CONSISTENCY OF TIMBRE PATTERNS IN EXPRESSIVE MUSIC PERFORMANCE

Mathieu Barthet, Richard Kronland-Martinet, Sølvi Ystad

CNRS - Laboratoire de Mécanique et d'Acoustique
31, chemin Joseph Aiguier
13402 Marseille Cedex 20, France
{barthet|kronland|ystad}@lma.cnrs-mrs.fr

ABSTRACT

Musical interpretation is an intricate process due to the interaction of the musician's gesture and the physical possibilities of the instrument. From a perceptual point of view, these elements induce variations in rhythm, acoustical energy and timbre. This study aims at showing the importance of timbre variations as an important attribute of musical interpretation. For this purpose, a general protocol aiming at emphasizing specific timbre patterns from the analysis of recorded musical sequences is proposed. An example of the results obtained by analyzing clarinet sequences is presented, showing stable timbre variations and their correlations with both rhythm and energy deviations.

1. INTRODUCTION

This article is part of a larger project aiming at analyzing and modelling expressive music performance. To follow the classification made by Widmer and Goebl in [1], we use an "Analysis-by-measurement" approach the first step of which is to define the performer's expressive patterns during the interpretation. Various approaches to identify performance rules have been proposed. Amongst these, the "Analysis-by-synthesis" approach developed at the KTH [2] [3] which relies on musical theory knowledge has led to the establishment of context-based performance rules. They mainly take into account the tempo and the intensity of musical notes or phrases, either to emphasize their similarity (grouping rules), or to stress their differences (differentiation rules). Another approach has been proposed by Tobudic and al. [4], leading to a quantative model of expressive performance based on artificial intelligence to reproduce the tempo and dynamic curves obtained from performances played by musicians. All these studies have mainly focused on rhythm and intensity variations.

In the present study, an investigation on the consistency of timbre expressive variations in music performance is proposed. A comparison between timbre, rhythmic and intensity expressive variations is also made, since the correlations between these parameters are probably strong. For this purpose, a professional clarinettist was asked to play a short piece of music (the beginning of a Bach's Cello Suite) twenty times. The choice of the instrument was mainly related to the fact that it is self-sustained and that the performer easily controls the sound event after note onset. In addition, earlier studies by Wanderley [5], report that the movements of a clarinettist are highly consistent for various music performances of the same piece. Since these movements seem to be closely linked to the interpretation, we also expect the expressive parameters to be highly consistent. In a previous study [6], the investigation of the performance parameters of a physically modelled clarinet indicates that timbre is involved in musical expressivity and seems to be governed by performance rules. In this study, we aim at checking if timbre also follows systematic variations on natural clarinet sounds.

We shall first describe a general methodology developed to analyze and compare recorded musical performances in order to point out consistency of timbre, rhythmic and intensity patterns in expressive music performance. An application of this methodology to twenty recorded musical sequences of the same clarinettist is then given. Eventually, we show that timbre, as rhythm and intensity, follows systematic variations, and that correlations exist between these parameters of the expressivity.

2. METHODOLOGY

In this section, we describe a general methodology to analyze and compare musical performances from recorded monophonic sequences.

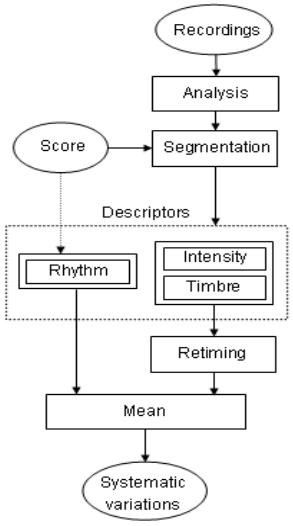

Figure 1: *Methodology.*

The hypothesis we want to verify is that when a performer plays several times a piece with the same musical intention, patterns of rhythm, intensity, and timbre over the course of the piece, show a high consistency. For that purpose, we derive from the recorded sequences some performance descriptors characterizing the musical expressivity of a performer at a note-level. We then calculate the mean of the performance descriptors to determine if

their variations are systematic. Figure 1 sums up the different steps of the methodology.

2.1. Sound corpus

If the expressive variations introduced by the musician resist an averaging over a large amount of performances played with the same musical intention, they can be considered as systematic. We thus need a large number of recordings of the same musical piece performed as similarly as possible to identify the consistency of musical expressivity patterns.

To avoid influence from room acoustics, the recordings of these performances have to take place in a non-reverberant acoustical environment.

In the following, we will note N, the number of notes of the musical melody, and n will refer to the n^{th} note played. We will note P the number of recorded performances, and p will refer to the p^{th} one.

2.2. Note segmentation

Note segmentation is an intricate task and is slowed down by difficulties such as the detection of two successive notes having the same pitch, or silences between musical phrases. In [7], the author describes a way to determine the timing of the note onsets from musical audio signals. Here the task can be facilitated by the a priori knowledge of the score giving an estimation of the fundamental frequencies. The note segmentation process is composed of two parts, the pitch tracking, consisting in estimating the fundamental frequencies of the recorded sequences, and the segmentation.

2.2.1. Pitch tracking

A lot of studies have been carried out on this subject. A review can be found, for instance, in [8].

In our case, we use the software LEA from the Genesis company to generate filtered sequences from the original recordings which only contain the fundamental frequencies of the notes played during the performances. Since these new sequences only contain a single frequency-varying sinusoidal component, it is pertinent to calculate their analytic signals $Z^p(t)$. Finally, we obtain the instantaneous fundamental frequencies $F0(t)$ due to the following relation:

$$F0^p(t) = \frac{1}{2.\pi}\frac{d\phi^p(t)}{dt} \tag{1}$$

where $\phi^p(t)$ is the phase of $Z^p(t)$.

2.2.2. Segmentation

As we have a large number of recordings, we built an automatic note segmentation method. It is also important that the process remains identical for each sequence in order to segment each note in the same way before the averaging of the performance descriptors.

Our method is based on the analysis of the fundamental frequency variations $F0(t)$. As a matter of fact, it presents instabilities at the transitions between notes. A detection of these instabilities gives the timing of the transitions. Hence, we obtain the note timings T_n^p for each note n and for each performance p.

2.3. Performance descriptors

Rhythm descriptors are obtained from the rhythm indications of the score and from data obtained after the note segmentation part. Intensity and timbre performance descriptors are high-level descriptors derived from a time/frequency representation of the recorded sequences.

2.3.1. Rhythm descriptors

We obtain the note durations D_n^p of each performance p from the note timings T_n^p. The rhythm deviation descriptor ΔD_n^p is defined as the difference between the note durations given by the score D_n^{score} (called nominal durations) and the durations of the notes played during the performances D_n^p (called effective durations):

$$\Delta D_n^p = D_n^p - D_n^{score} \tag{2}$$

It is a discrete time function calculated for each note.

2.3.2. Intensity and timbre descriptors

We derive these descriptors from a time/frequency analysis of the recorded sequences. They are also discrete functions of the time, but depending on the time bins defined by the analysis. In the following, $d^p(t)$ will refer to the descriptors calculated over the entire course of the performance p, and $d_n^p(t)$ will refer to the values of $d^p(t)$ restricted to the duration of the note n.

2.4. Retiming of the performance descriptors

To verify our hypothesis, we have to calculate the average of the performance descriptors $d^p(t)$ over all recorded sequences. As the performances are played by a human musician, the durations D_n^p of the notes are slightly different. In order to synchronize all these performance descriptors, a retiming process is thus necessary. This retiming consists in temporal contractions or dilations. We will denote by Γ these transformations. In our case, we do not need to realize an audio time-stretching keeping the frequency content of the signal as it is described for instance in [9], since the descriptors we derive from the signals are not going to be heard.

The dilation coefficient α_n^p will be chosen so as to adjust the duration D_n^p of the descriptors $d_n^p(t)$ to the mean duration $\overline{D_n}$ of the notes over all the recorded performances. Thus, we will alter the performance descriptors as little as possible. If $\alpha_n^p > 1$, Γ is a dilation, and if $\alpha_n^p < 1$, Γ is a contraction.

The mean note duration $\overline{D_n}$ is given by:

$$\overline{D_n} = \frac{1}{P}\sum_{p=1}^{P} D_n^p \tag{3}$$

The dilation coefficient α_n^p is then given by:

$$\alpha_n^p = \frac{\overline{D_n}}{D_n^p} \tag{4}$$

Finally, the retiming transformations Γ applied on the note performance descriptors $d_n^p(t)$ can be written as:

$$\Gamma : d_n^p(t) \longmapsto \Gamma[d_n^p(t)] = d_n^p(\alpha_n^p(t)) \tag{5}$$

2.5. Systematic and random variations of the descriptors

Once the synchronization of the note performance descriptors is realized, we calculate their mean to point out systematic behavior, and their standard deviation to characterize random fluctuations. The mean note descriptors $\overline{d_n(t)}$ over all the recorded performances are given by:

$$\overline{d_n(t)} = \frac{1}{P}\sum_{p=1}^{P} d_n^p(\alpha_n^p(t)) \tag{6}$$

Random fluctuations of the descriptors are characterized by their standard deviation $\sigma_{\overline{d_n}}(t)$.

Hence, if the behavior of the performance descriptors $d_n^p(t)$ is systematic over all the performances, they will be strongly correlated with their mean value, and the standard deviation will be rather low. Furthermore, the mean will be a smoothed version of the descriptors, loosing the random fluctuations. On the contrary, if the behavior of the descriptors is not systematic, then their mean will differ from the descriptors, and the standard deviation will be high.

We also evaluate the consistency of the performance descriptors by calculating the correlation coefficients $r^2(\Gamma[d])$ of the retimed observation p of the descriptor d and the $P-1$ others. The mean of these correlation coefficients $\overline{r^2(\Gamma[d])}$ measures the strength of the correlations.

3. AN APPLICATION TO THE CLARINET

3.1. Sound corpus

We asked the professional clarinettist C. Crousier to play the same excerpt of an Allemande of Bach (see Figure 2) twenty times with the same musical intention. This excerpt is destined to be played rather slowly and expressively (its score indication is "Lourd et expressif"). A 48 bpm reference pulsation was given to the musician by a metronome before the recordings. It was then stopped during the performance to give the player the freedom to accelerate or slow down. The reference pulsation allowed us to calculate the notes' nominal durations given by the score D_n^{score} and thus evaluate the performer's rhythmic deviations.

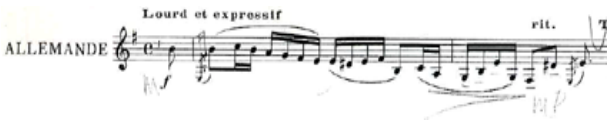

Figure 2: *Excerpt of Bach's Suite II B.W.V. 1007 (Allemande).*

The recordings of the clarinet were made in an anechoic chamber with a 44100 Hz sample frequency. We used SD System clarinet microphones fixed on the body and the bell of the instrument, avoiding recording problems due to the movements of the instrumentalist while playing. Both microphones have a flat frequency response (+/- 2,5 dB) in the frequency range where the timbre descriptors are calculated [100 - 8000 Hz].

3.2. Performance descriptors extraction

We applied the Short Time Fourier Transform (STFT) on each recorded musical sequences. Hanning windows of 1024 samples and 75 % of overlap have been used for this purpose. Timbre descriptors were calculated considering $N_{harm} = 15$ harmonics.

3.2.1. Rhythm descriptor

We normalized the rhythm descriptors ΔD_n^p given by the equation (2) according to the notes' nominal durations. Its mean expressed as a deviation percentage is hence given by:

$$\overline{\Delta D_n}(\%) = 100.\frac{\overline{\Delta D_n}}{D_n^{score}} \tag{7}$$

3.3. Intensity variations

We characterize intensity variations by the Root Mean Square envelopes of the recorded sequences.

3.4. Timbre variations

Three timbre descriptors adapted to clarinet sounds have been chosen to describe the timbre variation during musical performance: the spectral centroid, which can be regarded as the mean frequency of the spectrum, the spectral irregularity correlated to the differences between odd and even harmonics, and the odd and the even descriptors, correlated to the energy of odd and even harmonics in the spectrum. We will present a particular case showing that these timbre descriptors contain complementary information.

3.4.1. The Spectral Centroid

The definition we use for the spectral centroid SCB is the one given by Beauchamp in [10]. It differs from the classical definition by the presence of a term b_0 that forces the centroid to decrease when the energy in the signal is low, avoiding an increase of the spectral centroid at the end of the notes. It has been shown that the spectral centroid is correlated to the brightness of a sound and correlates with the main control parameters of the clarinettist, i.e. the mouth pressure and the reed aperture [11] [12]. It is defined by:

$$SCB_n^p(t) = \frac{\sum_{k=1}^{N_{sup}} k.A_k(t)}{b_0 + \sum_{k=1}^{N_{sup}} A_k(t)} \tag{8}$$

where the $A_k(t)$ are the modulus of the STFT considered up to the frequency bin N_{sup}. The term b_0 is given by:

$$b_0 = \max[A_k(t)], \;\; k = 1, .., N_{sup} \tag{9}$$

3.4.2. The Spectral Irregularity

Krimphoff has pointed out the importance of the spectral irregularity [13]. We here derived a new definition from the one Jensen gave in [14], including a term b_1 in the denominator for the same reason as for the spectral centroid. The spectral irregularity $IRRB$ can then be defined by:

$$IRRB_n^p(t) = \frac{\sum_{h=1}^{N_{harm}-1}(A_{h+1}(t) - A_h(t))^2}{b_1 + \sum_{h=1}^{N_{harm}} A_h(t)^2} \tag{10}$$

where:

$$b_1 = (\max[A_h(t)])^2, \;\; h = 1, .., N_{harm} \tag{11}$$

3.4.3. The Odd and Even descriptors

The lack of even harmonics compared to odd ones is characteristic of the clarinet timbre (see for instance [15]), but their energy increases as the breath pressure increases (see [12]). A measure of odd and even harmonics energy compared to the overall energy is given by the Odd and Even descriptors defined below. We will show a particular case where they explain subtle timbre variations of the clarinet.

$$Odd_n^p(t) = \frac{\sum_{h=0}^{N_{odd}-1} A_{2h+1}(t)}{\sum_{h=1}^{N_{harm}} A_h(t)} \quad (12)$$

$$Even_n^p(t) = \frac{\sum_{h=1}^{N_{even}} A_{2h}(t)}{\sum_{h=1}^{N_{harm}} A_h(t)} \quad (13)$$

where N_{odd} is the number of odd harmonics, and N_{even} the number of even harmonics.

4. CONSISTENCY OF THE PERFORMANCE DESCRIPTORS

4.1. Strong correlations between the performances

The mean correlation coefficients of the retimed performance descriptors are given in table 1. The high values of $\overline{r^2(\Gamma[d])}$ point out a strong consistency of the rhythm descriptor ΔD, the intensity descriptor RMS, and the timbre descriptors SCB, $IRRB$, Odd and $Even$, over the various performances.

d	ΔD	RMS	SCB	IRRB	Odd	Even
$\overline{r^2(\Gamma[d])}$	0.76	0.89	0.84	0.71	0.74	0.74

Table 1: *Mean correlations of the performance descriptors*

4.2. Rhythmic patterns

Figure 3 both shows the fundamental frequencies and durations of the notes to be played by the performer as indicated on the score and the mean of the measured fundamental frequencies and durations of the notes for the 20 performances. It points out that the total duration of the performances is on average longer than the nominal one (about 1s difference). In order to play expressively, the performer effects rhythmic deviations compared to the rhythm indicated on the score. These rhythmic deviations lead to local accelerandi or descelerandi. In general, certain notes tend to be shortened by the performer (case where $\overline{\Delta D_n} < 0$, see for example notes 10 and 20), whereas certain notes tend to be lengthened (case where $\overline{\Delta D_n} > 0$, see for example notes 5, 11 and 12). From 7s to the end, almost all the notes are played longer, up to twice their nominal durations for some of them. This reveals a slowing down of the tempo by the performer which is very common in endings of musical phrases. These results are in agreement with the "Duration Contrast" and "Final Retard" rules defined by Friberg and colleagues, which model the two rhythmic principles indicated above [2].

4.3. Intensity patterns

As can be seen on figure 5, the phrase begins forte and then there is a progressive decrescendo until the end of the phrase. The energy

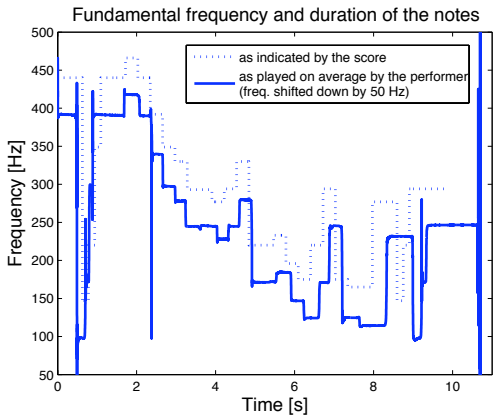

Figure 3: *Fundamental frequency and duration of the notes as indicated by the score (dotted) and as played on average by the performer (solid). The measured fundamental frequency (solid) has been shifted down by 50 Hz to point out the rhythmic differences between the nominal and the effective note durations.*

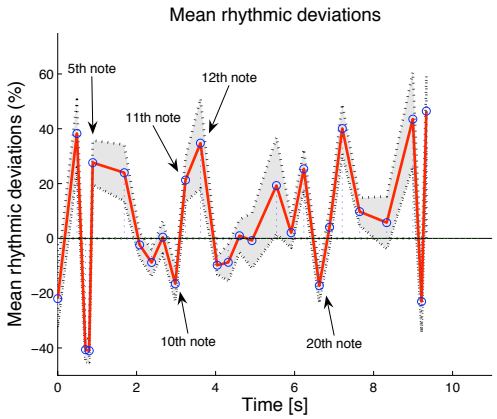

Figure 4: *Mean rhythmic deviations (bold), +/- the standard deviation (dotted). The timing of the notes are indicated by circles.*

peak at time bin 1600 may be due to the fact that the note played has a very low frequency (147 Hz) and is more radiated by the clarinet.

4.4. Timbre patterns

Figure 6 represents the mean spectral centroid and its standard deviation. It is strongly correlated to the mean intensity variations in a monotonous and increasing way ($r = 0.80$). Further, within the duration of notes, strong changes of spectral centroid can be observed for all the performances. This can easily be seen for the fifth note (around time bin 200), for which the difference between the lowest and the highest value of the mean spectral centroid (at the note onset and close to the note offset), is about 500 Hz. A neat change in the note's timbre is audible (sounds are given at http://w3lma.cnrs-mrs.fr/~kronland/DAFx06). It is worth noticing that after the attack of this note the values of the odd descriptor decrease and the values of the even descriptor increase (figure 8) so that the global energy of even harmonics grows faster than the

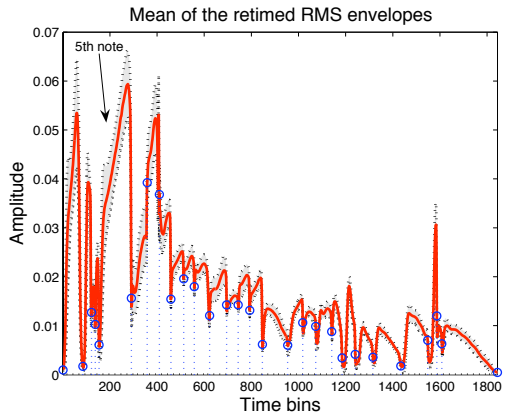

Figure 5: *Mean RMS envelope (bold), +/- the standard deviation (dotted). The note transitions are indicated by circles.*

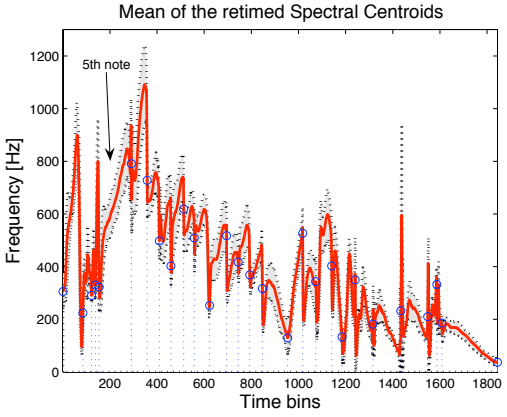

Figure 6: *Mean Spectral Centroid (bold), +/- the standard deviation (dotted). The notes transitions are indicated by circles.*

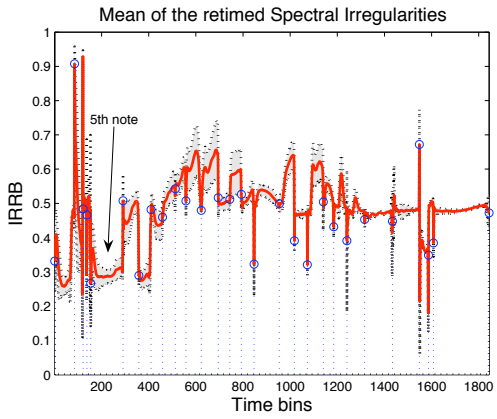

Figure 7: *Mean Spectral Irregularity (bold), +/- standard deviation (dotted). The note transitions are indicated by circles.*

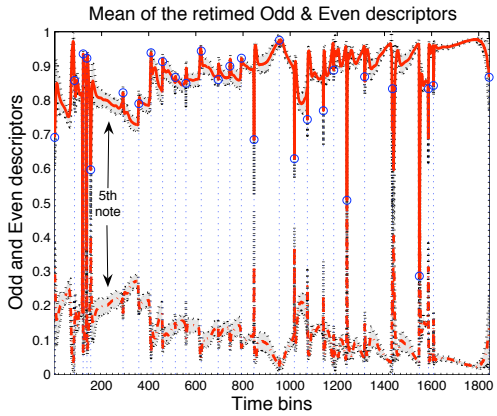

Figure 8: *Mean Odd and Even descriptors (bold), +/- the standard deviation (dotted). The note transitions are indicated by circles.*

global energy of odd harmonics. This does not mean that the phenomenon is equally distributed over the whole spectrum. If it was so, the spectral irregularity would decrease (the energy of even harmonics getting closer to the odd one) but this is not the case (see figure 7). The spectral irregularity remains quite stable within the duration of the fifth note after its attack. This is probably due to the fact that even harmonics grow faster than odd ones in a narrow frequency area. Indeed, we showed in the case of synthetic clarinet sounds that an increase of the breath pressure induces an energy increase of high-order harmonics and more particularly for even harmonics around the reed resonance frequency [12]. This is due to the non-linear coupling between the excitor (the reed) and the resonator (the bore) and explains the increase of the brightness of the sound.

4.5. Timbre and Intensity correlation

Figure 5 and 6 show that there is a strong correlation between the spectral centroid and the envelope. Nevertheless, the spectral centroid of a note depends on its fundamental frequency and this biases the observation. Hence, we have normalized the spectral centroid according to the mean instantaneous fundamental frequency as follows:

$$\overline{SCB'(t)} = \frac{\overline{SCB(t)}}{\overline{F0(t)}} \tag{14}$$

Figure 9 represents the normalized spectral centroid $\overline{SCB'}$ as a function of the normalized mean RMS envelope for two categories of notes, the short and piano ones, and the long and forte ones. It is worth noticing that these two categories of notes seem to follow different kinds of trajectories. Indeed, the spectral centroids of the short and piano notes increases very quickly compared to the envelope, whereas the spectral centroids of the long and forte notes seems to increase less rapidly than the envelope. This should be verified on a longer excerpt including a greater number of long and forte notes to cover a wider range of pitches, since the two trajectories below the diagonal correspond to the same pitch. The correlations we made are only qualitative but proves that there is a link between the variations of intensity, timbre and rhythm during the playing.

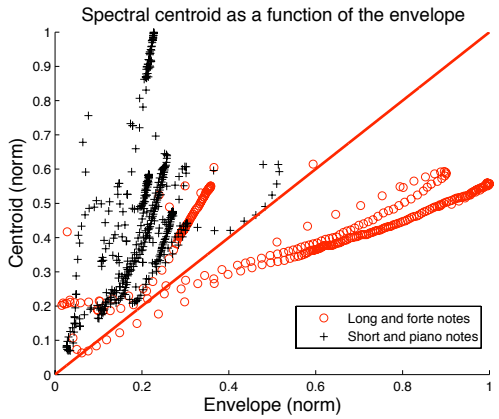

Figure 9: *Spectral centroid as a function of the RMS envelope.*

5. CONCLUSION AND FURTHER WORKS

Analysis and comparisons of recorded clarinet performances of an excerpt of a Bach Suite played several times by the same performer with the same musical intention, pointed out systematic and random variations of certain timbre descriptors. The present study confirmed former results obtained from performances played on a digital clarinet [6]. Strong correlations have been observed between the spectral centroid and the RMS envelope during the performances. In spite of these correlations, the spectral irregularity seems to be less correlated to the intensity, indicating that timbre changes are not only a consequence of intensity variations. More precise investigations on this topic are however needed to clarify the relations between timbre and intensity. Qualitative results show that timbre and intensity patterns also seem to be related to rhythmic deviations over the course of the musical piece. Multidimensional analysis are to be conducted to better understand the links between timbre, rhythm and intensity variations.

This study represents a first step to show the importance of timbre variations in expressive music performances. We plan in the future to use signal processing techniques to alter the interpretation and to evaluate the importance of variations in rhythm, intensity and timbre by psychoacoustic tests.

6. ACKNOWLEDGEMENTS

We would like to thank C. Crousier for his excellent advice and participation in this project. We are also grateful to the GENESIS company for providing the LEA software.

7. REFERENCES

[1] G. Widmer and W. Goebl, "Computational models of expressive music performance: The state of the art," *J. New Music Research*, vol. 33, no. 3, pp. 203–216, 2004.

[2] A. Friberg, "A quantative rule system for musical performance," Ph.D. dissertation, Department of Speech, Music and Hearing, Royal Institute of Technology, Stockholm, 1995.

[3] J. Sundberg, *Integrated Human Brain Science: Theory, Method Application (Music)*. Elsevier Science B.V., 2000, ch. Grouping and differentiation two main principles in the performance of music, pp. 299–314.

[4] A. Tobudic and G. Widmer, "Playing Mozart phrase by phrase," ÖFAI-TR-2003-02, Tech. Rep., 2003.

[5] M. Wanderley, *Gesture and Sign Language in Human-Computer Interaction: International Gesture Workshop*. Springer Berlin, Heidelberg, 2002, ch. Quantitative analysis of non-obvious performer gestures, p. 241.

[6] S. Farner, R. Kronland-Martinet, T. Voinier, and S. Ystad, "Timbre variations as an attribute of naturalness in clarinet play," in *Proc. 3rd Computer Music Modelling and Retrieval Conf. (CMMR05)*, Pisa, Italy, 2005, pp. 45–53.

[7] S. Dixon, "On the analysis of musical expression in audio signals," *Storage and Retrieval for Media Databases, SPIE-IS&T Electronic Imaging*, vol. 5021, pp. 122–132, 2003.

[8] E. Gomez, "Melodic description of audio signals for music content processing," Pompeu Fabra University, Barcelona, Tech. Rep., 2002, [Online] http://www.iua.upf.es/mtg/publications/Phd-2002-Emilia-Gomez.pdf.

[9] G. Pallone, "Dilatation et transposition sous contraintes perceptives des signaux audio: application au transfert cinema-video," Ph.D. dissertation, Aix-Marseille II University, Marseille, 2003.

[10] J. W. Beauchamp, "Synthesis by spectral amplitude and "brightness" matching of analyzed musical instrument tones," *J. Audio Eng. Soc.*, vol. 30, no. 6, pp. 396–406, 1982.

[11] P. Guillemain, R. T. Helland, R. Kronland-Martinet, and S. Ystad, "The clarinet timbre as an attribute of expressiveness," in *Proc. 2nd Computer Music Modelling and Retrieval Conf. (CMMR04)*, Esbjerg, Denmark, 2004, pp. 246–259.

[12] M. Barthet, P. Guillemain, R. Kronland-Martinet, and S. Ystad, "On the relative influence of even and odd harmonics in clarinet timbre," in *Proc. Int. Comp. Music Conf. (ICMC'05)*, Barcelona, Spain, 2005, pp. 351–354.

[13] J. Krimphoff, S. McAdams, and S. Winsberg, "Caractérisation du timbre des sons complexes, II Analyses acoustiques et quantification psychophysique," *Journal de Physique IV, Colloque C5*, vol. 4, 1994.

[14] K. Jensen, "Timbre models of musical sounds," Ph.D. dissertation, Department of Computer Science, University of Copenhagen, 1999.

[15] A. Benade and S. Kouzoupis, "The clarinet spectrum: Theory and experiment," *J. Acoust. Soc. Am.*, vol. 83, no. 1, pp. 292–304, Jan. 1988.