More info at
singing.style/transfer

Singing Style Transfer

CSE 481i
Winter 2019

Joseph Zhong, Andrew Li, Ollin Boer Bohan

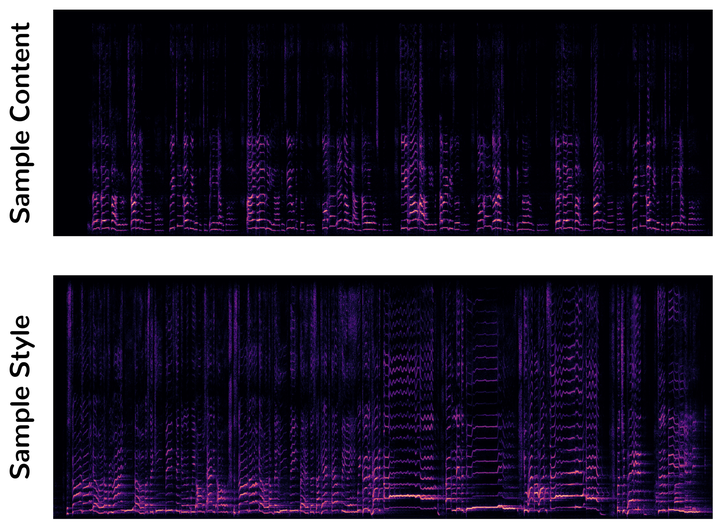
Problem & Use Case

The problem of image style transfer has been widely studied. We develop a method to perform style transfer for monophonic singing audio.

The target use case is music production—for example, stylizing an unprofessional singer to match a professional recording.

Data Representation

We represent audio as linear-frequency spectrograms:



The majority of available singing data is non-parallel (and we use this for training), but we have also prepared a small amount of parallel data for evaluating our model.

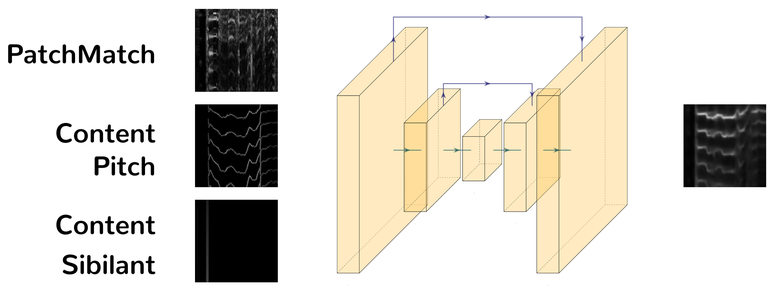
Featurization

We use the phoneme classifier from *Speaking like Kate Winslet* to featurize the input and style audio.

Spectral PatchMatch

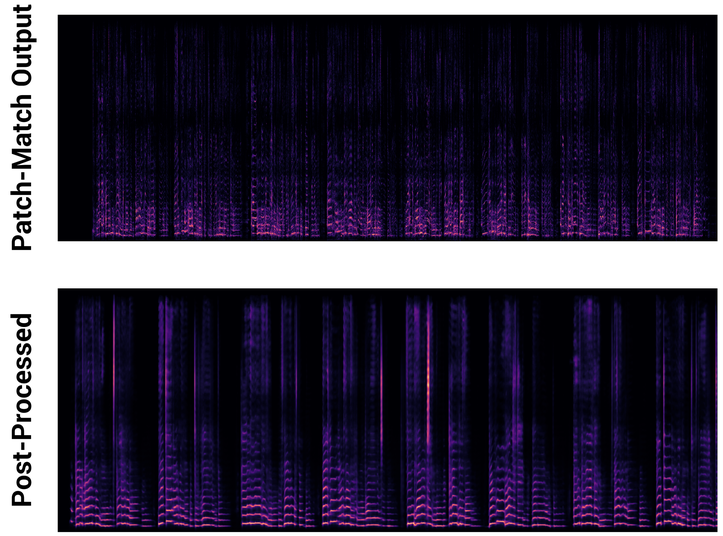
Inspired by *Deep Image Analogy*, our patch-matching component reconstructs the content using pitched slices of the style. Patch correspondences are selected using PatchMatch on the featurizer output.

Post-Processing Network



We use a simple U-net style post-processor to improve smoothness of the output spectrograms. The network takes in the PatchMatched spectrogram as well as channels indicating pitch curves and sibilant locations, and produces a smoothed output spectrogram.

Results



Future Work

- **Spectral PatchMatch**: we use a simple version of PatchMatch based on the original paper. *Deep Image Analogy* uses a multi-scale bidirectional PatchMatch with content blending, which should be used instead.
- **Post Processing & Phase**: the post-processor should be trained to predict plausible phase as well as amplitude, and should be capable of producing sharper output spectrograms.