TimbreTron

Mahshid Alinoori

Table of Contents

Timbre

“The psychoacoustician's multidimensional wastebasket category for everything that cannot be labeled pitch or loudness ”

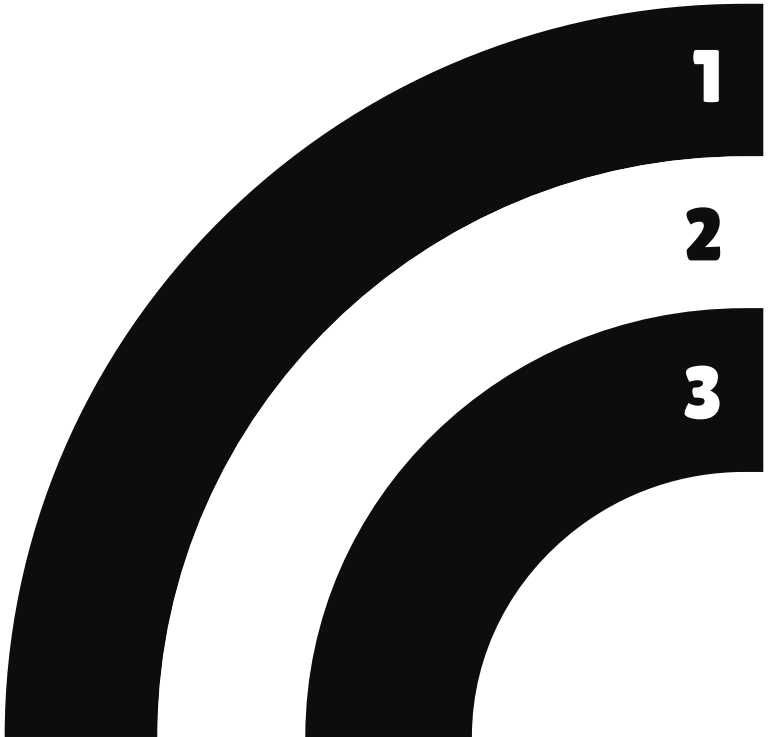
—McAdams and Bregman

Timbre Transfer

Transforming the timbre of a musical recording to match a set of reference recordings while preserving other musical content, such as pitch and loudness.

Inspired by Image style transfer.

TimbreTron Steps



1. Computing the log-magnitude of the CQT spectrogram
2. Performing timbre transfer using CycleGAN
3. Converting the generated log-CQT to a waveform using WaveNet synthesizer

Time–frequency Analysis

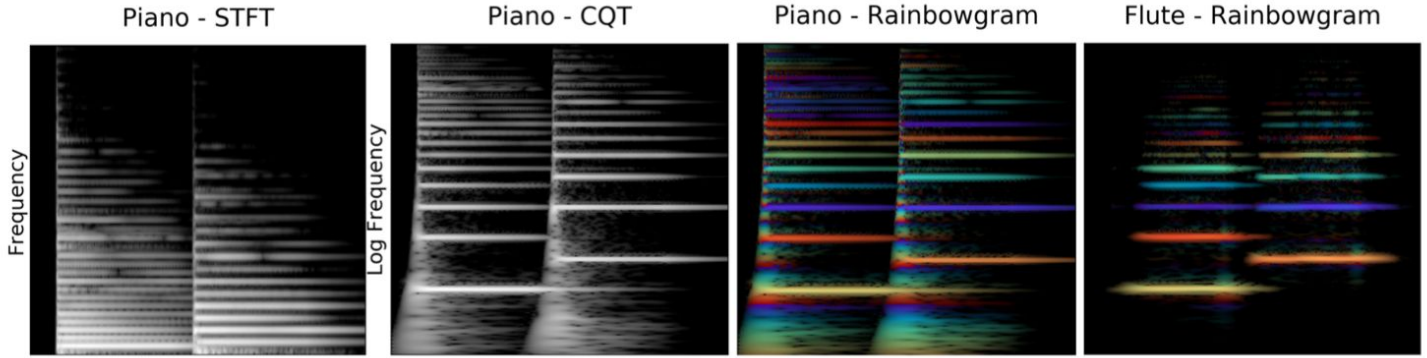
STFT

The discrete Fourier transform of the windowed signal

$$STFT\{x[n]\}(m,\omega_k) = \sum_{n=-\infty}^{\infty} x[n]w[n-m]e^{-j\omega_k n}$$

CQT

Geometrically spaced centre frequencies and resembles our auditory system.
Rainbowgram is a visualization with phase info encoded as color



Waveform Reconstruction

STFT: Griffin Lim

An algorithm for generating phase from STFT magnitude by randomly guessing the phase values and iteratively refining them.

CQT: ?

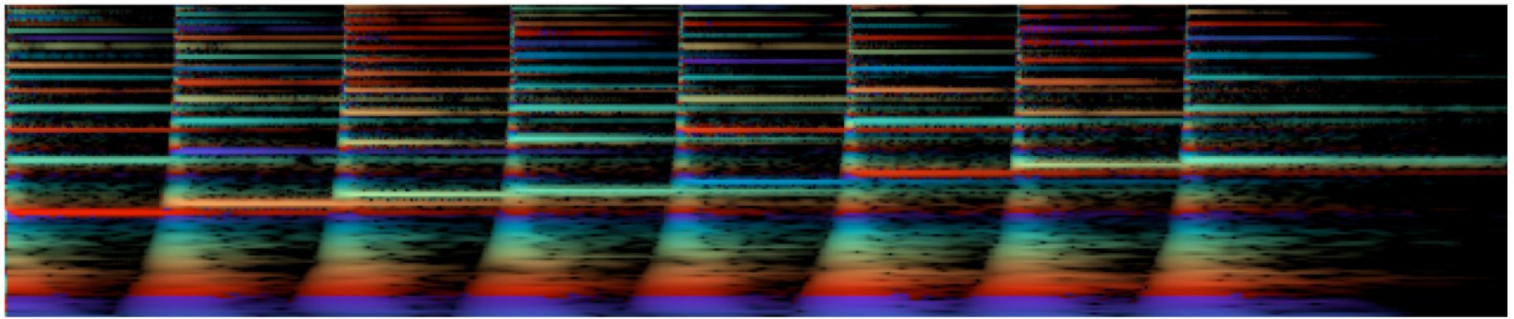
Existing Inverse algorithms need both phase and magnitude information.
In TimbreTron wavenet is used as the waveform synthesizer.

Why CQT?

Higher frequency resolution for lower frequencies and higher time resolution for higher frequencies

Approximate pitch equivariance: a pitch shift corresponds to a vertical translation. Pitch and Tempo **disentangling** is simpler

Harmonic are **approximately** integer multiples of f0 and each instrument has its own spectral signature



Why WaveNet?

The phase info is discarded due to difficulties in prediction and we finally need to infer the phase to recover the waveform

Waveform as an autoregressive generative model can generate high-quality waveforms conditioned on the generated CQT

Beam search and **inverse generation** are applied to deal with the difficulties that WaveNet has such as generated artifacts and missing or doubled attacks

CycleGAN

- Image-to-image translator without paired data.
- Using two discriminators and two generators.
- Adversarial loss and cycle consistency loss.

1. Replacing the deconvolution to remove checkerboard artifacts
2. Replacing the patch discriminator to achieve consistency in the spectrogram
3. Adding gradient penalty to balance the discriminator and generator
4. Applying Identity loss to preserve music content

Experiments

Dataset

Unrelated recordings of Piano, Flute, Violin, and Harpsichord

MIDI

Training the CycleGAN on MIDI data and testing it on real world data

Subjective Evaluation

Similarities between the generated piece and both the original and target pieces

STFT vs CQT

Human study to compare the similarity while using STFT or CQT

Conclusion

- High quality timbre transfer
- Working in CQT domain
- Using Wavenet as the waveform synthesizer
- Using CycleGAN as the translator

Thanks

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.