STUDY OF THE TIMBRE DYNAMICS OF AN EXPRESSIVE PERFORMANCE USING PRINCIPAL COMPONENT ANALYSIS OF SPECTRAL PARAMETERS

Hugo B. de Paula *- hugobp@cpdee.ufmg.br*
CEFALA – Programa de Pós-graduação em Engenharia Elétrica – UFMG

Hani C. Yehia *- hani@cpdee.ufmg.br*
CEFALA - Departamento de Engenharia Eletrônica – UFMG

Maurício A. Loureiro *- mauricio@musica.ufmg.br*
CEFALA / CPMC - Escola de Música – UFMG

CEFALA – Center for Research on Speech, Acoustics, Language and Music
www.cpdee.ufmg.br/~cefala
Universidade Federal de Minas Gerais
Av. Pres. Antonio Carlos, 6627
Belo Horizonte, MG, Brasil - 31270-010

Abstract

Expressive content of performed musical sounds involves the behavior of a wide range of parameters. Timbre dynamics can be responsible for much of the perceived expressiveness of long-note phrases played on wind and bowed instruments. In this paper, Principal Components Analysis (PCA) is used to represent dimensions of spectral dynamics, in order to reveal intentional expressive timbre changes within a gesture realization in a musical performance on the clarinet. Amplitude and frequency time-varying curves of the partials were measured. Spectral envelopes showed a high correlation both in frequency and time domain. PCA has been proved effective in identifying variation patterns in the spectral distribution. The PCA enabled data reduction and transformation making possible to build relations between the principal components and perceptual characteristics of sound. This approach seems to reveal important properties of these parameters that can possibly be characterized as “timbre variation”. Results showed that three principal components were enough to represent the “timbre variation” of the samples and the trajectories of these components elucidated some main characteristics of timbre development.

Introduction

The concept of timbre refers to the color or the quality of the sound and is defined by the ASA (American Standard Association) as “that attribute of the auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar.” (Risset and Wessel, 1982). This vague definition of timbre is related to its inherent multidimensional nature. This sound attribute cannot be easily scaled, unlike intensity and pitch, of which different time-varying levels can be classified by a piano-forte or a low-high one-dimensional scale and therefore quantitatively expressed by the traditional music notation system. A musical timbre is commonly defined in terms of grouping all sounds produced by a single musical instrument, even though the quality of these sounds can be quite different according to its intensity, pitch or duration. Most research on timbre characterization has investigated with auditory perception of musical instruments (Luce, 1963; Risset, 1991; McAdams e Bregman, 1979; McAdams, 1987; Gordon e Grey, 1977, 1978; Grey, 1975, 1978; Barrière, 1991). In the late 60's and in the 70's, several research works in musical instrument analysis/synthesis methods demonstrated that the dynamic spectral energy distribution supplies the acoustic determinant for our perception of sound quality and therefore could be well represented by the time-varying amplitude and frequency functions of its partials (Risset, 1965, 1991; Risset and Mathews, 1969; Strong and Clark, 1967; Gordon and Grey, 1977, 1978; Grey, 1975, 1978; Wessel, 1979). Strong e Clark used additive synthesis to construct wind instrument sounds, in which the amplitude of each partial was controlled by a single envelope. Later they used varying envelopes as a function of the partial frequency (Strong e Clark, 1967a e 1967b). Risset, 1965 analyzed individual trumpet sounds and obtained curves of amplitude and frequency for each partial. From the analyzed data he synthesized trumpet sounds using the MUSIC V synthesis system. Through auditory tests he was able to conclude that some characteristics of the trumpet timbre are more related to variations in the spectrum than to the spectrum structure itself (Risset, 1965). The motivation for these studies is certainly not limited to the purpose of duplicating these sounds electronically. Some of these authors have sought methods to measure and even to parameterize the timbre, in order to understand better how we hear the music produced with these instruments.

Objectives of the Study

Traditional acoustic instruments offer the possibility of producing and to accurately controlling a wide variety of timbres, depending on the pitch and on the way they are played. Most of the studies accomplished on timbre of musical instruments has been restricted to analyzing isolated musical notes, in general comparing sounds of different instruments, sampled outside any musical context, focusing the perceptual discrimination between instruments. Very little research was done in relation to the timbre variation across sounds produced by the same instrument or even along the duration of one single note. These variations are significant to the perceptive mechanisms responsible for the notion of expressiveness in a performed musical phrase. The present project intends to investigate the timbre variation of the clarinet along selected performed musical phrases. Instead of treating isolated and almost-static notes, this work seeks the meaning of physical parameters that might be responsible for the timbre variation along a musical performance. Although the current opinion is that the timbre of a musical instrument also depends upon the dynamic attributes in the sound attack, this study is focused on the time-varying spectral energy distribution during the sustained portion of the sound.

Data Acquisition

Selection of the Samples

The selection of the analyzed samples tried to characterize timbristic aspects of the instrument and to include phrases with great expressive content, such as the melody played by the clarinet in the opening of the ***Clarinet Quintet*** op. 115 in B minor for clarinet and string quartet by Johannes Brahms. Phrase constructs and the score organization in this passage explores well the potentials of timbre differentiation of the clarinet. A good example is the F sharp played by the clarinet on the sixth measure, which had its duration triplicated in relation to the same note played by the violins five measures before, when this theme is stated for the first time. Dynamic marks ( ), the harmonic pulse and the fast ascending arpeggios by the cello in this passage indicate to the performer an accentuated dynamic variation of timbre and intensity (Loureiro, 1996). This specific note is used in this article to demonstrate the analysis methods utilized in this investigation. The *Adagio* of this Quintet also provided suitable passages for this study, as well as the ***Clarinet Quintet*** in A Major (KV. 588) by W. A. Mozart and the ***Grand Quintetto*** in B flat Major op. 34 by Karl Maria von Weber, both written for the same instrumentation. The samples were recorded at the studios of the School of Music of the Universidade Federal de Minas Gerais, performed by clarinetist Maurício Loureiro.

Time-varying amplitude and frequency

In this study, the signal is represented by its spectrogram, assuming that the sampled sound is perfectly harmonic and that an isolated note has only one well defined fundamental frequency. This deterministic analysis was done using a STFT (Short Time Fourier Transform) of 65536 points using a Hamming window of 2048 samples. The 2048 samples windowed signal was zero-padded. Each frame has an "instantaneous" resolution of about 45 ms. A frame overlap of 1024 samples was used, increasing the overall time resolution to about 23 ms (1024 samples).

The "instantaneous" frequency of each harmonic was obtained by peak detection of the magnitude spectrum from the DFT (Discrete Fourier Transform). The use of a large DFT gave a frequency resolution of 0.67 Hertz. The effects of windowing and zero-padding in spectrum estimation are well explained in Masri et al., 1997. For each frame, the analysis algorithm calculated the magnitude peak, phase and frequency of the first 35 harmonics of the signals.

Data simplification

The magnitude curves of the spectrum showed small variations in short periods of time. Several researches try to validate simplification methods in order to smooth the spectrum. In general, these works drove to conclusions that small amplitude and frequency fluctuations are not noticeable and can be simplified by linear segments, making it possible to achieve a data compression from 20:1 to 50:1 without significant loss of information (Grey, 1975).

The first data reduction was performed selecting the number of harmonics. Although the analysis calculated the first 35 harmonics of each signal, all partials with its log-magnitude peak 40 dB below the overall log-magnitude peak were discarded. After this spectrum reduction, a 6^{th} order ***butterworth*** low-pass filter with cut-off frequency of 10 Hertz was used to smooth the magnitude curves (Beauchamp, J. and Horner, A. 1997). A bi-directional filtering was performed to avoid phase distortion. The signal reconstruction from these simplified data showed the little auditory significance of these variations and auditory tests were enough to validate this simplification.

Analysis of the data

Since the 60's, Statistical Multivariate Analysis methods have been used in musical timbre research, towards the development of multidimensional scaling of timbre. Among them, the Principal Components Analysis (PCA), has been widely applied to a variety of perceptual data. The first experiences with PCA directly applied to physical parameters of sound were made in 1995 (Sandell and Martens, 1995; Charbonneau et al, 1997a, 1997b). In that study, PCA was applied to measurements of physical parameters, seeking their relationship with perceptual characteristics.

Principal Components Analysis

PCA is the optimum transform in terms of concentrating the most energy in the fewest transform coefficients, and is also known as Karhunen-Loeve Transform (KLT). In PCA, vectors of data are represented by linear combinations of orthonormal basis functions (or vectors). This basis is determined by the directions of maximum variance in the space defined by the vectors of data. It represents a change to a new system of coordinates, in which the "principal components" of a group of vectors are defined by their axes (Johnson & Wichern, 1998; Rencher, 1995). Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \cdots \mathbf{x}_n]$ be the data matrix. The covariance matrix $\mathbf{C}_{XX}$ of $\mathbf{X}$ is given by:

$$\mathbf{C}_{XX} = \mathrm{E}[\mathbf{x}_0 \mathbf{x}_0{}^t] \qquad \text{where, } \mathbf{x}_0 = \mathbf{x} - m(\mathbf{x}) \text{ and, } \qquad m(\mathbf{x}) = \mathrm{E}[\mathbf{x}]$$

$\mathbf{C}_{XX}$ is calculated by the estimated covariance matrix $\overline{\mathbf{C}_{XX}}$.

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

where $\mathbf{x}_n$ are sample vectors and $\overline{m}$ the estimated mean of the sample set. The principal components (PCs) are the components of a transform matrix **U** determined by the singular value decomposition (SVD) of $\mathbf{C}_{XX}$:

$$\mathbf{C}_{XX} = \mathbf{U}\mathbf{S}\mathbf{U}^t$$

where ***S*** is a diagonal matrix whose non null elements are the eigenvalues of $\mathbf{C}_{XX}$ and **U** is an unitary transform matrix whose columns are the normalized eigenvectors (unitary Euclidean norm) associated to the eigenvalues of $\mathbf{C}_{XX}$. The representation of **X** in PCs is given by:

$$\mathbf{Y} = \mathbf{U}^t \mathbf{X}$$

and $\mathbf{X}$ can be recovered as:

$$\mathbf{X} = \mathbf{U}\mathbf{Y}$$

For compression and simplification purposes, it is possible to reconstruct the original data matrix using fewer PCs. The percentage of total variance of the signal "described" by a subset of PCs is given by the accumulated sum of the eigenvalues associated to the eigenvectors used in the reconstruction.

Influence of scaling to PCA

It is important to note that the way the PCA input data is scaled plays a relevant role on the analysis results. Linear scaled spectra (directly estimated values) have normally magnitude values within a range from 0 to 10^4. In this case, the lower partials present values about one hundred times greater than the upper partials. Consequently, any variation on the lower partials is much more important for the PCA. In other words, the PCA becomes more "interested in" reproducing the variations of lower partials. For values logarithmically scaled (dB), therefore limited into a range of 80 dB, the spectral representation will be flattened. In this case upper partial variations will play a much more important role for the PCA estimation. Fig. 1 shows two PCAs performed in the same sound, using these two scales.

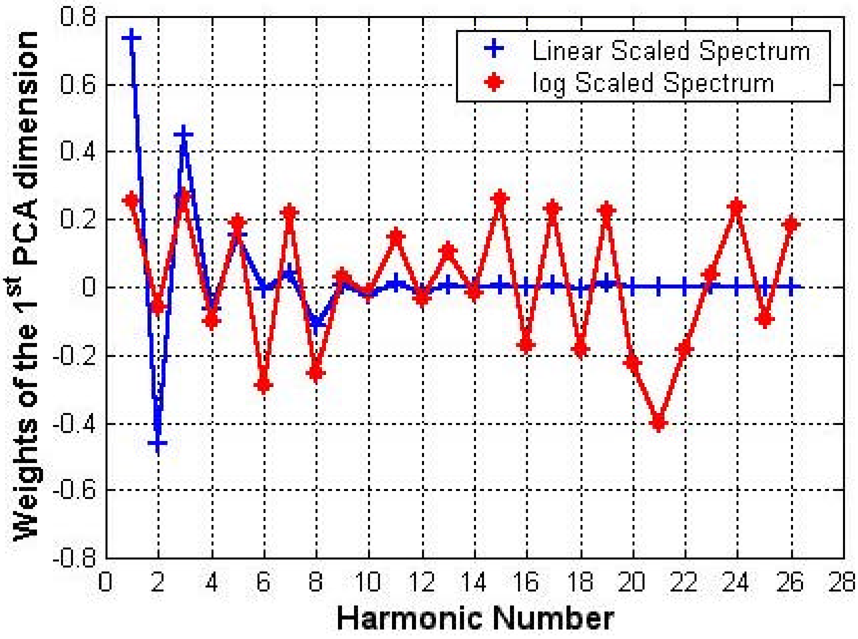

Figure 1: The weights of the 1st PCA eigenvector for the cases of linear and logarithmic analysis of the same sound.

Signal reconstruction from data obtained by PCA performed on log-magnitude spectra of sounds with noisy upper partials, such as notes from lower register from the clarinet, showed an unwanted *vibrato* in the lower partial. This results from the insertion of upper partial characteristics in the lower reconstructed partials. On the other hand, sound with a great expressive content with slowing varying upper partials may loose its "expressivity" when analyzed in a linear scale. The trade off between the number of PCs that is needed to represent a sound and the timbre characterization is highly dependent upon the scale used. In this study linear scale was used for all analyses.

Signal Reconstruction

Signal reconstruction from measured and processed spectral data was made by additive synthesis, in which each partial is represented by a sine wave. Note that the signal was assumed to be harmonic with no variation in frequency. No transients or noise components were considered, so additive synthesis can be calculated by (Serra, 1997):

$$s(t) = \sum_{r=1}^{R} A_r(t) \cos[2\pi f_r + \theta_r(0)]$$

where s is the reconstructed signal and $A_r(t)$, f_r and $\theta_r(0)$ are the amplitude, frequency and initial phase of partial r, respectively. Fig. 2 illustrates the analysis procedure.

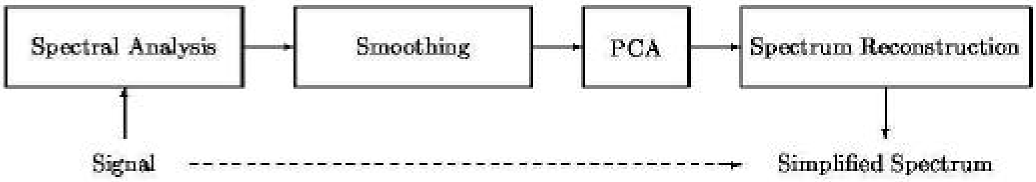

Figure 2: Diagram of the sound analysis

Results

This section presents an analysis of the results obtained with two samples. These samples have the same duration (3666 ms) and same frequency (749 Hz), both sampled at a rate of 44.1 kHz, in 16 bits, PCM linear mono. The first of these examples (expressive note) corresponds to the F sharp (tuned at A=445 Hz) sustained by the clarinet along the 8th and the first half of the 9th measures of Brahms Clarinet Quintet op. 115. In this example matrix **X** contains the amplitude values of the first 12 partials in 156 time frames, each one with a duration of approximately 23 ms. Fig. 3 shows the magnitude spectrum (matrix **X**) of the *expressive note.*

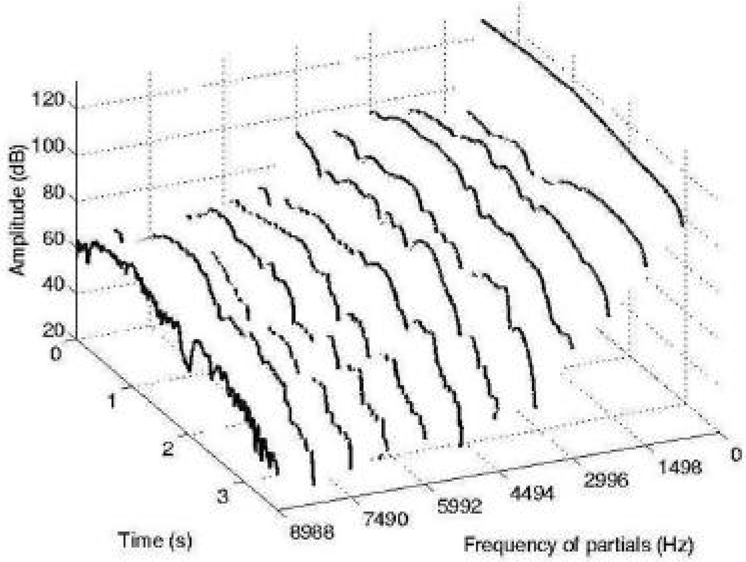

Figure 3: Spectrogram with the first 12 partials of an F sharp (749 Hz) played by the clarinet at the 6th measure of Brahms Clarinet Quintet op. 115 - *expressive note*.

The first 3 PCs recovered a total variance (likelihood) of 90% for this signal. Auditory comparison tests showed that the signal could be well represented by this reconstruction, as no significant difference could be perceived between both original and reconstructed sounds. Fig. 4 shows the spectrogram containing the first 12 harmonics of the *expressive note* reconstructed from the first 3 PCs.

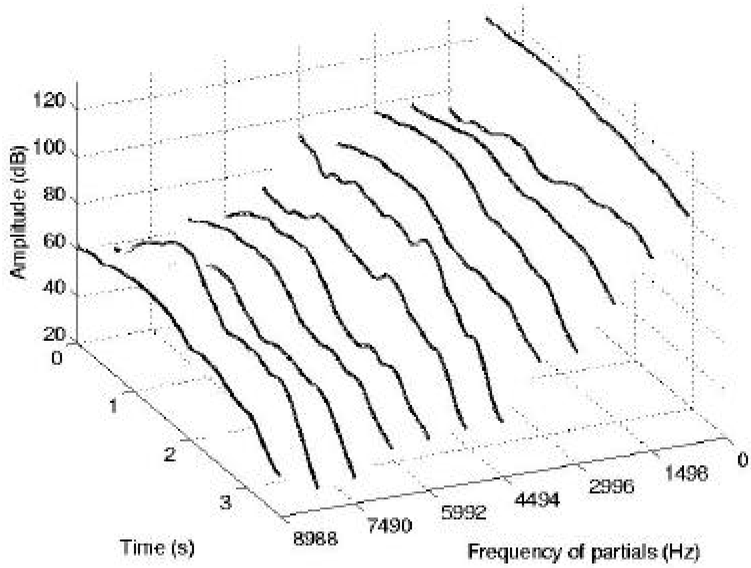

Figure 4: Spectrogram of the reconstruction of the *expressive note* from Figure 2, using 3 PCs.

The other example (*plain note*) corresponds to the same note played with a minimum of intentional dynamics variation. Fig. 5 shows the spectrogram containing the first 12 partials of the *plain note* (original) and Fig. 6 shows its reconstruction from 3 PCs.

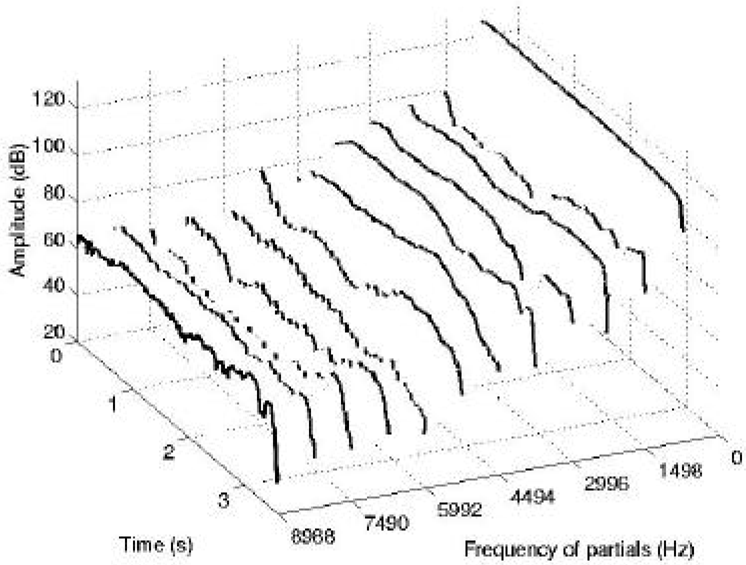

Figure 5: Spectrogram with the first 12 partials of the same F sharp (749 Hz) played on the clarinet with little dynamic variation – *plain note*.

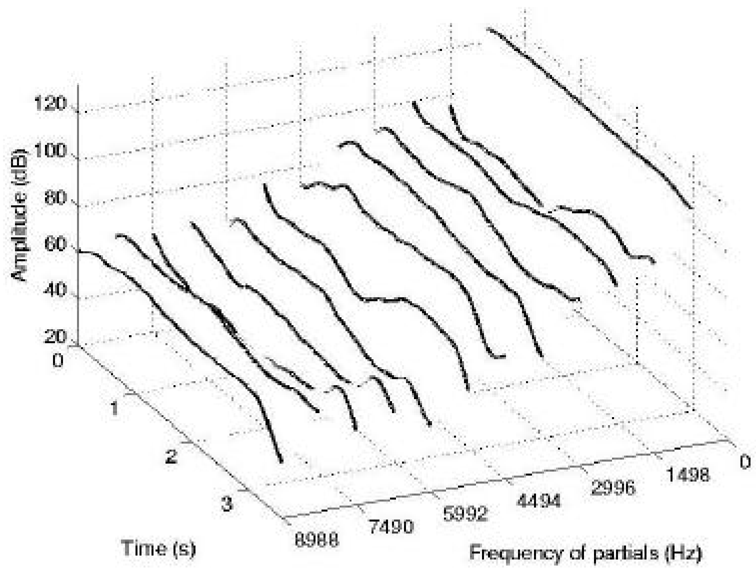

Figure 6: Spectrogram of the reconstruction of the *plain note* from Fig. 4, using 3 PCs.

Analysis of the Meaning of the Principal Components

Fig. 7 shows the 2nd PC as a function of the 1st for the *expressive note.* The plot corresponds to projecting the data measured in 12 dimensions (amplitudes of each of the first 12 partials) on a plane, which contain their largest variation. These 2 dimensions accumulated 90% of the signal variance. The curve shows the time evolution of the note where each time frame is represented by a star (★). These points are linked in chronological order. The begining and the end of the sound are marked on the graph. Since all time frames have the same length, the distance between two points represents the degree of time variation, i. e. the closer the points, the smaller the variation occurred between them.

Larger timbre variations can be observed nearly the end of the sound, precisely in the last 25 points (from point D through the end), corresponding to the last 575 ms. This is related to the accentuated *diminuendo* occurring at the end of the note. Note that this *diminuendo* begins approximately at 2 seconds from the end, but the analysis shows that larger timbre variations take place only on the last half second. While the starting point of this variation could be heard, more systematic auditory tests shall be made in order to validate this relationship.

From the beginning of the sound through point B of the expressive note, it was verified that the points were much closer to each other, which indicates the occurrence of softer timbre variations if compared to the above mentioned. This is where the *crescendo* takes place and, therefore, timbre variation appears to be slower along the *crescendi* if compared to the *diminuendi*.

The 1st PC shows a more accentuated variation. Examining the time-varying curves of each component, it can be seen that the 1st PC follows the sound amplitude envelope, indicating a narrow relationship with the signal total energy. The weights applied to this PC appear to shape each partial with a certain degree of variation related to the global sound amplitude envelope. It is equivalent to saying that the weight of the first PC measures the tendency of a specific partial to follow the global amplitude of the sound during the time. The variations in the 2nd PC seem to be related to located configurations of the spectrum that, possibly, represent timbristic nuances of the sound.

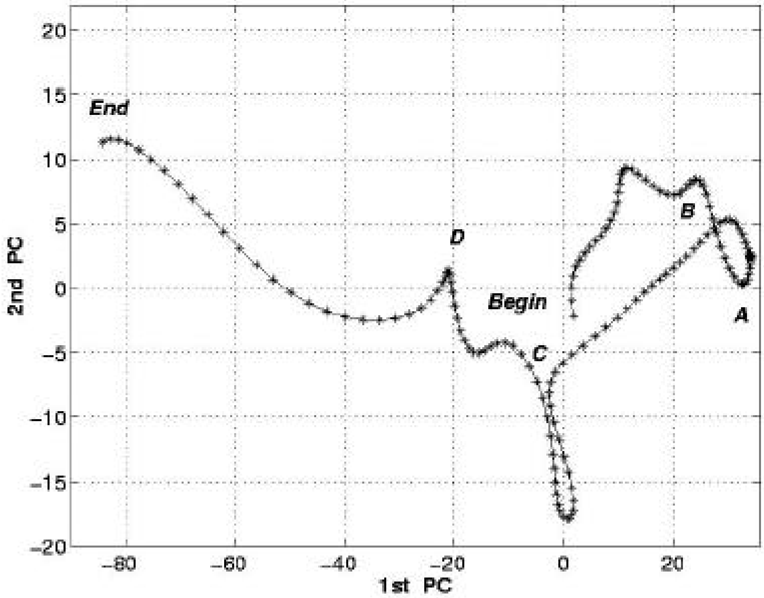

Figure 7: First 2 PCs for the *expressive* F sharp (749 Hz).

Fig. 8 shows the 2nd PC as a function of the 1st PC for the *plain note*. Smaller variation of the 1st component is evident here, confirming the statement that the 1st PC is related to the variation of the sound intensity (energy). The variation of the 2nd PC occurs practically in two movements: negatively during the first three quarters of the sound and positively after that. This reveals a more ordered behavior of the 2nd PC, which seems to represent a smaller complexity of timbre development for plain sounds. This factor is still under investigation by comparing notes from different regions of the instrument.

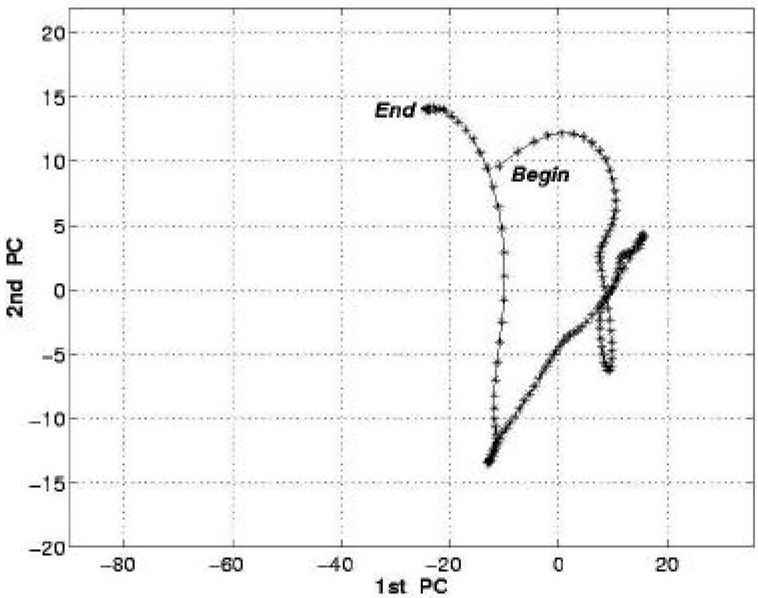

Figure 8: First 2 PCs for the *plain* F sharp (749 Hz).

It is also verified that the accumulated variance of the reconstruction is much smaller for the plain note (74%) than for the expressive note (90%), but they grow much faster, almost equaling their values when 6 PCs are used. This fact appears to be related to short duration fluctuations that assume less importance for PCs of lower order for the case of signals with larger dynamic variation. Further tests will be made with PCs of higher order.

Analysis made in different registers of the clarinet showed that notes of different timbre characteristics follow that same overall behavior of timbre variation here discussed. Fig. 9 shows the 2nd PC as a function of the 1st for an F sharp two octaves below (187 Hz), which was taken from the same passage form Brahms Quintet, 11 eleven measures ahead (15^{th} measure). Figure 10 shows the same plot for the same note played with a minimum of intentional dynamics variation.

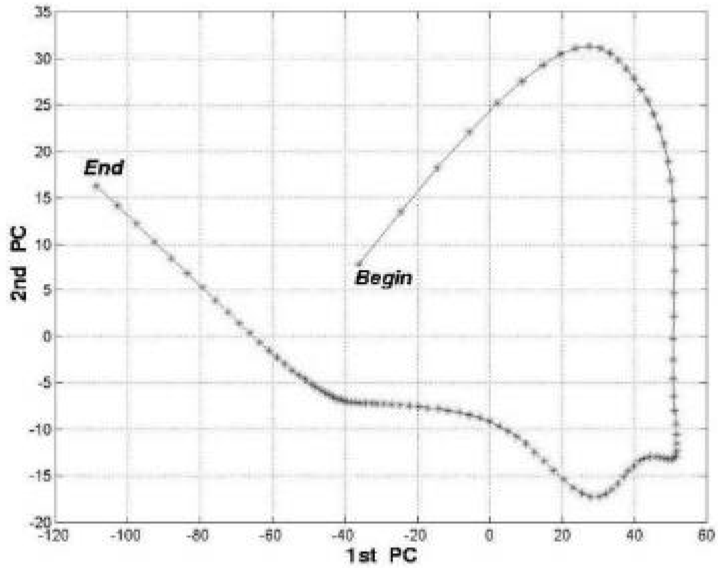

Figure 9: First 2 PCs for the *expressive* F sharp (187 Hz).

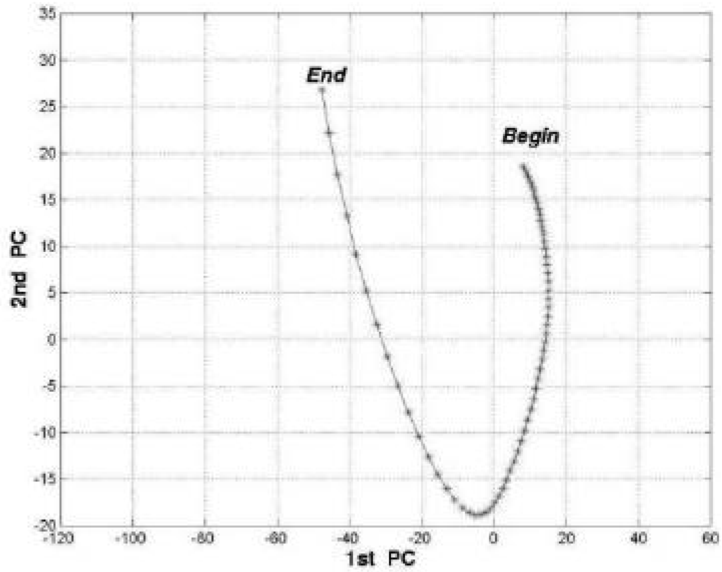

Figure 10: First 2 PCs for the *plain* F sharp (187 Hz).

Conclusion

Each PCA discussed was performed individually for each sound. This kind of analysis is especially useful for data reduction, since an optimal basis for each note is obtained. This makes possible to have most of the sound reconstructed by few PCs. The sound representation obtained after this data reduction was able to reveal many aspects of spectral parameter variations related to dynamic timbre evolution due to intentional expressive inflexions given by the player in a musical performance. However, a systematic observation on these timbre characteristics comparing different notes of the instrument was not facilitated due to the fact that each sound was represented on its own basis. A future work is to perform only one PCA across several concatenated spectra from different notes covering the whole extension of the clarinet. The basis generated by this kind of analysis explains all the notes at once, creating what can be called a *Clarinet Timbre Space*. In such a space, each note will have its region of variation, and notes with similar timbre characteristics will lie closer within the clarinet space.

The importance and need of this study elapses from the own nature of timbre, which doesn't have yet a precise and logical definition, even though the perceptual mechanisms involved in its detection are perhaps the most precise of our perceptual system. The comprehension of the dynamic control that the performer retains on the sound he/she produces on his/her instrument and how this control is perceived by the listener contributes not only to our understanding of the phenomenon “music”, but also drives us to formulations of platforms that might offer other types of control for other types of musical structures: “*...the future of the live performance depends on new instruments.*” (Smalley, 1986).

Acknowledgements

This paper presents partial results of a project support by CNPq, Brazilian funding agency for technology and science development. We also would like to thank Maurílio Nunes Vieira from the Physics Department (UFMG), for his valuable contribution and Sérgio Freire from the School of Music of (UFMG) for his expert assistance on the recording of the samples used in this study.

References

Barrière, J.-B. 1991. Le Timbre, Métaphore pour la Composition. Paris: IRCAM e Cristian Bourgois.

Beauchamp, J. and Horner, A. 1997. “Spectral Modelling and timbre hybridisation programs for computer music.” Organised Sound, vol. 2, no. 3, pp. 253-258

Chabornneau, G.; Hourdin. C.; Moussa, T. 1997a. “A Multidimensional Scaling Analysis of Musical Instrument's Time-Varying Spectra,” Computer Music Journal, vol. 21, no. 2, pp. 40-55.

Chabornneau, G.; Hourdin. C.; Moussa, T. 1997b. “A Sound Synthesis Technique Based on Multidimensional Scaling of Spectra,” Computer Music Journal, vol. 21, no. 2, pp. 56-68.

Gordon , J. and Grey , J. M. 1977. “Perception of spectral modifications on orchestral instrument tones.” Computer Music Journal, vol. 2 no. 1, pp. 24-31.

Gordon , J. and Grey , J. M. 1978. “Perceptual effects of spectral modifications on music timbres.” Journal of Acoustics Society of America, vol. 63, pp. 1493-1500.

Grey, J. M. 1975. An Exploration of Musical Timbre. Ph.D. Dissertation, Department of Psychology, Stanford University, Department of Music Report STAN-M-2, Palo Alto, CA.

Grey, J. M. 1978. “Timbre discrimination in musical patterns.” Journal of Acoustics Society of America, vol. 64, pp. 467-472.

Jonhnson, R. . and Wichern, D. W. 1998. Applied Multivariate Statistical Analysis. Upper Sadlle, NJ: Prentice-Hall.

Loureiro, M. A. 1996. “Ilustrando na Clarineta a Variação e o Controle do timbre na Realização do Pensamento Musical,” in Ulhôa, M. T, ed. Anais do IX Encontro Anual da ANPPOM. Rio de Janeiro, pp. 276-283.

Luce, D. A. 1963. Physical Correlates of Nonpercurssive Musical Instrument Tones. Ph. D. Dissertation, Department of Physics, MIT, Cambrigde, MA.

Masri, P., Bateman, A. and Canagarajah, N. 1997. “A review of time-frequency representations, with application to sound/music analysis-resynthesis.” Organised Sound, vol. 2, no. 3, pp. 195-205

McAdams, S. e Bregman, A. 1979. “Hearing Musical Streams.” Computer Music Journal, vol. 3, no. 4, pp. 26-44. Reimpresso em C. Roads e J. Strawn, eds. Foundations of Computer Music. Cambridge, Massachussets: MIT Press, pp. 658-698.

McAdams, S. 1987. “Music: A Science of Mind?” Contemporary Music Review, vol. 2. no. 1, pp. 1-61.

Rabiner, Lawrence, R. e Schafer, Ronald W. 1978 Digital Processing of Speech Signals. Prentice-Hall, Inc. Englewood Cliffs, New Jersey 07632.

Rencher, A. C. 1995. Methods of Multivariate Analysis. New York: John Wiley & sons: New York.

Risset, J. C. and M. V. Mathews, 1969. “Analysis of musical instrument tones.” Physics Today, vol. 22, no. 2, pp. 23-40.

Risset, J. C. 1965. “Computer Study of Trumpet Tones.” Journal of the Acoustics Society of America. (Abstracts), vol. 38, pp. 912. 1966. Bell Laboratories Report, Murray Hill, NJ.

Risset, J. C. 1991. “Timbre analysis by synthesis: Representations, imitations and variants for musical composition.” In A. Piccialli G. de Poli and C. Roads, editors, Representation of Musical Signals, pp. 7-43. MIT Press, Cambridge, Massachussets.

Risset, J. C., and D. Wessel. 1982. “Exploration of Timbre by Analysis and Synthesis.” In D. Deutsch, ed. Psychology of Music. San Diego, California: Academic Press, pp. 25-58.

Sandell, G. J. and Martens, W. 1995. “Perceptual Evaluation of Principal-Component-Based Synthesis of Musical Timbres.” Journal of the Audio Engeneering Society. vol. 43, no. 12, pp. 1013-1028.

Serra, Xavier. 1997. “Musical Sound Modeling with Sinusoids plus Noise.” In A. Piccialli; C. Roads and S. Pope eds. Musical Signal Processing. Swets & Zeitlinger Publishers.

Smalley, D. 1986. “Spectro-morphology and Structuring Processes.” in S. Emmerson, ed. The Language of Electroacoustic Music. London: The Macmillan Press. Pp. 61-93.

Strong, W. and Clark, M. 1967a. “Synthesis of Wind-Instrument Tones.” Journal of the Acoustics Society of America. vol. 41, no. 1, pp. 39-52.

Strong, W. and Clark, M. 1967b. “Pertubations of Synthetic Orchestral Wind-Instrument Tones.” Journal of the Acoustics Society of America. vol. 41, no. 2, pp. 277-285.

Wessel, D. 1979. “Timbre space as a musical control structure.” Computer Music Journal, vol. 3, no. 2, pp. 45-52.