Applications of FFT for timbral characterization in woodwind instruments

Yubiry Gonzalez[1], Ronaldo C. Prati[1]

[1]Center of Mathematics, Computer Science and Cognition - CMCC

Federal University of ABC – UFABC. Av. Dos Estados, 5001, Bangú Santo André, 09210-580, SP, Brazil

Yubiry.gonzalez.17@gmail.com, ronaldo.prati@ufabc.edu.br

Abstract

The conceptualization of the musical timbre, which allows its quantitative evaluation in an audio record, is still an open-ended issue. This paper presents a set of dimensionless descriptors to assess the musical timbre of woodwind instruments in recordings of the fourth octave of the tempered musical scale. These descriptors are calculated from the Fast Fourier Transform (FFT) spectra using the Python Programming Language, specifically the SciPy library. The characteristic spectral signature of the clarinet, bassoon, transverse flute, and oboe are obtained in the fourth musical octave, observing the presence of degeneration for some musical sounds, that is, two given different aerophones may present the same harmonics. It is concluded that the proposed descriptors are sufficient to differentiate the aerophones studied, allowing their recognition, even in the case that there present the same set of harmonic frequencies.

1. Introduction

To characterize sound in general, and musical sound in particular, it is necessary to know the attributes of pitch, intensity, duration, and timbre. The first three characteristics of sound correspond respectively to the acoustic magnitudes of frequency, intensity and time, and they are direct, measurable magnitudes. Timbre, however, is a multi-dimensional and briefly defined attribute that makes it possible to distinguish between different sounds even when they have the same intensity, duration, and pitch. It allows discriminating sounds of different musical instruments even when it is the same musical note, with the same duration and intensity.

The problem of the univocal characterization of musical timbre raises the need to develop descriptors (derived magnitudes, coefficients or functional) that evaluate timbre from digital audio records [1]. Since musical timbre is a phenomenon of auditory perception, many of the investigations are developed in the line of psychoacoustics to evaluate verbal descriptors that reveal measurable attributes of musical timbre [2,3].

Although the psychoacoustic perception of musical timbre cannot be ignored, it must be recognized that the main timbral characteristics must be, in some way, inscribed within its spectral content [4]. Assume that an FFT calculated from a musical audio record does not contain some significant timbral characteristics. In that case, the deconvolved audio (reverse convolution) could not be distinguished timbrically. However, as we can differentiate between musical instruments after the deconvolution, the initial assumption is not valid.

Therefore, the FFT contains all the essential tonal characteristics. Other research focuses on the presentation of an exhaustive collection of acoustic descriptors in the form of coefficients or functional (Timbre ToolBox, Librosa) that can be computationally extracted from the statistical analysis in the digitization of the spectrum (FFT) and the digital signal itself (spectrogram). These coefficients usually refer to the statistical and mathematical characterization of the maxima in the FFT and the spectrograms, such as the mean value in frequency (centroid) and amplitude, standard deviation, kurtosis, roots or poles of the distribution, arithmetic sequences and geometric in frequency, mean and mean square values of the amplitudes among others [5-9].

There is no consensus on which and how many are the acoustic descriptors' for musical timbre. However, many of them are derivatives or combinations of others and thus highly correlated [5]. It is worth asking which is the minimum set of timbre descriptors that differentiate the musical instruments in a given audio recording.

Despite recent advances in verbal descriptors (psychoacoustic) such as statistical or mathematical descriptors (Timbre Toolbox [5], Librosa [9,10]) of musical timbre, it is still far from being characterized, and it constitutes a challenge for the automated identification of musical instruments.

An alternative approach is to construct the timbral descriptors from that of musical acoustics, using an analogy to the analysis and digital processing of spectra in other areas of knowledge. The FFT represents the two-dimensional plane of the frequencies and intensities (alternatively energy-frequency) present in a signal.

This paper aims to propose a set of acoustically motivated timbral descriptors (mathematical functional or oversized coefficients) that allow the computational extraction of timbral information, for the automated identification of woodwind musical instruments. For this, a preliminary study of the timbral characteristics in digital musical sounds is presented; using the FFT in the 4th tempered musical Octave of a sample of woodwind instruments: Clarinet, Bassoon, Transverse Flute and Oboe.

In Section 2, the methodology for obtaining the spectrum and their quantification of the Audio records are presented. In Section 3, the coefficients or descriptors and their valuation for the selected aerophones are shown. In Section 4, the chromatogram for the 4th octave of the tempered musical scale is evaluated, and the spectral signature (“fingerprint”) that characterizes each studied aerophone is obtained. Finally, the conclusions and future work are presented in the last section.

2. Methodology

The audio records of the aerophones were obtained from the TinySOL open-source sound library [11], which contains recordings of individual sounds, played in an ordinary way, at a level of dynamic/intensity mezzo forte in a WAV audio format that minimizes information losses, sampled at 44.1 kHz in single-channel (mono) at 16-bit depth. From this library, we have restricted our analysis to only some instruments of the woodwind family of a typical western symphonic music orchestra: Clarinet, Bassoon, Transverse Flute and Oboe.

On these audios, the module for calculating the FFT available in the SciPy library in Python [12] was used to obtain the frequency spectrum; then for the identification of the local maxima, the find_peaks function in Scipy.signal was used.

Next, the tables of the maximum frequencies, expressed in Hz, with relative amplitudes, normalized considering the maximum amplitude value achieved in each spectrum of the FFT, were obtained in order to compare the various audio spectra with different power levels (Watt) or relative power (dB).

The calculation of the coefficients was performed on the frequency spectrum of the fourth octave of the tempered musical scale (~261 Hz to ~493 Hz) corresponding to musical sounds between C4 to B4 in musical notation. This octave was initially considered as it is the lowest common octave in the tessitura of the selected aerophones; therefore, it is the one that would present the most significant amount of harmonics for each instrument and musical sound.

3. Timbral Coefficients

Musical frequencies make up a set of only 12 different values in the 4th musical octave. Therefore, in a first approximation, musical timbre can be characterized by a limited set of dimensionless quantities related to frequencies and amplitudes in the spectrum. They are related to the fundamental frequency and the partials that arise from the FFT. In the timbre analysis, the centroid is usually used; its use does not allow a specific acoustic interpretation since the frequency does not correspond to any of the parcials frequencies of the sounds studied. The centroid is defined as the average frequency. Where f_i is the harmonic frequencies for each spectrum.

$$\bar{f} = \frac{\sum_{1}^{N} a_i f_i}{\sum_{1}^{N} a_i} \qquad (1)$$

In Figure 1, the FFT spectra of the musical sound E4 are shown (329.6 Hz). Note that the centroids do not correspond to any audible signal or related to the musical note E4.

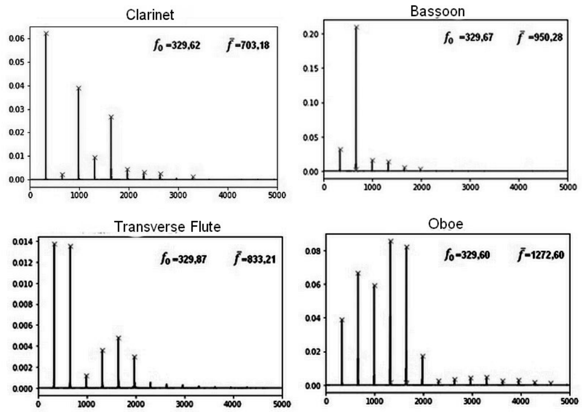

Figure 1: FFT spectra of the musical sound E4 for the studied aerophones, nominal frequency 329.6 Hz.

For all aerophones studied, Figure 2 shows that, in general, the centroid presents variations of less than 15% with respect to its mean value of the 4th octave. Also, Figure 2 shows that, in some sounds, the centroid values are similar for several instruments. This similarity does not allow their differentiation w.r.t. the type of instrument.

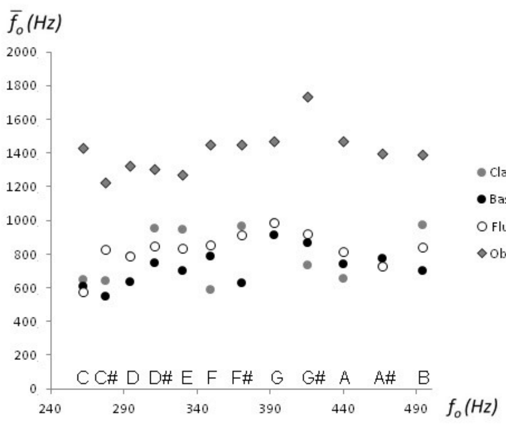

Figure 2: Centroid in 4th Octave of the musical scale

The fundamental frequency f_0 would only coincide with the Centroid $\overline{f}$ if there are no other partial frequencies. This fact is impossible in real musical instruments due to the harmonics, the superposition and the beats inside the resonator tubes of the aerophones. The separation between the fundamental frequency and the mean value of the frequencies would be important in musical acoustics. The Affinity coefficient describes how far the spectrum is from the ideal case, that is, how far the maximum Principal f_0 is from the mean value in frequency or Centroid ($\overline{f}$)

$$A \equiv \frac{\overline{f}}{f_0} \qquad (2)$$

The affinity coefficient allows discriminating different sounds with the same Centroid (figure 3). Figure 4 shows the values of *A* for the 4th octave.

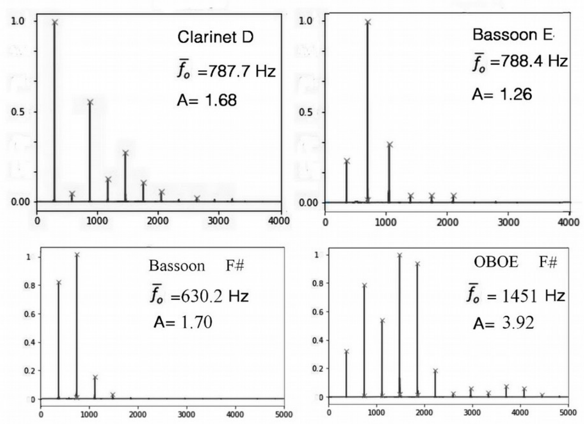

Figure 3: Fourier spectrum with different coefficients. Top: Examples with very similar centroids. Botton: Examples with Equal musical sounds.

It can be observed in Figure 1, that different musical instruments have different amplitudes in their fundamental frequency. This motivates another descriptor: the Sharpness coefficient (*S*).

$$S \equiv \frac{a_0}{\sum_{1}^{N} a_i} \qquad (3)$$

Ideally, *S*=1 in a "pure" sound that would have a single maximum without secondary frequencies. By construction, $S \leq 1$. Figure 5 exemplifies Sharpness in different musical instruments and for the same musical instrument with different sounds.

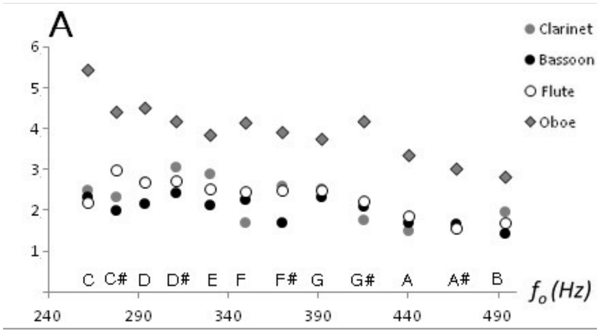

Figure 4: The values of affinity (A) for the 4th octave

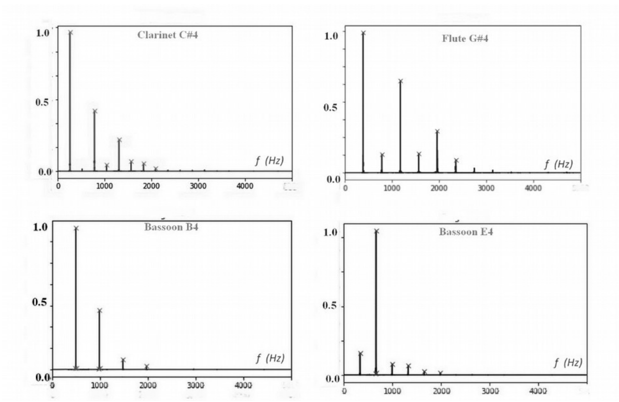

Figure 5: Timbral coefficient *S*. Above: The FFTs of the Clarinet in C# (277.2 Hz) and of the Transverse Flute in G# (415.3 Hz) both with S = 0.55. Bottom: Bassoon FFT in B4 and E4, S = 0.98 and S = 0.11 respectively

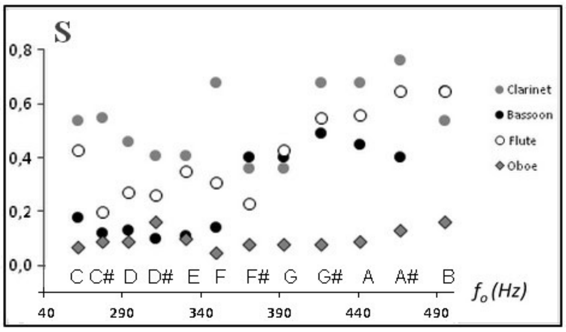

Figure 6. Sharpness (S) in the 4th octave.

Figure 6 shows the variation of the Sharpness of the aerophones in the 4th Octave. It is observed that, in general (except for F#), the clarinet is sharper than the transverse flute, and is sharper than the bassoon. In any case, the Oboe is the least shaper of all. In other words, musical sounds have better sharpness on the clarinet and worse sharpness on the Oboe.

4. Spectral Signature

In general, the identification of musical instruments and of the wave emitting source can be made through the analysis of their spectrum. In the case of electromagnetic radiation, the specific profile of intensities versus wavelength (absorption or emission) allows the identification of the emitter (is it a gas, a reflective surface or an absorbing medium) and is called Spectral Signature. By analogy, we understand as "Spectral signature of a musical instrument", the distribution of frequencies in the Fourier Transform Spectrum of musical sounds (tempered scale) that allow their identification.

Suppose we restrict ourselves to the aerophones studied, which are only a small sample of all common musical instruments. In that case, a discriminatory identification between them can be made from the spectra of the presented FFT spectra for the 4th octave records. (Figure 7).

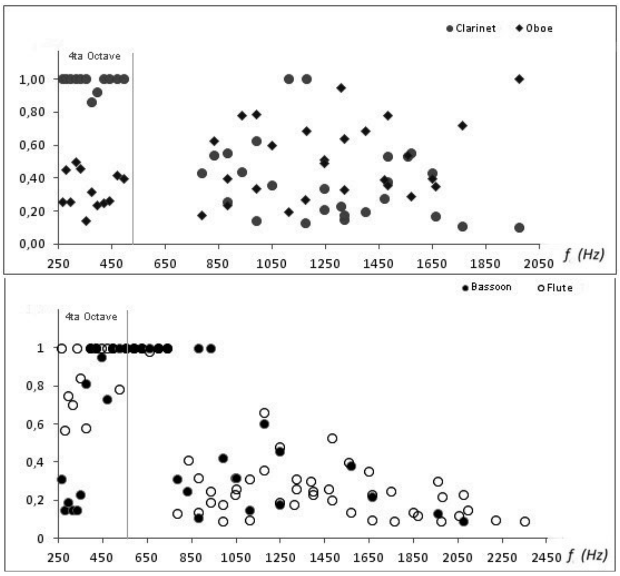

Figure 7. Spectral Signatures in the 4th Octave for Aerophones: Top: Oboe and Clarinet. Bottom: Bassoon and Transverse Flute.

If all the harmonics of a certain musical sound are present, this would allow, to discriminate the sounds between each instrument (Table 1). The harmonic frequencies () are related to the fundamental frequency through the integer multiplicity (n = 2,3, 4,…):

$$f_n = n f_0 \qquad (4)$$

This discrimination, regarding which harmonics are present or absent, is not univocal. It presents degeneration, in the sense that for some musical sounds, two different aerophones present the same harmonics (highlighted with the symbol * in table 1).

Musical sound and frequency (Hz)	Characteristic Harmonic Frequencies			
	Clarinet	Bassoon	Flute	Oboe
C4 : 261,6	{3,…,8}	{2,…,8}	{2,…,5,}	{2, …,9}
C4#: 277,2	{3,…,7}	{2,…,6}	{2,…,8}	{2, …,9}
D4 : 293,7	{2,…,7,9}	{2,…,7}*	{2,…,7}*	{2,…,8, 10,…,15}
D4#: 311,1	{2,…,10}	{2,…,7}*	{2,…,7}*	{2, …,12}
E4 : 329,6	{2,…,9}	{2,…,6}*	{2,…,6}*	{2, …,13}
F4 : 329,67	{2,…,5}	{2,…,6}*	{2,…,6}*	{2, …,13}
F4#: 370	{3,…,7}	{2,…,4}	{2,…,5}	{2,…,12}
G4 : 494	{2,…,5}	{2,…,5,8}	{2,…,6}	{2,…,12}
G4#: 415,3	{2,…,8}	{2,…,6}	{3,…,6}	{2,…,13}
A4 : 440	{2,…,5}	{2,…,7}	{2,…,6}	{2,…,11}
A4#: 466,2	{2,…,8}	{2,…,6}	{2,…,5}	{2,…,9}
B4 : 494	{2,…,6}*	{2,…,5}	{2,…,6}*	{2,…,9}

Table 1: Harmonic set as a function the sound and musical instrument of the fourth octave.

(*) Denotes degeneracy: two different instruments with the same set of harmonics.

However, as the FFTs are not the same between any two musical instruments, the different relative intensities, through the Affinity and Sharpness coefficients, would allow completing of the spectral signature and the univocal identification of the instrument (Figure 8).

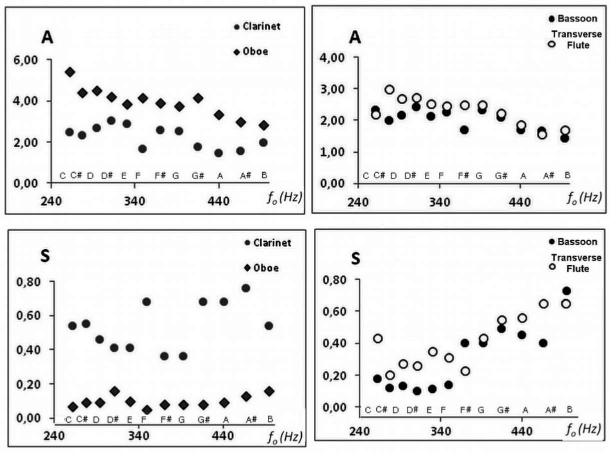

Figure 8. Timbral Affinity and Sharpness

The variation of these coefficients for musical sounds in the 4th octave, allows the spectral signature to be uniquely identified without degeneration.

On the other hand, in a given spectrum, there are always partial or secondary frequencies to f_0, that can be counted. However, it can also happen that a secondary frequency is not strictly an integer multiple of the fundamental frequency but whose frequency is very close to an integer multiple of f_0.

In a way, the sound would be more "harmonic" than if its frequency value were very different from the multiplicity of the fundamental frequency. We have defined the coefficients from acoustics' perspective where an non-oscillating oscillator in harmonic motion is known as an anharmonic oscillator. The difference is more than mere semantics.

This was made by design as Inharmonicic coefficient defined by Peters [14] quantifies only the frequencies that are multiples of the fundamental (integer k values). However, in our definition of the Harmonicity coefficient, all frequencies of the Fourier spectrum are quantified. To describe this property, the timbral coefficient of Harmonicity (H) is proposed and defined as:

$$H \equiv \sum_{1}^{N} \left(\frac{f_j}{f_0} - \left[\frac{f_j}{f_0} \right] \right) \qquad (5)$$

Where the [ ] denotes the integer part.

This timbral coefficient evaluates how harmonic the partial or secondary frequencies ($f_1, f_2, f_3 ... f_j$) in the FFT spectrum. The idea is that any frequency f is a harmonic of f_j if the quotient between them is an integer. Figure 9 shows this variation in examples of FFT spectra with different harmonicity.

Our definition is very different from the inharmonicity descriptors [14,15]. First of all, we do not weigh the amplitudes, for the purposes of multiplicity in frequency. Secondly, note that if f_k is not an integer multiple of f_0, it will not appear in the summation nor will it be considered in Bullock's expression [15], but it will influence the harmony of the set and, therefore, the expression shown in Equation (5).

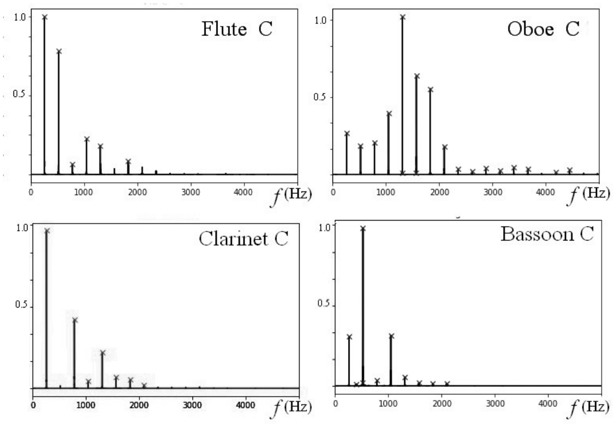

Figure 9. Comparison of partial or secondary frequencies in the sounds C, between the Transverse Flute (*H*= 0), Oboe (*H* = 16.9), Clarinet (*H*= 2.00) and Bassoon (*H*= 7.48).

Figure 9 shows a comparison of secondary frequencies in the sound C4 for the instruments we are analysing. Analysing the figure, we observe that the oboe is more inharmonic than the flute, with harmonicity coefficients of H = 0 and H = 16.9, respectively. Comparing it with the coefficient proposed by Peters [14], the same value is obtained for both instruments (Zero) and, as a consequence, it does not describe the harmonicity that we define.

The C4 sound of the Flute is much more harmonic than the Oboe, even though the latter has many more secondary frequencies, some of which can be harmonics of f_0, as is the case of the second, third and fourth maximum. Furthermore, if the secondary frequencies are all harmonics of f_0 then H=0. Every time there is one or more frequencies that are not harmonics of f_0, the j-th term, in the sum, will be non-zero, and H increases. The maximums of the spectrum distribution of the Flute in Figure 9 are all integer multiples of the fundamental and therefore it is highly harmonic (H = 0).

For the fourth octave, the results of the harmonicity assessment are shown in figure 10. The Transverse Flute has the highest harmonicity (H near zero), that is, its secondary frequencies are usually integer multiples of the corresponding fundamental sound, while the Oboe is the most anharmonic (large H). We note that the quality of harmonicity varies with the musical note and instrument, so for the characteristic sound of reference A= 440 Hz the Bassoon and Clarinet are completely harmonic, and the Transverse Flute is completely hamonic also for C, D, E, and A#.

The Harmonicity coefficient (H) allows, through the FFT, to quantify the number of harmonics present for a given fundamental frequency. Therefore, in a first approximation, we can affirm that it is related to "color" or "sonority" of these instruments [13].

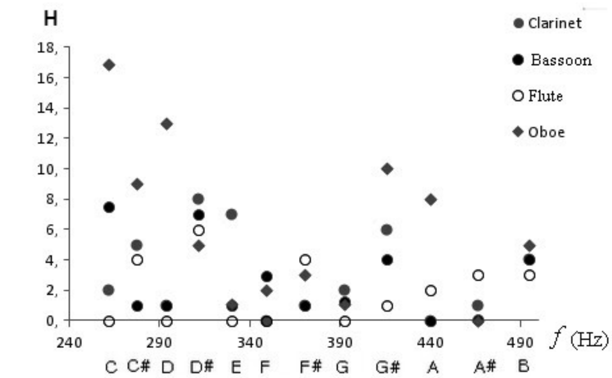

Figure 10. Harmonicity (H) in the 4th Octave.

6. Conclusions

The Fast Fourier Transform (FFT) allows analyzing digital audio records on the tempered musical scale, generating a unique spectrum that characterizes it. As a finite, bounded and countable collection of discrete frequencies with different amplitudes, this spectrum provides two-dimensional arrangements of frequency and amplitudes for the musical sounds of aerophones. Moreover, they can be analyzed using the usual automated signal analysis techniques. Therefore, the characteristics of the audio records, including the timbre, are susceptible to being automated by artificial intelligence techniques.

It is possible to calculate oversized coefficients that reflect timbral characteristics of the fundamental frequency and harmonic distribution in the FFTs of simple 4th-octave monophonic music registers for woodwind instruments. Thus, the proposed timbral descriptors, Affinity and Sharpness, allow distinguishing the timbre of Clarinet, Bassoon, Transverse Flute, and Oboe.

The Spectral Signature of the wooden aerophones, which allows the univocal identification of them, can be obtained through the FFT by the distribution of harmonics present outside the loudness range of the 4th octave. The number of harmonics present and their succession does not always characterize the timbre between different aerophones. There may be degeneration (two or more instruments with equal harmonics). However, the proposed coefficients allow a unique timbral identification in each musical sound studied.

We conclude that the proposed descriptors allow describing timbral characteristics of the instruments studied, in addition to differentiating them, allowing their recognition from the FFTs.

Future work will implement the FFT techniques and the described timbric coefficients until completing the tessitura of the selected aerophones. Later the study will be extended to incorporate the temporal variations of the timbre coefficients in the case of melodic fragments that are important for quantising the timbre in the harmony and the musical composition.

References

[1] M. Adeli, , J. Rouat, , S. Wood, , S. Molotchnikoff, , and E. Plourde, A flexible bio-inspired hierarchical model for analyzing musical timbre. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 24(5): 875-889. 2016.

[2] W. Jiang,, J. Liu, X. Zhang, S. Wang and Y. Jiang, Analysis and Modeling of Timbre Perception Features in Musical Sounds. Applied Sciences, 10 (3): 789. 2020.

[3] D. Fourer, Jean-Luc Rouas, P. Hanna, M. Robine. Automatic Timbre Classification of Ethnomusicological Audio Recordings. International Society for Music Information Retrieval Conference (ISMIR 2014), Oct 2014, Taipei, Taiwan.

[4] M. A. Loureiro, and H. B. D. Paula. Timbre de um instrumento musical: caracterização e representação. Per Musi, Vol 14, pp. 57-81. 2006.

[5] G. Peeters, B. L. Giordano, P. Susini, N. Misdariis, and S. McAdams. The timbre toolbox: Extracting audio descriptors from musical signals. Journal of the Acoustical Society of America, 130 (5): 2902-2916. 2011.

[6] J. Garcia and G. Tzanetakis,. Musical Instrument Classification Using Individual Partials in IEEE Transactions on Audio Speech and Language Processing · February 2011.

[7] S. Joshi and A. Chitre. Identification of Indian musical instruments by feature analysis with different classifiers. Proceedings of the Sixth International Conference on Computer and Communication Technology, pp. 110-114, 2015.

[8] H. Ezzaidi, M. Bahoura and G. E. Hall. Towards a characterization of musical timbre based on chroma contours. International Conference on Advanced Machine Learning Technologies and Applications, Springer, Berlin, Heidelberg, pp. 162-171. 2012

[9] S. Böck, F. Korzeniowski, J. Schlüter, F. Krebs and G. Widmer. Madmom: A new python audio and music signal processing library. In Proceedings of the 24th ACM international conference on Multimedia. pp. 1174-1178. 2016.

[10] B. McFee, C. Raffel, D. Liang, D. P. Ellis, M. McVicar, E. Battenberg, and O. Nieto. Librosa: Audio and music signal analysis in python. In Proceedings of the 14th python in science conference Vol. 8, pp. 18-25. 2015.

[11] C. E. Cella, D. Ghisi, V. Lostanlen, F. Lévy, J., Fineberg, and Y.Maresz,. OrchideaSOL: a dataset of extended instrumental techniques for computer-aided orchestration. arXiv preprint arXiv:2007.00763. 2020.

[12] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau and P. van Mulbregt,. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature methods, 17(3), 261-272. 2020.

[13] Y. González and R. C. Prati. Characterization of the sonority associated with woodwinds instruments through spectral analysis. Conference: Brazilian Symposium on Computer Music: São João Del Rei, Minas Gerais, Brazil, September 2019.

[14] G. Peeters. A large set of audio features for sound description (similarity and classification) in the CUIDADO project. CUIDADO Ist Project Report, 54(0), 1-25. 2004.

[15] J. Bullock. Implementing audio feature extraction in live electronic music (Doctoral dissertation, PhD thesis, University of Birmingham). 2008.