Timbre Transformation

Yu-Chen Huang
Carnegie Mellon University
California, CA 94035
yuchenhu@andrew.cmu.edu

Hung-Kuang Han
Carnegie Mellon University
California, CA 94035
hungkuah@andrew.cmu.edu

Yi Wang
Carnegie Mellon University
Pittsburgh, PA 15213
wangyi@andrew.cmu.edu

1 Research problem statement

Timbre is the characteristic of instruments. People use timbre to identify different musical instruments. Sometimes, people may want to hear a piece with different instrumentation. Timbre transformation would expand the scope of sounds of any existing music, for instance, we could transform a piece of classical guitar performance into one performed with electric guitar without having the musician playing it all over. Together with the technique of instrumentation separation, we could apply timbre transformation on each individual instrument and yield a high-quality remix of any existing piece of music.

2 Literature research

In the previous work of [1], Settel, et al.(1994) use FFT/IFFT in real time to conduct digital signal processing in Max programming environment, which requires no compilation for digital signal processing(DSP). They use what's called overlap-add technique, including the following steps: (1) windowing input signal (2) transformation of the input signals into the spectral domain using FFT (3) operate on signal's spectra (4) resynthesis of modified spectra using IFFT (5) windowing the output signal. Their operation in the spectral domain includes convolution, addition, square root. We want to apply similar procedures for our timbre transformation project on data from Megenta's NSynth.

3 Proposed solution

Currently, our research mainly focused on extracting timbre information from frequency domain. Similar to the procedure we stated above, after we obtain the spectra with STFT (Short-time Fourier transform), we then would build a model to find the transformation matrix between the features of different timbre. In this stage, we would experiment multiple approaches of transformation on instruments of different styles but belong to the same instrument family.

For example, we would choose a set of synthesized guitar sounds as unprocessed source and a set of acoustic guitar sounds as target source, both from the guitar family. Obtain their spectrum and then arbitrarily apply operations such as convolution, addition, subtraction on some combinations of frequencies of their spectrum and recompile those spectres into music samples. Finally use both our arithmetic evaluation metrics and intuition of our hearing to judge whether we have obtained a high-quality sound of similar timbre within the target source set.

Ideally, this process would give us some guidelines and rules for high quality timbre transformation. If such experiment does not generalize well on properties for timbre transformation, we could always conduct comparison experiments on frequency spectra of same instrument family and same styles to find out characteristics of their frequency spectra for each sound source. Then use those characteristics as guideline for operation we could conduct on our unprocessed source to obtain target source. On the other hand, if such experiments managed to deliver valuable information in one instrument family, we would probably take it to other instrument families, hoping it would give

Preprint. Under review.

Table 1: Timeline and division of work

Task	ETA	Members
Search baseline model	10/10	Huang, Han
Design training model	10/17	Huang, Han, Wang
Transform feature for baseline model	10/24	Huang, Han
Transform feature for training model	10/24	Wang
Experiment on baseline model	11/07	Huang, Han
Implement training model	11/07	Wang
Finish midterm report	11/10	Huang, Han, Wang
Experiment on training model	11/14	Huang, Han
Improve feature transformation	11/21	Wang
Improve training model	11/28	Huang, Han
Experiment on training model	12/05	Wang
Finish final report	12/13	Huang, Han, Wang

information on timbre-transformation of different styles sound of same instrument family. After feature transformation, we we will use IFFT to re-synthesize the feature back to the signals.

However, frequency-domain methods could run into phase reconstruction issues. If those issues block us from getting a high-quality timbre transformation, we would switch to time-domain methods.

4 Dataset

We will use The NSynth Dataset, a large-scale and high-quality dataset of annotated musical notes. The NSynth Dataset was collected and published by Google, which contains 305,979 musical notes with different sound production methods (i.e., acoustic, electronic, and synthetic) and different instrument families (e.g., bass, guitar, string).

5 Evaluation metrics

We will use KL(Kullback–Leibler) divergence as evaluation metrics to calculate the difference between the spectrum matrix of, for example, the transformed guitar sound set and the acoustic guitar sound set.

6 Timeline and Division of work

See Table 1.

References

[1] Settel, Z., & Lippe, C. (1994). Real-time timbral transformation: FFT-based resynthesis. Contemporary Music Review, 10 (2), 171-179.

[2] Wakabayashi, Y., Fukumori, T., Nakayama, M., Nishiura, T., & Yamashita, Y. (2017, March). Phase reconstruction method based on time-frequency domain harmonic structure for speech enhancement. In 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 5560-5564). IEEE.

[3] Yoshii, K., Tomioka, R., Mochihashi, D., & Goto, M. (2013, November). Beyond NMF: Time-Domain Audio Source Separation without Phase Reconstruction. In ISMIR (pp. 369-374).