocals from instrumental audio
- Epoch Data from training used to improve and iterate on U-net parameters
 - Specifically, the SDR (Signal Distortion Ratio) and Loss factors
- Formatting of file conversion pipeline was also improved upon
- XLSX to .npy arrays for spectrogram storage
- Rectangular Complex STFT to Log Magnitude STFT
- Different loss functions to improve metrics



The left side of the slide features a series of overlapping, thin, light-brown geometric lines that form various polygons and triangles, creating a dynamic, abstract pattern.

VIDEO DEMONSTRATION

Project

ENSC429_Group3

data

Samples

sunflower-street-drumloop.mp3

Spectrograms

temp

EasyWayOut10sClip.npy

frances-forever-space-girl.npy

output.npy

Documents

ENSC429_Group3_Project_Proposal.pdf

scripts

audio_to_npy_spectrogram.py

EasyWayOut10sClip_reconstructed.wav

frances-forever-space-girl_reconstructed.w

isolate_vocals.py

npv_spectrogram_to_audio.py

.gitignore

EasyWayOut10sClip.wav

frances-forever-space-girl.wav

README.md

spectrogramCNN.ipynb

vocal_isolator.pth

External Libraries

Scratches and Consoles

audio_to_npy_spectrogram.py

isolate_vocals.py

spectrogramCNN.ipynb

Python 3.12.10: auto-start

25

26

27

28

29

30

31

32

33

model=model,

train_loader=train_data_loader,

val_loader=val_data_loader,

criterion=criterion,

optimizer=optimizer,

device=device,

num_epochs=epochs,

checkpoint_path='vocal_isolator.pth'

)

20.070000000000014

20.100000000000016

20.110000000000017

20.120000000000002

20.130000000000002

20.140000000000022

20.150000000000023

20.160000000000025

20.170000000000027

20.180000000000028

20.19000000000003

20.20000000000003

Validation: Loss = 0.0903, MAE = 0.0903, SDR = 6.4394 dB

Saved prediction to predicted_epoch_19.npy

Training complete, model saved at vocal_isolator.pth

Error running console

Error: Cannot run program "C:\Users\chis\AppData\Local\Programs\Python\Python313\python.exe" (in directory "C:\Users\chis\Documents\GitHub\ENSC429_Group3"): CreateProcess error=2, The system cannot find the file specified

ENSC429_Group3

spectrogramCNN.ipynb

261:1

CRLF

UTF-8

4 spaces

Python 3.12 (spleeter-pytorch-master)

10:56 PM

2025-08-04