# TIMBRE TRANSFORMATION

**Claire Wenner**
University of Rochester, Department of
Electrical & Computer Engineering
claire.wenner@rochester.edu

## ABSTRACT

Timbre is the quality of a sound that makes it identifiable as a certain instrument, sometimes referred to as the quality of a sound. 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 through manipulation of the timbre. This paper presents a method for changing the timbre of an instrument in a monophonic recording, while maintaining the pitch, timing, and expressiveness of the original recording

## 1. INTRODUCTION

When we listen to music, the type of instrument being used has a large influence on how we experience the piece. In a classical orchestra setting, for instance, the instrumentation plays a significant role. However, even when a certain instrument is desired, we don't always have access to that instrument or are unable to ourselves play it. In the case of composition, for instance, we want to be able to create any instrument sound from some user input. One way to do this, as this approach looks at, is to transform the timbre of an input melody. In this way, a composer could theoretically play all the different instrument parts on a single instrument, and then have the ability to choose which instrument should "play" each line.

There are other methods of creating instrument sounds with different timbres, which are largely synthesis-based or sample-based. However, the main drawback of these methods is they can sound mechanical and not like a musical performance. By taking an already expressive input melody and using a transformative method as opposed to a generative method, we can preserve that expression and convert only the instrumentation (timbre) of the melody.

The timbre of an instrument itself is not easy to quantify. It is often defined by what it is not, as in the ASA definition: "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]. As some studies have shown, however, there are certain aspects of sound that are more influential in the human perception of instrument classification [3], [4]. Attributes of sound including harmonic structure, onset, and vibrato contribute to the timbre of a sound. In the presented method, the harmonic structures and onsets are considered.

## 2. MODEL CREATION

In order to impose a timbre of an instrument onto a melody, we must first define a model for the timbre of that instrument. We are interested in modeling the harmonic structure and the onsets for each of the instruments. Samples from the good-sounds database are used [5]. Originally created to define measures of how "good" a sound played by a musician is, this dataset is ideal for our purposes as it contains recordings of single notes played on many different classical instruments. For simplicity, this project looks at creating models for the trumpet, clarinet, and violin. We use the "reference" recordings for these instruments from the dataset and use the recordings that were made with the Neumann U87 microphone.

### 2.1 Harmonic Structure

The harmonic structure for each instrument is modeled by a vector of 11 weights, one for the fundamental and each of the first 10 harmonics. The audio file is normalized to have a maximum amplitude of 1, and then transformed into the frequency domain using the fast Fourier Transform (FFT), as in Equation 1. The location of the fundamental frequency is given, and the frequencies of the harmonics are calculated as integer multiples of the fundamental. To account for slight errors in the FFT, we find the maximum magnitude value within 20Hz of the predicted harmonic frequency. The energy is computed by squaring the magnitude, and this is normalized by dividing it by the sum of the energy, and then saved as the energy for that harmonic.

$$X(n) = \sum_{k=0}^{N-1} x(k) e^{(-j2\pi nk)/N} \quad (1)$$

This process of extracting the harmonic energies is repeated for a mid-range of frequencies for each instrument. After this process is completed for a particular instrument, the magnitude values are averaged for each harmonic. This is used as the model of the harmonic structure.

### 2.2 Onsets

Since onsets are difficult to model as they contain a significant amount of inharmonic energy, we directly extract

the onset for each note. We use an energy-based onset detection method to find the location of the onset. For each window, the signal envelope energy is calculated in Equation 2:

$$E_w(n) = \sum_{m=-M}^{M} \left| x(n+m)w(m) \right|^2 \qquad (2)$$

where n is the sample index, x is the input signal, M is the window length, and w is the window function. Then, the envelope derivative is calculated as follows in Equation 3:

$$\Delta E(n) = \left| E_w(n+1) - E_w(n) \right| \qquad (3)$$

Finally, the energy is normalized, and thresholding and peak picking are used to determine the onset location. The onset is taken to be half of one frame prior to and two frames after the onset location.

## 3. TRANSFORMATION OF TIMBRE

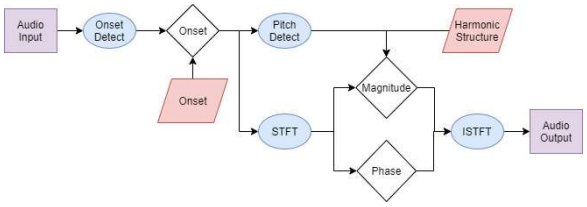


*Figure 1: Diagram of the timbre transformation process*

### 3.1 Onset Transformation

The overall process for transforming the timbre of an inputted melody is shown in Figure 1 above. The first step in the process of transforming an inputted melody from one timbre to another is the onset detection and replacement. The same energy-based onset detection as in Section 2.2 is used, and the information from half of one frame before the onset and two frames after the onset are replaced with a normalized version of the model onset. It is normalized to have the same maximum value as the original onset. Then, the transition between the model onset and the input data is smoothed using a basic moving average method to mask the change in data.

### 3.2 Pitch Detection

A necessary step in the process of timbre transformation is pitch detection. This is the most important aspect of the original audio to preserve. For determining the pitches of the harmonics of the note, the fundamental pitch is also necessary. In this project, the YIN method for pitch estimation is used [CITATION]. This is adequate for our purposes as we are only considering a monophonic melody as the input.

The YIN method starts by computing the difference function for a specified time window. The next step computes the cumulative mean normalized difference function and chooses dips in the function that exceed a defined threshold. Finally, parabolic interpolation is used to obtain the exact locations of these dips. Calculating the distance in samples between adjacent dips gives the fundamental period for this window.

### 3.3 Harmonic Structure Transformation

To obtain the magnitudes of each frequency over time, we use the short-time Fourier transform (STFT) on the signal with the adjusted onset. The STFT calculates the FFT over a defined window length, producing both magnitude and phase information over time. For each window of the STFT, the fundamental frequency is found using the YIN method in Section 3.2, and the frequencies of the harmonics are found as integer multiples of the fundamental frequency. At each of these frequencies, the value of the harmonic structure model is multiplied by the total energy at that window, and the square root of this is taken to get the magnitude value of that harmonic. By relating the model harmonic energy to the energy of the total frame, dynamics should be preserved as the energy of the frame is preserved. The new magnitude value is saved in a new matrix which begins normalized at zero.

### 3.4 Recombination

To create an audio file of the new instrumentation, the new magnitude information is recombined with the original phase data using Equation 4.

$$S = |S| e^{j\theta} \qquad (4)$$

where |S| is the magnitude of S and $\theta$ is the phase of S. Then, the inverse STFT (ISTFT) is used with the same window and hop length as the STFT in Section 3.3. The real components of this and the original sampling frequency are used to create the output audio file.

## 4. RESULTS

In its current state, this method has only been tested on singular notes from clarinet, trumpet, and violin recordings. An example of the resulting magnitude spectrum of a violin transformed to a clarinet can be seen in Figure 4, with the input violin spectrum in Figure 2 and the model harmonic amplitudes of the clarinet in Figure 3. As we can see, the method does a good job of transforming the harmonic peaks; however, there is also non-negligible noise at other frequencies. When listening to the output audio, this noise can be heard and makes the sustain of the note sound slightly less convincing.

## 5. CONCLUSION

We have presented a technique for transforming the timbre of one instrument to that of another, chosen by the user. The method considers the harmonic structure and onsets as the main influential characteristics of timbre. It also preserves the pitch and dynamics of the input note.

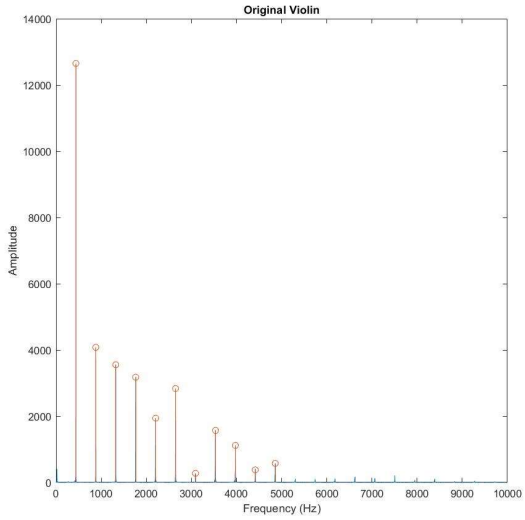


*Figure 2: Spectrum of input violin note. Harmonics are highlighted with orange circles.*

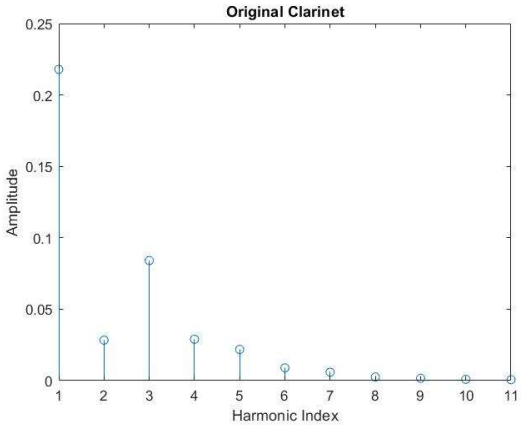


*Figure 3: Model of clarinet harmonic amplitudes*

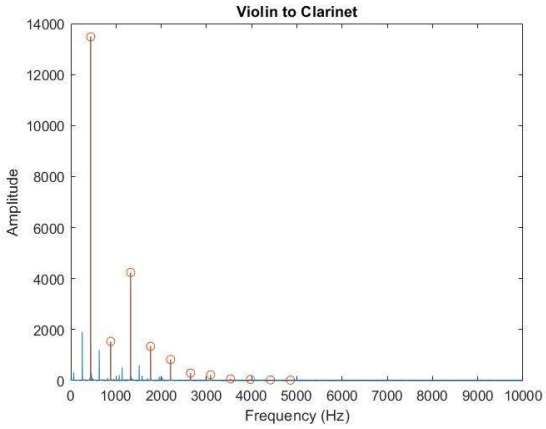


*Figure 4: The spectrogram of the transformed audio, with orange circles at harmonics.*

### 5.1 Future Improvements

Though the results show that the method produces an output closer in timbre to the desired instrument, they also show that it is not perfect. Two main problems in our approach could be contributing to this.

The first problem is the potential over-simplification of the harmonic structure model. Though it is true that the spectrums of classical instruments are mostly made up of harmonics, there is also some energy at other frequencies than the harmonics, which can be modeled as a noise signal. If we were to model this noise and use it to add some energy at other frequencies than the harmonics, this could potentially produce a better outcome.

The second potential problem in our method is that the phase is not manipulated at all and could be causing some of the issues with the output audio. Some phase generation methods could be implemented, or phase smearing around the onset could be implemented. Both approaches could potentially improve our output.

### 5.2 Limitations

Beyond the issues in the current outputs, there are also limitations to our current models. Firstly, this method has not yet been tested on a full melody. An issue which may arise with applying our method to a full melody is with onsets. If false onsets are detected, this will disrupt the output audio. Also, if pitch changes which were originally slurred but are detected as onsets, there will be an articulated note where it was not intended. Another limitation is the speed at which the YIN algorithm can detect pitch change. On passages with faster note transitions, if this algorithm fails then the wrong pitches will be output.

Another limitation comes in the limited frequency range of the model. This constraint was chosen as the timbre of an instrument itself changes with frequency. However, it limits the model to only work well in this frequency range. This can be rectified by creating models for certain frequency ranges for each instrument, and smooth between these ranges.

### 5.3 Future Work

In addition to the proposed improvements in Section 5.1, further improvements could be made by considerations of other timbre features. One such feature is vibrato which is present in some instruments (i.e. saxophone, violin) but not others (i.e. clarinet, piano). Therefore, vibrato could be taken out of a signal that is being transformed into an instrument that does not typically produce vibrato, and perhaps added to a signal being transformed into an instrument that does use vibrato. Another method that could be applied to this problem is image processing methods such as edge detection. We could use these methods on the instrument we are modeling as well as our input audio, since we can easily visualize the magnitude of the STFT of data as a spectrogram and then use image processing techniques on this spectrogram.

## 6. REFERENCES


[1] A. D. Cheveigné and H. Kawahara, “YIN, a fundamental frequency estimator for speech and music,” *The Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1917–1930, 2002.

[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*.

[3] E. L. Saldanha and J. F. Corso, “Timbre Cues and the Identification of Musical Instruments,” *The Journal of the Acoustical Society of America*, vol. 36, no. 11, pp. 2021–2026, 1964.

[4] J. M. Grey and J. W. Gordon, “Perceptual effects of spectral modifications on musical timbres,” *The Journal of the Acoustical Society of America*, vol. 63, no. 5, pp. 1493–1500, 1978.

[5] O. Romani Picas, H. Parra Rodriguez, D. Dabiri and X. Serra, (2017). *Good-sounds dataset,* Zenodo. [Data set]. Available: http://doi.org/10.5281/zenodo.820937