Harmonic Profiles of Orchestral Instruments using the Short-Time Fourier Transform

A Preprint

Yiming Song
Dulwich College Beijing
Beijing, China 101300
yiming.song22@stu.dulwich.org

Abstract

The timbre of melodic instruments is partially determined by the strength of various harmonic overtones relative to the base frequency. No prior work has produced a standard reference for the weightings associated with concert instruments. Using audio samples from the University of Rochester Multi-Modal Music Performance dataset associated with known instruments, we use the short-time Fourier transform (STFT) to compute the relative strengths of each harmonic for each instrument (eight orchestral instruments). We use mean squared error to evaluate the distinguishability of different instruments. We show that we can distinguish between instruments of different families such as trombone and oboe, though instruments belonging to the same family, such as violin and cello, are harmonically very similar.

Keywords Harmonic profiles · Timbre · Short-time Fourier transform

1 Introduction

Musical timbre, or tone, refers to the perceived quality of a sound. It allows listeners to distinguish between instruments, even when playing the same note with the same fundamental frequency. Physical characteristics of the sound, including the strengths of frequencies emitted, as well as the envelope (how the note changes over time), determine an instrument's timbre (Patterson [2010]). This paper focuses on the former and aims to provide a reference for the relative strengths of each harmonic of an instrument, with harmonics referring to any integer multiples of the fundamental frequency that resonate together with the fundamental. For this paper, the strengths of each harmonic, relative to the base frequency, will make up what is known as the harmonic profile of an instrument. The results may be useful for sound synthesis, audio transcription, and instrument detection purposes.

Because the harmonic profile of a instrument depends on how it is being played (eg. a violin playing *sul ponticello* versus *sul tasto*) and may also change over time, it is a difficult problem to create a reliable and standardized set of profiles for instruments. Literature is sparse on the topic: Petersen compared harmonic strengths for a few instruments (flute, oboe, violin) playing a single pitch and Livshin [2007] compared the first 20 harmonics of a violin, flute, and trumpet using the discrete Fourier transform.

2 Background

2.1 Fourier Transform

The Fourier transform (FT) is a mathematical transform that can map an input function of time into a complex-valued function of frequency, whose magnitude represents the strength of the frequency present in the original function and argument represents the phase of the frequency. It is useful when decomposing an audio signal into its constituent frequencies. The continuous FT (CFT), $\hat{f}$, of an continuous integrable function $f : \mathbb{R} \mapsto \mathbb{C}$ is given by:

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(t) \cdot e^{-2\pi i t \xi}\, dt \tag{1}$$

where t represents time, and the variable ξ represents frequency (Weisstein). With audio signals however, the input signal is usually temporally discrete, a set of equally spaced (based on the sample rate of the audio signal) data points. In this case, the discrete Fourier Transform (DFT) is more practical. The DFT is a function which takes an input sequence of N real numbers $x \coloneqq x_0, x_1, ..., x_{N-1}$ and maps it onto another sequence of complex N numbers, $X \coloneqq X_0, X_1, ..., X_{N-1}$, where each element in the new set is given by:

$$[h]X_k = \sum_{n=0}^{N-1} x_n \cdot e^{\frac{-2\pi i}{N} kn} \tag{2}$$

In this paper, we use a specific type of DFT, known as the short-time Fourier transform (STFT) to conduct audio analysis. It is created by performing multiple DFTs over short overlapping windows, serving as a way to extract frequency and amplitude information out from the audio signal in short chunks of time, known as the window length of the STFT (Smith [accessed 15/11/21]). Plotting the frequencies and amplitudes over time, one can create a spectrogram, which can be used to visualize an audio signal, as shown below, where time is on the x-axis, frequency on the y-axis, and strengths of frequencies given by the colors. The vertical bands going down the spectrogram are the individual time windows according to which the DFTs are computed, while the horizontal columns give the frequency bands.

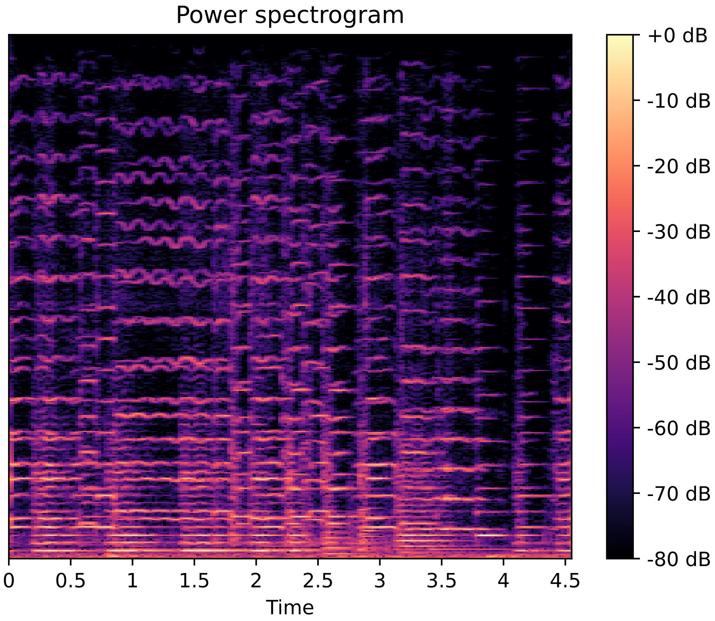


Figure 1: Example spectrogram.

A well known problem with the STFT is the Gabor limit, which describes the STFT's fixed frequency-time resolution, where increasing the window length will yield better frequency resolution, but a narrow window length will yield better time resolution at the expense of frequency resolution. This is because signals do not contain precise time and frequency information; a way to work around this limit will be described in the method.

For this paper, *librosa* (McFee et al. [2015]), a Python package for music and audio analysis, was used to compute the STFTs.

2.2 Harmonics

Most instruments emit harmonics, oscillations (overtones) with frequencies that are integer multiples of and occur at the same time as the fundamental frequency. Each instrument has a different harmonic profile, emitting different strengths

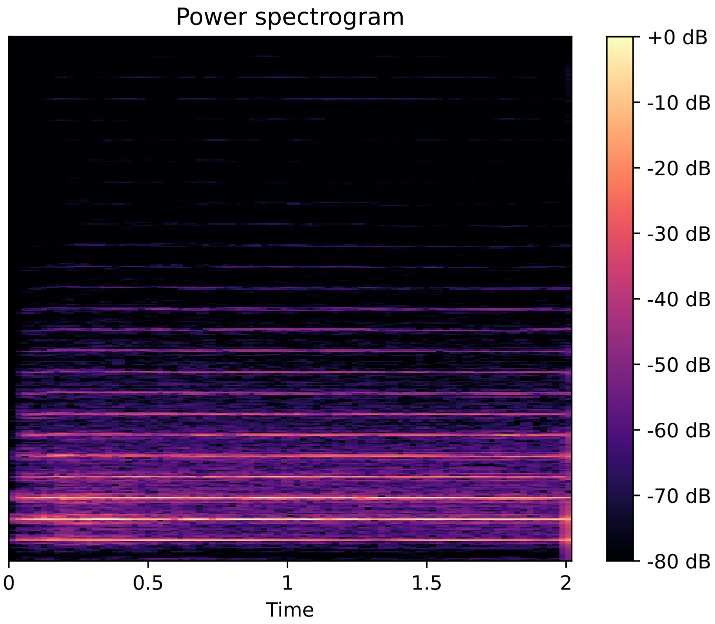


Figure 2: Spectrogram visualization of a trumpet playing A=440Hz.

of each harmonic. By taking in the different relative levels of the harmonics, one is able to perceive different tones, or timbres, of instruments. Seen in a spectrogram, the harmonics are represented by the decreasingly loud, equally spaced (on a log-scale frequency axis) bands that have similar onset/offset times:

Harmonic analysis involves recognizing the different harmonic profiles of these instruments. Obtaining harmonic profiles is much more difficult for percussive instruments (drums, cymbals) and the voice. This is because drums are not harmonic instruments; not only is it difficult to determine a “fundamental frequency” of a drum sound, but their overtones are also not integer multiples of any supposed fundamental, making it difficult to create definitive harmonic profiles. Similarly, with voices, different syllables and vowels that are sung can lead to boosts or dips of certain harmonics, making it difficult to create a definitive harmonic profile. Consonants will also affect the frequency distribution, with fricatives like [s] and [f] producing high frequency noise, while plosives like [p] and [b] contain lower-frequency information (Fry [1979]). Thus, considering these variations, this paper will focus on the harmonic profiles of melodic instruments belonging to the string, brass, and woodwind families.

2.3 Dataset

In this paper, the University of Rochester Multi-Modal Musical Performance (URMP) dataset was used [Li et al., 2018]. It contains separately recorded stereo audio tracks and includes stems from aforementioned instruments. It also includes ground-truths identifying the starting times, frequencies, and duration of each individual note played by each instrument, which allows easy extraction of audio samples of individual notes by different instruments. The recordings were independently created by musicians in different recording rooms, meaning that even with the same instrument there should be some degree of variation in recordings. In total, the harmonic profiles of eight instruments were analysed (violin, cello, trumpet, trombone, clarinet, flute, oboe, alto saxophone).

3 Method

3.1 Preliminary Data Processing

For each individual audio stem in the dataset, the STFT is computed, with a sample rate of $22050Hz$ and windowed signal length (given by the parameter 'n_fft' in *librosa*) of 2^{12}. This results in $1 + n_{fft}/2 = 1025$ frequency bands of the STFT, splitting the frequency range into bands spaced $11025/1025 \simeq 10.75Hz$ apart. This value of n_fft allows the lower-frequency notes still to be distinguished while minimizing the program's runtime. This STFT outputs an array containing the complex-valued Fourier coefficients of the audio signal, with time on the first axis and frequency on the second axis. Taking the absolute value of this array disregards the phase information in the audio and preserves the magnitude (strength) information. For example, to obtain the magnitude of an audio signal X at time n and frequency band m:

```
signal, sr = lbsa.load(X, sr=22050, mono=True)
stft = lbsa.stft(y, n_fft=2^12)
strength = np.abs(stft)[n][m]
```

Before computing all STFTs, a high-pass filter starting at $75Hz$ was applied onto all audio signals to reduce low frequency noise and reduce error in the weightings from the algorithm used below. Samples were also all converted into mono. Then, each STFT is segmented time-wise according to the start/end times of notes in the sample, provided by the datasets. Then, according to the base frequency, a list of nine overtone frequencies for that specific note are generated.

3.2 Frequency Snapping

Before all the strengths of the harmonics are extracted from the STFTs, one important consideration must be made. Instrument pitch refers to how high or low a sound is perceived, and is determined by the frequency of a sound wave, though not linearly. Western musical tuning conventions split a an interval known as an octave into twelve notes (these divisions are known as semitones). At lower pitches (eg. $A_2 = 110Hz$ and $A\#_2 = 116.5Hz$), consecutive notes are closer spaced frequency-wise compared to higher pitches (eg. $A_4 = 440Hz$ and $A\#_4 = 466.16Hz$). This is because to obtain a pitch that is one octave higher than an initial pitch, one must double the frequency; therefore, consecutive octaves span larger ranges on the frequency spaces. Thus, it is useful to take a log-scale frequency when looking at spectrograms, as it would make the harmonics of notes equally spaced rather, as seen previously in fig. 2. A dilemma arises because the frequency bands of an STFT are spaced linearly, not in log-space. Furthermore, because, instruments often play with vibrato as well as glissando, the frequency of one note does not always remain constant, and may oscillate. Consequently, the harmonics will also oscillate, and since their frequencies are multiples of the base frequency, a small oscillation in the base note could cause larger oscillations in the harmonic that span multiple STFT bands.

To combat all of these problems, the following methodology was used. For each STFT band, we compute the mean squared distance between the log of the lowest frequency of the band and log the harmonic, and assign each band to a harmonic. This method accounts for the logarithmic scale of pitch as well as vibrato and glissando, since adjacent bands are most likely to be assigned to the same harmonic, so notes that are in multiple bands will be grouped together. After the bands are assigned, we obtain a frequency spectrum that is evenly partitioned into n sections that correspond to n harmonics.

3.3 Computing and evaluating harmonic profiles

With the bands partitioned, the magnitudes of all the STFT bands of one harmonic are summed up to obtain the absolute weight for that specific harmonic. This process is repeated for all harmonics, after which the weights are normalized by dividing each weight by the sum of all weights. Then, we take the average weights computed for all notes of one instrument to obtain the final harmonic profile. The standard deviations are also taken.

To compare harmonic profiles of instruments and evaluate whether there exists distinctive qualities between the harmonics of different instruments, we compute the pair-wise Euclidean distances between all given weight values for individual harmonics to obtain an intra-instrument measurement, then compute pair-wise distances between weight values of two instruments for individual harmonics to obtain an inter-instrument measurement. Computation is performed with the *sklearn* "mse" package (Pedregosa et al. [2011]). These values can then be compared; if inter-instrument distances exceed the intra-instrument distances, then there should be a detectable difference between the weights of one instrument and the other.

4 Results

4.1 Profiles

The histogram plots for the normalized mean weights of are shown below, as well as the number of notes that were analyzed. For most instruments, the weight of the harmonic 10 is inaccurate, with the snapping algorithm taking the sum of all weights of the frequencies above it. The standard deviations are shown by the black bars.

As can be seen, the standard deviations are extremely large, suggesting that the average harmonic profile of an instrument does not remain constant across different pitches or may shift over time. However, the mean weights still show a

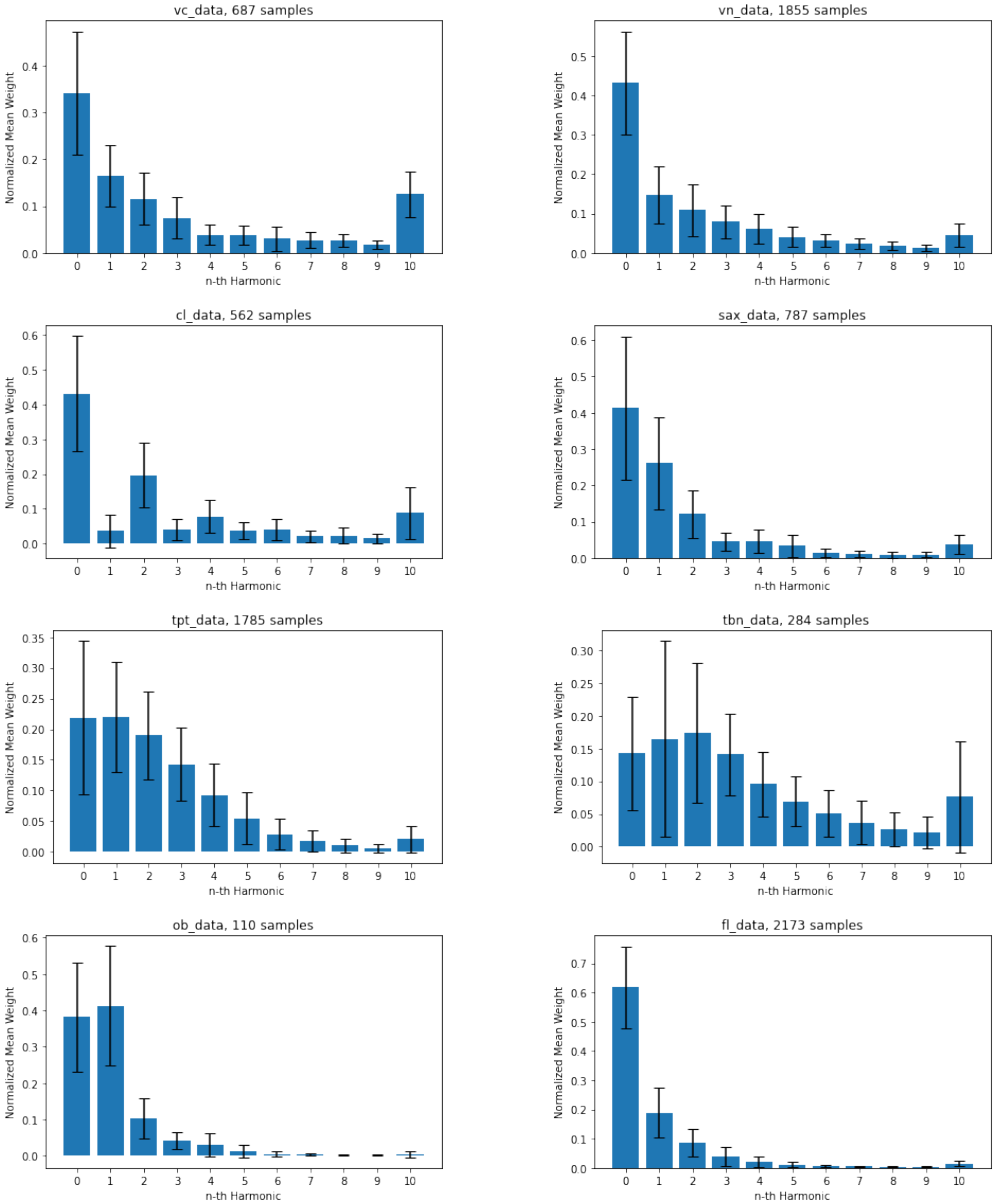


Figure 3: Obtained harmonic profiles

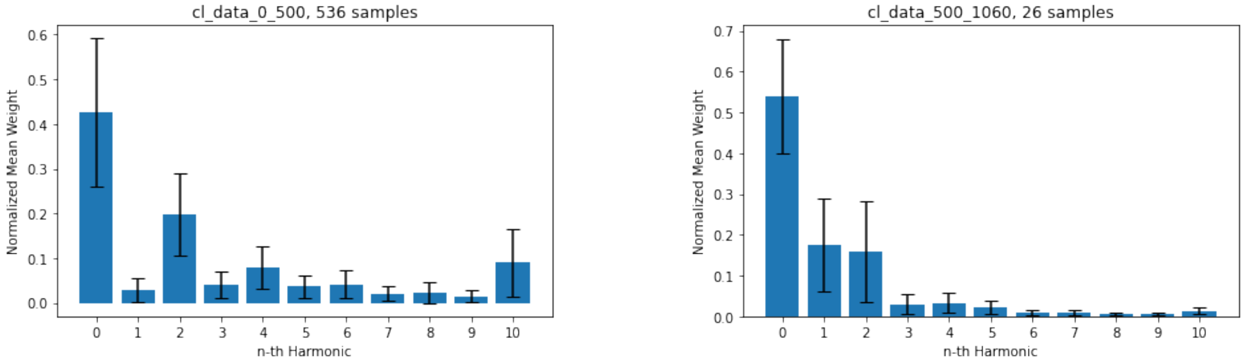


Figure 4: Differences in harmonic profiles between clarinet registers

significant difference between families of instruments. Brass instruments are fuller harmonically while the oboe and flute have weak upper harmonics. The vioiln and cello profiles are very similar. The clarinet, notably, has an extremely weak second harmonic.

There are a few reasons for the large standard deviations. When brass instruments play in different registers, the sounds become noticeably more 'hoarse' as the air flow increases in speed. The clarinet has three main registers: the lowest *chalumeau* register, which has a dark tone, the *clarion* register, which has a brighter and sweeter tone, and then upper *altissimo* register, which is shrill (Lowry [1985]). Each of these registers would have different harmonic profiles.

Given the large standard deviations, we recompute the profiles on certain frequency ranges for different instruments.

For clarinet, we check two registers, the *chalumeau* register, which approximately ranges from 0 to $500Hz$ and the *clarion* register, which approximately ranges from 500 to $1060Hz$. The dataset contains no samples of notes in the third register. The number of notes used that the profiles were extracted from is also shown:

For the

4.2 Distances

The histograms for the distances of each instrument are shown below, with the mean shown as a dotted line:

We provide a few examples of inter-instrument distances for instruments belonging to the same/different families. The blue histogram and blue line correspond to the intra-instrument distances of the first instrument in the title, while the orange histogram red line corresponds to the inter-instrument distances:

The large overlap between the distances for the violin and cello suggest that the main discriminating factor between the two instruments is simply the pitch they are being played at. Surprisingly, the intra-instrument distances between oboe and saxophone are overlapping; while the mean harmonic profiles are different, the amount of variation among the saxophone's notes is comparable to the amount of variation between the notes of the saxophone and oboe.

5 Conclusion

In this paper, we present harmonic profiles for eight orchestral instruments. We show that it is possible to use these profiles to distinguish between instruments of differing families (eg. oboe and trombone) while instruments belonging to the same family such as violin and cello are, expectantly, harmonically similar and thus harder to tell apart by analyzing their harmonics. More metrics other than MSE could be used to compare the profiles of different instruments. The scope of this work is limited as we only analyze up to nine harmonics with a limited dataset containing only eight instruments and 17 pieces in total. With more datasets containing a more diverse range of musical material, a wider range of frequencies could be covered by the instruments, allowing for further analysis of how profiles shift at different instrument registers. Future directions include applying these harmonic profiles to aid in instrument source separation (such work could also consider other sonic properties of an instrument in order to achieve accurate separation, especially in cases of harmonically similar instruments).

These results can also be applied to instrument synthesis; for example, using the Ableton operator synth, relative strengths of harmonics can be inputted to create a digital instrument. However, more factors, such as differing profiles over different registers, how the harmonic profile may change over time, as well as transient features (attack strength,

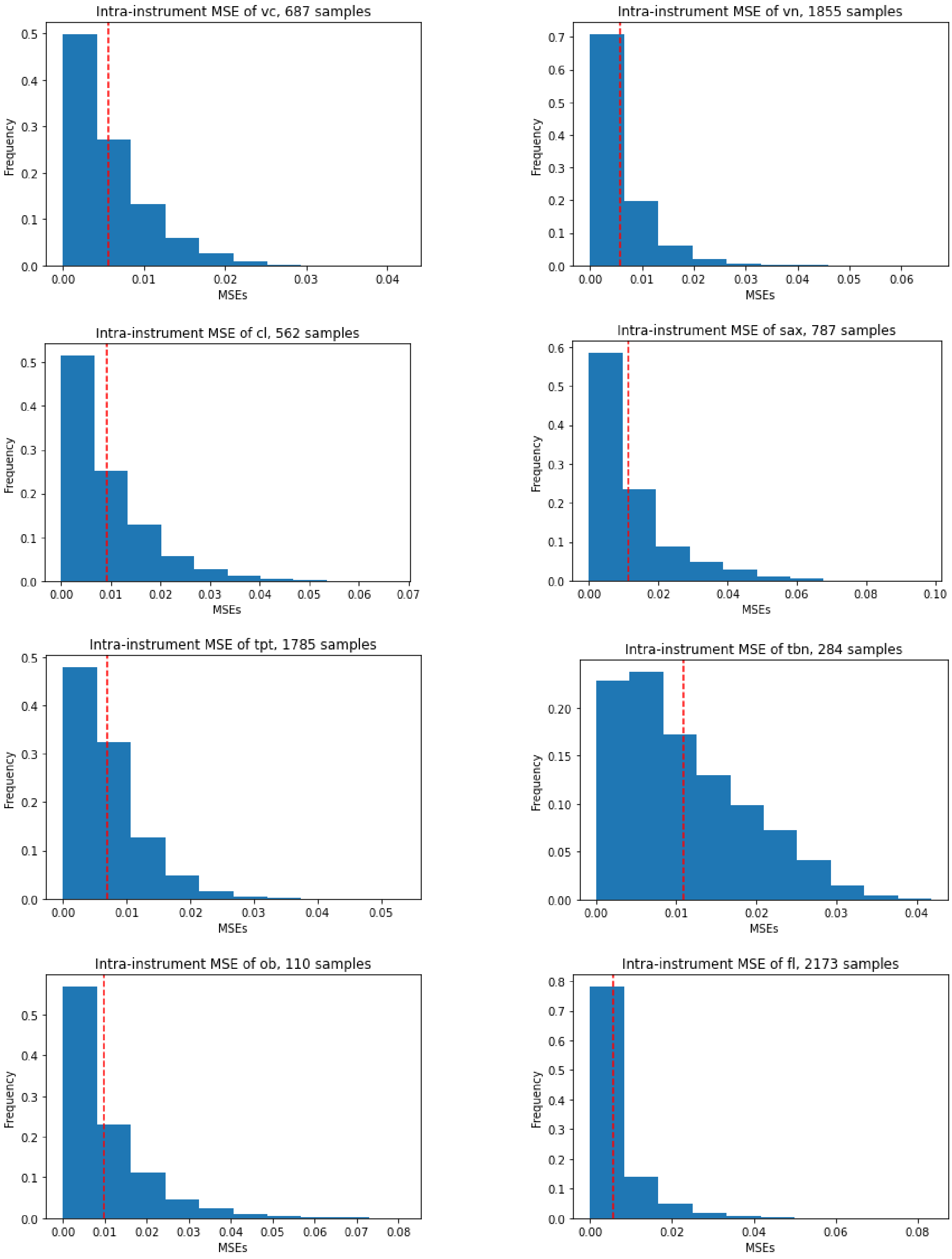
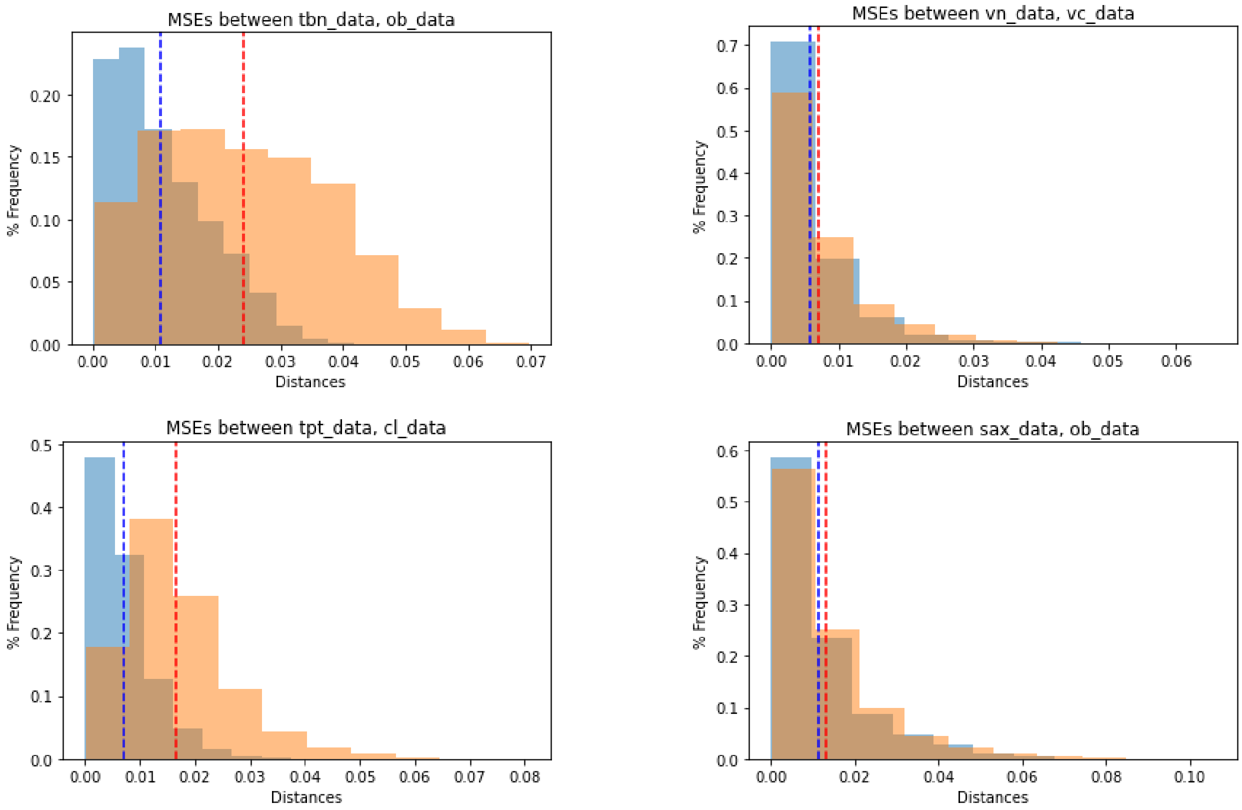


Figure 5: Euclidean distances between samples of each instrument



sustain, etc.) must also be considered to create realistic instrument syntheses. Unique profiles can also be generated (by maximizing intra-instrument MSE when comparing to existing instruments) to create entirely new instruments, enabling new forms of creative musical expression.

6 Bibliography

References

Gaudrain E. Walters T. C. Patterson, R. D. The perception of family and register in musical tones. In *Music Perception*, pages 13–50. Springer Science + Business Media, 2010.

Mark Petersen. Mathematical harmonies. URL `https://amath.colorado.edu/pub/matlab/music/MathMusic.pdf`. Visited on 07/11/21.

Arie Livshin. Automatic musical instrument recognition and related topics. 12 2007.

Eric W. Weisstein. Fourier transform. URL `https://mathworld.wolfram.com/FourierTransform.html`. Visited on 15/11/21.

Julius O. Smith. *Spectral Audio Signal Processing*. `http://ccrma.stanford.edu/~jos/sasp/`, accessed 15/11/21. online book, 2011 edition.

Brian McFee, Colin Raffel, Dawen Liang, Daniel PW Ellis, Matt McVicar, Eric Battenberg, and Oriol Nieto. librosa: Audio and music signal analysis in python. In *Proceedings of the 14th python in science conference*, volume 8, 2015.

D. B. Fry. *The Physics of Speech*. Cambridge Textbooks in Linguistics. Cambridge University Press, 1979. doi:10.1017/CBO9781139165747.

Bochen Li, Xinzhao Liu, Karthik Dinesh, Zhiyao Duan, and Gaurav Sharma. Creating a multitrack classical music performance dataset for multimodal music analysis: Challenges, insights, and applications. *IEEE Transactions on Multimedia*, 21(2):522–535, 2018.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Robert Lowry. *Practical hints on playing the b-flat clarinet*. Warner Bros Pubns, Alfred Music, 1985. ISBN 978-0-7692-2409-1.