TIMBRE CLASSIFICATION OF A SINGLE MUSICAL INSTRUMENT

Mauricio A. Loureiro
School of Music

Hugo B. de Paula
Grad. Program in Electrical Engineering

Hani C. Yehia
Dept. of Electronic Engineering

CEFALA – Center for Research on Speech, Acoustics Language and Music
UFMG - Federal University of Minas Gerais - Brazil

ABSTRACT

In order to map the spectral characteristics of the large 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 and of grouping them 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 strong relation between the perceived timbre and the cluster label to which the notes were assigned. Self-Organizing Maps lead to results similar to those obtained by PCA representation and K-means clustering algorithm.

1. INTRODUCTION

Representation of a musical instrument involves the estimation of the physical parameters that contribute to the perception of pitch, 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. Unlike timbre, 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. 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. 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 multidimensionality. "Timbre values" of different instruments were positioned on low-dimensional *timbre space* according to their similarity/dissimilarity responses between pairs of distinct timbres, providing a quantification of a relatively complex structure upon quite simple data. More recent studies were able to relate measurable physical parameters with the dimensions shared by the timbre represented in these spaces, establishing correlations between purely perceptive factors related to timbre and acoustic measurements extracted directly from sound [4, 10]. A historical review of the development of research on musical timbre is found in [8].

Most studies on musical timbre research 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 the instrument.

2. TIMBRE SET SPECIFICATION

Timbre representation here investigated was built upon spectral parameters extracted from samples of sounds performed along the entire pitch range of the instrument. 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 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"; (ii) 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 of the sound, the second simplification limits the investigation to the perception of slow variation of timbre, which commonly happens along longer notes.

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 of 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: *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 these 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) through A5 (880 Hz), played at the four levels of intensity defined above, with an average duration of 3 seconds.

3. SPECTRAL BASES FOR TIMBRE CHARACTERIZATION

3.1. Spectral parameters estimation

Amplitude curves of partials 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 frames ("peak continuation"), associating these values to instantaneous frequency and amplitude values of harmonic components [9, 12]. It was assumed that all sampled sounds could be represented by a weighted sum of sinusoids without abrupt variations of amplitude and frequency values. Components with intensity more than 60 dB below the maximum level were discarded and 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 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) [5]. PCA calculates an orthogonal basis determined by the directions of maximum variance of a set of multidimensional variables. 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. Recent works presented the effectiveness of this kind of representation of musical timbre [1-3, 11].

3.3. Physical Timbre space trajectories

In order to represent the spectral distributions of the tested sounds, spectral basis were calculated using as input data the concatenation of the four samples of each

note, *pp*, *mp*, *mf* and *ff*, as defined in Section 2. Samples were normalized in amplitude and duration, with 75 frames (approx. 870 ms) each, taken from the center of the note. The spectral basis thus obtained constitutes a *timbre space* for these notes, where each sound occupies a unique position, according to its spectral configuration. Figure 1 shows three-dimensional trajectories of four intensity levels of four contiguous notes, A3 (220 Hz), Bb3 (233 Hz), B3 (247 Hz) and C4 (262 Hz), represented in the principal component space defined by them.

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, while the 2nd and 3rd PCs vary differently in different directions for each 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.
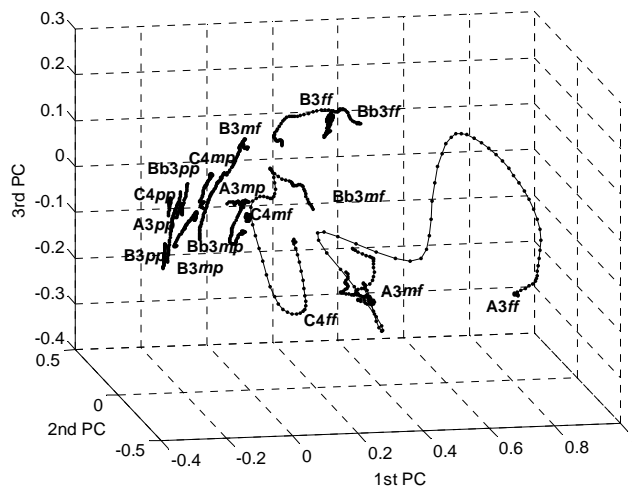

Figure 1: Three-dimensional trajectories of notes A3 (220 Hz), Bb3 (233 Hz), B3 (247 Hz) and C4 (262 Hz) in the spectral space defined by them.

4. TIMBRE CLASSIFICATION

4.1. 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 [6]. 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. 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 2 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 space*. This figure highlights the correlation of the cluster to intensity level and shows that intensity level variation spreads the sounds more than pitch variation. Informal auditory tests showed strong coupling between perceived brightness and clusters assignment. Due to the known relationship of spectral centroid to the perception of brightness, cluster labels were ordered according to the mean 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. Some notes of higher pitch in *mf* and *ff* were also assigned to those clusters. While higher pitched notes were grouped more tightly into these clusters, the four last clusters contain almost only *mf* and *ff* notes of the lower octave, showing the tendency of timbre variation reduction as pitch increases.
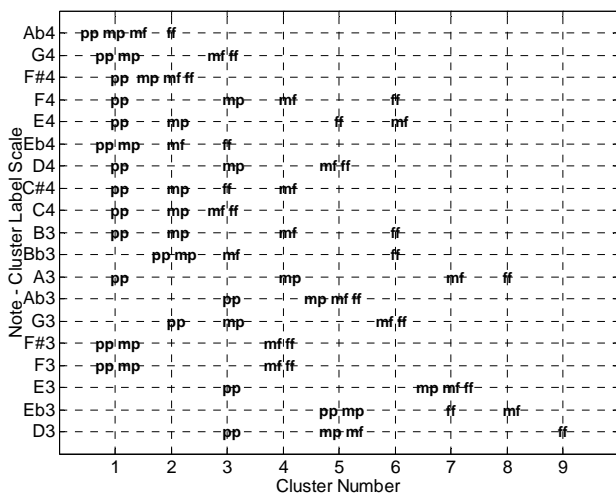

Figure 2: Cluster Label of the 19 notes of the low register of the clarinet, D3 (147 Hz) through Ab4 (415 Hz). Notes are ordered by pitch and cluster labels by the mean of the spectral centroids.

4.2. 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 [7]. Toiviainen [13] compared the efficiency of musical timbre representations in spaces built by topological distances

calculated by SOM to subjective measurements of similarity, proving a high correlation degree between the two domains. De Poli [3], Cosi and colleagues [2] developed studies on classification of musical timbre using SOM. This paper used a Matlab Toolbox from Versanto and colleagues [14].

An hexagonal SOM of size 16-by-9 was used to map the 76 sounds (19 notes) of the low register of the clarinet. Figure 3 shows the relation of this mapping to sound intensity levels. *pp* and *mp* sounds were more tightly clustered than *mf* and *ff* sounds, which can be verified by the distance metrics distribution of the SOM shown on the graph on the right side of Figure 3, in which distances between hexagons represent distances between map cells.

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.
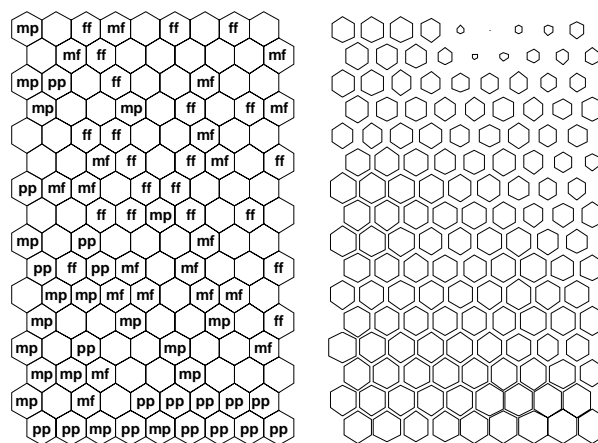

Figure 3: SOM mapping of intensity levels (left) of the 76 sounds (19 notes) of the low register of the clarinet, (D3 to Ab4); distribution of the distance between map cells (right). Larger cells mean closer matching units.

Figure 4 shows the trajectories of six notes of the lower octave of the instrument: D3, Eb3, F3, F#3, Ab3 and A3. Despite being closer in pitch, they were mapped onto two distinct groups on opposite sides. Comparing Figures 2 and 4 we observe that notes mapped on the upper left corner (D3, Eb3 and Ab3) were assigned to the same clusters by the K-means as were also the notes mapped on the lower right corner (F3, F#3 and A3). 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.

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 clustering techniques are used to categorize the different
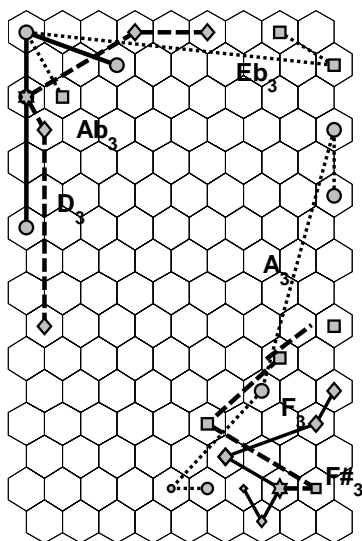
Figure 4: SOM mapping of the trajectories of notes
D3, Eb3, F3, F#3 Ab3 and A3.

timbres produced. 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 PC *timbre space*. 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 *timbre space*. It was clearly verified across all the results presented in this study that timbre classes tend to be divided as a function of spectral brightness, which is known to be correlated to intensity level in wind instruments. It was also noted that the lower pitched notes of the clarinet exhibit in general much more richness of timbre variation and spectral brightness than higher pitched notes.

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

7. REFERENCES

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

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

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

[4] Hajda, J. M., R. A. Kendall, E. C. Carterette, and M. L. Harshberger, "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] Johnson, R. and D. W. Wichern, Applied Multivariate Statistical Analysis. Upper Sadlle, New Jersey, 1998.

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

[7] Kohonen, T., Self-Organing Maps - Springer Series in Information Sciences, vol. 30. Berlim: Springer Verlag, 1995.

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

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

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

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

[12] 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.

[13] Toiviainen, P., M. Kaipainen, and J. Louhivuori, "Musical Timbre: Similarity Ratings Correlate with Computational Feature Space Distances," Journal of New Music Research, vol. 24, pp. 282-298, 1995.

[14] Versanto, J., J. Himberg, E. Alhoniemi, and J. Parhankangas, Self-Organizing Map in Matlab: the SOM Toolbox. Helsinky, Finland,: Helsinky University of Technology, 2000.