Modeling Issues for Fine Musical Timbre Characterization

Mauricio A. Loureiro[1], Hugo B. de Paula[2], Hani C. Yehia[3]

CEFALA – Center for Research on Speech, Acoustics Language and Music

[1]Escola de Música - Universidade Federal de Minas Gerais (UFMG)
[3]Departamento de Eletrônica – Universidade Federal de Minas Gerais (UFMG)
[2]Instituto de Informática – Pontifícia Universidade Católica de Minas Gerais (PUCMINAS)

mauricioloureiro@ufmg.br, hugo@pucminas.br, hani@ufmg.br

Abstract. *In order to represent the variety of sounds a musical instrument may produce, it is necessary to find a model that can cope with sound features independent from its scale. In this work, several models of timbre characterization were applied to sample notes in several intensity levels across the whole extension of a clarinet. These models were based on amplitude and frequency time-varying curves of partials, which were measured by Discrete Fourier Transform. Principal Component Analysis techniques and Genetic Algorithms were used to define spectral sub-spaces capable of representing all tested sounds and of grouping them. The K-means clustering algorithm was used to infer timbre classes. Self-Organizing Maps lead to results similar to those obtained by PCA representation and K-means algorithms.*

Resumo. *Para se representar a variedade de sons que um instrumento musical é capaz de produzir, é necessário se utilizar modelos que podem lidar com um conjunto de parâmetros, independentemente de sua dimensão ou escala. Neste trabalho, vários modelos para caracterização do timbre foram aplicados a amostras de notas executadas em várias intensidades, cobrindo toda a extensão do clarinete. Estes modelos se basearam nas curvas de amplitude e freqüência dos parciais, obtidos através da transformada discreta de Fourier. Análise por Componentes Principais e Algoritmos Genéticos foram utilizados para se definir um subespaço espectral capaz de representar e agrupar estes sons. O algoritmos de agrupamento das K-médias foi utilizado para se determinar classes de timbre neste subespaço. Mapas Auto organizáveis produziram resultados similares aos produzidos pela PCA e K-médias.*

1. Introduction

Representation of a musical instrument involves the estimation of the physical parameters that contribute to the perception of pitches, intensity levels and timbres 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. Timbre 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.

In one of the most classic studies on musical timbre, (Grey, 1975) measured subjective judgment of similarity between pairs of timbres from 16 different musical instruments, submitted them to multidimensional scaling (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 (Hajda, Kendall *et al.*, 1997; Misdariis, Smith *et al.*, 1998). 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 (Mcadams, Winsberg *et al.*, 1995).

A technique commonly used in research on musical timbre nowadays is Principal Component Analysis (PCA). Recent works applying PCA to time-varying amplitude and frequency curves of harmonic components have produced similar results with similar sets of sounds (Cosi, De Poli *et al.*, 1994; Sandell e Martens, 1995; Charbonneau, Hourdin *et al.*, 1997; De Poli e Prandoni, 1997; Rochebois e Charbonneau, 1997; Beauchamp e Horner, 1998). Multiple Wavetable synthesis (MWS) and Genetic Algorithms have also been commonly used to build orthogonal bases similar to the PCA. However, MWS have the advantage to allow a better understanding of timbre dynamics and its relation to the acoustics of the instrument (Horner, Beauchamp *et al.*, 1993; Horner, 1995a; b; Horner e Ngai-Man, 1996).

The above mentioned studies on timbre analysis have approached comparisons among different musical instruments outside any musical context. This study investigates methods for representing the variety of sonorities produced by one single musical instrument, searching for models that are able to describe fine timbre attributes that are found in a single instrument timbre class.

1.1. Timbre modeling procedure

There are several ways of producing sound features for timbre modeling. (Tzanetakis, 2005) suggests a Music Information Retrieval (MIR) pipeline composed by the following steps: data acquisition, parameter extraction, feature estimation and information processing (classification, synthesis, and so on). The following sections will describe methods for the construction of a MIR pipeline for fine timbre characterization.

2. Timbre Set Specification

For the data acquisition step, it is important to have a data set properly defined. Since the purpose of this study is to show ways of representing 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. The timbre set used in this study was limited to the sound palette commonly produced on musical instruments in traditional classical 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" and only the sustained part of relatively long sounds was considered, excluding attack, decay and transitions between consecutive notes.

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, different timbres were sampled for each note in the following intensity levels: *pianissimo* (*pp*), *piano (p)*, *mezzo-piano* (*mp*), *mezzo-forte* (*mf*), *forte (f)* and *fortissimo* (*ff*).

3. Spectral Bases for Timbre Characterization

3.1. Spectral Parameters Estimation

The amplitude curves of the harmonic components were used as the initial parameters of the model, and were extracted using the short-time Fourier transform, according to McAulay and Quatieri's method (Mcaulay e Quatieri, 1986; Serra, 1997). In order to reduce the complexity of the data amplitude curves were smoothed by a low pass filter with cut-off frequency of 10 Hz.

3.2. Features measurement by Principal Component Analysis

The parameter space defined by the spectral analysis has as many dimensions as the number of partials extracted. The high correlation of these 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) (Johnson e Wichern, 1998). 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.

3.3. Features measurement by Multiple Wavetable Synthesis and Genetic Algorithms

Musical sounds can be efficiently synthesized using an automatic genetic algorithm (GA) that decomposes the sounds into a group of wavetables (usually 3-5). The decomposition process consists of finding the optimal group of tables that reconstructs a signal with minimum distortion. An approximation of the sound can be then constructed by Multiple Wavetable Synthesis (MWS) as a linear combination of these tables (Horner, Beauchamp *et al.*, 1993). In this work, GA was used for the feature estimation procedure using wavetable components that were sampled from spectral components of the input data itself. This was possible because it was assumed that the recorded database comprises the timbre universe of the clarinet. In this novel approach, the population of the GA is constructed from real measured spectra, instead of being randomly generated, which makes it specially useful for analysis purposes. See (De Paula, Loureiro *et al.*, 2004) for a detailed description of the procedure.

Figure 1 shows the results obtained by the use of PCA and GA for feature estimation of the Timbre space. These bases were calculated using a concatenated group of four notes (16 sounds): B3, C4, C#4 and D4, each one played in four intensity levels: *pianissimo, piano, forte* and *fortissimo*.

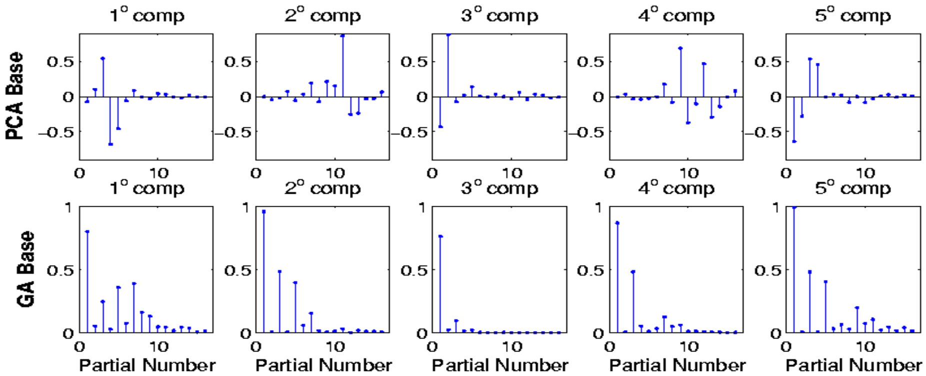

Figure 1: Comparison between the 5-dimensional PCA and GA base. The bases were calculated for the concatenated group of 4 notes (16 sounds): B3 (220 Hz), C4 (233 Hz), C#4 (247 Hz) and D4 (262 Hz).

4 The feature space of the clarinet timbre

The reduction in dimensionality resulted from PCA and GA made it possible the representation of the spectral distribution on low dimensional spaces. Figure 2 shows three-dimensional trajectories of the four notes of Figure 1. The first dimension of the PCA has a high correlation with the intensity level of the sounds and the PC value increases as the sound intensity increases. Although there is no visible correlation for the first component of the GA space, there is a clear intensity grouping and organization along the second and third dimensions of the GA. Since the spectral matching is an optimization process based on randomly generated populations, there is no real meaning in the order of the GA dimensions, and the amount of data each one explains is not ordered in the same way as the PCA. It can also be observed that the range of variation decreases as the dimension number increases in the PCA space, showing a clear hierarchy of dimensions. The same does not happen in the GA space, where all dimensions keep similar ranges of variation.

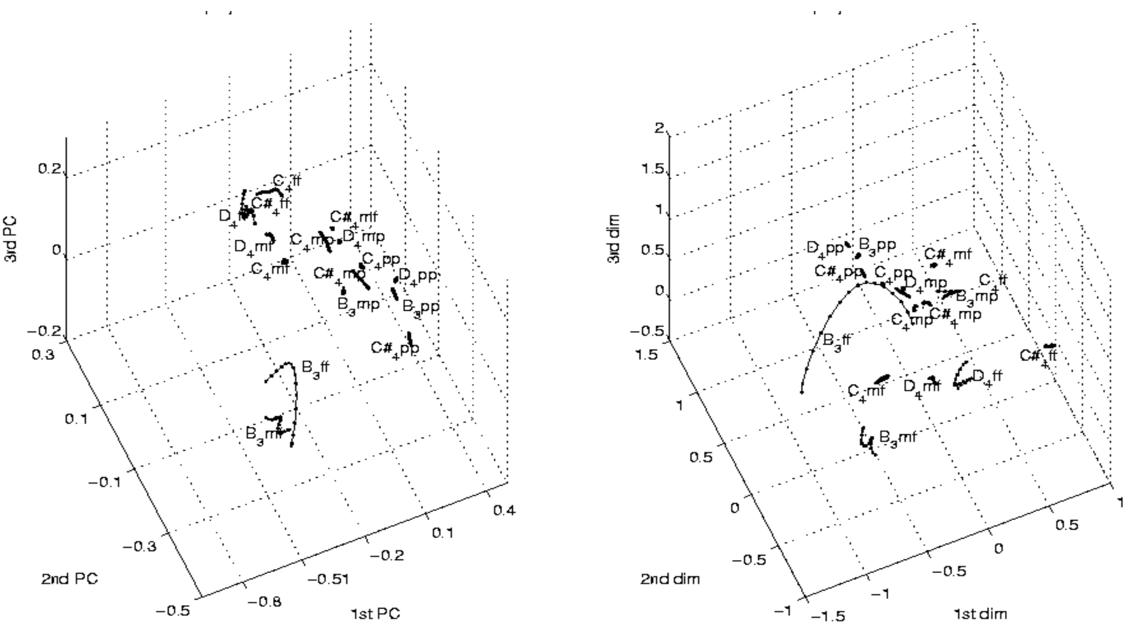

Figure 2: Three-dimensional trajectories of the four notes B3 (220 Hz), C4 (233 Hz), C#4 (247 Hz) and D4 (262 Hz) on PCA (left) and GA (right) spaces.

5. Timbre Classification

5.1. K-means Cluster Analysis

An attempt to investigate the timbre distribution along the entire instrument was made with Cluster Analysis, using the K-means algorithm (Kaufman e Rousseeuw, 1989). In the present analysis, the variance of a cluster was calculated using the Squared Euclidian Distance, although other types of distance were tested giving similar results. To avoid local minima, the K-means was run 40 times and the best solution was chosen. Comparison of timbre parameters among notes of different pitch becomes more complex, as timbre may vary significantly as a function of the note played, 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. At first, a cluster analysis was performed using the 19 notes (76 sounds) from the low register of the clarinet, from D3 (147 Hz) through Ab4 (415 Hz). Nine clusters provided the best correlation between auditory tests and the classification obtained for this set of sounds. A new clustering analysis was then performed using the 19 notes (76 sounds) from the low register of the clarinet. Nine clusters provided the best correlation between auditory tests and the obtained classification for this set of sounds (Loureiro, De Paula *et al.*, 2004). Very few of these sounds had their principal component coordinates 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 3 shows the 11 lower notes of the clarinet (from D3 through 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 (D3 to C4), 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 be also observed in Figure 4, which orders all 76 sounds of the low register of the clarinet by pitch and shows the cluster to which each one was assigned. Each sound is represented by the location of its central frame on the *low register timbre spac*e. This figure highlights the correlation of the cluster to intensity level and shows that intensity level variation spreads the sounds more than pitch variation. Auditory tests showed strong coupling of perceived brightness to clusters 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, It can be seen 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.

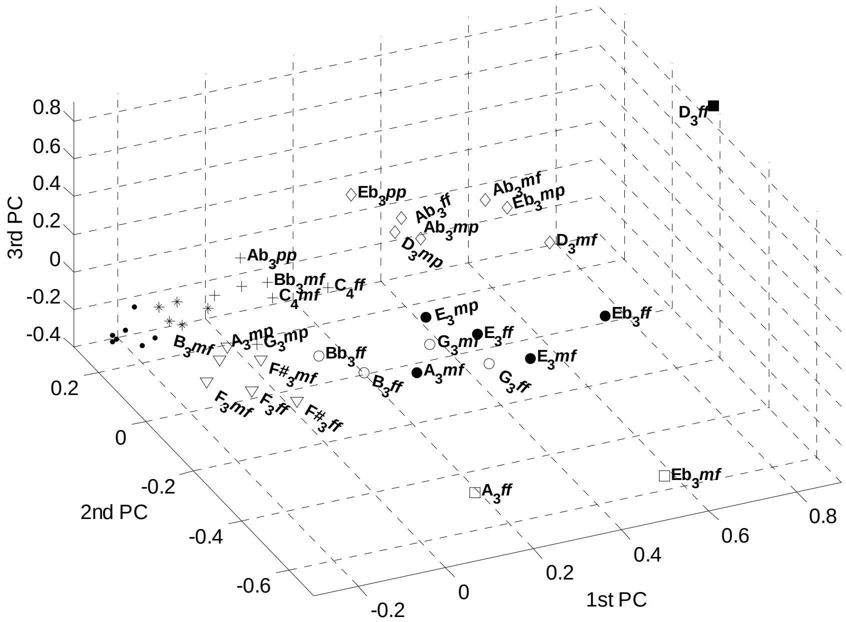

Figure 3: Three-dimensional trajectories of the four sounds of each of the 11 lowest notes of the clarinet, from D3 through C4, in the spectral space defined by the entire low register (points not labeled on the left of the figure correspond to notes: D3 *pp*, E3 *pp*, F3 *pp*, F3 *mp*, F#3 *pp*, F#3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

A new clustering analysis was then performed using the 33 notes (132 sounds) from the first and second register of the clarinet. Twelve clusters were found to be more adequate for this sound set. Figure 5 shows a similar plot to Figure 4, including now the 14 notes of the second register, A4 through Bb5. With the exception of one *mf* and four *ff* sounds (Bb *mf*, Bb4 *ff*, C# *ff*, B4 *ff* and Bb5 *ff*), all sounds of the second register lied in the five first clusters, together with *pp* and *mp* notes from the lowest register. This shows the 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.

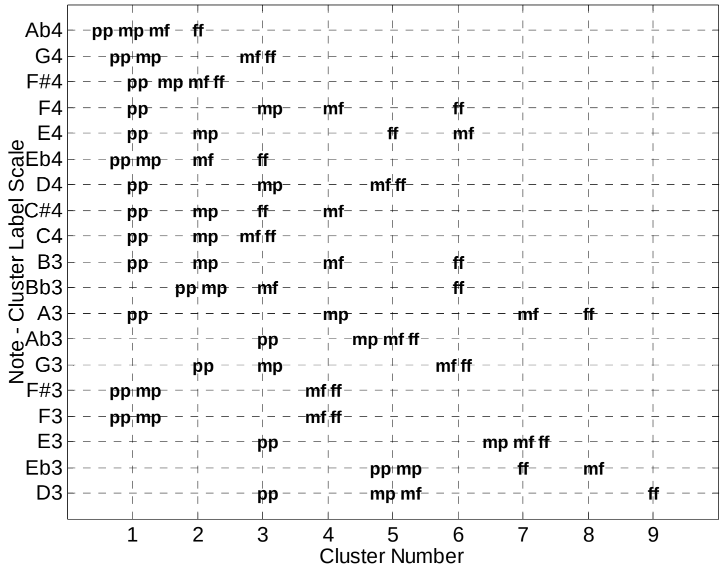

Figure 4: Cluster Label of the 19 notes of the low register of the clarinet. Notes are ordered by pitch and cluster labels by the mean of the spectral centroids.

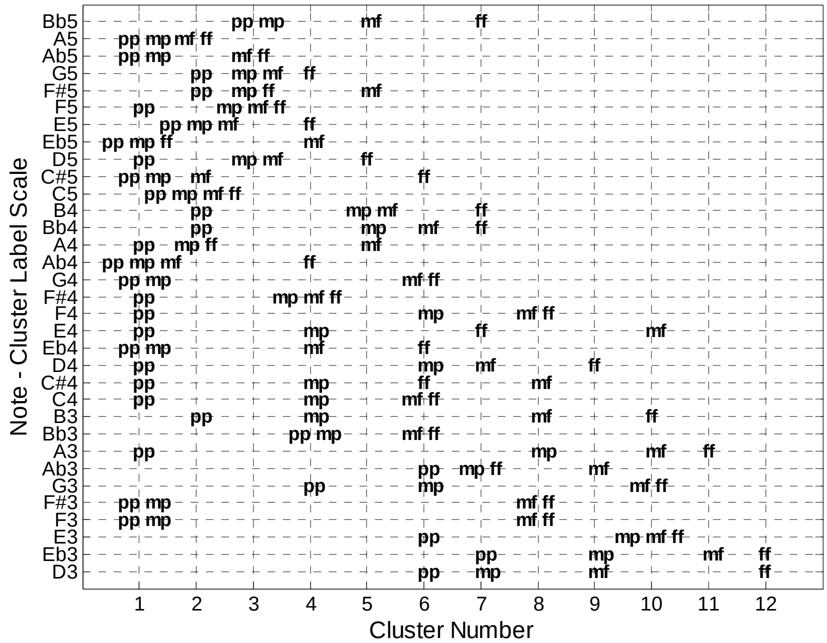

Figure 5: Cluster Label of the 4 four sounds of each of the 33 notes of the first two registers of the clarinet. Notes are ordered by pitch and cluster labels by the mean of the spectral centroids.

5.2. Self-Organizing Maps

Self-Organizing Maps are algorithms formalized by Kohonen for non supervised neural nets, capable of mapping input data of large dimensions into lower dimensional spaces, preserving the essential topological relationships of the original data (Kohonen, 1995).

Because it is not based on *a priori* suppositions about the characteristics of the analyzed data, SOM is a powerful tool to analyze complex and non-linear data, such as musical sounds. Leman (Leman, 1994) proposed a comparison between timbre mapping obtained by SOM and those spaces built by MDS starting from psychological measurements, as a basic reference for cognitive research in music. Toiviainen (Toiviainen, Kaipainen *et al.*, 1995) compared the efficiency of musical timbre representations in spaces built by topological distances calculated by SOM to subjective measurements of similarity. The results of that work proved a high correlation degree between the two domains, suggesting an adaptation of the model of Kohonen to project multidimensional perceptive complexes in this kind of representation. De Poli (De Poli e Prandoni, 1997), Cosi and colleagues (Cosi, De Poli *et al.*, 1994) developed studies on classification of musical timbre using SOM. Faiten (Faiten e Gunzel, 1994) obtained timbre specialization by processing pre-processed spectral parameters with SOM. This paper used a Matlab Toolbox from Versanto and colleagues (Versanto, Himberg *et al.*, 2000).

An hexagonal SOM of size 16-by-9 was used to map the 76 sounds (19 notes) from the low register of the clarinet. Figure 6 shows the relation of this mapping to sound intensity levels. As in the classification with k-means, *ff* and *mf* sounds tend to be grouped together. Moreover, *pp* and *mp* sounds were also tighter clustered than *mf* and *ff* sounds, as already observed on the spectral *timbre space* of Figure 3. This can be verified by the distance metrics distribution of the SOM shown on the graph on the right side of Figure 6, in which distances between hexagons represent distances between map cells.

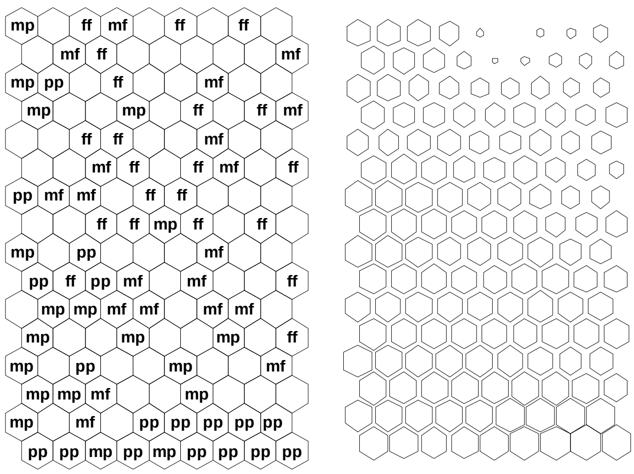

Figure 5: SOM mapping of intensity levels (left) of the 76 sounds (19 notes) of the low register of the clarinet and the distribution of the distance between map cells (right). Closer (larger) cells mean closer matching units.

Although SOM mapping is projected onto two dimensions, some consistency between both representations was identified. Like the K-means, SOM was able to map together every frame of a single sound into one or at most two cells. Figure 7(a) shows the mapping of the sounds of two contiguous notes of the lower octave of the clarinet, F3 and F#3. The labels indicate also the percentage of hits of that sound into the respective cell. Note also that the sounds of these notes were all packed together at the bottom left corner of the map. A similar clustering could be also observed in the spectral space representation. This fact suggests that the sounds of these notes do not vary considerably with intensity level. Figure 7(b) shows the hits of two other contiguous notes, Ab3 and A3. Despite being closer in pitch, they were widely spread in the map, a fact that can also be confirmed by their trajectories in Figure 3. Comparing Figures 4 and 7(b), it is possible to note the high correlation between the K-means analysis and the Kohonen map. In both classifications, the Ab3 was assigned to higher order clusters (top cells in the map) and its sounds were tighter together, while the A3 was wider spread over clusters and cells.

Figure 6(c) shows the trajectories of six notes of the lower octave of the instrument, including the notes plotted in figures 7(a) and 7(b): D3, Eb3, F3, F#3 and Ab3. Despite being closer in pitch, they were mapped onto two distinct groups on opposite sides, D3, Eb3 and Ab3 on the upper left corner and F3, F#3 and A3 on the lower right corner. Comparing Figures 4 and 7(c) we observe that notes of the upper left corner were assigned to the same clusters by the K-means as were also the notes mapped on the lower right corner. They were also spread into opposite sides on the spectral *timbre space* (Figure 3). Moreover, in both classifications Ab3 sounds were positioned tightly together, while A3 sounds were widely spread over clusters and cells, corroborating the high correlation between the K-means and the Kohonen map.

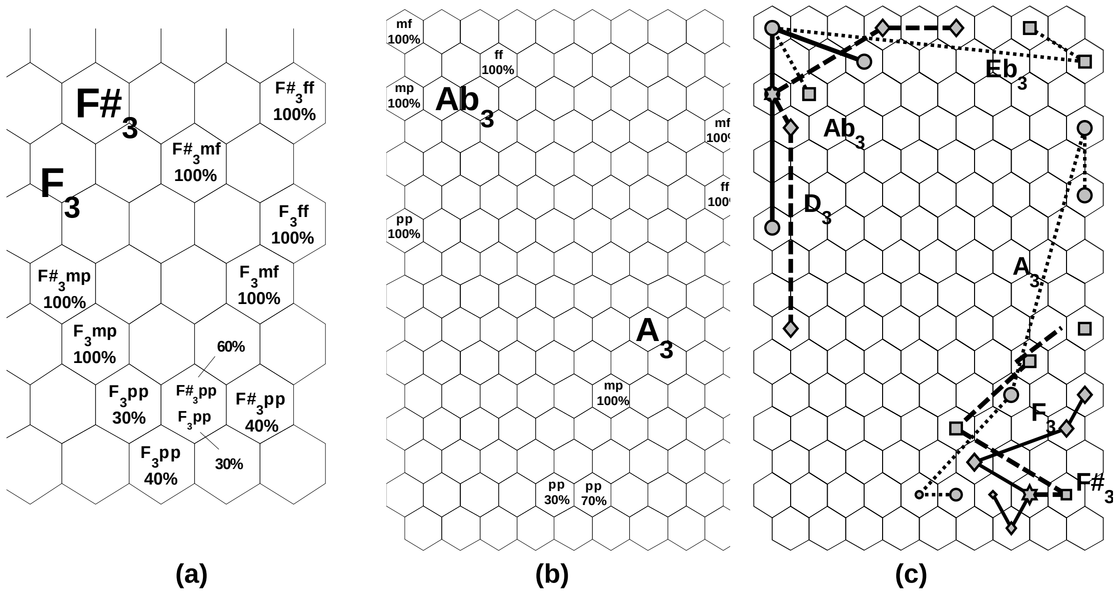

Figure 6: SOM mapping of : (a) notes F3 and F#3; (b) notes A3 and Ab3; (c) trajectories of notes D3, Eb3, F3, F#3 and Ab3.

6. Conclusion

PCA and GA were presented as adequate models for a lower dimension timbre space definition. Auditory tests of discrimination with resynthesized sounds with normalized pitch showed the effectiveness of the representation model presented in this paper, showing a clear relation between the perceived timbre and the cluster label to which the notes were assigned. 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. Both K-means and Self-Organized Maps provided a descriptive comparison of the dynamic variation of timbre. These representations and clustering techniques showed a strong matching, as they are mapping data from the same spectral *timbre space*. 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 exhibit in general much more richness of timbre differentiation than higher pitched notes; (iii) the highest octave of the mid register (from C5 up), 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 is to extend the investigation to shorter sounds, like staccati and pizzicati, as well as attack, decay and transition between notes, for which auditory models seems to be an adequate analysis tool.

7. 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.

8. References

Beauchamp, J. W. e A. Horner. Spectral Modelling and Timbre Hybridisation Programs for Computer Music. Organised Sound, v.2, n.3, p.253-258. 1998.

Charbonneau, G., C. Hourdin, et al. A Multidimensional Scaling Analysis of Musical Instrument's Time-Varying Spectra. Computer Music Journal, v.21, n.2, p.40-55. 1997.

Cosi, P., G. De Poli, et al. Auditory Modelling and Self-Organizing Neural Networks for Timbre Classification. Journal of New Music Research, v.23, p.71-98. 1994.

De Paula, H. B., M. A. Loureiro, et al. Representing Timbre Dynamics of a Musical Instrument: comparison between GA and PCA. IEEE International Workshop on Machine Learning for Signal Processing. São Luis, Brazil. October, 2004. 381-390 p.

De Poli, G. e P. Prandoni. Sonological Models for Timbre Characterization. Journal of New Music Research, v.26, p.170-197. 1997.

Faiten, B. e S. Gunzel. Automatic Indexing of a Sound Database Using Self-Organizing Neural Nets. Computer Music Journal, v.18, n.3, p.53-65. 1994.

Grey, J. M. An exploration of musical timbre. CCRMA, Dept. of Music Stanford University. Stanford, Calif., p.133. 1975. (STAN-M-2).

Hajda, J. M., R. A. Kendall, et al. Methodological Issues in Timbre Research. In: I. Deliège e J. Sloboda (Ed.). Perception and Cognition of Music. Hove: Psychology Press, 1997. Methodological Issues in Timbre Research, p.253-306

Horner, A. Wavetable Interpolation Synthesis Based on Time-Variant Spectral Analysis of Musical Sounds. Journal of The Audio Engineering Society. 1995a.

______. Wavetable Matching of Dynamic Instruments with Genetic Algorithms. Journal of The Audio Engineering Society, v.43, n.11, p.916-931. 1995b.

Horner, A., J. W. Beauchamp, et al. Methods for Multiple Wavetable Synthesis of Musical Instrument Tones. Journal of The Audio Engineering Society, v.41, n.5, p.336-356. 1993.

Horner, A. e Ngai-Man. Group Synthesis with Genetic algorithms. Journal of The Audio Engineering Society, v.44, n.3, p.130-147. 1996.

Johnson, R. e D. W. Wichern. Applied Multivariate Statistical Analysis. Upper Sadlle, New Jersey. 1998

Kaufman, L. e P. J. Rousseeuw. Finding Groups in Data: An Introduction to Cluster Analysis. New York: John Wiley & Sons. 1989. 111-116 p.

Kohonen, T. Self-Organing Maps - Springer Series in Information Sciences. Berlim: Springer Verlag, v.30. 1995

Leman, M. Schema-Based Tone Center Recognition of Musical Signals. Journal of New Music Research, v.23, n.2, p.169-204. 1994.

Loureiro, M. A., H. B. De Paula, et al. Timbre Classification Of A Single Musical Instrument. 5th International Seminar on Music Information Retrieval. Barcelona, Spain. November, 2004. 546-549 p.

Mcadams, S., S. Winsberg, et al. Perceptual Scaling of Synthesized Musical Timbres: Common Dimensions, Specificities and Latent Subject Classes. Psychological Research, v.58, p.177-192. 1995.

Mcaulay, R. J. e T. F. Quatieri. Speech Analysis/Synthesis Based on a Sinusoidal Representation. IEEE Transactions on Acoustics, Speech, and Signal Processing, v.34, n.4, p.744-754. 1986.

Misdariis, N. R., B. K. Smith, et al. Validation of a Multidimensional Distance Model for Perceptual Dissimilarities Among Musical Timbres. Proceediongs of the 16th International Congress on Acoustics. Woodbury, New York: ASA - The Acoustical Society of America, 1998. p.

Rochebois, T. e G. Charbonneau. Cross-Synthesis Using Interverted Principal Harmonic Sub-Spaces. In: M. Leman (Ed.). Music, Gestalt and Computing: Studies in Cognitive and Systematic Musicology. Berlin-Heidelberg: Springer Verlag, 1997. Cross-Synthesis Using Interverted Principal Harmonic Sub-Spaces, p.375-385

Sandell, G. J. e W. Martens. Perceptual Evaluation of Principal-Component-Based Synthesis of Musical Timbres. Journal of The Audio Engineering Society, v.43, n.12, p.1013-1028. 1995.

Serra, X. Musical Sound Modeling with Sinusoids plus Noise. In: A. Piccialli, C. Roads, et al (Ed.). Musical Signal Processing: Swets & Zeitlinger Publishers, 1997. Musical Sound Modeling with Sinusoids plus Noise

Toiviainen, P., M. Kaipainen, et al. Musical Timbre: Similarity Ratings Correlate with Computational Feature Space Distances. Journal of New Music Research, v.24, p.282-298. 1995.

Tzanetakis, G. A personal history of Music Information Retrieval. Multimedia and Mathematics. Banff, Canada, 2005. p.

Versanto, J., J. Himberg, et al. Self-Organizing Map in Matlab: the SOM Toolbox. Helsinky, Finland,: Helsinky University of Technology 2000.