A Computational Exploration of Musical Timbre

Nichole Voegeli
Science, Technology & Mathematics
Lincoln University
Jefferson City, Missouri, USA
nichole.voegeli723@my.lincolnu.edu

David Heise
Science, Technology & Mathematics
Lincoln University
Jefferson City, Missouri, USA
heised@lincolnu.edu

***Abstract*—This extended abstract describes an investigation into musical timbre and the automatic discrimination of musical instrument sources. Arising from the first author's undergraduate honors thesis, this research begins with exploration of fundamental concepts before investigating perceptual correlates of timbre. The work described showcases how one might approach the problem of musical instrument classification through computational analysis. It is hoped that this work may lend ideas to others interested in exploring musical timbre, or likewise inspire others to explore a topic of interest in an unfamiliar domain.**

***Index Terms*—timbre, musical instrument classification, audio signal processing**

I. Introduction

Humans have the ability to recognize sounds and sound sources through a process known as auditory scene analysis. The auditory system creates a meaningful representation of the environment by deciding which sound signals should be grouped together or separated [1]. It is this system which can decipher sounds and their sources, for example, while in a city park where one can hear vehicles passing, children playing, ambient sounds of birds, insects, and other animals, and people talking. Auditory streaming groups similar sounds, received over an interval of time, into streams, facilitating the formation of conceptual units that may be processed at a higher cognitive level [2]. An example of auditory streaming occurs while listening to music; when a melodic line is heard, the auditory system groups those notes together into a sequence to create that melody (or countermelody) in one's mind [1].

Timbre is defined as "the characteristic quality of a sound, independent of pitch and loudness, from which its source ... can be inferred" [3]. Along with auditory streaming, the auditory system can differentiate sounds by their timbre. An example of differentiating sounds based on their timbre is noticing the difference in sound between a flute and trumpet. Both instruments can play notes that have the same frequency, but each has a different quality of sound that aids the auditory system in identifying what we are hearing. As an example, one can "pick-out" flute sounds from songs featured in Disney's Fantasia such as "Rite of Spring" or "Dance of the Flutes".

The innate ability of humans to differentiate between sounds produced by different sources motivates exploration of how one might automate the process of doing so computationally. The next section describes the overarching approach, data, experimentation, and results of the present work.

This work is supported by the National Science Foundation, award 2100874.

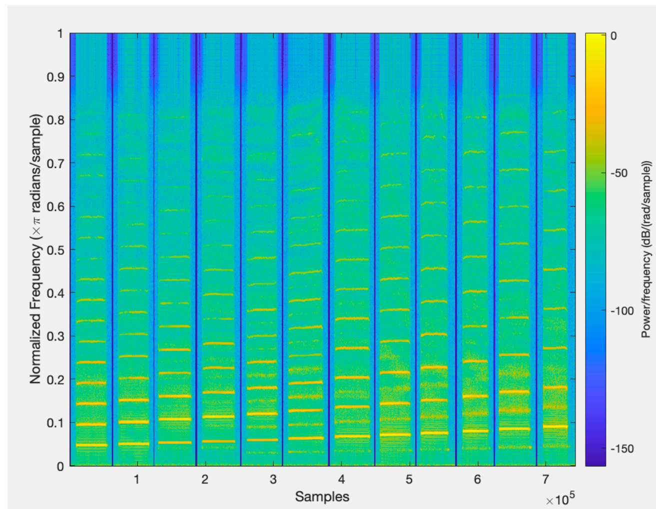

Fig. 1. Spectrogram of audio file containing 12 flute tones, ranging chromatically from C5 to B5. Sample rate is 22,050 Hz.

II. Approach

A. Perceptual correlates of timbre

Timbre has long eluded simple characterization in terms of quantifiable parameters. In 1951, Licklider commented that "until careful scientific work has been done on the subject, it can hardly be possible to say more about timbre than it is a 'multidimensional' dimension" [4]. In 1977, Grey conducted an influential experiment that mapped musical instrument tones into a three-dimensional space based on perceived similarities between pairs of instruments [5]. Axes I and II of Grey's three-dimensional space correlate with spectral energy distribution and spectral fluctuation, respectively. Spectral energy and spectral flux can be readily computed from an audio signal, thus suggesting that one could automatically discriminate between musical instruments by computing and comparing the spectral energy and spectral flux of isolated musical tones.

B. Data

Data for this study were taken from musical instrument samples available from the University of Iowa Electronic Music Studios [6]. Six files were chosen, representing flute, Eb clarinet, and trumpet, all played fortissimo without vibrato. The first set of audio files (for the three instruments) each

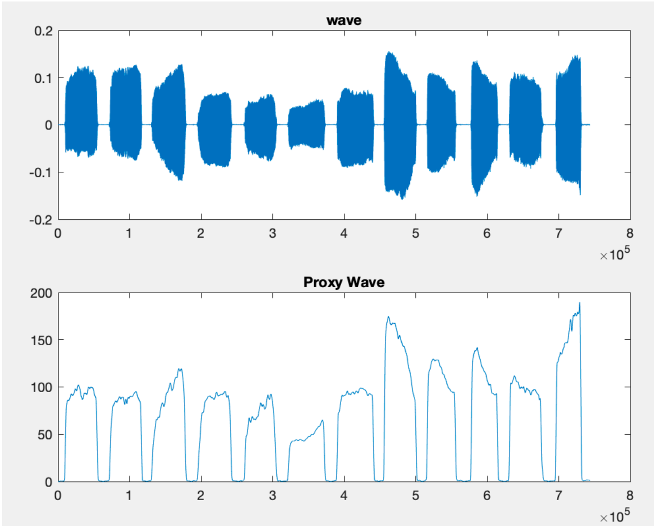

Fig. 2. Top: time-domain signal of 12 flute tones, played chromatically from C5 to B5. Bottom: a proxy for energy within the top signal, computed as the sum of the absolute value of samples over 100 ms sliding windows. The x-axis is in samples (22,050 Hz); the y-axis is in arbitrary units of amplitude.

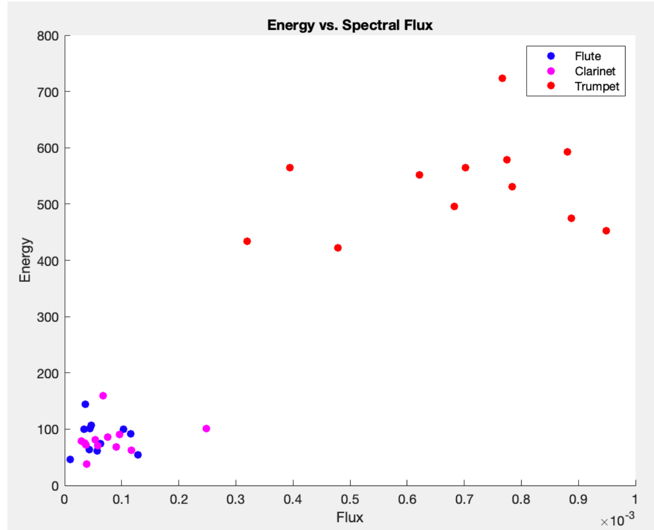

Fig. 3. Scatterplot of energy versus spectral flux for flute, clarinet, and trumpet tones. Units are based on arbitrary units of amplitude from the original audio signals.

contained twelve chromatic notes ranging from C5 to B5. The second set of audio files contained a sequence of various notes; only the first three notes of each file in this second set were analyzed. Audio was downsampled and filtered (using MATLAB function `decimate`) to 22,050 samples/second.

C. Isolation of tones, feature extraction

To compute the necessary parameters for timbre classification, each tone contained within an audio file must be isolated. All data processing/rendering was performed using MATLAB. Each audio file was plotted as shown by the example in Fig. 2. The bottom subplots (from the complete set generated) reveal that the pauses between each tone has a windowed energy of less than 5 units (arbitrary units of amplitude). A MATLAB script was written to detect the temporal boundaries of tones within each audio file, using a threshold of 5 units to detect onsets and offsets from the windowed energy signal. The matrix of detected onsets and offsets, comprising a list of tones in a given audio file, were used to extract features (maximum energy, spectral flux, and pitch) for each note. Energy was computed as the sum of the absolute value of the samples over a 100 millisecond sliding window. (Note that energy would typically be computed as the sum of the square of the samples over a window, hence referring to the "proxy for energy" in Fig. 2.) Spectral flux is calculated using the MATLAB function `spectralFlux`, computed over the first 100 milliseconds of each tone. Pitch (not used for timbre classification) is detected using the MATLAB function `pitch`.

D. Results

Fig. 3 shows a scatterplot of the analyzed notes, plotting maximum energy against spectral flux for each note. One can readily see that the trumpet notes have higher energy and spectral flux (as indicated by the red points in the upper half of the figure). As the energy of a signal is directly dependent upon the dynamic (loudness) used by the musician, spectral flux was chosen to separate woodwind tones from trumpet tones. A threshold of 3×10^{-4} was chosen to implement automatic discrimination between woodwind and trumpet notes, and this proved satisfactory for all data tested in this work. Flute and clarinet were not readily separated by this analysis; incidentally, [5] indicates that flute and clarinet are best separated on Axis III, which is not addressed here.

III. Conclusion

This exploration demonstrates the potential to classify musical instruments using features that correlate to timbre. This work, completed as part of the first author's undergraduate honors thesis [7], may be extended to discriminate between the two woodwind instruments (flute and clarinet); in particular, one might compute and use features from the notes corresponding to Axis III in [5]. Timbre remains an understudied and incredibly interesting topic for future research. It is hoped that this work will inspire others – especially undergraduates – to explore a topic of interest and engage in research, whether it be musical timbre or any other domain.

References

[1] A. Bernier-Robert, "Timbre space," Accessed: May 7, 2024. [Online.] Available: https://timbreandorchestration.org/writings/timbre-lingo/timbre-space.
[2] S. McAdams and E. Bigand, *Thinking in Sound: The Cognitive Psychology of Human Audition*, Oxford University Press, 04 1993.
[3] "Timbre definition & meaning," Accessed: May 8, 2024. [Online.] Available: https://www.dictionary.com/browse/timbre.
[4] D. B. Lindsley and S. Stevens, "Handbook of experimental psychology," 1951.
[5] J. M. Grey, "Multidimensional perceptual scaling of musical timbres," *The Journal of the Acoustical Society of America*, vol. 61, no. 5, pp. 1270–1277, 05 1977.
[6] L. Fritts and U. of Iowa Electronic Music Studios, "Musical instrument samples," Available: https://theremin.music.uiowa.edu/MIS.html.
[7] N. Voegeli, "Identifying musical instruments in audio," B.S. honors thesis, Lincoln University, 2024.