

Singing Voice Conversion: CycleGAN- VC

This presentation introduces CycleGAN-VC, a novel non-parallel voice conversion method leveraging CycleGANs to learn mappings between source and target singers without relying on parallel data. This approach allows for high-quality voice conversion in a general-purpose manner, paving the way for novel applications and advancements in speech processing and augmentation.

21ucc129 Ayush Bajaj

21ucs226 Vamsi Krishna

21ucs196 Shreshtha Gupta

21ucs217 Swayam Bhatt

Objective



Source Singer: Arijit Singh



Target Singer: Kishore Kumar

Our implementation aims to convert Arijit Singh's voice to sound like Kishore Kumar's.

Problem: Parallel Data Reliance

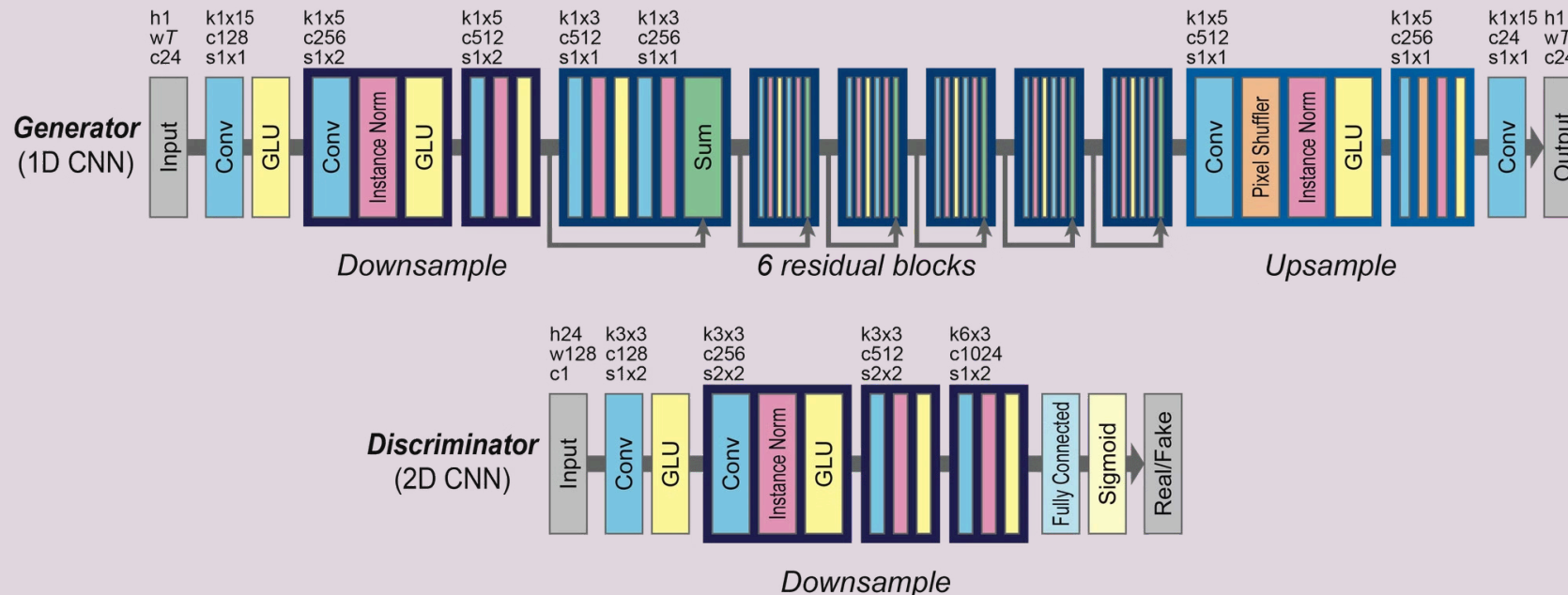
Limitations of Parallel Data

Traditional VC methods necessitate paired source and target speech, often unavailable for various voices.

CycleGAN-VC's Innovation

This method overcomes this limitation by utilizing a CycleGAN, enabling learning without parallel data.

CycleGAN-VC Architecture



1

Learning the Voice of someone

One of the important characteristics of speech is that it has sequential and hierarchical structures. It is important for the model to learn the fundamental characteristics of the singers' voices.

2

Gated Convolutional Neural Networks

Instead of using commonly used RNNs for sequential data this uses Gated CNNs which enhance the model's ability to capture complex speech features and relationships.

3

Gated Linear Unit

In a gated CNN, gated linear units (GLUs) are used as an activation function. A GLU is a data-driven activation function, and the gated mechanism allows the information to be selectively propagated depending on the previous layer states.



The Key Components for learning

1

Adversarial Loss

It guides the generator to produce converted speech that is indistinguishable from the target speech, minimizing the difference in distributions.

2

Cycle-Consistency Loss

It enforces the mappings to be bijective, ensuring that the converted speech can be converted back to the original source.

3

Identity-Mapping Loss

This loss encourages the generator to maintain linguistic information, preserving the identity of the source speaker in the converted speech.

Dataset Preparation



Song Collection

Gather songs from both the source and target singers, ensuring diversity in their styles.



Vocal Separation

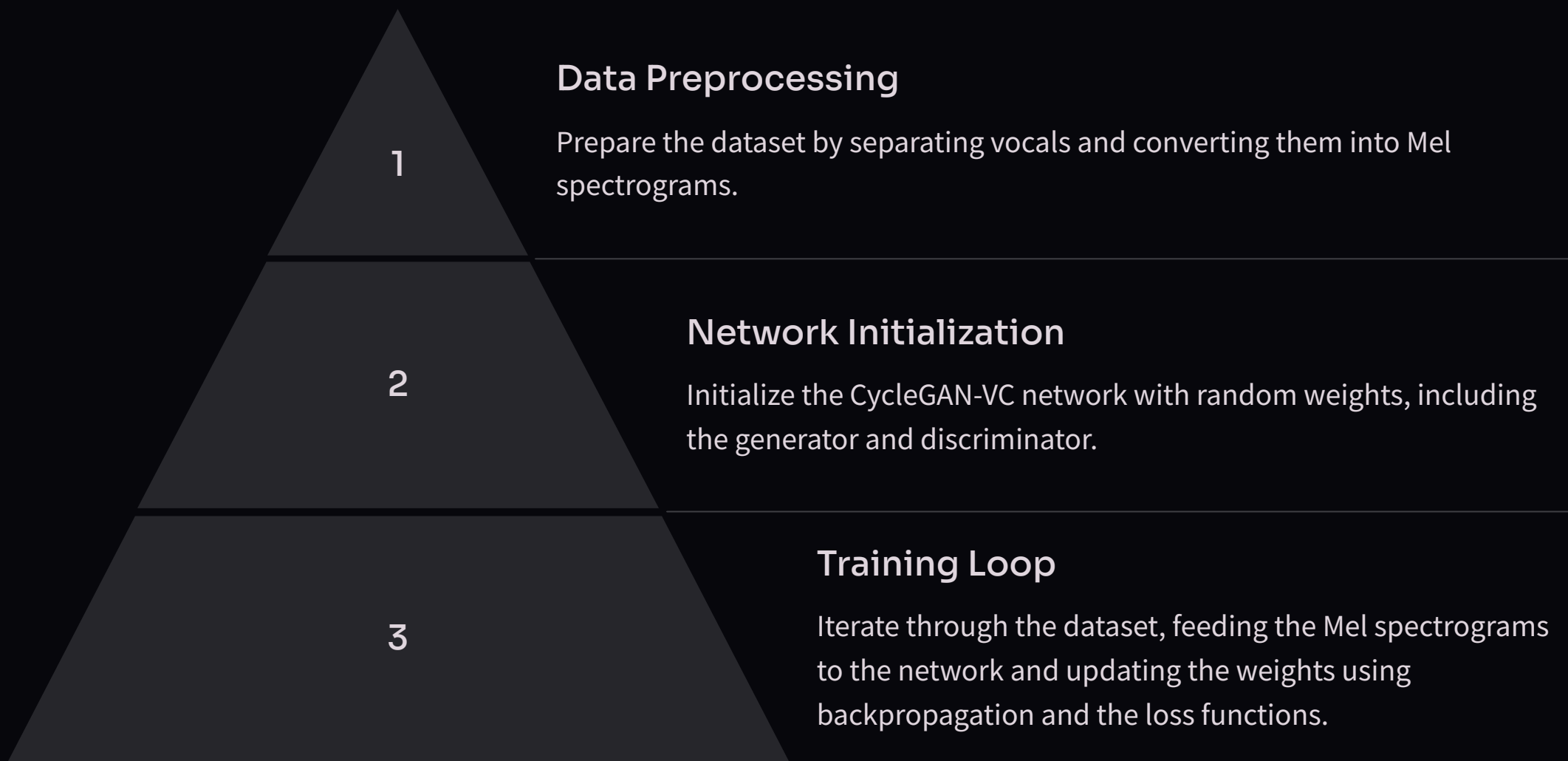
Isolate the vocal tracks from the background music using techniques like Non-negative Matrix Factorization (NMF) with Spleeter library.



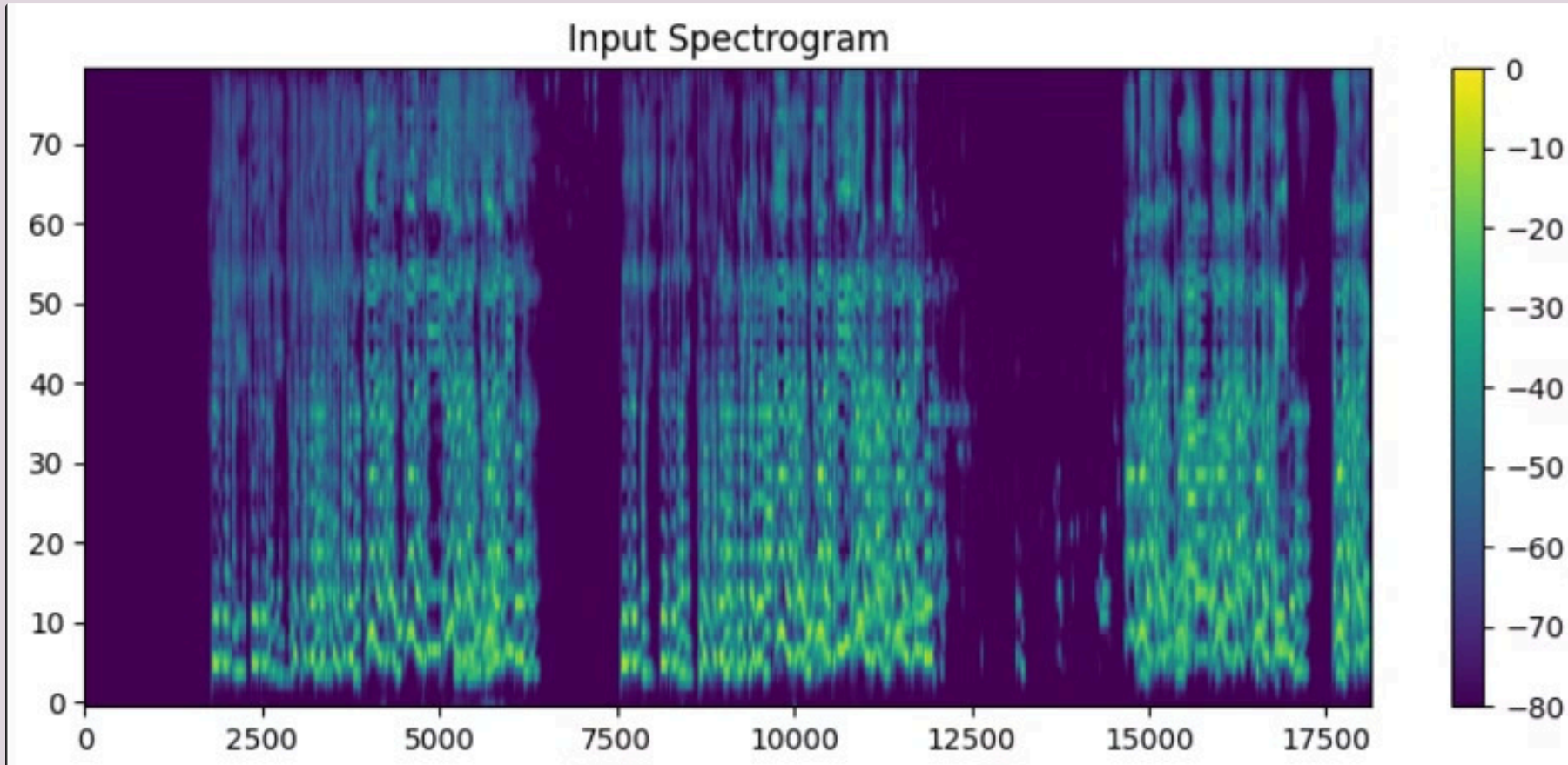
Spectrogram Conversion

Transform the vocal tracks into Mel spectrograms using Librosa library, which capture important frequency information for voice conversion.

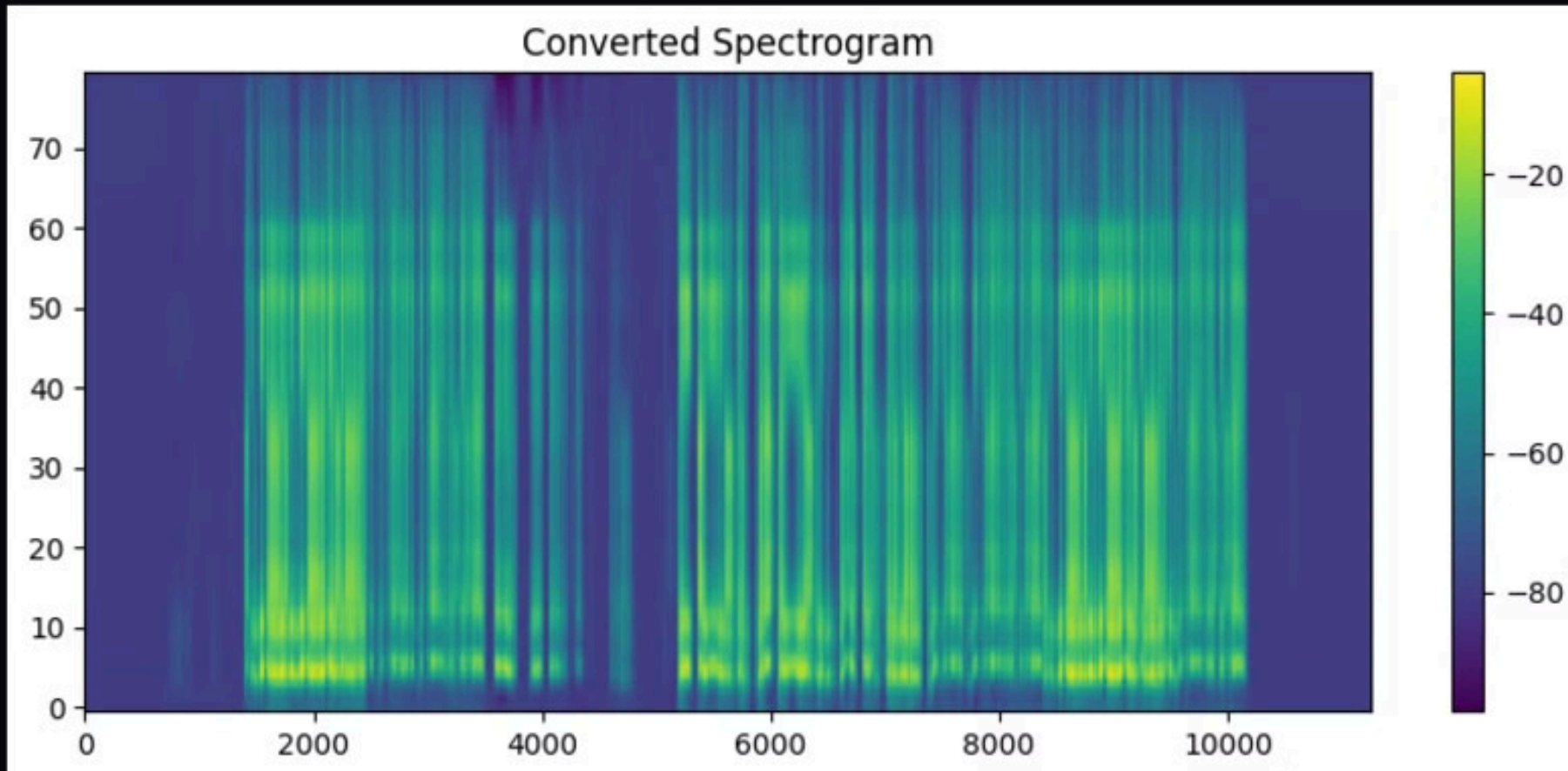
Training Process



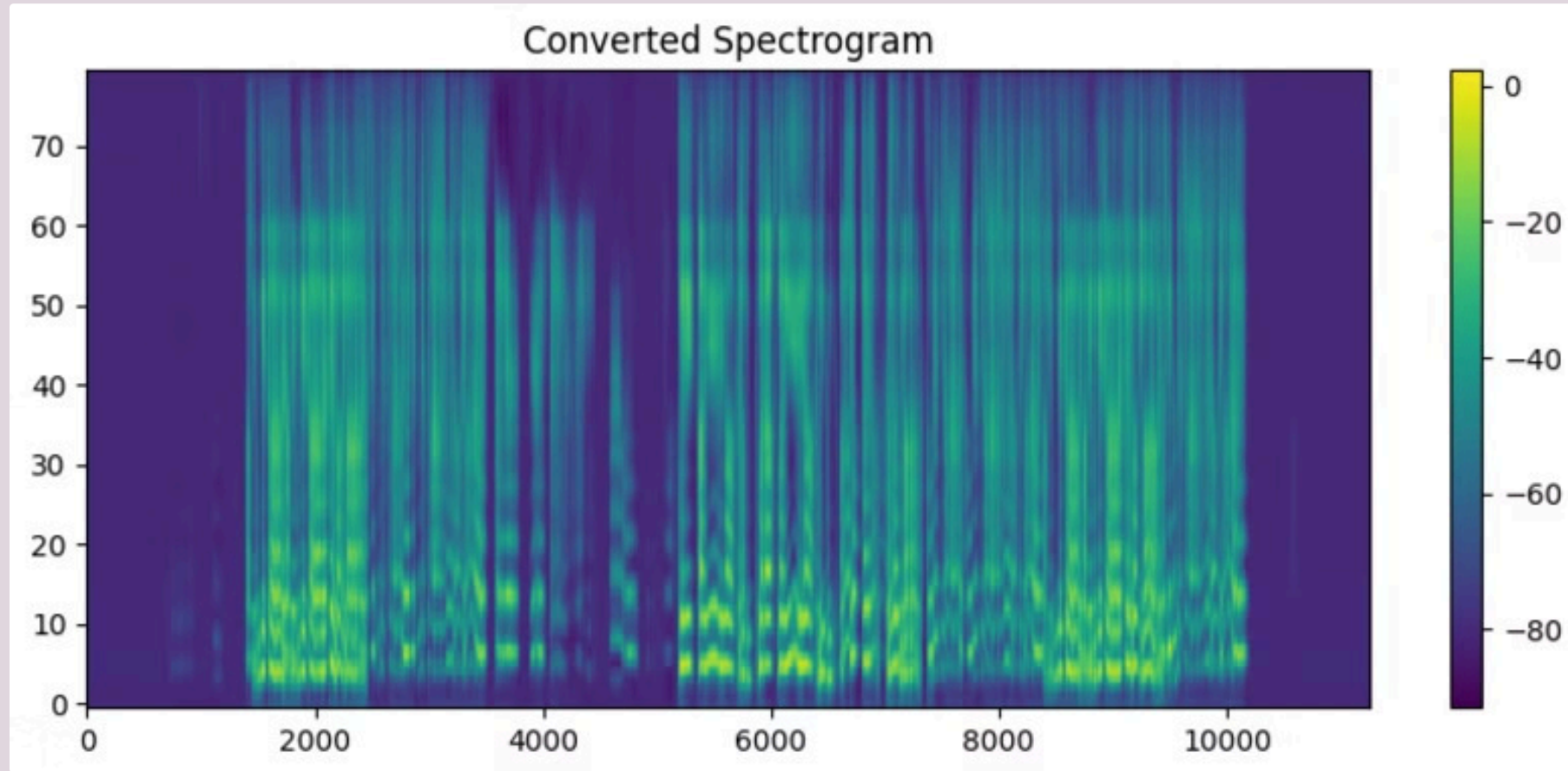
Eg- Input spectrogram of "Ye fitoor mera...."



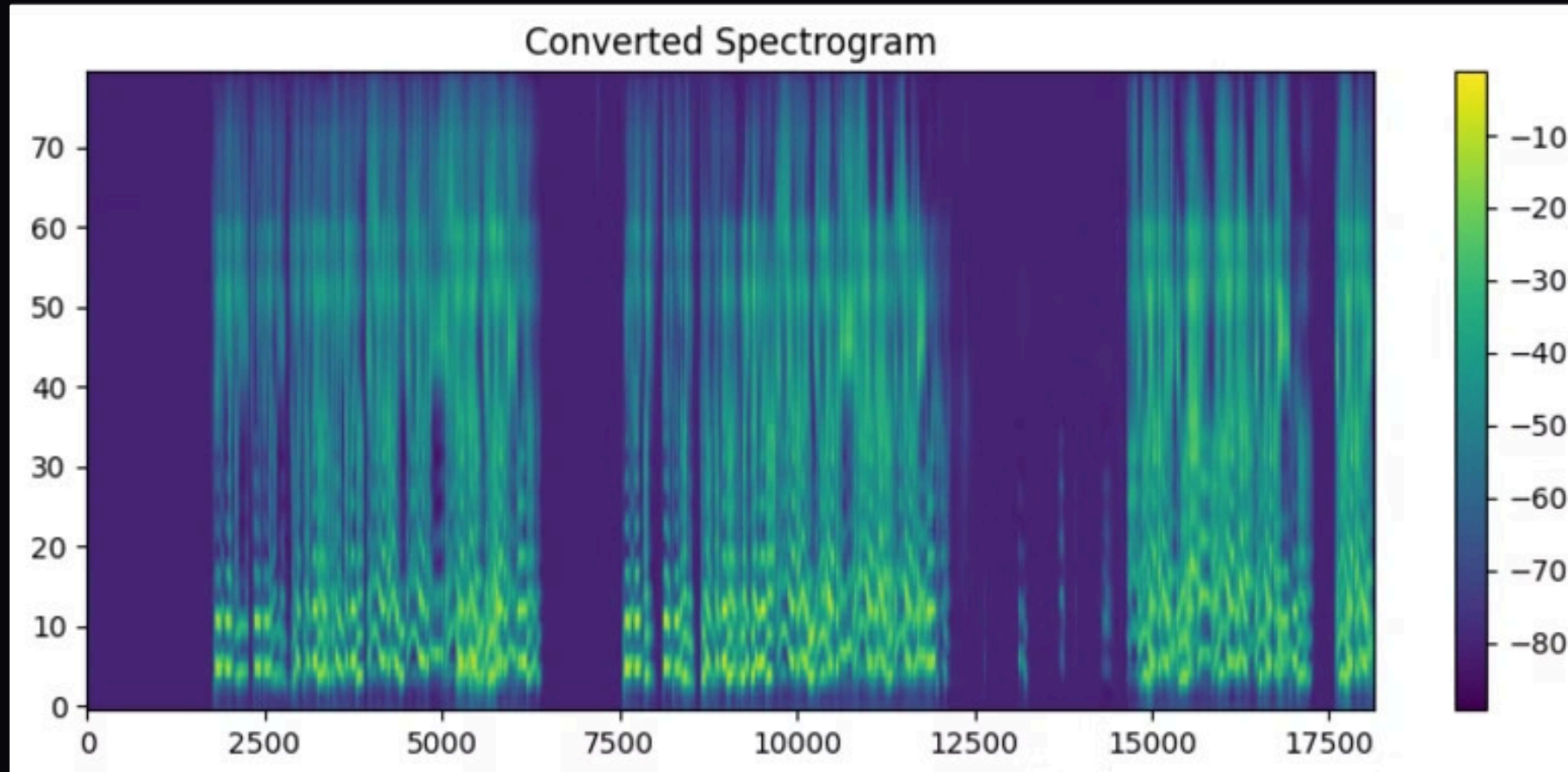
E-294




E-838





E-1355






Input Audio



SoundCloud
Yeh Fitoor Mera by Ayush Bajaj
Listen to Yeh Fitoor Mera by Ayush Bajaj #np on #SoundCloud

Output Audio



SoundCloud
arijit_to_kishore_epoch1355_...
Listen to arijit_to_kishore_epoch1355_ye_fito...

CycleGAN-VC: Benefits

Generalizability

The model can be applied to convert voices of different genders, accents, and languages, making it highly versatile.

High Quality

CycleGAN-VC achieves high-quality voice conversions, preserving the linguistic information and producing natural-sounding speech.

Simplicity

The method is computationally efficient and does not require complex alignment procedures, reducing computational overhead.

Future Directions

1

Multi-Speaker Conversion

Extending the model to handle simultaneous conversions between multiple source and target speakers.

2

Emotion Transfer

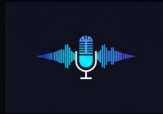
Developing techniques to transfer emotional qualities from one speaker's voice to another, enhancing expressiveness.

3

Real-Time Conversion

Optimizing the model for real-time voice conversion, enabling live applications in communication and entertainment.

Conclusion: A New Era in Voice Conversion



CycleGAN-VC presents a significant breakthrough in voice conversion by offering high-quality, general-purpose conversion without parallel data. The model holds immense potential for diverse applications in entertainment, accessibility, education, and privacy, ushering in a new era of voice manipulation and customization.