Timbre Transformation

Claire Wenner

Department of Electrical & Computer Engineering, University of Rochester

HAJIM SCHOOL OF ENGINEERING & APPLIED SCIENCES
UNIVERSITY of ROCHESTER

Abstract

Timbre is the quality of an instrument which makes it identifiable from other instruments. Sometimes, it is desirable to change a melody played by one instrument to that same melody played by a different instrument. We can achieve this with manipulation of the timbre. This paper presents a method for changing the timbre of an instrument in a monophonic recording. Models for clarinet, trumpet, and violin were created.

Goal

This project aims to create models for the characteristics of timbre for certain instruments which can be used to modify the timbre of an instrumental recording without affecting any other parameters of the performance. More specifically, we look to preserve the expressiveness of the performance while manipulation the timbre. While there do exist many audio synthesizers which achieve good results, this project looks more specifically at transforming the timbre of a sound rather than simply generating it.

Background

Timbre itself is difficult to define in quantitative terms. It is often defined as what it is not: "that attribute of auditory sensation which enables a listener to judge that two nonidentical sounds, similarly presented and having the same loudness and pitch, are dissimilar"[2]. Certain temporal and spectral aspects of sound are important in instrument recognition, such as harmonic amplitudes, onset, envelope, and vibrato. This project looks at harmonic amplitudes and onsets as the main indicators of timbre of classical instruments.

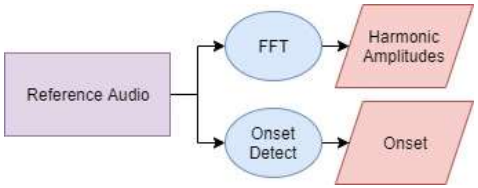
Methods

Dataset

To create the model, we use samples from the good-sounds dataset[1]. The dataset is mainly comprised of separate recordings of singular notes. The reference recordings for violin, clarinet, and trumpet are used in this project.

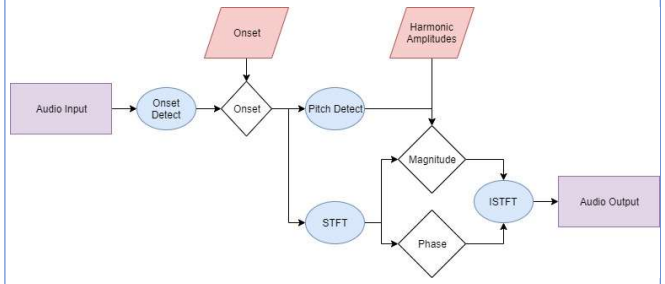
Model Creation

The models used in this project were created using direct extraction of features. For each audio recording, the fast Fourier transform (FFT) was computed and the amplitudes for the fundamental and the first 10 harmonics were calculated. The harmonic amplitudes were averaged across instrument..

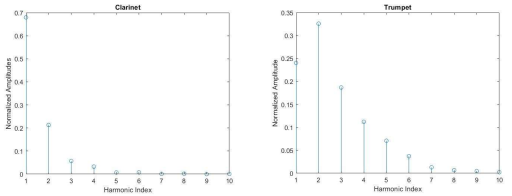


The onsets for each instrument were extracted directly from the time domain signals.

Timbre Transformation



Models



Results

The model appears to be accurately extracting the harmonic amplitudes and onsets from the dataset. However, when the transformed instrument is transformed back to the time domain and listened to, it doesn't sound quite right. There is perhaps a mistake in the recombination with the phase, or with the ISTFT method used.

Future Work

While this is a good baseline for modeling and transforming timbre, there is obviously a need for more work in the application of a model to an audio signal. Further improvements could be made by considerations of other timbre features, such as vibrato which is present in some instrument (i.e. saxophone, violin) but not others (i.e. clarinet, piano). Some image processing methods such as edge detection could also be implemented on the spectrogram of the sound.

References:
[1] Romani Picas O. Dabiri D., Serra X. **"A real-time system for measuring sound goodness in instrumental sounds"** 138th Audio Engineering Society Convention, Warsaw, 2015
[2] Acoustical Society of America Standards Secretariat (1994). "Acoustical Terminology ANSI S1.1–1994 (ASA 111-1994)". American National Standard. ANSI / Acoustical Society of America.