

Article

Acoustic Descriptors for Characterization of Musical Timbre Using the Fast Fourier Transform

Yubiry Gonzalez *  and Ronaldo C. Prati 

Center of Mathematics, Computer Science and Cognition at Federal University of ABC, Av. Dos Estados 5001, Santo André 09210-580, SP, Brazil; ronaldo.prati@ufabc.edu.br

* Correspondence: yubiry.gonzalez.17@gmail.com

Abstract: The quantitative assessment of the musical timbre in an audio record is still an open-ended issue. Evaluating the musical timbre allows not only to establish precise musical parameters but also the recognition, classification of musical instruments, and assessment of the musical quality of a sound record. In this paper, we present a minimum set of dimensionless descriptors, motivated by musical acoustics, using the spectra obtained by the Fast Fourier Transform (FFT), which allows describing the timbre of wooden aerophones (Bassoon, Clarinet, Transverse Flute, and Oboe) using individual sound recordings of the musical tempered scale. We postulate that the proposed descriptors are sufficient to describe the timbral characteristics in the aerophones studied, allowing their recognition using the acoustic spectral signature. We believe that this approach can be further extended to use multidimensional unsupervised machine learning techniques, such as clustering, to obtain new insights into timbre characterization.

Keywords: musical timbre; Digital Signal Processing; wooden aerophones; audio analysis; FFT



Citation: Gonzalez, Y.; Prati, R.C. Acoustic Descriptors for Characterization of Musical Timbre Using the Fast Fourier Transform. *Electronics* **2022**, *11*, 1405. <https://doi.org/10.3390/electronics11091405>

Academic Editor: Daniele Salvati

Received: 4 April 2022

Accepted: 26 April 2022

Published: 27 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The implementation of computational methods for the digitization of sound, through the Fast Fourier transform (FFT), has allowed important advances in Music Information Retrieval Systems [1,2], in the recognition and identification of Musical Instruments [3,4] and the characterization of musical audio records.

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 directly measurable quantities. Timbre, however, is a multi-dimensional and shallowly defined attribute that allows one to distinguish between different sounds even when they have the same intensity, duration, and tone; that is, it allows discriminating sounds of different musical instruments even when it is the same musical note, with the same duration and the same intensity. The timbre of a specific musical sound (with a fundamental frequency in the tempered musical scale) is mainly related to the attack, sustain, and decay lapses of that frequency, the tonal variation, and the harmonics present [5,6].

The problem of univocal characterization of musical timbre raises the need to elaborate descriptors (derived magnitudes, coefficients, or functional) that evaluate timbre from digital audio records [7]. This is important for problems such as Automatic Music Transcription (AMT) [8]. It has been identified that, for audio recordings, it is necessary to have efficient systems that identify the different musical timbres with high precision and quantitatively [9].

Since musical timbre is a phenomenon of auditory perception, many of the investigations are developed in the line of psychoacoustics, aiming to evaluate verbal descriptors that reveal measurable attributes of musical timbre [10–13], timbre attributes of musical sounds, and visual colors through subjective perception evaluation experiments [14]. Other,

more recent studies focus on the analysis of timbre similarity between auditory images of various types of musical instruments and the linguistic-cognitive dimensions of timbres through a model where the timbre representation is contained in 20 dimensions [15,16].

Although the psychoacoustic perception of musical timbre cannot be ignored, it must be recognized that the main timbral characteristics must be somehow inscribed within the FFT that enables the recording and subsequent reproduction of musical sound. For the sake of argument, suppose any significant timbral characteristics that are not contained in the FFT performed on a musical audio record. Then, the deconvolved audio (inverse convolution) of the reproduced digital record could not be distinguished timbrally. However, this fact does not happen in musical digitization as we could distinguish timbral aspects from the deconvolved audio. Therefore, the FFT contains all the significant timbral characteristics.

Other research focuses on the presentation of an exhaustive collection of acoustic descriptors in the form of coefficients or functional (Timbre Toolbox, Librosa, etc.) that can be computationally extracted from the statistical analysis in the digitization of the spectrum (FFT). This focuses mainly on the statistical and mathematical characterization of the maximums in the FFT, such as the mean value in frequency (centroid) and amplitude, standard deviation, kurtosis, roots or poles of the distribution, arithmetic and geometric sequences in frequency, mean values and mean-squares of the amplitudes, among others [1,9,16–21]. Currently, there is no consensus on which and how many acoustic descriptors are to characterize musical timbre. However, it is recognized that many of them are derivatives or combinations of others and that in general they are correlated with each other. We wonder, then, what would be the minimum timbral descriptors that allow such a characterization of the musical timbre.

Peeters and collaborators [22] have studied the 34 most common statistical and mathematical descriptors and have found that only 9 descriptors are relatively independent with a correlation of 60%, and only 7 are so with a 50% correlation, which suggests that future studies on musical timbre should consider as a minimum: (1) a measure of the central tendency of spectral descriptors that vary in time; (2) a measure of the temporal spread of time-varying spectral descriptors; (3) a descriptor of the energy content of the sound signals and of the time envelope of the energy; (4) a periodicity descriptor. Peeters and collaborators [22] add: "It should be noted that this minimum requirement characterizes few or none of the previous studies on the perception of musical timbre". Since we are studying only musical monophonic sounds of aerophones, we have been left with a minimum set of timbre descriptors from a single FFT analysis, excluding psychoacoustic and perceptual characteristics for quantification.

Despite recent advances using verbal (psychoacoustic) descriptors as statistical or mathematical descriptive analysis of musical timbre, this is still an open-ended problem. Taffeta et al. [23] state that most acoustic descriptors fail to examine all sound parameters that could be important when attempting to analyze timbral dimensionality in musical sounds and it constitutes a challenge for the automated identification of musical instruments and the quantitative evaluation of the "quality" of the sound in melodic, simple or compound fragments (polyphonic music).

An alternative approach is to construct the timbral descriptors from musical acoustics, using by analogy the analysis and digital processing of spectra in other areas of knowledge. FFT provides a representation in the two-dimensional plane of the frequencies and intensities (alternatively energy-frequency) present in a signal. Spectral analysis techniques consist of describing and interpreting the statistical distribution of the amplitudes for different frequencies. In that sense, it does not depend on the wavelength or frequency analyzed. Therefore, an infrared spectrum (Chemistry), visible (Astronomy), signals from ultrasound, Electroencephalograms and electrocardiograms (Medicine), seismic (Geology) and sonic (Ultrasound for failure analysis, or mechanical vibrations in acoustics and Engineering) are equivalent. From the physical point of view, and also from the mathematical description, an audio FFT spectrum does not differ from the spectra obtained in Mechanical Engineer-

ing for the analysis of vibrations, or from the infrared (IR) spectra in Chemistry for the rotational and vibrational oscillations of molecules.

The objective of this work is to propose a minimum set of acoustically motivated timbral descriptors (functional mathematical or dimensionless coefficients) that allow the computational extraction of timbral information from musical records using the FFT. To evaluate the effectiveness of these timbral descriptors, an empirical study of the timbral characteristics in digital musical sounds of a sample of wooden aerophones is presented: Clarinet, Bassoon, Transverse Flute, and Oboe.

This paper is organized as follows: the next section presents the methodology for obtaining the spectra and their quantification of the audio records. Then, the Timbral coefficients section shows the coefficients or descriptors and their evaluation for the selected aerophones. In the Spectral Signatures section, the chromatogram for the fourth octave of the tempered musical scale and the spectral signature (“fingerprint”) that characterizes each aerophone studied are evaluated. Finally, conclusions and future work directions are presented in the last section.

2. Methodology

The aerophone audio records were obtained from the TinySOL open-source sound library [24] which contains recordings of individual sounds, played ordinarily, at a level of dynamic/intensity mezzo-forte, in a WAV audio format that minimizes information losses, sampled at 44.1 kHz on a 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, these are the Transverse Flute (Fl), Oboe (Ob), Clarinet (Cl), and Bassoon (Bn).

The selection of the family of wooden aerophones is justified by having the greatest timbral diversity and more melodic possibilities for functioning as a solo instrument within a symphony orchestra but also presenting well-differentiated timbral characteristics when they function as a set of instruments [25].

Although Pons, Jordi, et al. [26] make the timbre analysis based on spectrograms, we propose in this work a simpler alternative way studying the FFTs since these must contain the same relevant information of the musical timbre, applied to monophonic sounds and considering the proposed set of instruments. This simpler alternative considers coefficients extracted from the full FFT spectra, instead of the spectrogram obtained using the short-time FFT.

To calculate the FFT and obtain the frequency spectra, the SciPy library module in Python [27] is used. 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 to compare the various audio spectra with different power levels (Watt) or relative power (dB). Due to the FFT used, which operates on the entire function, information on size, overlap, etc. is not considered in this study.

The calculation of the coefficients was performed on the frequency spectrum of the common range of the instruments studied, corresponding to musical sounds between B3 to D#5 (~246 Hz to ~623 Hz).

The study of acoustic phenomena affirms that all musical instruments consist of a resonator (string, membrane, or resonant tube) whose geometry and materials will determine, for a given impulse, the natural modes of vibration of the generated stationary wave. Thus, the nature of the component materials of the musical instrument, and its geometric shape determine the secondary frequencies and harmonics possible of a mode of vibration or impulse associated with a certain musical note. This is what allows the reproducibility of the sounds for each musical note and consequently of the melodies.

The values of the amplitude of each vibration mode, both for the fundamental frequency and its harmonics and other component partial frequencies, will depend on several factors such as the energy associated with the initial impulse, the losses due to transmission

and absorption, and the combination of standing waves (resonances, constructive and destructive interferences due to multiple reflections, reverberation, etc.). However, the distribution of the set of harmonics and the values of the fundamental frequencies and the partial frequencies do not change, even when the amplitudes of the waves are attenuated or intensified in the propagation of the waves inside the resonant cavity (aerophones), along the vibrating string (rubbed string instruments and plucked strings) or on the struck surface (flexible membrane or vibrating surface).

Then, the sound of a musical instrument will consist essentially of a succession of waves, which are characterized by the magnitudes of amplitude and phase. At each instant, the sound is then described by amplitudes (magnitude measured on any scale: intensities, decibels, energy, etc.) and frequencies. Consequently, a Fourier spectrum of a monophonic audio signal is essentially a finite, paired collection of numbers associated, respectively, with the component amplitudes and frequencies.

The discrete distribution of such points in a configuration space can be described in terms of (i) the maximum of the acoustically associated distribution with the sound of the corresponding musical note (fundamental frequency and fundamental amplitude); (ii) the general form of homogeneity and range of variability of the amplitudes of the partial frequencies that accompany the fundamental frequency of the musical sound considered; and (iii) the statistical measurements of the mean amplitudes and middle frequencies.

Quantitatively assessing the musical timbre then involves weighing these three aspects: the fundamental frequency and amplitude, a measure of the secondary or harmonic frequencies present in number and relative importance (assessable through their intensity relative to the fundamental frequency), and a measurement of the grouping of the frequencies present (cadence, monotony, sequence of harmonics, etc.). Two magnitudes (amplitude and frequency) are used for these three aspects (musical sound, shape of the distribution, mean value of the partial sounds). A minimum of six functionals are then required to describe the monophonic musical timbre of individual sounds, in terms of the normalized amplitudes and frequencies present.

3. Results: Timbral Coefficients

It should be considered that, unlike the timbral study of speech and environmental sounds, musical frequencies make up a finite, countable, and discrete set of only 12 different values in each musical octave, for a total of 96 possible fundamental frequencies and their integer multiples that are in the audible range: from 20 Hz to 20 kHz. Therefore, the musical timbre can be characterized by a limited set of timbral coefficients, which are dimensionless quantities related to the frequencies and amplitudes in the Fourier spectrum of the audio records. Motivated by musical acoustics, these coefficients are tonal descriptors and, in essence, functionally describe the discrete distribution of normalized frequencies and amplitudes. As the amplitudes of the spectra of the FFTs are normalized (using the quotient of the amplitude of each partial frequency between the amplitude greater measured in each spectrum) it is possible to compare the relative amplitudes among them. They can be grouped into descriptors of the fundamental frequency (musical scale, 96 possible frequencies) and descriptors of the rest of the partial frequencies that arise when performing the FFT of the audio under analysis (descriptors of the shape of the distribution and statistics frequency distribution), these proposed descriptors are dimensionless coefficients.

The FFT values are essentially a discrete collection of pairs of different amplitudes and frequencies; therefore, they can be summarized by the following dimensionless parameters:

1. Measurement of the fundamental frequency in relation to the average frequency, (Affinity A).
2. Measurement of the frequency distribution with respect to the average frequency (Mean Affinity MA).
3. Quantification of the amplitude of f_0 with respect to the collection of amplitudes (Sharpness S).
4. Quantification of the average amplitude of the pulse collection (Mean Contrast MC).

5. Descriptor of how close the secondary pulses are to being integer multiples of the fundamental frequency (Harmonicity H).
6. Descriptor of the envelope through the average slope in the collection of pulses (Monotony M).

The details of these dimensionless timbral coefficients are described next and exemplified for the FFTs of monophonic woodwind sounds.

3.1. Fundamental Frequency Descriptors

In the analysis of acoustic signals, the centroid is usually used to describe the properties of the sound; however, in the timbral characterization, its use entails the difficulty of its acoustic interpretation, since given an audio record, the value of the average frequency (centroid) does not correspond to the envelope, nor to any of the fundamental and harmonic frequencies of the sounds studied. Its interpretation in musical acoustics is difficult because it cannot be correlated with the set of discrete frequencies of musical sounds in the Tempered Scale used in Western music (96 characteristic frequencies). The centroid is defined as the average frequency.

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

where a_i are the normalized amplitudes of the frequencies f_i of the Fourier spectrum. The normalized amplitudes are obtained by dividing the amplitudes values by the amplitude of the fundamental frequency a_0 .

Figure 1 shows the Fourier Transform Spectra of the (4) wooden aerophones studied, particularly for the audio records of the musical sound E4 that corresponds to the nominal frequency of 329.6 Hz. The centroid or mean frequencies of the Fourier spectrum do not correspond to any of the harmonics of the analyzed sound.

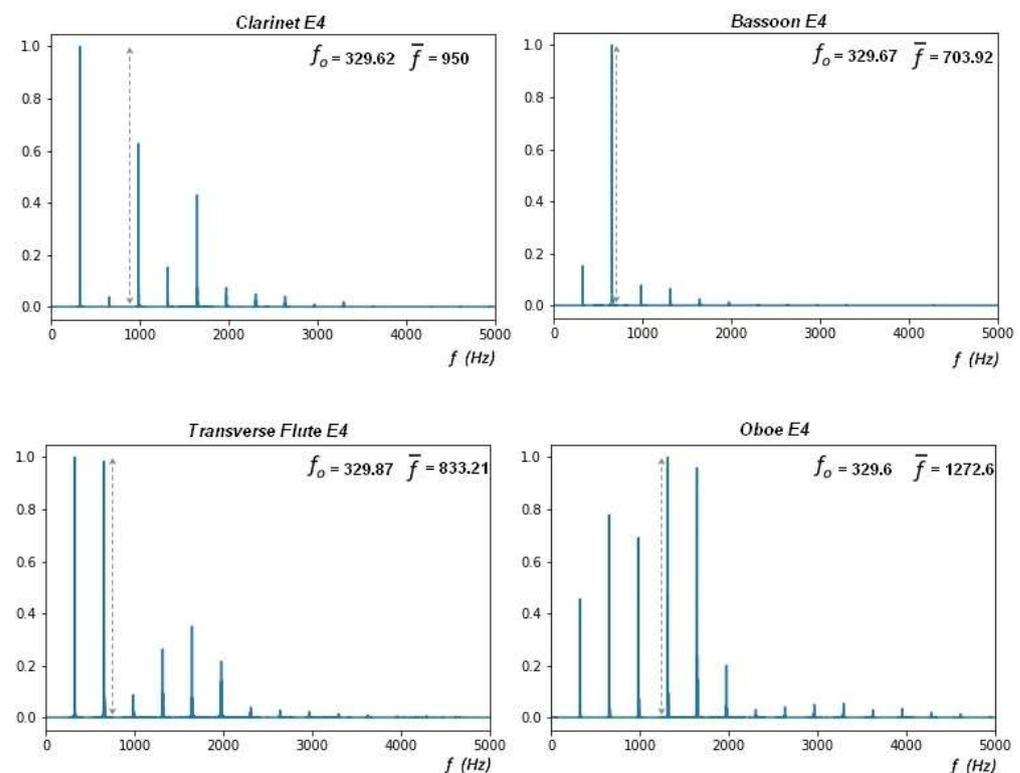


Figure 1. FFT spectra for the aerophones studied. Note that in each instrument the centroids (dotted line) do not coincide with any of the frequencies in the distribution and are not directly linked to the musical sound E4 of nominal frequency 329.6 Hz.

Figure 2 shows the centroid values concerning the musical sounds of the common tessitura of the selected aerophones. It can be seen that the centroid values are almost constant for the frequency range considered.

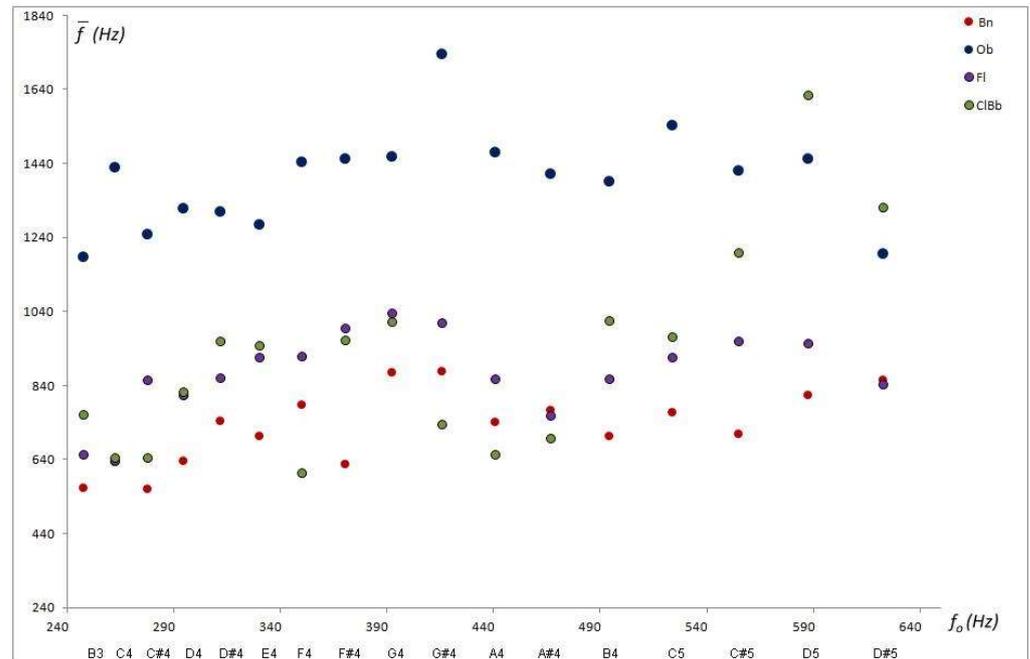


Figure 2. Variation of the centroid concerning the common tessitura of the wooden aerophones studied: Bassoon (Bn), Oboe (Ob), Transverse Flute (Fl), Clarinet in Bb (ClBb). For comparison purposes, the location of the musical notes of the scale below the abscissa axis is also included.

The centroid, in any case, indicates the presence of frequencies other than the fundamental one, with a distribution of intensities such that the average frequency (centroid) has a magnitude much greater than the frequency prescribed for each musical sound. However, it does not report the harmonicity of the sounds and does not correlate with the “natural” sequence of the musical scale in any of the instruments in the sample.

The fundamental frequency f_0 only coincides with the Centroid \bar{f} of the distribution if there is no harmonic or partial frequency in the sound. However, this is impossible in musical instruments whose resonators are exposed to multiple deflections, superpositions, and beats given by the geometry of the resonant tubes and openings, which characterize aerophones.

The separation between the fundamental frequency and the mean value of the frequencies would be important from the point of view of musical acoustics. This quantity can be evaluated through a dimensionless descriptor (Timbral Coefficient) that we call Affinity (A).

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 (\bar{f}):

$$A \equiv \frac{\bar{f}}{f_0} \tag{2}$$

If f_0 and \bar{f} are close there is more “Affinity” with the centroid and A tends to be one. In Figure 3, two Fourier spectra with very different coefficients of A are compared. Moreover, in the upper panel, it is noted that different sounds in different instruments may have the same centroid values. However, their affinity, defined by Equation (2), will be different.

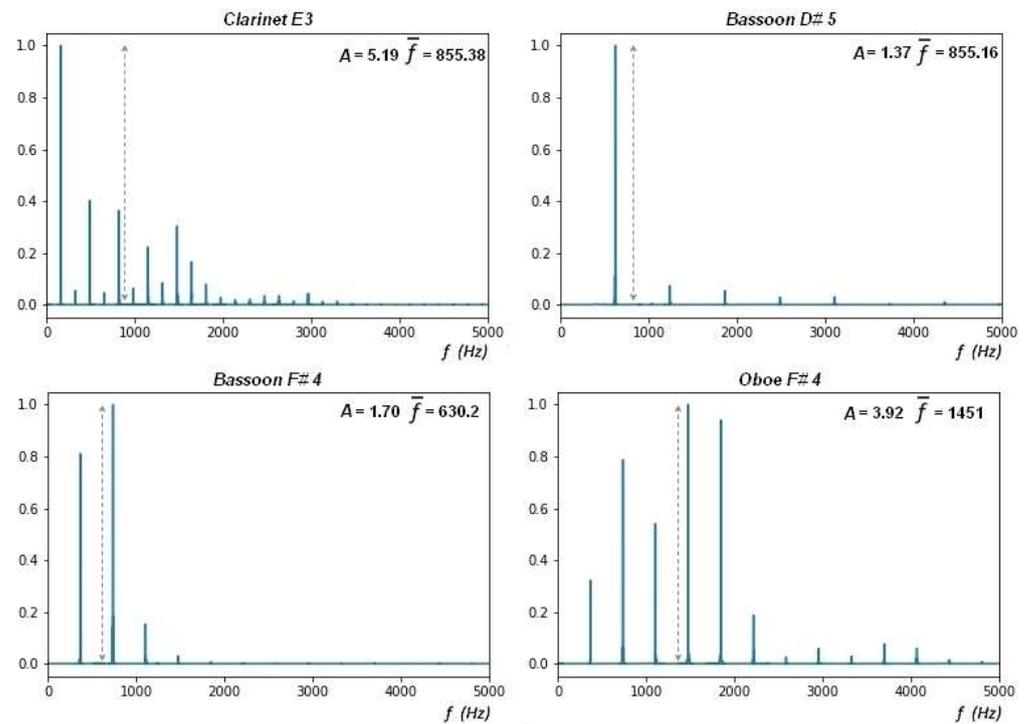


Figure 3. Examples of Fourier spectra for sounds with different Affinity coefficients and similar centroids (top panel) and the same sound in different aerophones. The dashed line indicates the relative location of the centroid.

The assessments of Affinity A for the audio records of the common tessitura (sounds B3 to D#5) for the aerophones studied are shown in Figure 4.

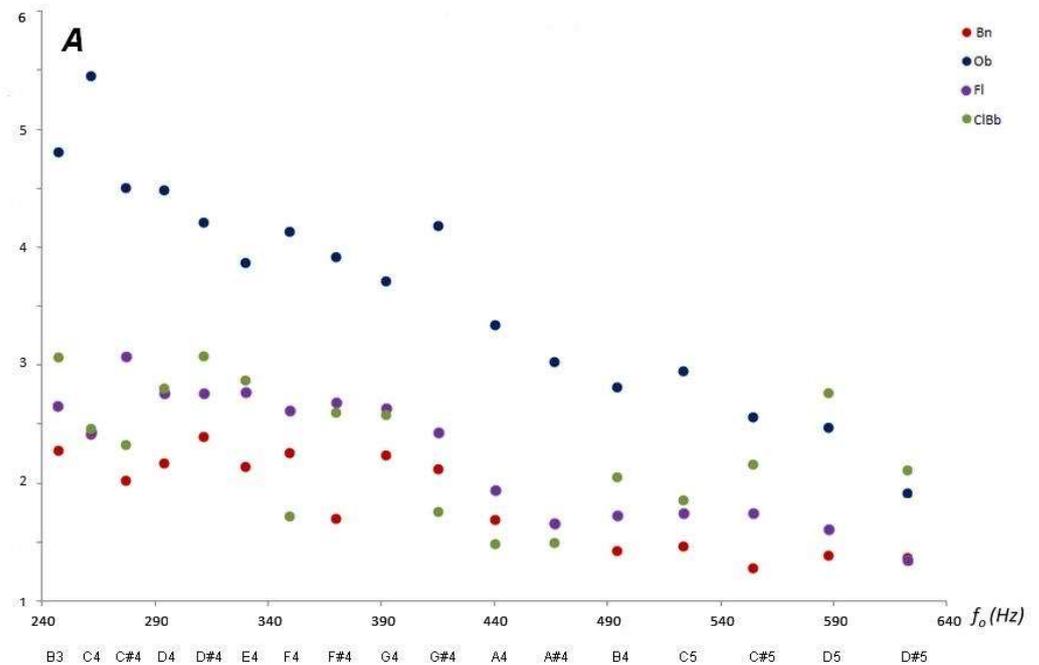


Figure 4. Variation of Affinity A in the selected sample of aerophones.

In Figure 4, it is observed that the Oboe has less Affinity with the fundamental frequency of the expected musical sounds (high values of A) and, at the other extreme, the

Bassoon has a mean frequency closer to the expected musical sound (values of A closer to the unit).

If $A \cong 2$ then the median or centroid frequency is twice the fundamental, as in the case of the Bassoon. For the Transverse Flute, the average frequency is of the order of $5/2f_0$ for the musical notes between B3 and F4, and of the order of $2f_0$ for the notes A#4 to D#5. Similarly, the Oboe behaves with more extreme values in both cases.

On the other hand, it can be observed in Figure 1 that various musical instruments have different amplitudes in their fundamental frequency for normalized Fourier spectra (that is, regardless of the intensity of the recording sound). Furthermore, for the same instrument, the amplitude of the fundamental frequency can, on occasions, remain constant between different musical sounds, or vary in amplitude for different frequencies.

One way to evaluate this variation of the amplitude of the fundamental frequency, concerning the amplitudes of the other frequencies of the spectrum, is by utilizing another dimensionless descriptor or Sharpness coefficient (S). Defined by:

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

Thus, the Sharpness (S) is a measure of the “height”, amplitude, or relative intensity of the fundamental frequency f_0 concerning the distribution, and its acoustic interpretation is immediate: the greater the Sharpness of a musical sound, the easier it will be due to its relative intensity to perceive the fundamental or nominal note that you want to play with the instrument. Ideally, a “pure” sound would be one that would have a single maximum without secondary frequencies $S = 1$. By construction $S \leq 1$. Figure 5 shows how spectra of different musical instruments, and different fundamental frequencies, can have the same Sharpness. In addition, that the same musical instrument can, depending on the frequency, present very different Sharpness.

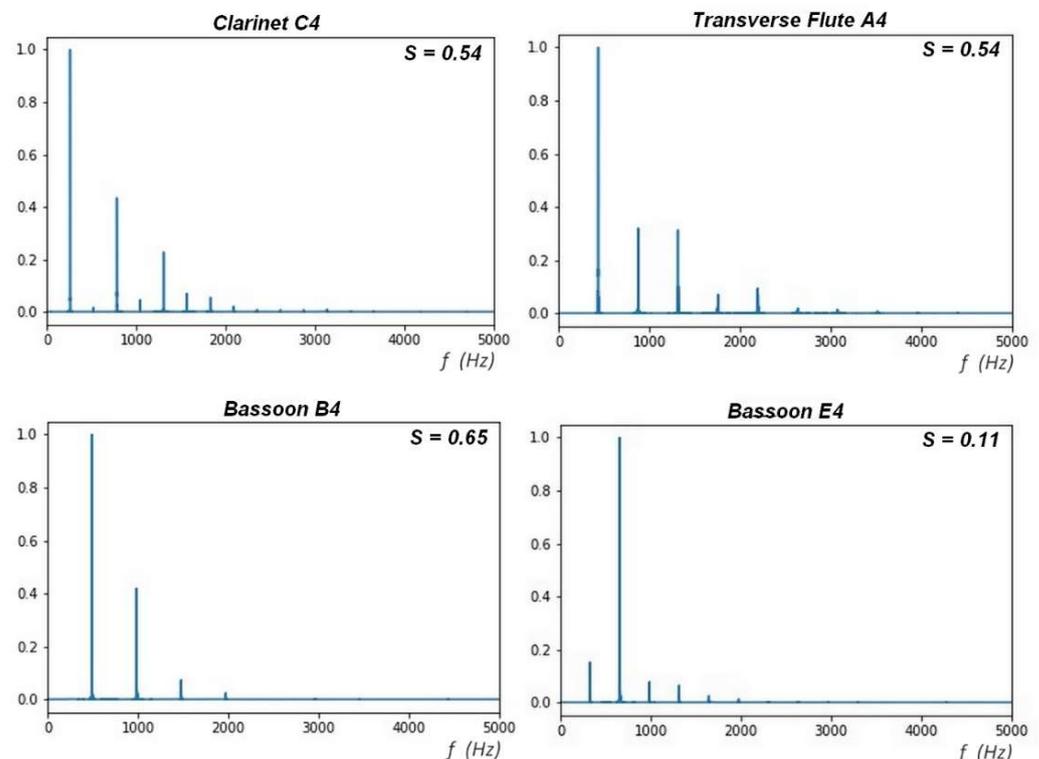


Figure 5. Example of Timbral Coefficient S . Above: The spectra of the Clarinet for C#4 and of the Transverse Flute for A4 show the same Sharpness $S = 0.54$. Bottom: Bassoon spectra in B4 and E4, with coefficients of $S = 0.65$ and $S = 0.11$ respectively.

Figure 6 shows the variation in the Sharpness of aerophones. It is observed that, in general, (except for the F#4), the Clarinet is sharper than the Transverse Flute and this is sharper than the Bassoon; in any case, the Oboe is the least sharp of all. That is to say, the musical sounds are better defined in the Clarinet and poorly defined in the Oboe.

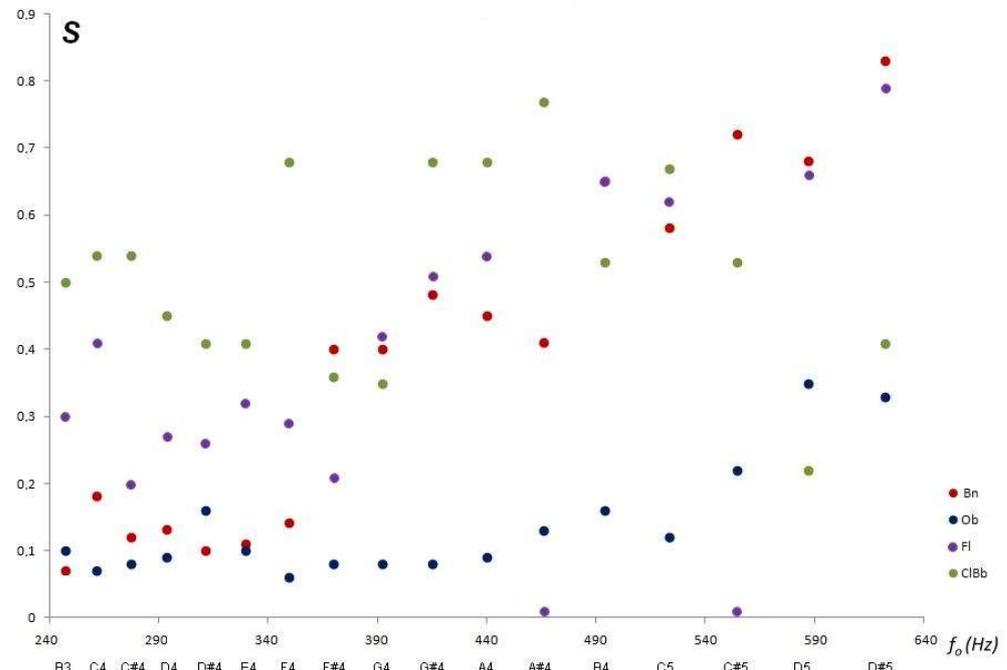


Figure 6. Variation of Sharpness (S) concerning common ranges.

3.2. Descriptors of Frequency Distribution

It is also necessary to characterize the distribution of partial frequencies that appear associated with the fundamental frequency f_0 in the spectra of the FFTs of the audio records. Many authors use statistical descriptors of partial or secondary maximums, evaluating the kurtosis, the regular sequences when they occur, and statistical parameters.

However, it is possible to construct descriptors with a more direct acoustic meaning. In a given spectrum, there are always partial or secondary frequencies to f_0 and can or cannot be harmonics. Harmonics 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 , and in a way, the sound would be more “harmonic” than if its frequency value were very different from the multiplicity of the fundamental frequency. To describe this property, the timbral coefficient of Harmonicity (H) is proposed and defined as:

$$H \equiv \sum_{j=1}^N \left(\frac{f_j}{f_0} - \left[\frac{f_j}{f_0} \right] \right) \tag{4}$$

where the symbol $[]$ denotes the integer part. This timbral function evaluates how harmonic the partial or secondary frequencies ($f_1, f_2, f_3 \dots f_j$) of the FFT spectrum are. The idea is that any frequency f_j is a harmonic of f_0 if the quotient between them is an integer. Figure 7 shows this variation in examples of FFT spectra with different Harmonicity.

The Fourier spectra of the upper panel of Figure 7 show that 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 may be harmonics of f_0 , as is the case of second, third and fourth maximum. If the secondary frequencies are all harmonics of f_0 then $H = 0$. Every time there are 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

Transverse Flute in Figure 7 are all integer multiples of the fundamental and, therefore, it is highly harmonic ($H = 0$).

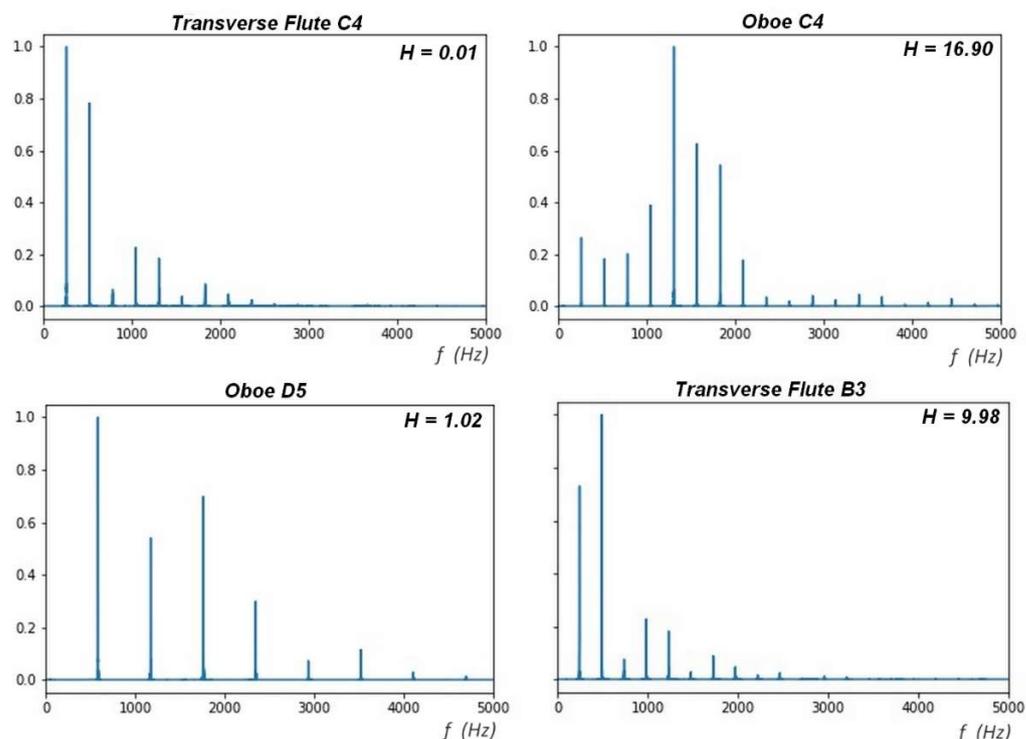


Figure 7. Comparison of Harmonicity H for different sounds and instruments. Top panel: C4 Flute and Oboe sound comparison. Lower panel comparison of different sounds for the same instrument (Left: Transverse Flute, right Oboe).

The same instrument can present very different values of harmonicity, depending on the musical sound considered, as observed for the Oboe on the right side of Figure 7, in sounds C4 ($H \cong 17$) and D5 ($H \cong 1$). Even the same instrument with a qualitatively similar Fourier spectrum can present a very different harmonicity coefficient, as shown for the Transverse Flute in the left panel of Figure 7; sounds C4 (very harmonic: $H \cong 0$) and B3 (very little harmonic: $H \cong 10$).

For common ranges, the results of the harmonicity assessment are shown in Figure 8.

The Transverse Flute has the highest harmonicity (H close to 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 with the instrument, so for the characteristic sound of reference $A = 440$ Hz the Bassoon and Clarinet are completely harmonic, and the Transverse Flute is so for musical sounds C4, D4, E4, and A#4.

Another outstanding aspect of the distribution of maximums in the FFTs of the audio records is the variability of the amplitudes as a function of frequency. After the fundamental frequency f_0 , the following and successive maxima may have amplitudes that increase or decrease, that is, they could decrease in amplitude (decreasing Monotony) or increase (increasing monotony), or vary in growth with some tendency to increase or decrease.

Figure 9 shows records with different behaviors in the monotony of the secondary or partial frequencies (whether they are harmonic or not).

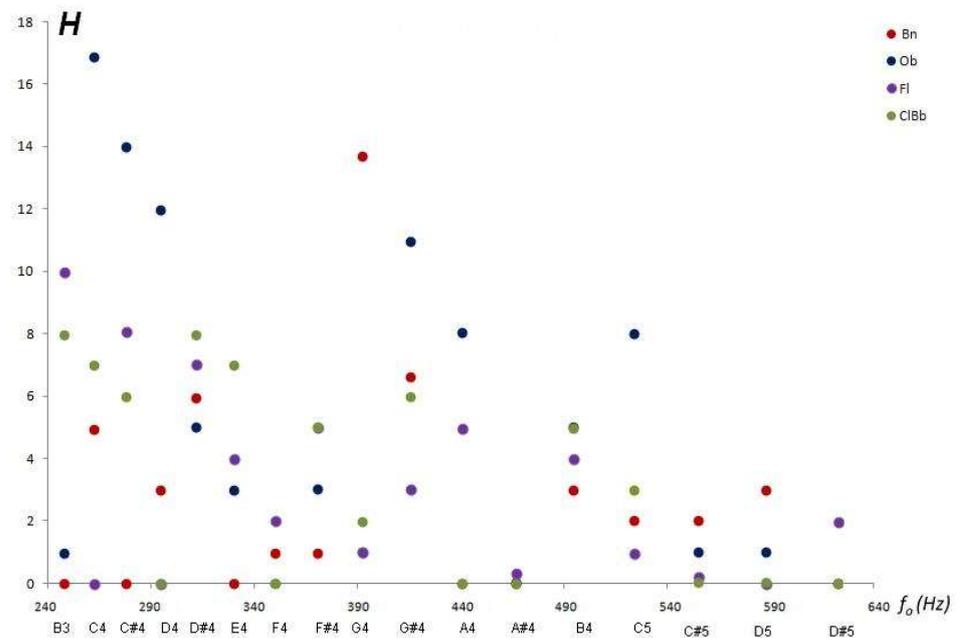


Figure 8. Variation of the Harmonicity H of aerophones in the range of sounds B3–D#5.

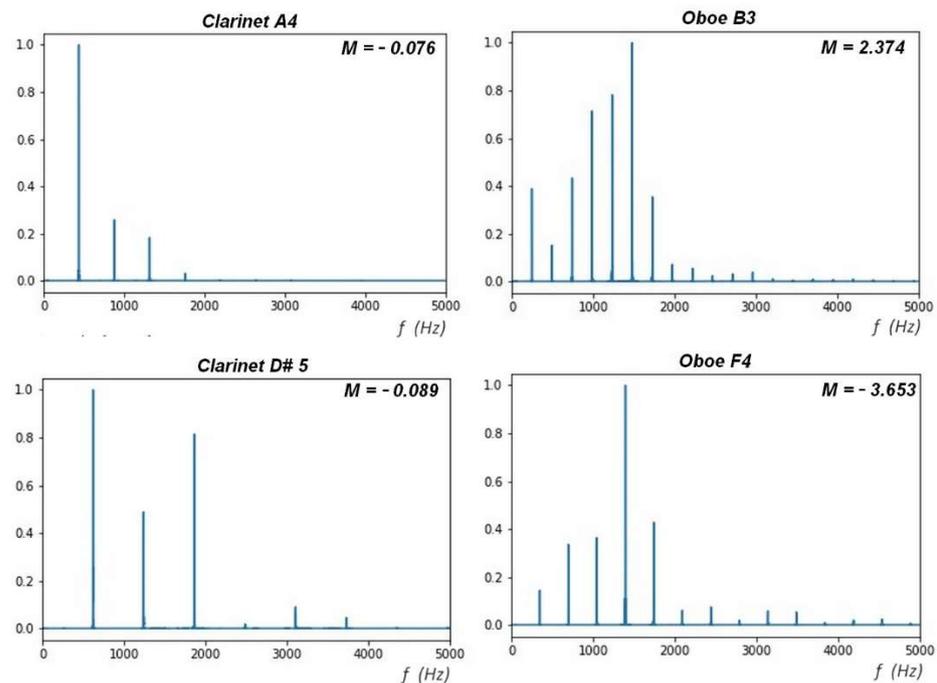


Figure 9. Effect of monotony on FFT spectra. Decreasing sound A4 Clarinet (left), increasing sound B3 Oboe and alternates, Clarinet D#5, and F#4 Oboe.

The timbral coefficient of Monotony (M) quantifies the regularity of the distribution of secondary frequencies (harmonic or not). It is defined as the average of the variations between two successive maxima:

$$M \equiv \frac{f_0}{N} \sum_{j=1}^N \left(\frac{a_{j+1} - a_j}{f_{j+1} - f_j} \right) \tag{5}$$

Essentially, monotony informs whether the harmonics appear in increasing (positive M) or decreasing (negative M) succession after the fundamental frequency. It can be

said that this coefficient approximates and discretely quantifies the first derivative of the frequency distribution.

Essentially monotony informs whether the harmonics appear in mostly increasing (positive M) or decreasing (negative M) succession after the fundamental frequency. It can be said that this coefficient approximately and discretely quantifies the first derivative of the frequency distribution.

Figure 10 shows the assessment of the monotony of the common tessitura for the musical instruments studied. The Clarinet has negative values very close to zero and, on average, its frequencies decay in amplitude, which is to say that the sound attenuates very slightly after the fundamental frequency. The same homogeneity is appreciated in the sound of the Transverse Flute. On the contrary, the Bassoon shows high positive and negative values in several musical notes, that is, the bassoon has more sound alternation after the fundamental frequency. The Oboe has a certain alternation with a decreasing trend and the very negative M value of it reports a large alternation with a trend or decreasing envelope for several of the notes.

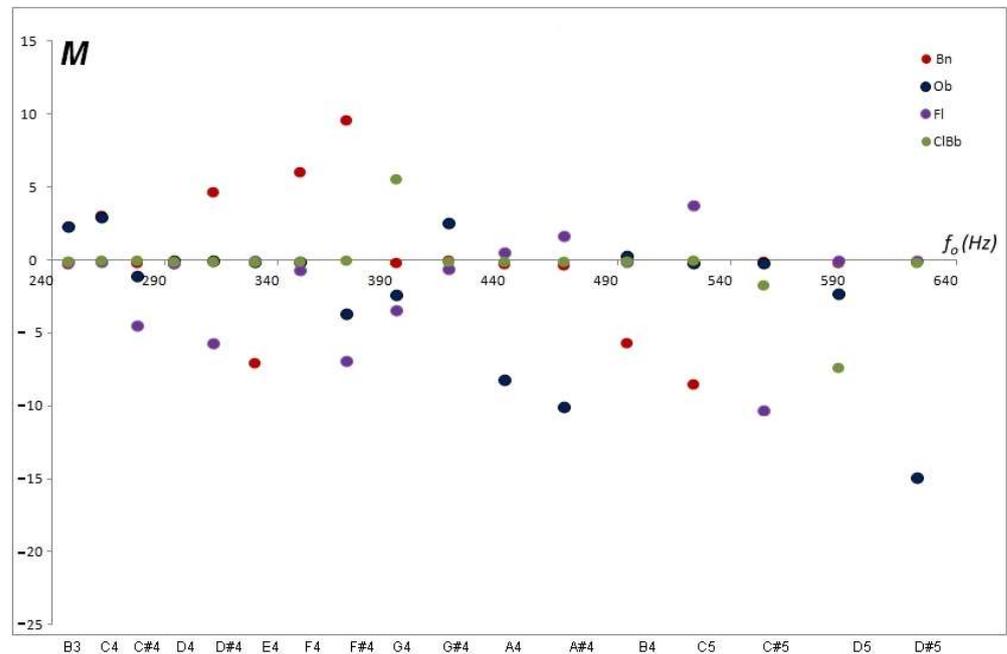


Figure 10. Monotony (M) of aerophones for common tessitura.

3.3. Distribution Statistics

Another aspect that is perceived in the FFTs of the audio records for different aerophones is that the secondary or partial frequencies occupy a different frequency range, that is, the partial frequencies (harmonic or not) are sometimes very close together or are grouped, and, in others, they are more spatially distributed or more separated. If the secondary frequencies are very close to each other or close to the fundamental, the sounds may appear more “compact or thick”, while the spatial separation of the frequencies gives more diaphanous or transparency to the fundamental frequency. Transparency is the timbral coefficient defined as the Mean Affinity (MA). It quantifies the frequency extent of the distribution and evaluates the separation of the partial frequencies concerning the centroid.

$$MA \equiv \frac{1}{Nf_0} \sum_{i=1}^N |f_i - \bar{f}| \tag{6}$$

The MA coefficient attempts to evaluate how compact the frequency distribution is concerning that mean value (including the fundamental frequency f_0). Note that the subscript i has been used instead of j to denote that the sum also includes f_0 . In other

coefficients the sum over j does not include f_0 because it refers to the secondary maxima, whether harmonic or not. Figure 11 shows how Mean Affinity (MA) varies in various FFT spectra. Figure 11 shows that the same sound can have different MA values in different instruments. For example, in the upper panel the C4 sound of the Clarinet ($MA = 2.51$) and the Transverse Flute ($MA = 2.98$) both have very similar affinity coefficients, as well as, in the lower panel, the D5 sound in Oboe ($MA = 2.44$) and Bassoon ($MA = 1.56$). This holds even for sounds and instruments with the same Affinity ($A = 2.47$) such as the C4 Clarinet ($MA = 2.51$) and the D5 Oboe ($MA = 2.44$). Two sounds with equal affinity values are observed in Figure 11 wherein both registers have the same Affinity $A \cong 2.4$ but their average Affinity are different ($MA = 2.51$ Clarinet and $MA = 2.98$ Transverse Flute).

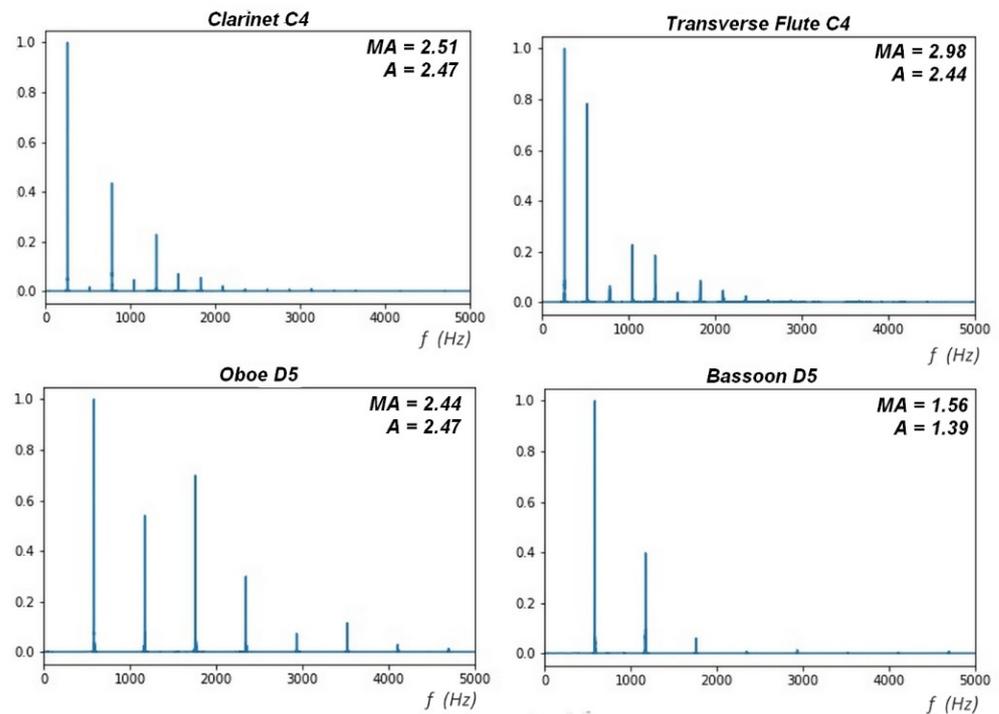


Figure 11. Variation of the mean amplitude of Fourier spectra. Sound C4 top panel: Clarinet ($MA = 2.80$) and Transverse Flute ($MA = 1.30$). Lower panel sound D5 Oboe ($MA = 2.44$) and Bassoon ($MA = 1.56$).

Lower MA values refer to compact distributions of secondary sounds close to the centroid. Thus, the Oboe has a lower average Affinity (high A value) for the range of sounds studied, which is equivalent to saying that the secondary frequencies (different from the fundamental one) cover a greater range of frequencies. The opposite is the case with the Bassoon, whose secondary sounds are more compact for each of the nominal frequencies of the scale, as shown in Figure 12. It is important to note that, by its definition, the MA value multiplied by f_0 gives us the effective mean frequency.

Finally, it should be noted that two FFT spectra can have the same number of secondary or partial frequencies, even if in both cases they are harmonic, but if one of these partial spectra has higher amplitude compared to the other spectrum, they will appear to us with greater variability in frequencies (Figure 13).

The Mean Contrast (MC) is the coefficient that measures the amplitude of the partial frequencies (tones or “colors”) concerning the fundamental frequency.

$$MC \equiv \frac{1}{N} \sum_{j=1}^N |a_0 - a_j| \tag{7}$$

MC increases as the intensity of partial frequencies. The MC values for the common tessitura of the aerophones studied are shown in Figure 14.

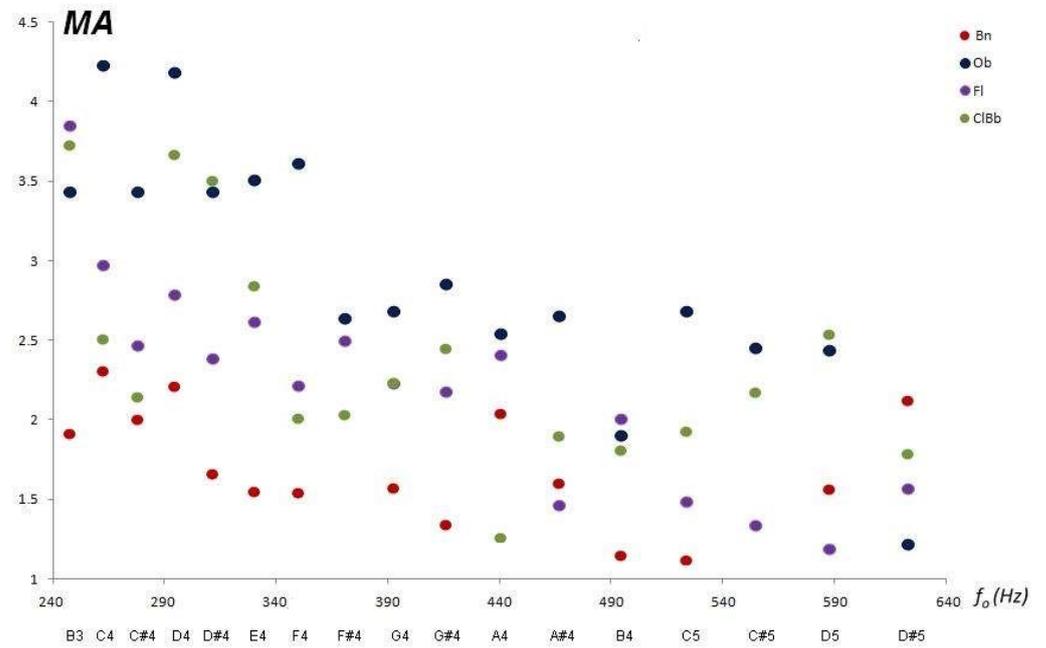


Figure 12. Variation of Mean Affinity MA in the range of sounds B3–D#5.

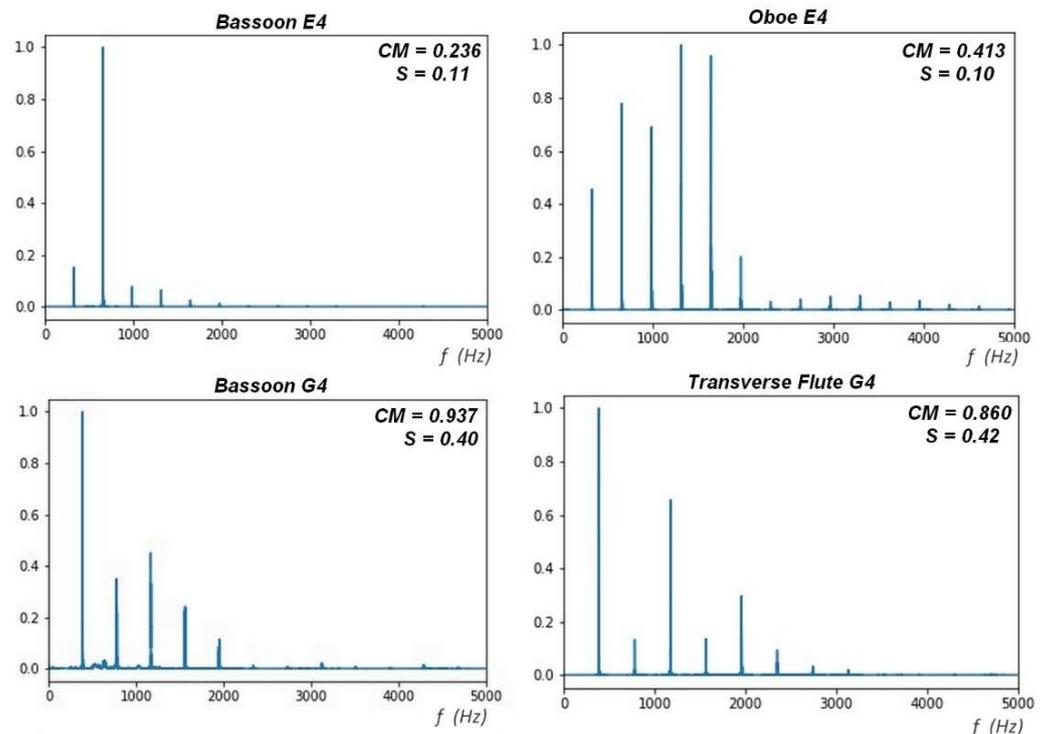


Figure 13. Average contrast in the FFT spectra of the audio records. Top panel: E4 sound for Bassoon ($CM = 0.236$) and Oboe ($CM = 0.413$), both with similar brilliance ($S \approx 0.1$). Bottom panel: G4 sound for Bassoon and Flute both with similar medium contrast ($CM \approx 0.9$) and brightness ($S \approx 0.4$).

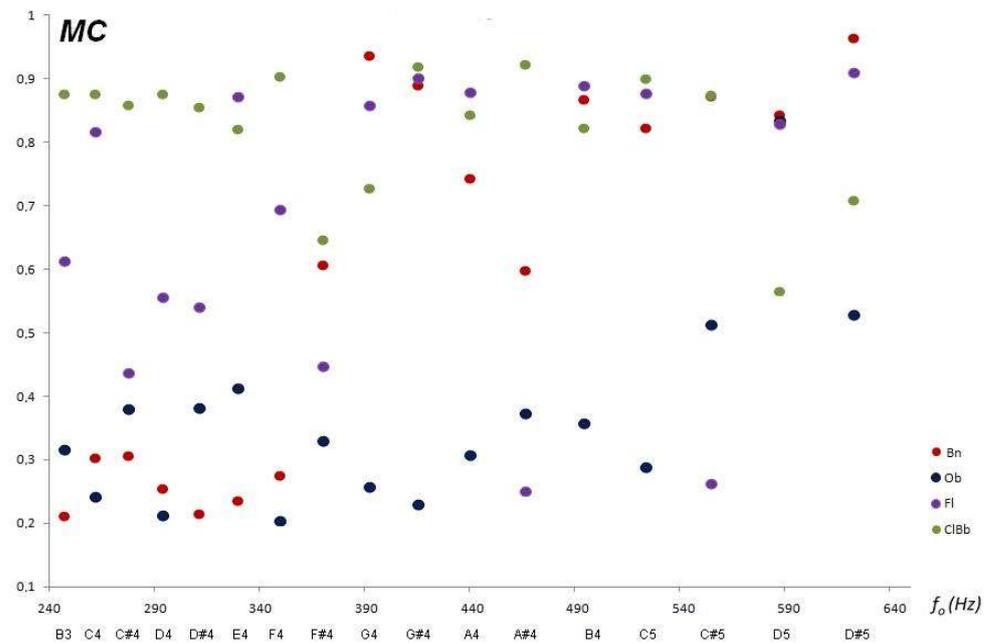


Figure 14. Mean contrast (MC) of the instruments studied for common tessituras.

The Bassoon has a greater diversity of partial frequencies for the first half of the musical scale considered because the variations in amplitudes concerning amplitude a_0 of the fundamental frequency are greater when compared to other woodwind instruments. The same can be said of the Oboe for the other upper half of the musical scale concerning the other woodwind instruments. The Transverse Flute and the Clarinet exhibit approximately the same contrast of the amplitudes of their secondary frequencies related to the amplitude of the fundamental frequency.

In the appendix, the tables of the coefficients for all the tessitura of the four aerophones studied are presented. In these tables, the first column corresponds to the nominal frequency of the musical sound, the second column the Anglo-Saxon denomination of the musical sound, followed by the fundamental frequency and the centroid obtained utilizing the FFT of the audio records. Then, in the last columns, the values correspond to the coefficients of Affinity (A), Brightness (S), Harmonicity (H), Monotony (M), Mean Affinity (MA), and Mean Contrast (CM).

4. Discussion: Spectral Signatures

The identification of musical instruments, and in general of the wave emitting source, can be done 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 (be it a gas, a reflective surface, or an absorbing medium) and is called Spectral Signature. By analogy, we understand it as an “acoustic spectral signature” as the distribution of frequencies in the Fourier Transform Spectrum of musical sounds (tempered scale) that allow their univocal identification. If we restrict ourselves to the studied aerophones, which are only a small sample of all common musical instruments, a discriminatory identification between them can be made from the presented FFT spectra for common records in the range of B3 sounds to D#5, as shown in Figure 15.

The harmonic frequencies (f_n) are those that are related to the fundamental frequency f_0 through the integer multiplicity ($n = 2, 3, 4, \dots$):

$$f_n = nf \tag{8}$$

It is the case that not all the harmonics of a certain musical sound are always present. This allows, in principle, to discriminate the sounds of one of the aerophones from the

others (Table 1). For a musical sound like B-3, the Oboe is distinguished from the other aerophones by presenting harmonics 2 to 13 inclusive. While the Bassoon lacks harmonic $n = 6$, the Flute contains harmonic $n = 6$ but lacks harmonic $n = 13$ that appears in the Oboe and the Clarinet, and the Clarinet lacks the harmonics $n = 11$ and $n = 12$ which distinguishes it from the Oboe and the Flute, respectively. From the collection of harmonics listed in Table 1 it is observed that one aerophone can be distinguished from another in most of the musical sounds of the common tessitura (in the range of sounds B3–D#5), so for example for C#4 the Bassoon only has up to the harmonic $n = 6$, the clarinet up to $n = 7$, the Flute reaches up to $n = 9$, and the Oboe even has harmonics of higher order.

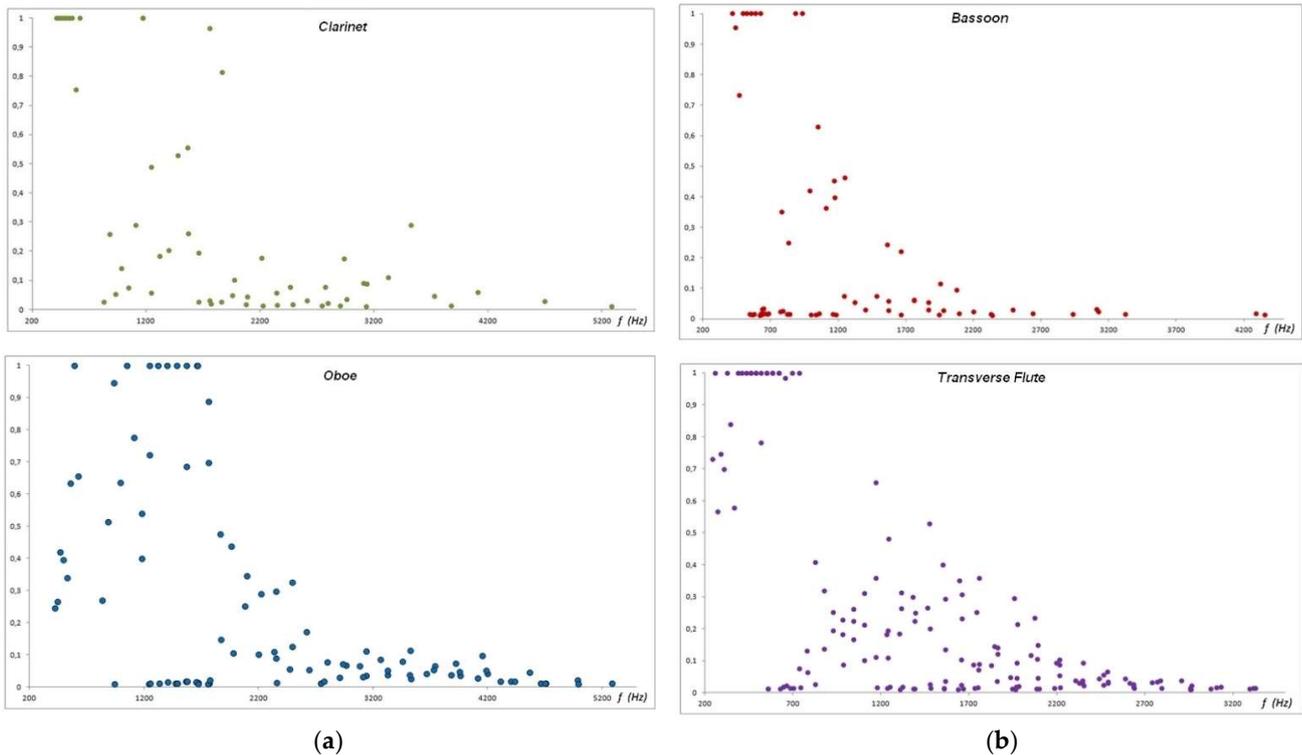


Figure 15. Spectral Signatures for Aerophones: (a): Clarinet and Oboe, (b): Bassoon and Transverse Flute.

Table 1. Set of harmonics depending on the musical sound and the musical instrument.

Sound	Number of Harmonics (n)				
	Note	f_0 (Hz)	CIBb	Ob	Bn
B3	246.94	{3, ... ,8}{10}{13}	{2, ... ,13}	{2, ... ,5}{7}	{2, ... ,10}{12}
C4	261.6	{3, ... ,8} *	{2, ... ,14}{16,17}	{2, ... ,8} *	{2, ... ,9}
C#4	277.2	{3, ... ,7}	{2, ... ,9}{11,12}{14}{16}	{2}{4, ... ,6}	{2, ... ,9}
D4	293.7	{2, ... ,11}	{2, ... ,12}{15}	{2, ... ,7}	{2, ... ,9}
D#4	311.1	{2, ... ,11}	{2, ... ,13}	{2, ... ,5}{7}	{2, ... ,6}{8,9}
E4	329.6	{2, ... ,8}{10}	{2, ... ,14}	{2, ... ,6}	{2, ... ,10}
F4	329.67	{2, ... ,6} *	{2, ... ,10}{12,13}	{2, ... ,6} *	{2, ... ,8}
F#4	370	{2, ... ,7}	{2, ... ,12}	{2, ... ,4}	{2, ... ,9}
G4	494	{2, ... ,8} *	{2, ... ,11}	{2, ... ,6}{8}{11}	{2, ... ,8} *
G#4	415.3	{2, ... ,7}	{2, ... ,12}	{2, ... ,5}{8}	{2, ... ,8}
A4	440	{2, ... ,4}	{2, ... ,10}	{2, ... ,6}	{2, ... ,6}{8}
A#4	466.2	{2, ... ,5} *	{2, ... ,10}	{2, ... ,5} *	{2, ... ,5} *
B4	466.2	{2, ... ,6} *	{2, ... ,9}	{2, ... ,4}	{2, ... ,6} *
C5	523.25	{2, ... ,6}	{2, ... ,9}	{2, ... ,4}	{2, ... ,5}
C#5	554.37	{2, ... ,7}	{2, ... ,9}	{2,3}	{2, ... ,5}
D5	587.33	{2, ... ,9} *	{2, ... ,9} *	{2,3}{5}	{2, ... ,4}
D#5	622.25	{2, ... ,6}	{2, ... ,5}	{2, ... ,5}{7}	{2,3}{5}

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

It should be noted that the amplitudes of the obtained and normalized FFTs may present frequencies of negligible values compared to the frequency of greater amplitude. In any case, the harmonics counted in Table 1 and the calculation of coefficients are only those partial frequencies whose amplitude is greater than or equal to 1% of the maximum amplitude.

This discrimination, in terms of which harmonics are present and which are missing, 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). However, as the FFTs are not the same between any two aerophones, the different relative intensities, through the Affinity and Sharpness coefficients, would allow the completion of the spectral signature and univocal identification of which instrument it belongs (Figure 16).

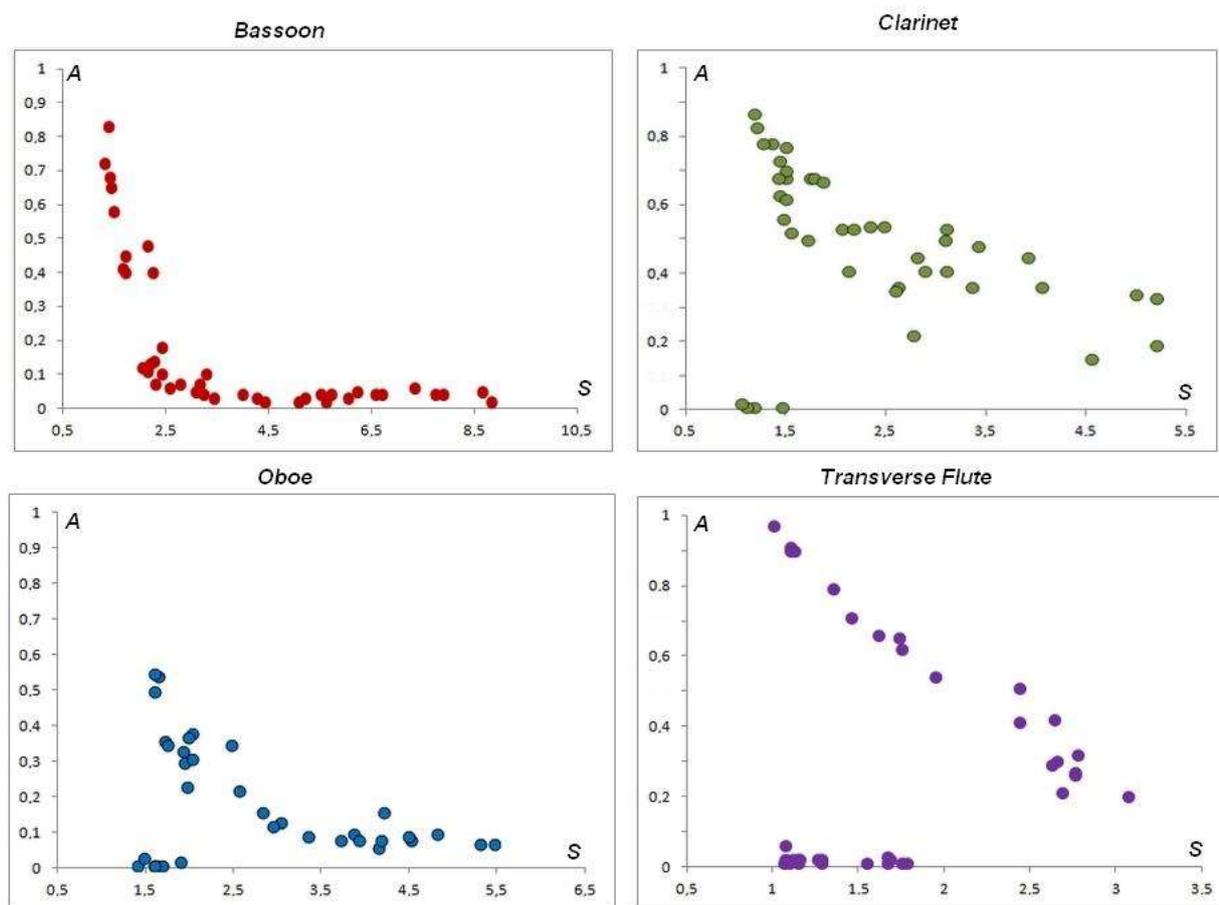


Figure 16. Two-dimensional representation in the A - S plane of the timbral space of the coefficients for the selected aerophones.

The degeneracy of the harmonic frequencies of the $C4$ sound between the Clarinet and the Bassoon is not resolved with the centroid value (Figure 2) which is the same for both, but with the coefficient of brilliants S , for example (Figure 6). Similarly, all other degenerations can be solved using one or more of the six proposed timbral coefficients.

Thus, for example, the triple apparent degeneration in sound $A\#4$ is solved by using the coefficient of Mean Affinity AM (Figure 14) or Mean Contrast (Figure 16) which allows the musical instrument to be uniquely identified. On the other hand, it should be noted that the number of harmonic frequencies present does not necessarily mean greater harmonicity in the musical sound. Note, for example, that the $C\#4$ sound of the Oboe has 12 harmonics (Table 1) and yet the value of $H = 14$ (Figure 8) shows great anharmonicity, while the Bassoon with only 4 harmonics exhibits great harmonicity ($H = 0$) in the same musical sound. All this illustrates that counting the harmonics present is not enough to determine the harmony of musical sounds.

From the preceding analysis, it is inferred that the timbral characteristics of monophonic musical sounds can be represented by a set of six dimensional coefficients. That is to say that each sound of the tempered scale of a musical instrument corresponds to one and only one point in an abstract space of six dimensions, with coordinates (A, S, H, M, MA, MC) . For certain musical instruments, the collection of points in that timbral space represents a bounded region, which characterizes it. Figures 16–18 illustrate this grouping using the two-dimensional projections of the planes $A-S$, $H-M$, $MA-MC$ respectively of the sounds of the complete tessitura (Tables A1–A4 of the Appendix A) in the aerophones studied. Any monophonic musical sound can be located in said space through the timbral coefficients already defined from the FFT of the audio register and its Euclidean distance from the delimited regions. Therefore, the proposed coefficients would allow, in principle, the evaluation of the timbral characteristics using that can be applied to large dimensional spaces. In future work, we plan to exploit unsupervised machine learning techniques, such as clustering algorithms, to identify patterns from the timbral characteristics.

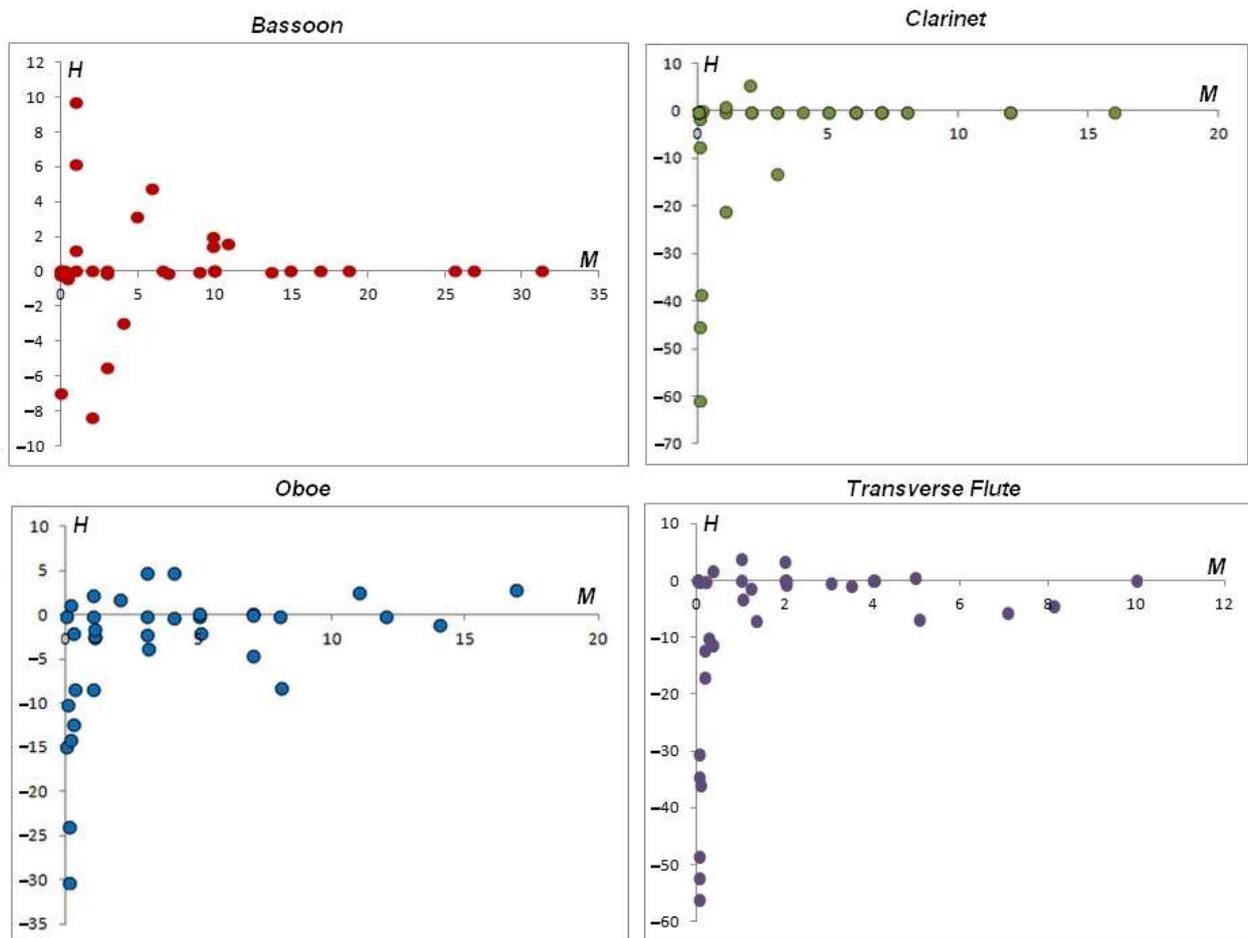


Figure 17. Two-dimensional representation in the $H-M$ plane of the timbral space of the coefficients for the selected aerophones.

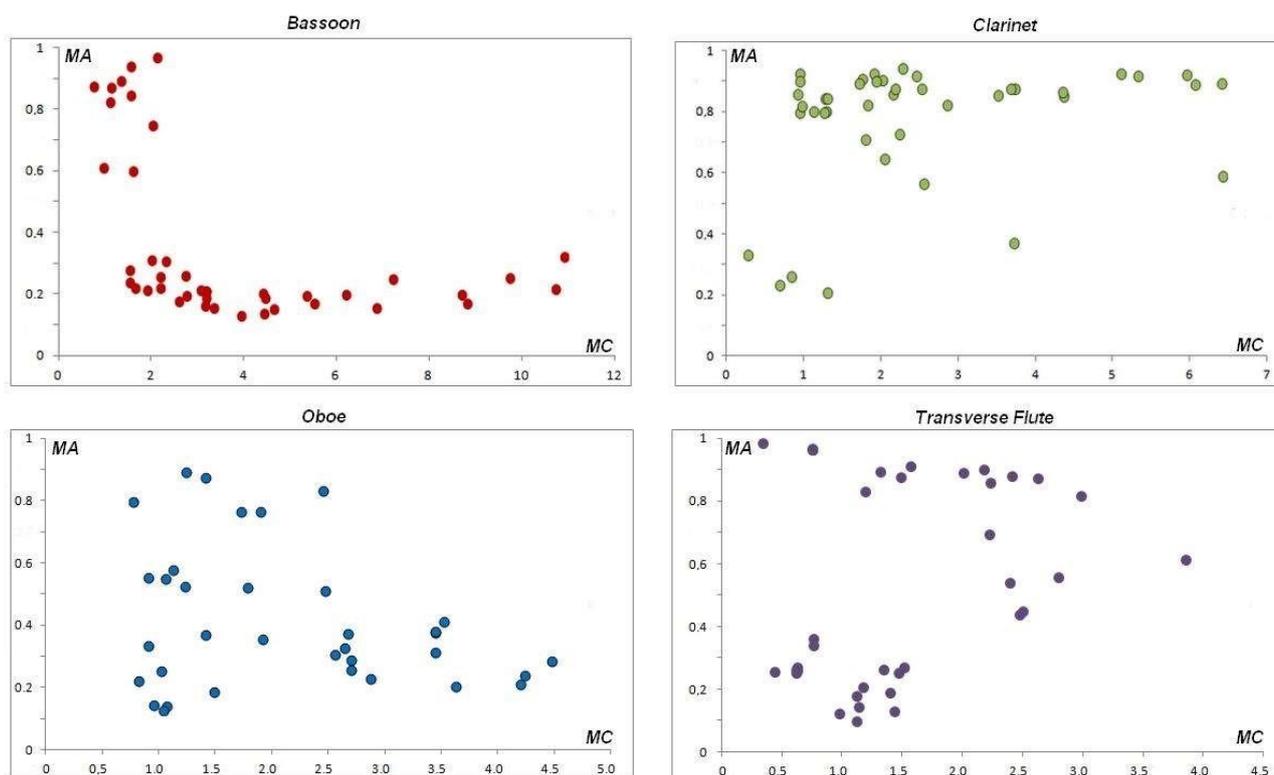


Figure 18. Two-dimensional representation in the MA–MC plane of the timbral space of coefficients for the selected aerophones.

5. Conclusions

Given the discrete nature of the frequency spectrum of musical sounds, it is possible to construct dimensionless statistical descriptors, which relate the timbral characteristics of the fundamental frequency and the harmonic distribution in the FFTs of simple monophonic musical registers for wooden aerophones. Thus, the proposed timbral descriptors, Affinity and Sharpness, allow distinguishing the timbre of Clarinet, Bassoon, Transverse Flute, and Oboe.

The coefficients of Harmonicity, Monotony, Mean Affinity, and Mean Contrast are sufficient to describe the distribution of harmonics in the FFTs of the aerophones studied, observing distinctions and variabilities between the different sounds of the musical scale for the same instrument.

The Spectral Signature of the wooden aerophones, which allows the identification of the timbre for the instruments studied, can be obtained through the FFT by the distribution of harmonics present outside the loudness range of the common tessitura (B3–D#5). The number of harmonics present and their succession does not always characterize the timbre between different aerophones, and there may be degeneration (two or more instruments with equal harmonics); however, with the proposed coefficients, the musical timbre, by monophonic musical sound, is uniquely identified in the studied aerophones.

The FFT coefficients proposed in this work allow analyzing digital audio records of musical sounds, generating a unique spectrum that characterizes them. This spectrum is a finite, bounded, and countable collection of discrete frequencies with different amplitudes. We verify that, with the six proposed coefficients, we are able to discriminate between timbre differences and similarities for the set of test audios.

Their visualization in two-dimensional arrangements reveals some grouping patterns for specific instruments. We believe that this approach can be further extended to use multidimensional unsupervised that can be analyzed using Machine Learning techniques, such as clustering. Therefore, the characterization of the musical timbre is susceptible to being extended and automated.

The proposed timbral coefficients allow us to reduce the problem of the timbral characterization of musical instruments to a grouping problem in an abstract space of six dimensions for monophonic musical sounds. Therefore, in future work, we plan to perform a correlation analysis with the proposed coefficients, in addition to using unsupervised machine learning techniques.

To sum up, it is concluded that the proposed descriptors sufficiently describe the timbral characteristics in the aerophones studied and differentiate them, allowing their recognition and classification from the FFTs of the audio records. Finally, it is important to highlight that the proposed timbral coefficients can be extended to any of the melodic musical instruments, since the used methods are indifferent to the instruments studied.

Author Contributions: Conceptualization, Y.G. and R.C.P.; methodology, Y.G. and R.C.P.; software, Y.G. and R.C.P.; validation, Y.G. and R.C.P.; formal analysis, Y.G. and R.C.P.; investigation, Y.G. and R.C.P.; resources, Y.G. and R.C.P.; data curation, Y.G. and R.C.P.; writing—original draft preparation, Y.G.; writing—review and editing, Y.G. and R.C.P.; visualization, Y.G.; supervision, R.C.P.; project administration, Y.G. and R.C.P.; funding acquisition, Y.G. and R.C.P.; All authors have read and agreed to the published version of the manuscript.

Funding: This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The sounds used in this work are available at the following link: <https://zenodo.org/record/3685367#.XnFp5i2h1IU%22>, accessed date 6 June 2021).

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Timbral coefficients for the tessitura of the Aerophones.

Table A1. Bassoon (Bn).

Sound (Hz)	Sound	f_0 (Hz)	Centroid (Hz)	A	S	H	M	MA	MC
58.27	A#1	58.24	504.01	8.65	0.05	0.30	−0.018	10.91	0.318
61.74	B1	61.81	546.07	8.83	0.02	31.29	−0.012	10.73	0.214
65.41	C2	65.37	514.66	7.87	0.04	0.41	−0.222	9.75	0.250
69.30	C#2	69.25	535.29	7.73	0.04	0.48	−0.427	7.22	0.245
73.42	D2	73.44	538.17	7.33	0.06	26.90	−0.011	8.83	0.166
77.78	D#2	77.90	521.51	6.69	0.04	25.65	−0.013	8.71	0.196
82.41	E2	82.34	497.30	6.04	0.03	0.23	−0.330	4.41	0.199
87.31	F2	87.40	542.59	6.21	0.05	18.80	−0.020	6.21	0.195
92.50	F#2	92.46	606.37	6.56	0.04	0.10	−0.002	6.86	0.153
98.00	G2	98.02	558.48	5.70	0.04	14.96	−0.005	5.36	0.192
103.83	G#2	103.81	582.73	5.61	0.03	1.01	0.001	3.94	0.126
110.00	A2	110.02	605.65	5.50	0.04	9.98	−0.001	4.43	0.135
116.54	A#2	116.60	606.79	5.20	0.03	16.91	−0.004	5.52	0.165
123.47	B2	123.45	692.23	5.61	0.02	3.02	−0.003	4.47	0.184
130.81	C3	130.81	663.11	5.07	0.02	0	−0.003	4.64	0.149
138.59	C#3	138.68	611.69	4.41	0.02	9.96	−0.011	3.16	0.160
146.83	D3	146.88	625.99	4.26	0.03	9.98	−0.012	3.20	0.205
155.56	D#3	155.53	617.10	3.97	0.04	0.95	1.200	3.18	0.205
164.81	E3	164.90	519.32	3.15	0.07	10.91	1.545	3.20	0.184
174.61	F3	174.63	571.59	3.27	0.10	9.93	1.399	2.75	0.258
185.00	F#3	184.97	631.07	3.41	0.03	0.01	−0.028	3.36	0.150
196.00	G3	196.00	599.45	3.06	0.05	9.00	−0.076	3.08	0.210

Table A1. *Cont.*

Sound (Hz)	Sound	f_0 (Hz)	Centroid (Hz)	A	S	H	M	MA	MC
207.65	G#3	207.71	669.87	3.23	0.04	9.94	1.949	2.76	0.190
220.00	A3	220.08	604.87	2.75	0.07	4.04	-3.006	2.19	0.218
233.08	A#3	233.16	598.01	2.56	0.06	6.99	-0.140	2.59	0.173
246.94	B3	246.89	563.73	2.28	0.07	0	-0.191	1.91	0.211
261.63	C4	261.64	633.21	2.42	0.18	4.96	3.141	2.31	0.302
277.18	C#4	277.12	560.82	2.02	0.12	0	-0.126	2.00	0.307
293.66	D4	293.66	636.99	2.17	0.13	3.00	-0.165	2.21	0.254
311.13	D#4	311.19	745.33	2.40	0.10	5.96	4.712	1.66	0.215
329.63	E4	329.62	703.92	2.14	0.11	0.02	-6.987	1.55	0.236
349.23	F4	349.23	788.65	2.26	0.14	0.97	6.084	1.54	0.276
369.99	F#4	369.96	629.77	1.70	0.40	0.98	9.711	0.97	0.606
392.00	G4	391.96	876.69	2.24	0.40	13.73	-0.091	1.57	0.937
415.30	G#4	415.34	879.21	2.12	0.48	6.63	0.031	1.34	0.890
440.00	A4	439.89	742.19	1.69	0.45	0	-0.197	2.04	0.743
466.16	A#4	466.06	774.81	1.66	0.41	0	-0.247	1.60	0.598
493.88	B4	493.93	704.56	1.43	0.65	3.02	-5.568	1.15	0.868
523.25	C5	523.32	767.90	1.47	0.58	2.02	-8.428	1.12	0.822
554.37	C#5	554.29	708.98	1.28	0.72	2.02	0.002	0.75	0.872
587.33	D5	587.28	816.11	1.39	0.68	3.00	-0.120	1.56	0.843
622.25	D#5	622.19	855.16	1.37	0.83	0.01	0.002	2.12	0.965

Table A2. Clarinet (ClBb).

Sound (Hz)	Sound	f_0 (Hz)	Centroid (Hz)	A	S	H	M	MA	MC
146.83	D3	146.90	667.22	4.54	0.15	11.96	-0.009	3.71	0.371
155.56	D#3	155.52	807.57	5.19	0.19	0.06	-0.007	6.41	0.590
164.81	E3	164.79	855.38	5.19	0.33	7.01	-0.002	6.40	0.895
174.61	F3	174.91	872.03	4.99	0.34	15.97	-0.004	6.06	0.890
185.00	F#3	184.96	723.79	3.91	0.45	0.02	-0.002	5.31	0.918
196.00	G3	196.01	606.96	3.10	0.53	6.00	-0.004	5.10	0.925
207.65	G#3	207.64	708.66	3.41	0.48	2.00	0	5.95	0.923
220.00	A3	220.06	736.50	3.35	0.36	11.97	0	4.35	0.851
233.08	A#3	233.01	943.53	4.05	0.36	0.03	0.001	4.34	0.865
246.94	B3	246.96	761.78	3.08	0.50	7.99	-0.038	3.73	0.876
261.63	C4	261.63	647.42	2.47	0.54	7.00	0.001	2.51	0.876
277.18	C#4	277.25	646.10	2.33	0.54	6.00	0.006	2.14	0.860
293.66	D4	293.62	824.39	2.81	0.45	0.01	-0.003	3.67	0.877
311.13	D#4	311.15	961.71	3.09	0.41	7.99	-0.004	3.50	0.856
329.63	E4	329.67	950.00	2.88	0.41	6.99	-0.001	2.84	0.822
349.23	F4	349.19	604.04	1.73	0.68	0	-0.002	2.01	0.905
369.99	F#4	369.98	964.48	2.61	0.36	5.00	0.001	2.03	0.646
392.00	G4	391.98	1013.86	2.59	0.35	1.98	5.655	2.23	0.728
415.30	G#4	415.35	736.15	1.77	0.68	6.00	-0.002	2.45	0.920
440.00	A4	439.97	654.32	1.49	0.68	0	-0.076	1.26	0.843
466.16	A#4	466.15	697.09	1.50	0.77	0	-0.007	1.90	0.924
493.88	B4	494.00	1015.80	2.06	0.53	4.99	-0.021	1.81	0.823
523.25	C5	523.20	971.92	1.86	0.67	3.00	0.002	1.93	0.901
554.37	C#5	554.27	1200.24	2.17	0.53	0.02	-1.662	2.17	0.875
587.33	D5	587.17	1625.55	2.77	0.22	0.03	-7.326	2.54	0.566
622.25	D#5	622.21	1321.24	2.12	0.41	0	-0.089	1.79	0.709
659.26	E5	659.32	943.24	1.43	0.73	3.00	-0.051	1.74	0.909
698.46	F5	698.49	944.36	1.35	0.78	4.00	-0.036	2.27	0.943
739.99	F#5	739.97	1103.77	1.49	0.70	1.00	-0.056	1.71	0.892
783.99	G5	784.01	1116.56	1.42	0.68	2.00	-0.102	1.29	0.845
830.61	G#5	830.60	1218.21	1.47	0.56	1.01	-20.909	1.12	0.801

Table A2. Cont.

Sound (Hz)	Sound	f_0 (Hz)	Centroid (Hz)	A	S	H	M	MA	MC
880.00	A5	880.04	1504.44	1.71	0.50	3.02	-13.201	0.93	0.799
932.33	A#5	932.27	1336.47	1.43	0.63	0	-0.166	1.28	0.802
987.77	B5	980.69	1427.35	1.46	0.01	0.14	0.148	1.29	0.210
1046.50	C6	1046.51	1615.13	1.54	0.52	1.00	1.061	0.97	0.819
1108.73	C#6	1108.68	1659.61	1.50	0.62	0	-0.135	1.25	0.800
1174.66	D6	1174.64	1386.54	1.18	0.87	0	-0.022	0.94	0.924
1244.51	D#6	1236.23	1464.12	1.18	0.01	0.07	-38.571	0.67	0.232
1318.51	E6	1318.41	1587.83	1.20	0.83	0	-0.053	0.93	0.900
1396.91	F6	1389.25	1533.03	1.10	0.01	0.05	-45.157	0.82	0.262
1479.98	F#6	1479.89	1864.28	1.26	0.78	0	-0.095	0.91	0.857
1567.98	G6	1559.80	1633.74	1.05	0.02	0.03	-60.715	0.27	0.332

Table A3. Oboe (Ob).

Sound (Hz)	Sound	f_0 (Hz)	Centroid (Hz)	A	S	H	M	MA	MC
233.08	A#3	233.08	1235.40	5.30	0.07	3.03	-2.085	4.47	0.288
246.94	B3	246.91	1187.76	4.81	0.10	0.98	2.374	3.44	0.317
261.63	C4	261.69	1428.42	5.46	0.07	16.90	2.964	4.23	0.242
277.18	C#4	277.28	1249.26	4.51	0.08	14.00	-1.018	3.44	0.380
293.66	D4	293.72	1319.32	4.49	0.09	11.98	-0.018	4.19	0.213
311.13	D#4	311.11	1308.45	4.21	0.16	5.00	-0.015	3.44	0.383
329.63	E4	329.60	1275.10	3.87	0.10	2.99	-0.049	3.51	0.413
349.23	F4	349.16	1444.03	4.14	0.06	0.01	-0.027	3.62	0.205
369.99	F#4	369.97	1451.34	3.92	0.08	3.05	-3.653	2.64	0.330
392.00	G4	391.96	1459.30	3.72	0.08	1.03	-2.354	2.69	0.258
415.30	G#4	415.32	1735.33	4.18	0.08	10.98	2.616	2.86	0.230
440.00	A4	440.06	1470.57	3.34	0.09	8.04	-8.114	2.55	0.309
466.16	A#4	466.15	1412.77	3.03	0.13	0.04	-9.982	2.66	0.374
493.88	B4	493.89	1391.55	2.82	0.16	5.00	0.322	1.91	0.359
523.25	C5	523.29	1543.43	2.95	0.12	8.00	-0.123	2.69	0.290
554.37	C#5	554.29	1421.10	2.56	0.22	1.00	-0.118	2.46	0.513
587.33	D5	587.28	1452.77	2.47	0.35	1.02	-2.225	2.44	0.834
622.25	D#5	622.19	1195.44	1.92	0.33	0.01	-14.883	1.22	0.529
659.26	E5	659.25	1278.22	1.94	0.30	2.00	1.814	1.05	0.553
698.46	F5	698.50	1409.19	2.02	0.38	7.00	0.266	1.89	0.768
739.99	F#5	740.02	1494.08	2.02	0.31	7.00	0.153	1.78	0.524
783.99	G5	783.99	1540.35	1.96	0.23	4.00	-0.145	1.40	0.371
830.61	G#5	830.63	1641.41	1.98	0.37	7.01	-4.476	1.72	0.766
880.00	A5	871.65	1388.28	1.59	0.01	0.28	-8.262	1.48	0.187
932.33	A#5	932.27	1494.20	1.60	0.50	1.01	-8.302	0.77	0.799
987.77	B5	977.48	1582.06	1.62	0.01	0.15	1.146	0.90	0.336
1046.50	C6	1046.55	1713.49	1.64	0.54	5.01	-1.974	1.40	0.876
1108.73	C#6	1108.78	1894.03	1.71	0.36	4.01	4.864	1.12	0.581
1174.66	D6	1174.60	1862.60	1.59	0.55	1.02	-1.488	1.23	0.896
1244.51	D#6	1235.06	1741.56	1.41	0.01	0.09	-30.128	1.01	0.255
1318.51	E6	1308.77	1936.83	1.48	0.03	0.16	-13.947	1.06	0.142
1396.91	F6	1396.94	2425.15	1.74	0.35	2.99	4.948	0.90	0.554
1479.98	F#6	1471.43	2780.54	1.89	0.02	0.26	-12.310	1.03	0.129
1567.98	G6	1552.42	2625.20	1.69	0.01	0.23	-1.907	0.95	0.147
1661.22	G#6	1651.26	2645.09	1.60	0.01	0.11	-23.845	0.81	0.224

Table A4. Transverse Flute (Fl).

Sound (Hz)	Sound	f_0 (Hz)	Centroid (Hz)	A	S	H	M	MA	MC
246.94	B3	247.00	656.41	2.66	0.30	9.98	−0.098	3.85	0.614
261.63	C4	261.75	638.71	2.44	0.41	0.01	−0.095	2.98	0.819
277.18	C#4	277.31	855.27	3.08	0.20	8.09	−4.423	2.47	0.439
293.66	D4	293.62	814.33	2.77	0.27	0.01	−0.123	2.79	0.558
311.13	D#4	311.17	862.68	2.77	0.26	7.04	−5.654	2.39	0.542
329.63	E4	329.87	918.06	2.78	0.32	4.02	0.013	2.62	0.874
349.23	F4	349.22	919.00	2.63	0.29	2.03	−0.665	2.22	0.696
369.99	F#4	370.02	996.03	2.69	0.21	5.04	−6.881	2.50	0.449
392.00	G4	391.93	1036.91	2.65	0.42	1.03	−3.377	2.23	0.860
415.30	G#4	415.33	1012.08	2.44	0.51	3.04	−0.560	2.18	0.903
440.00	A4	440.01	858.62	1.95	0.54	4.97	0.545	2.41	0.880
466.16	A#4	456.90	760.75	1.67	0.01	0.35	1.732	1.47	0.253
493.88	B4	493.95	858.77	1.74	0.65	4.00	−0.034	2.01	0.892
523.25	C5	523.24	918.83	1.76	0.62	0.99	3.765	1.49	0.879
554.37	C#5	546.30	962.23	1.76	0.01	0.25	−10.220	1.34	0.265
587.33	D5	587.32	954.20	1.62	0.66	0	−0.025	1.19	0.831
622.25	D#5	622.29	845.94	1.36	0.79	2.00	−0.010	1.57	0.911
659.26	E5	651.19	1006.10	1.55	0.01	0.19	−0.195	1.51	0.269
698.46	F5	698.53	1019.65	1.46	0.71	1.99	3.226	1.32	0.896
739.99	F#5	731.94	1238.05	1.69	0.02	1.35	−7.199	1.43	0.129
783.99	G5	781.35	1303.96	1.67	0.03	1.22	−1.479	1.14	0.143
830.61	G#5	821.02	1039.68	1.27	0.02	0.16	−12.339	1.12	0.180
880.00	A5	871.60	1120.36	1.29	0.02	0.16	−17.141	1.40	0.190
932.33	A#5	924.89	1064.71	1.15	0.01	0.05	−0.316	0.76	0.361
987.77	B5	987.83	1112.08	1.13	0.90	0.01	0.011	0.75	0.965
1046.50	C6	1038.96	1192.67	1.15	0.02	0.06	−30.581	0.63	0.270
1108.73	C#6	1100.58	1229.78	1.12	0.02	0.06	−34.671	0.63	0.259
1174.66	D6	1174.68	1303.52	1.11	0.91	1.01	0.008	0.75	0.968
1244.51	D#6	1244.57	2150.82	1.79	0.01	3.49	−0.914	0.98	0.123
1318.51	E6	1307.60	1377.56	1.11	0.90	2.01	0.004	0.75	0.965
1396.91	F6	1388.37	1406.13	1.08	0.06	0.05	−0.323	0.76	0.340
1479.98	F#6	1479.98	1615.88	1.16	0.02	0.08	−36.079	1.17	0.206
1567.98	G6	1560.07	1501.70	1.01	0.97	0.01	0.001	0.34	0.986
1661.22	G#6	1652.56	1675.73	1.07	0.01	0.04	−48.602	0.62	0.254
1760.00	A6	1752.18	1784.12	1.08	0.02	0.04	−52.271	0.44	0.258
1864.66	A#6	1856.44	1921.76	1.10	0.01	0.04	−56.204	0.62	0.264
1975.53	B6	1975.52	2392.19	1.29	0.01	0.37	−11.401	1.12	0.100

References

- Lartillot, O.; Toivianen, P.; Eerola, T. A Matlab Toolbox for music information retrieval. In *Data Analysis, Machine Learning and Applications*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 261–268.
- Li, H.; You, H.; Fei, X.; Yang, M.; Chao, K.M.; He, C. Automatic Note Recognition and Generation of MDL and MML using FFT. In *2018 IEEE 15th International Conference on e-Business Engineering (ICEBE)*; IEEE: Piscataway, NJ, USA, 2018; pp. 195–200.
- Nagawade, M.S.; Ratnaparkhe, V.R. Musical instrument identification using MFCC. In *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*; IEEE: Piscataway, NJ, USA, 2017; pp. 2198–2202.
- Chakraborty, S.S.; Parekh, R. Improved musical instrument classification using cepstral coefficients and neural networks. In *Methodologies and Application Issues of Contemporary Computing Framework*; Springer: Singapore, 2018; pp. 123–138.
- Łętowski, T. Timbre, tone color, and sound quality: Concepts and definitions. *Arch. Acoust.* **2014**, *17*, 17–30.
- Adeli, M.; Rouat, J.; Wood, S.; Molotchnikoff, S.; Plourde, E. A Flexible Bio-Inspired Hierarchical Model for Analyzing Musical Timbre. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2016**, *24*, 875–889. [[CrossRef](#)]
- Alías, F.; Socoró, J.C.; Sevillano, X. A Review of Physical and Perceptual Feature Extraction Techniques for Speech, Music and Environmental Sounds. *Appl. Sci.* **2016**, *6*, 143. [[CrossRef](#)]
- Benetos, E.; Dixon, S.; Duan, Z.; Ewert, S. Automatic Music Transcription: An Overview. *IEEE Signal Process. Mag.* **2018**, *36*, 20–30. [[CrossRef](#)]
- Hernandez-Olivan, C.; Pinilla, I.Z.; Hernandez-Lopez, C.; Beltran, J. A Comparison of Deep Learning Methods for Timbre Analysis in Polyphonic Automatic Music Transcription. *Electronics* **2021**, *10*, 810. [[CrossRef](#)]

10. Jiang, W.; Liu, J.; Zhang, X.; Wang, S.; Jiang, Y. Analysis and Modeling of Timbre Perception Features in Musical Sounds. *Appl. Sci.* **2020**, *10*, 789. [[CrossRef](#)]
11. Guven, E.; Ozbayoglu, A.M. Note and Timbre Classification by Local Features of Spectrogram. *Procedia Comput. Sci.* **2012**, *12*, 182–187. [[CrossRef](#)]
12. Fourer, D.; Rouas, J.L.; Hanna, P.; Robine, M. Automatic timbre classification of ethnomusicological audio recordings. In Proceedings of the International Society for Music Information Retrieval Conference (ISMIR 2014), Taipei, Taiwan, 27–31 October 2013.
13. McAdams, S. The perceptual representation of timbre. In *Timbre: Acoustics, Perception, and Cognition*; Springer: Cham, Switzerland, 2019; pp. 23–57.
14. Liu, J.; Zhao, A.; Wang, S.; Li, Y.; Ren, H. Research on the Correlation Between the Timbre Attributes of Musical Sound and Visual Color. *IEEE Access* **2021**, *9*, 97855–97877. [[CrossRef](#)]
15. Reymore, L.; Huron, D. Using auditory imagery tasks to map the cognitive linguistic dimensions of musical instrument timbre qualia. *Psychomusicol. Music Mind Brain* **2020**, *30*, 124–144. [[CrossRef](#)]
16. Reymore, L. Characterizing prototypical musical instrument timbres with Timbre Trait Profiles. *Music. Sci.* **2021**. [[CrossRef](#)]
17. Barbedo, J.G.A.; Tzanetakis, G. Musical Instrument Classification Using Individual Partial. *IEEE Trans. Audio Speech Lang. Process.* **2010**, *19*, 111–122. [[CrossRef](#)]
18. Joshi, S.; Chitre, A. Identification of Indian musical instruments by feature analysis with different classifiers. In Proceedings of the Sixth International Conference on Computer and Communication Technology; IEEE: Piscataway, NJ, USA, 2015; pp. 110–114.
19. Ezzaidi, H.; Bahoura, M.; Hall, G.E. Towards a Characterization of Musical Timbre Based on Chroma Contours. *Robotics* **2012**, *322*, 162–171. [[CrossRef](#)]
20. Böck, S.; Korzeniowski, F.; Schlüter, J.; Krebs, F.; Widmer, G. Madmom: A new python audio and music signal processing library. In Proceedings of the 24th ACM International Conference on Multimedia; IEEE: Piscataway, NJ, USA, 2016; pp. 1174–1178.
21. McFee, B.; Raffel, C.; Liang, D.; Ellis, D.P.; McVicar, M.; Battenberg, E.; Nieto, O. Librosa: Audio and music signal analysis in python. In Proceedings of the 14th Python in Science Conference, Piscataway, NJ, USA; 2015.
22. Peeters, G.; Giordano, B.L.; Susini, P.; Misdariis, N.; McAdams, S. The Timbre Toolbox: Extracting audio descriptors from musical signals. *J. Acoust. Soc. Am.* **2011**, *130*, 2902–2916. [[CrossRef](#)] [[PubMed](#)]
23. Elliott, T.M.; Hamilton, L.S.; Theunissen, F.E. Acoustic structure of the five perceptual dimensions of timbre in orchestral instrument tones. *J. Acoust. Soc. Am.* **2013**, *133*, 389–404. [[CrossRef](#)] [[PubMed](#)]
24. Cella, C.E.; Ghisi, D.; LOSTANLEN, V.; Lévy, F.; Fineberg, J.; Maresz, Y. 2020. OrchideaSOL: A Dataset of Extended Instrumental Techniques for Computer-Aided Orchestration. *arXiv* **2020**, arXiv:2007.00763. [[CrossRef](#)]
25. Adler, S.; Hesterman, P. *The Study of Orchestration*; WW Norton: New York, NY, USA, 1989.
26. Pons, J.; Slizovskaia, O.; Gong, R.; Gómez, E.; Serra, X. Timbre analysis of music audio signals with convolutional neural networks. In Proceedings of the 2017 25th European Signal Processing Conference (EUSIPCO), Kos, Greece, 28 August–2 September 2017; pp. 2744–2748. [[CrossRef](#)]
27. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0 Contributors. SciPy 1.0 Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [[CrossRef](#)] [[PubMed](#)]