Representation and Classification of the Timbre Space of a Single Musical Instrument

Hugo B. de Paula, Mauricio A. Loureiro, Hani C. Yehia

CEFALA – Center for Research on Speech, Acoustics Language and Music
UFMG - Federal University of Minas Gerais - Brazil
hugobp@cefala.org; mauriciololoureiro@ufmg.br; hani@cefala.org; http://www.cefala.org/

Abstract

In order to map the spectral characteristics of the great variety of sounds a musical instrument may produce, different notes were performed and sampled in several intensity levels across the whole extension of a clarinet. Amplitude and frequency time-varying curves of partials were measured by Discrete Fourier Transform. A limited set of orthogonal spectral bases was derived by Principal Component Analysis techniques. These bases defined spectral sub-spaces capable of representing all tested sounds, which were validated by auditory tests. Sub-spaces involving larger groups of notes were used to compare the sounds according to the distance metrics of the representation. A clustering algorithm was used to infer timbre classes. Preliminary tests with resynthesized sounds with normalized pitch showed a clear relation between the perceived timbre and the cluster label to which the notes were assigned.

1. Introduction

Representation of a musical instrument involves the estimation of the physical parameters that contribute to the perception of pitch, intensity level and timbre of all sounds the instrument is capable of producing. Of these attributes, timbre poses the greatest challenges to the measurement and specification of the parameters involved in its perception, due to its inherently multidimensional nature. Intensity and pitch time-varying levels can be classified according to soft/loud and low/high one-dimensional scales and are, hence, capable of being quantitatively expressed by the traditional music-notation system. On the other hand, timbre is not so easily scaled. It is perceived by means of the interaction of a variety of static and dynamic properties of sound grouped into a complex set of auditory attributes. Due to the multidimensionality of this attribute, the identification of the contribution of each one of these competitive factors has been the main subject of psychoacoustics research on timbre perception.

The introduction of the notion of "similarity rate" of hearing judgment responses together with *Multidimensional Scaling* (MDS) techniques allowed the reduction of this dimensionality and made it possible to investigate the complex structure of this attribute, which motivated the first studies on musical timbre upon perceptive data [1] and [2]. In one of the most classic studies on musical timbre, Grey [3] measured subjective judgment of similarity between pairs of timbres from 16 different musical instruments, submitted them to an MDS and built a three-dimensional *Timbre Space*, in which multidimensional "timbre values" of different instruments were positioned according to their similarity/ dissimilarity. Other than mapping geometrically the concept of acoustic similarity, that study also showed the capability of the method for providing a psychological quantification of a relatively complex structure upon quite simple data – similarity/dissimilarity responses between pairs of distinct timbres.

More recent studies were able to relate measurable physical parameters with the dimensions shared by the timbre represented in these spaces, combining quantitative models of perceptive relationships with psychophysical explanations of the identified parameters [4] and [5]. The possibility of establishing correlations between purely perceptive factors related to timbre and acoustic measurements extracted directly from sound, directed research on musical timbre towards more quantitative approaches. A historical review of the development of research on musical timbre is found in [6].

A technique commonly used in research on musical timbre is Principal Component Analysis (PCA), which also builds multidimensional data representation. However, while MDS representation relates built-in variables in data obtained from similarity judgment, PCA manipulates the variance of measured acoustic data. Recent works applying PCA to time-varying amplitude and frequency curves of harmonic components have produced similar results with similar sets of sounds [7], [8], [9], [10], [11], [12] and [13].

The above mentioned studies have approached comparisons among isolated notes of different musical instruments outside any musical context, focusing on the perceptive mechanism that discriminates a musical instrument from another. Little has been achieved regarding perceptive discrimination within the timbre palette produced by a single musical instrument, or even along the extent of a single note. Focused on the timbre of a single instrument, this study investigates methods for representing the variety of sonorities produced by one single musical instrument, sharing the same questions raised by recent research that investigates the contribution of acoustic parameters to the conveyance and perception of musical expressiveness.

2. Timbre set specification

The purpose of this study is to represent the timbre of a musical instrument upon spectral parameters extracted from samples of sounds performed on that instrument. An adequate set of sounds for such a representation should include as many as possible different timbres, performed along the instrument entire pitch range. Two major simplifications were considered in defining the timbre set used in this study: (i) it was limited to the sound palette commonly produced on musical instruments in traditional western music performance, excluding sonorities produced on the instrument on the context of other musical traditions, as well as those

regularly used in contemporary music known as "extended techniques" [14]; (ii) in order to facilitate the estimation of spectral parameters, only the sustained part of relatively long sounds was considered, excluding attack, decay and transitions between consecutive notes. Due to dependence of timbre on these parts, the second simplification limits the investigation to the perception of slow variation of musical timbre, which commonly happens along longer notes during a musical performance.

Intentional variations of timbre, together with fluctuations of intensity and duration are commonly used by the player, in order to convey his or her expressive intentions. Although timbre may vary independently from intensity and duration, its dependence on intensity is evident. This high level of correlation facilitates the sampling of different timbre "values" of the same note upon specification of intensity levels. Thus, four different timbres were sampled for each note by asking the player to perform each note in four different intensity levels, with minimal variation. Four levels were defined: *pianissimo* (*pp*), *mezzo-piano* (*mp*), *mezzo-forte* (*mf*) and *fortissimo* (*ff*). The performer was asked to establish the lowest and highest level limits as softer and louder as possible, respectively, within the range of commonly used timbres on western classical music. Intermediate levels were to be defined by comparison with the lowest and highest limits. Samples were obtained through high quality recordings of all notes of the two lowest registers of a B flat clarinet, ranging from D3 (147 Hz) to A5 (880 Hz), played at the four levels of intensity defined above, with an average duration of 3 seconds.

3. Principal Component spectral bases

3.1. Spectral parameter estimation

The amplitude curves of the harmonic components were estimated according to McAulay and Quatieri's method, which searches for maximum amplitude values ("peak detection") of a Fourier Transform and establishes a correspondence between the closest peak values in adjacent analysis frames ("peak continuation"), associating these values to instantaneous frequency and amplitude values of harmonic components [15, 16]. In order to reduce the complexity of the data, some simplifications were considered: (i) all sampled sounds can be represented by a weighted sum of sinusoids, whose amplitude and frequency values do not vary abruptly in the course of duration; (ii) components whose intensity were more than 60 dB below the maximum level were discarded; (iii) amplitude curves were smoothed by a low pass filter with cut-off frequency of 10 Hz.

3.2. Principal Component Analysis

The high correlation of spectral parameters, presented in both the frequency and time domains, which is a common characteristic of spectral distribution of sounds of musical instruments, allowed an efficient data reduction using Principal Component Analysis (PCA) [17]. Applied to a set of multidimensional variables, PCA calculates an orthogonal basis determined by the directions of maximum variance of the analyzed data. The projections of the original data on this basis, denominated *principal components* (PCs), follow trajectories that accumulate the maximum variance of the data in a decreasing order. This allows an approximate representation of the data, using only a reduced number of dimensions. Given the estimated covariance matrix of the analyzed data $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, ...\mathbf{x}_n]$:

$$\overline{\mathbf{C}_{XX}} = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n - \overline{\boldsymbol{\mu}_\mathbf{x}})(\mathbf{x}_n - \overline{\boldsymbol{\mu}_\mathbf{x}})^t, \quad (1)$$

the orthogonal basis **U** is determined by singular value decomposition (SVD):

$$\overline{\mathbf{C}_{XX}} = \mathbf{U}\mathbf{S}\mathbf{U}^t. \quad (2)$$

The projection of **X** in PCs is given by $\mathbf{Y} = \mathbf{U}^t\mathbf{X}$ and can be further recovered by $\mathbf{X} = \mathbf{U}\mathbf{Y}$. The original spectra can be reconstructed by adding the basis, properly weighted by the amplitudes of the corresponding trajectories.

3.3. Spectral basis of a single note

At first, a set of orthogonal spectral bases associated to amplitude envelopes was calculated for each sound. These envelopes combined with the corresponding basis, were able to render the sound with great precision. After that, spectral sub-spaces were built to represent the spectral distributions of all possible sounds of a single note by calculating a spectral basis using as input data the concatenation of the four samples of this note, *pp, mp, mf* and *ff*, as defined in Section 2. Samples were normalized in amplitude and duration, with 75 time frames each, equivalent to 870 ms, taken from the center of the note. Previous studies showed that the first 5 PCs were capable of reconstructing all the sounds without any perceptible loss of characteristics of timbre [18].

Table 1 shows the cumulative variance explained by the first five Principal Components in each individual execution of the note Bb3 (233 Hz) compared to the variance obtained when PCA is applied to all executions of this note. The first component explains, alone, no less than 74% of the total variance for every isolated sound, but only 68.7% if PCA is calculated for all four sounds. A reconstruction of 99% is achieved with 3 PCs for every isolated sound, but 5 PC are needed with the PCA applied to all four sounds.

PC	*pp*	*mp*	*mf*	*ff*	*pp-mp-mf- ff*
1	87.3	96.3	76.1	74.4	68.7
2	99.8	99.7	96.2	97.2	89.5
3	99.9	99.9	99.2	98.6	94.4
4	100.0	99.9	99.6	99.5	97.2
5	100.0	100.0	99.8	99.7	99.0

Table 1*: Cumulated Variance of the first 5 PCs of Bb3 (233 Hz) in four intensity levels (columns 2 to 5) and variance of PCA applied to all executions (last column).*

3.4. Spectral basis of a group of four notes

A set of spectral bases of a group of notes was then calculated by applying PCA to the concatenation of the sounds of each note. Four contiguous notes were chosen due to the perceptive similarity of their timbres: A3 (220 Hz), Bb3 (233 Hz), B3 (247 Hz) and C4 (262 Hz). The spectral basis thus obtained constitutes a *timbre space* for these four notes, where each

sound occupies a unique position, according to its spectral configuration. Figure 1 compares the 1st, 3rd, 5th and 7th harmonics of the original Bb3 with its resynthesized version using the spectral sub-space thus calculated. Comparison between the amplitude curves of the harmonics of the original sounds and their reconstructions generated from this spectral sub-space shows that the model is effective in the representation of harmonics with larger amplitude. Also, in this case, auditory tests showed no loss of perceptible characteristics of timbre [18].

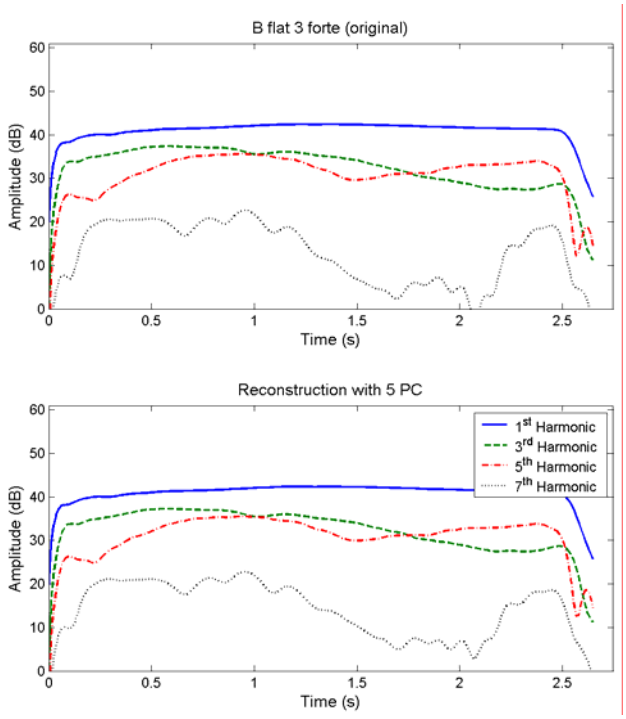

Figure 1: *1st, 3rd, 5th and 7th harmonics of the original Bb3 ff (top) and its resynthesized version (bottom) using the spectral sub-space of the notes A3 (220 Hz), Bb3 (233 Hz), B3 (247 Hz) and C4 (262 Hz) calculated with 5 principal components.*

PCs	A3 – C4 (4 notes)	Low Register D3 – Ab4 (19 notes)	Low+Mid Registers D3 – A5 (32 notes)
1	56.4	59.2	53.4
2	80.4	77.9	71.8
3	90.3	87.2	84.1
4	94.2	92.8	89.7
5	96.9	96.7	94.2

Table 2: *Cumulated Variance for different PCA bases: A3 (220 Hz) – C4 (262 Hz); D3 (147 Hz) – Ab4 (415 Hz); D3 (147 Hz) – A5 (880 Hz).*

Sets of spectral bases for larger groups of notes in all four intensity levels were then calculated, in order to represent a larger variety of spectral distributions. As expected, as the number of notes involved increases, the less efficient the representation becomes. Table 2 shows the cumulated variance of PCA bases calculated for three different groups of notes: (1) the four above mentioned contiguous notes (A3 - C4); (2) the 19 notes (76 samples) of the lowest register of the instrument, ranging from D3 (147 Hz) to Ab4 (415 Hz); (3) the two lowest registers of the instrument, from D3 (147 Hz) to A5 (880 Hz), encompassing 32 notes (128 samples).

4. The instrument physical timbre space

4.1. Trajectories

The reduction in dimensionality resulted from PCA made it possible the representation of the spectral distribution on low dimensional spaces. Figure 2 shows three-dimensional trajectories of the four sounds of note Bb3 along its own space. The correlation between intensity level and the first PC is evident as the spectral points belonging to each sound are separated in groups positioned in increasing order from *pp* to *ff* along the first PC dimension. Almost confined in their position along the first PC, the 2nd and 3rd PCs vary differently in different directions for each sound.

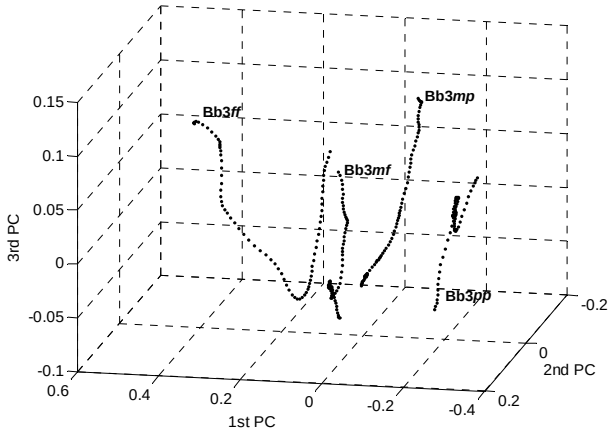

Figure 2: *Three-dimensional trajectories of the four sounds of note Bb3 in the Bb3 timbre space.*

Comparison of timbre parameters among notes of different pitch becomes more complex, as timbre may vary largely as a function of the note played (pitch), depending on the instrument. Clarinet sounds, as used in this study, present irregular variation of timbre from note to note, which can be very accentuated, depending on the region of the instrument, like the abrupt timbre change between the low and mid registers, a well known characteristic of the clarinet. Figure 3 shows the four contiguous notes above mentioned (A3, Bb3, B3 and C4), represented in the spectral space defined by them. The same correlation between intensity level and the first PC is present, as well as the grouping of all frames of a single sound. Moreover, we can identify clustering of different sounds from different notes: softer sounds (*pp* and *mp*) on the right side of the space and louder sounds on the left. We can also observe that louder sounds such as A3 *ff*, A3 *mf* and C4 *ff* have their trajectories more spread then softer sounds.

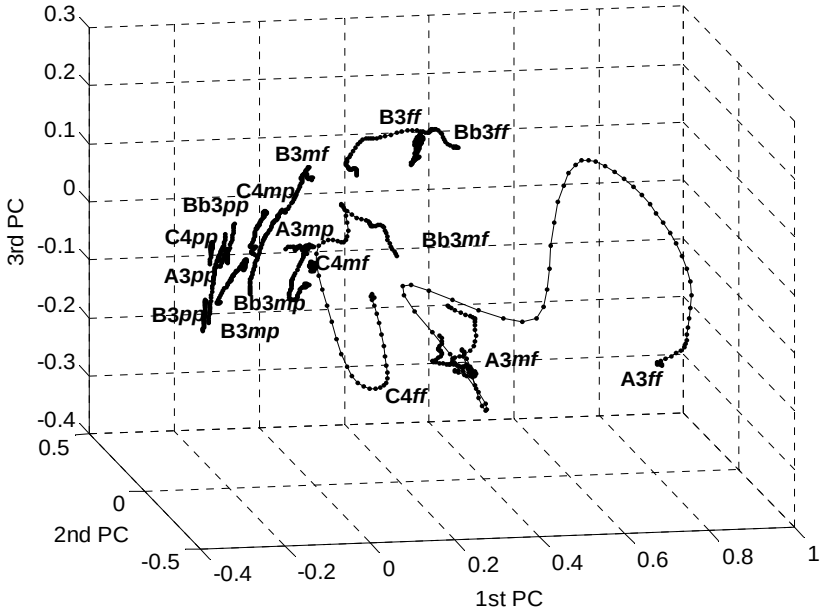

Figure 3*: Three-dimensional trajectories of all four sounds of notes A3, Bb3, B3 and C4 in the spectral space defined by them.*

4.2. Clustering

An attempt to investigate the timbre distribution along the entire instrument was made with Cluster Analysis, using the K-means algorithm [19]. In the K-means problem, given a set of N M-dimensional points $\mathbf{X_n}$, the goal is to arrange these points into K clusters, with each cluster having a representative point $\mathbf{Z_k}$, usually chosen as the centroid of the points in the cluster. Each cluster variance is defined by

$$\mathrm{ESS}(C_k) = \sum_{i \in C_k} (\mathbf{x}_i - \mathbf{c}_k)^2 , \tag{3}$$

where $\mathbf{c}_k$ contains the M-dimensional coordinates of Cluster $\boldsymbol{C_k}$, and $(\mathbf{x}_i - \mathbf{c}_k)$ can be any desired distance metric. The individual variances ESS($\boldsymbol{C_k}$) of Equation 3 is minimized in such a way that moving any single point to a different cluster increases the overall variance defined by

$$\mathrm{ESS} = \sum_{k=1}^{K} \mathrm{ESS}(C_k) . \tag{4}$$

Since the K-means can converge to a local optimum, the initial centroid values must be appropriately chosen; or the iterative algorithm must be run exhaustively with different random starting points. In the present analysis, the variance of a cluster was calculated using the Squared Euclidean Distance, although other types of distance were tested giving similar results. Every time an iteration produces an empty cluster, the algorithm creates a new cluster consisting of the points most distant from its centroid. To avoid local minima, the K-means was run 40 times and the best solution was chosen.

Initially, the 16 sounds from the four notes in Figure 3 were classified using the 75 point trajectories of the first 5 PCs as inputs. The cluster analysis distributed all 16 sounds in 6 clusters, where all the 75 spectral points of each sound lied in one single cluster. A new cluster analysis was then performed using the 19 notes (76 sounds) from the low register of the clarinet. Nine clusters provided the best correlation between preliminary auditory tests and the classification obtained for this set of sounds. Very few of these sounds had their spectral points split into different clusters and, when this happened, no more than 2 clusters were involved and the cluster assigned to the central part of the sound was always the cluster where the majority of points lied. Figure 4 shows the 11 lower notes of the clarinet (from D3 to C4), represented by the location of its central frame on the low register *timbre spac*e. The figure shows a large group of sounds clustered together close to the origin of the space (left), which includes the *pp* or *mp* version (or both) of every note of this set, except for the Eb3. Sounds *mf* and *ff* are more spread along all three dimensions, showing that intensity level differentiation spread the sounds more strongly than pitch differentiation.

This can also be observed in Figure 5, which orders all 76 sounds of the low register of the clarinet by pitch and shows the cluster to which each one was assigned. This figure is a projection of Figure 4, that highlights the correlation of the cluster with intensity level. Informal auditory tests showed a strong coupling of perceived brightness to cluster assignment. Due to the known relationship of spectral centroid to the perception of brightness, cluster labels were ordered according to the mean of the spectral centroid of the group of sounds assigned to it. Note that the first 3 clusters group almost every *pp* and *mp* sounds of the whole set. Moreover, notes of higher pitch in *mf* and *ff* were also assigned to these clusters. While higher pitched notes were grouped more tightly into these clusters, the four last clusters contain only *mf* and *ff* notes of the lower octave, except for the sounds E4 *mf* and F4 *ff*. The fact that only the amplitude spectra were used in the cluster analysis and no perceptual weighting was applied corroborated with these results.

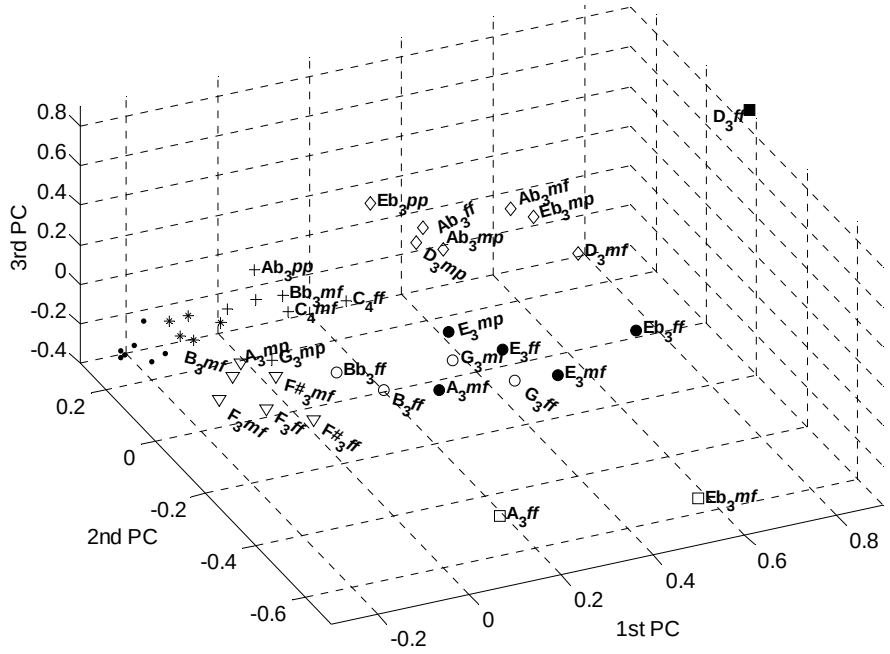

Figure 4: *Three-dimensional trajectories of the four sounds of each of the 11 lowest notes of the clarinet, from D3 to C4, in the spectral space defined by the entire low register (points not labeled on the left of the figure correspond to notes D3 pp, E3 pp, F3 pp, F3 mp, F#3 pp, F#3 mp, G3 pp, Ab3 pp, A3 pp, Bb3 pp, Bb3 mp, Bb3 mf , B3 pp, B3 mp, C4 pp and C4 mp).*

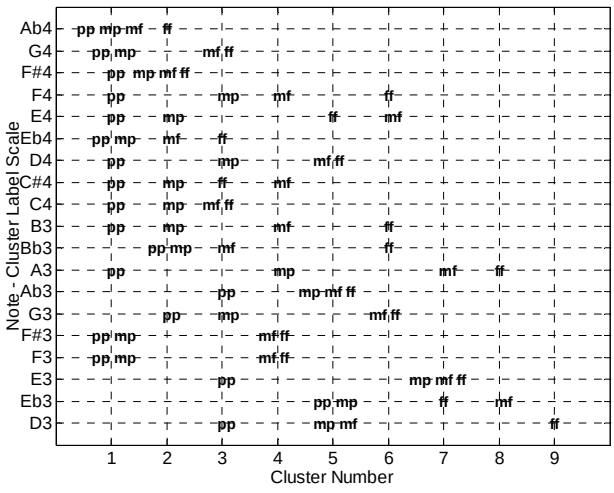

Figure 5: *Cluster Label of the 19 notes of the low register of the clarinet. Notes are ordered by pitch and cluster labels by the mean of the spectral centroids.*

Figure 6 shows the cluster trajectory of the four sounds of each of the three lowest and three highest notes of the low register. Notes are ordered by pitch and clusters are scaled onto vertical axis. It represents the individual lines of Figure 5 in 2 dimensions, highlighting that notes of lower pitch tend to have a wider variation in timbre and that timbre becomes more stable and concentrated in lower clusters as the pitch increases.

A new cluster analysis was then performed using the 33 notes (132 sounds) including the first and second registers (D3 to Bb5) of the clarinet. Twelve clusters were found to be more adequate for this sound set. Except for a few cases, the sounds of the second register (A4 to Bb5) lied in the five first clusters, together with other *pp* and *mp* notes from the lowest register. This corroborates the above observed tendency of higher notes to be clustered tighter together, reinforcing the correlation of the classification to the variation of the intensity level, as well as the diminishing of timbre variation as pitch increases.

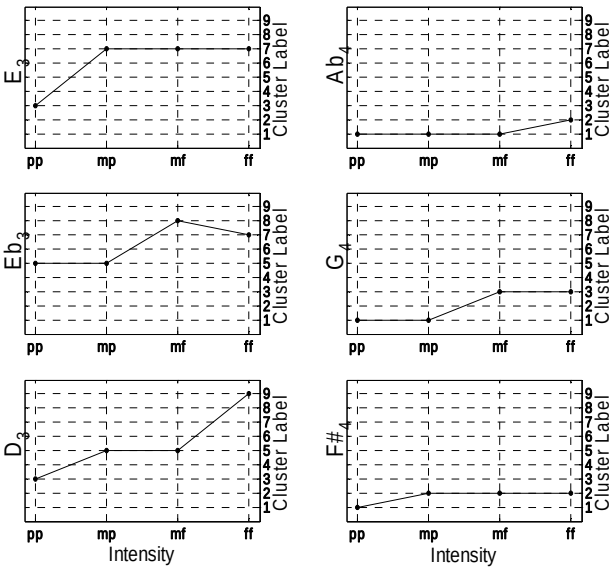

Figure 6*: Cluster trajectory of the 3 lowest and the 3 highest notes of the low register of the clarinet, D3, Db3, E3, F#4, G4, Ab4. Notes are ordered by pitch and cluster labels by the mean of the spectral centroids.*

5. Conclusion

This study carries out timbre representation of a musical instrument based on spectral parameters extracted from sounds performed on that instrument. Principal component analysis is used for dimensionality reduction, and a clustering technique is used to categorize the different timbres produced. Auditory tests of discrimination with resynthesized sounds with normalized pitch showed the effectiveness of this representation model, showing a clear relation between the perceived timbre and the cluster label to which the notes were assigned. Contiguous notes presented individual spectral bases with similar characteristics and were mapped to closer sub-spaces. This similarity allowed expansion of the size of these sub-spaces, facilitating representation of larger groups of notes. The construction of spectral sub-spaces involving all possible sounds produced by the instrument made it possible a compact representation of the whole timbre palette of the instrument. This unified representation allowed a timbre classification according to the distance metrics of the PCA *timbre space* by cluster analysis techniques, providing a descriptive comparison of the dynamic variation of timbre.

Summarizing, it could be clearly verified across all the results presented in this study that: (i) timbre classes tend to be divided as a function of spectral brightness, which is known to be correlated to intensity level in wind instruments; (ii) the lowest octave of the clarinet exhibits in general much more richness of timbre differentiation than higher pitched notes; (iii) higher notes exhibit less spectral brightness and less timbre differentiation.

The results of this study applied to wider dynamic timbre variation will facilitate the investigation of the use of intentional timbre differentiation by the performer to convey musical expressiveness. Other perspectives for this project are to extend the investigation to shorter sounds, like staccati and pizzicati, as well as attack, decay and transition between notes, for which auditory models seem to be an adequate analysis tool.

6. Acknowledgments

This work was supported in part by CAPES (Brazilian Higher Education funding agency), and by CNPq (National Council for Scientific and Technological Development), Brazil.

7. References

[1] Plomp, R., "Timbre as a Multidimensional Attribute of Complex Tones," in Frequency Analysis and Periodicity Detection in Hearing, R. Plomb and G. F. Smoorenburg, Eds. Leiden: A. W. Sijthoff, 1970.

[2] Wessel, D. L., "Timbre Space as a Musical Control Structure," in Foundations of Computer Music, C. Roads and J. Strawn, Eds. Cambridge, Massachusetts: MIT Press, 1979, pp. 640-657.

[3] Grey, J. M., "An exploration of musical timbre," CCRMA, Dept. of Music Stanford University, Stanford, Calif., Report STAN-M-2, 1975.

[4] Hajda, J. M., Kendall, R. A., Carterette, E. C. and Harshberger, M. L., "Methodological Issues in Timbre Research," in Perception and Cognition of Music, I. Deliège and J. Sloboda, Eds. Hove: Psychology Press, 1997, pp. 253-306.

[5] Misdariis, N. R., Smith, B. K., Pressnitzer, D., Susini, P. and McAdams, S., "Validation of a Multidimensional Distance Model for Perceptual Dissimilarities Among Musical Timbres," presented at Proceediongs of the 16th International Congress on Acoustics, Woodbury, New York, 1998.

[6] McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G. and Krimphoff, J., "Perceptual Scaling of Synthesized Musical Timbres: Common Dimensions, Specificities and Latent Subject Classes," Psychological Research, vol. 58, pp. 177-192, 1995.

[7] Baliello, S., De Poli, G. and Nobili, R., "The Color of Music: Spectral Characterization of Musical Sounds Filtered by a Cochlear Model," Journal of New Music Research, vol. 27, 1998.

[8] Beauchamp, J. W. and Horner, A., "Spectral Modelling and Timbre Hybridisation Programs for Computer Music," Organised Sound, vol. 2, pp. 253-258, 1998.

[9] Charbonneau, G., Hourdin, C. and Moussa, T., "A Multidimensional Scaling Analysis of Musical Instrument's Time-Varying Spectra," Computer Music Journal, vol. 21, pp. 40-55, 1997.

[10] Cosi, P., De Poli, G. and Lauzzana, G., "Auditory Modelling and Self-Organizing Neural Networks for Timbre Classification," Journal of New Music Research, vol. 23, pp. 71-98, 1994.

[11] De Poli, G. and Prandoni, P., "Sonological Models for Timbre Characterization," Journal of New Music Research, vol. 26, pp. 170-197, 1997.

[12] Rochebois, T. and Charbonneau, G., "Cross-Synthesis Using Interverted Principal Harmonic Sub-Spaces," in Music, Gestalt and Computing: Studies in Cognitive and Systematic Musicology, M. Leman, Ed. Berlin-Heidelberg: Springer Verlag, 1997, pp. 375-385.

[13] Sandell, G. J. and Martens, W., "Perceptual Evaluation of Principal-Component-Based Synthesis of Musical Timbres," Journal of The Audio Engineering Society, vol. 43, pp. 1013-1028, 1995.

[14] Bartolozzi, B., New Sound for Woodwind. London: Oxford University Press, 1967.

[15] McAulay, R. J. and Quatieri, T. F., "Speech Analysis/Synthesis Based on a Sinusoidal Representation," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 34, pp. 744-754, 1986.

[16] Serra, X., "Musical Sound Modeling with Sinusoids plus Noise," in Musical Signal Processing, A. Piccialli, C. Roads, and S. T. Pope, Eds.: Swets & Zeitlinger Publishers, 1997.

[17] Johnson, R. and Wichern, D. W., Applied Multivariate Statistical Analysis. Upper Sadlle, New Jersey, 1998.

[18] de Paula, H. B. (2000). Análise e Re-síntese de Som Natural de Clarineta Utilizando Análise por Componentes Principais. Master Dissertation, Department of Electrical Engineering, Universidade Federal de Minas Gerais, Belo Horizonte.

[19] Kaufman, L. and Rousseeuw, P. J., Finding Groups in Data: An Introduction to Cluster Analysis. New York: John Wiley & Sons, 1989.