Improving Musical Expressiveness by Time-Varying Brightness Shaping

Mathieu Barthet, Richard Kronland-Martinet, and 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. Former studies show that removing brightness temporal variations of tones in an expressive clarinet performance induces a decrease in the appreciation judgments among listeners. This study aims to investigate the reciprocal process where a time-varying brightness pattern is extracted from a clarinet performance and added to MIDI sound sequences generated by sampler-based instruments. Performances with and without brightness temporal variations have hereby been generated using samples from three sustained instruments (piccolo flute, clarinet and cello) and two non-sustained instruments (guitar and sitar). 20 listeners were asked to rate the dissimilarities between the performances in terms of interpretation, and then to select the performance they preferred. Results show that the application of the brightness temporal variations extracted from the clarinet performance increases the appreciation judgment of the performances for almost all the instruments used in the experiment.

Key words: Musical expressiveness, timbre, analysis-synthesis, clarinet

1 Introduction

Music Performance and Timbre Since the beginning of the twentieth century, many experts with various backgrounds such as musicology, psychology and acoustics have worked on the characterization of expressive music performance either focusing on the control of the instrument or the produced sound [1]. The intricate process of interpretation in music is defined in [2] as "the act of performance with the implication that in this act the performer's judgment and personality necessarily have their share". As a matter of fact, the analysis of musical interpretations by measurements revealed that performers generally deviate from a regular transcription of the nominal information given by the score when playing expressively. Although studies on psychology of music and expressive music performance modeling have often dealt with timing and dynamics aspects (see for instance [3], [4], [5], [6], [7]), the influence of timbre, topic of the present work, has been much less considered. This may not be unrelated to the

fact that timbre is a complex notion which does not yet have a proper definition and is hardly measurable. Moreover, most of the fore mentioned studies deal with traditional Western tonal music known to be governed by the structure of pitches and rhythms, hiding the importance of timbre, at least at first glance. Lerdhal explains that the secondary role of timbre in traditional music is due to the fact that, as opposed to pitches and rhythms, timbre does not possess a true hierarchical organization [8]. This lack of hierarchical organization has effectively proved to be an obstacle for memorization and musical perception [9]. Although timbre does not own specific notations in the traditional Western music system and remains an ambiguous notion, it is nonetheless a very important aspect of musical sounds; already in 1938, Seashore [10] pointed out that "timbre as a fourth attribute of tone is by far the most important aspect of tone and introduces the largest number of problems and variables".

On the Definition of Timbre The notion of timbre strongly depends on the context. A composer willingly speaks about the timbres by reference to the instruments of the orchestra. This facet of timbre is called *timbre-identity* in [11] and expresses the link of timbre to the physical cause generating the sound. This primitive causal link with the sound source probably comes from the origin of the word timbre, initially used to design an instrument, a sort of drum composed of tensed strings which gave a specific "color" to the resulting sound [12]. On the other hand, musicians are often identified by their timbre (e.g. "the timbre of Coltrane"), which here refers to the specific sound that took them years to establish and which often allows them to be distinguished from other performers. This notion is designed by *timbre-individuality* in [11] but seems similar to the *timbre-identity* notion apart from the fact that the categories of sounds are different (in one case it identifies the instruments, in the other the musicians). Finally, psychoacousticians often refer to timbre as the character of the auditory sensation which permits the distinction of sounds having the same pitch, loudness and durations, an aspect called *timbre-quality* in [11]. The incoherences of the standard definitions (given in [13] and [14]) have often been reported in the literature (see for instance [15]). Schaeffer, one of the pioneers of timbre research raised one major paradox with timbre which states that each instrument has to possess its proper timbre while it generates sounds which all have their own timbres [16]! Such a phenomenon can be illustrated by transposing a low register piano tone into a higher pitched tone without taking care of the decrease of the harmonics energy as the pitch increases. The resulting sound no more corresponds to a piano tone showing that timbre also depends on the instrument range. A better understanding of the nature of timbre has therefore to deal with two complementary visions: one related to its morphology (study of its structure and form) and the other related to its typology (systematic classification of types that have common characteristics or invariants).

The Quest for Timbre Descriptors and Timbre Control Some authors looked for acoustic correlates of timbre aiming at identifying different musical

sounds and understanding their perceptual discrimination (see [17], [18], [19], [20], etc.). Such studies unveiled the multidimensional nature of timbre usually represented in a three dimensional space where each dimension is linked to some spectral, temporal and spectro-temporal aspects of sounds. These descriptors are often derived from the spectral envelope (its irregularity is a strong characteristic for cylindrical bore instruments) and amplitude envelopes of the harmonic components of the sound. Amongst the timbre descriptors, the brightness, which is linked to the center of gravity of the sound's spectrum (so-called Spectral Centroid [21] [22]), has often proved to be a relevant acoustic signature of the musical sounds. Other authors investigated ways to control timbre from these descriptors [23]. This process is often referred to as feature-based synthesis [24].

Timbre and Time: the Notion of Timbre Temporal Shape As remarked in [25], most of the explorations of timbre by analysis and synthesis have focused on isolated tones. Nevertheless, when analyzing timbre in a musical context, the notion of time is of great importance as music usually involves developments and variations used to communicate the musical ideas through the course of the composition. Timbre can then no more solely be viewed as a global characteristic of a sound but also as a process that belongs to time. In music composition, Schoenberg is one of the first who used progressions of timbres instead of the traditional progressions of pitches by developing the concept of "color" melodies (*Klangfarbenmelodie*). Given these considerations, the notions of material and shape introduced by Schaeffer can be useful to describe the morphological evolutions of timbre in time. What we would capture supposing that we could hear the timbre of a musical sound at a given instant will be called the Instantaneous Timbre Quality (ITQ). If the same musical sound is heard as a whole, one may feel a trajectory that shapes the sound material in time. This trajectory will be designed by the Timbre Temporal Shape (TTS), i.e. the way the Instantaneous Timbre Quality evolves with time.

The Perceptual Effects of Timbre Temporal Shape The focus of this study is precisely to analyze the perceptual effects of Timbre Temporal Shapes when listening to a musical interpretation. We showed in a former experiment that removing the initial brightness temporal variations of the tones of an expressive clarinet performance induced a decrease of the appreciation judgments among listeners [26]. This study investigates the influence on perception of the reciprocal process, consisting in adding brightness temporal variations issued from an expressive playing to performances played without expressivity. We namely study the consequences of applying a brightness evolution of a specific sustained instrument, the clarinet, on performances generated on other instruments either they belong to the same family or not (sustained or non-sustained instruments).

In the first part, we will explore some of the characteristics of the clarinet timbre with references to sound examples both issued from natural clarinet playing and physics-based synthesis. We will also show why the brightness is well

adapted to describe the clarinet timbre and recall the main results of the former experiment. The second part will be devoted to the methodological aspects of the present study. First, the time-varying brightness shaping procedure developed to design the stimuli is exposed. Details regarding the procedure of the perceptual test based on dissimilarity and preference judgments for paired comparisons are then given, and finally, the statistical techniques used to analyze the data (multidimensional scaling and hierarchical cluster analyses) are described. In the third part, a discussion of the results is presented, as well as an analysis of the questionnaire given to the participant to explain some of their strategies. The final part exposes the conclusions and outlooks of the present work. The various sound examples to which we refer in the text can be found at `http://www.lma.cnrs-mrs.fr/~kronland/cmmr2007/cmmr2007.html`

2 Background

2.1 Links between Control, Dynamics and Timbre

Physics-based synthesis can be of great help to uncover connections between the control of the instrument and the resulting timbre. Guillemain *et al.* developed a simplified physical synthesis model of clarinet [27]. Besides the length of the bore that acts on the fundamental frequency of the tones, the main control parameters of the model are the dimensionless mouth pressure γ and the dimensionless reed aperture ζ. A hundred clarinet tones were generated with the same fundamental frequency ($f0 \approx 170Hz$) and durations ($1s$) but varying mouth pressure (10 values in the range $0.4 \leq \gamma \leq 0.5$) and reed aperture (10 values in the range $0.2 \leq \zeta \leq 0.5$). The time-varying Spectral Centroids of the clarinet tones were computed according to the formula given by [22]. Figure 1(a) shows the mean values of the time-varying Spectral Centroids calculated over the second half of the tones in order to take into account the steady-state regime for all sounds. The loudness of the tones, computed according to the algorithm defined by Zwicker and Fastl [28] is shown in figure 1(b).

As the dimensionless mouth pressure (γ) and the dimensionless reed aperture (ζ) increase, the mean Spectral Centroids increase in a quasi-monotonous way. This puts forward that a large variety of brightness values can be obtained for the same pitch (cf. Sound Example 1 [29] to hear a series of tones corresponding to increasing values of the mouth pressure and fixed reed aperture). Hence, in addition to be a recurrent acoustic correlate of timbre-space dimensions in the case of timbres generated by different instruments [20], the Spectral Centroid also seems to well explain the discrimination of timbres produced by a single instrument, here a digital clarinet. In the case of natural clarinet sounds, the Spectral Centroid has also proved to be a relevant descriptor [30].

As can be seen in figure 1(b), the loudness of the tones also increases in a monotonous way with increasing pressure and reed aperture. Hence, considering this category of synthetic clarinet tones, the correlation between mean Spectral Centroids and perceived level is strong (loud tones are brighter than soft ones). As a matter of fact, for most instruments, a natural correlation exists between
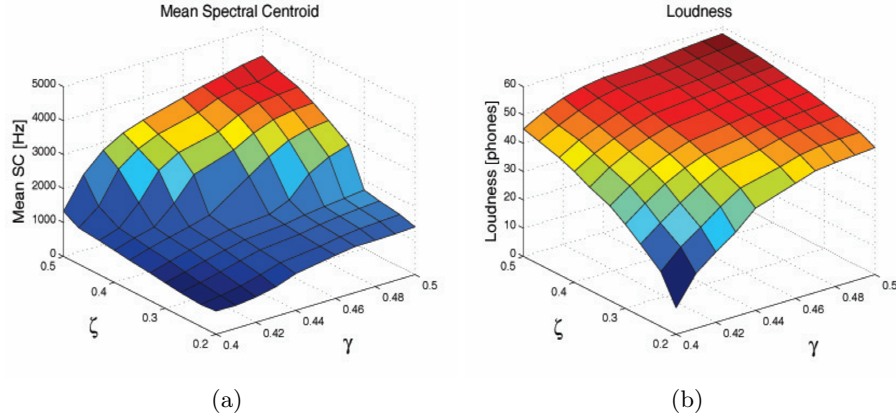
Fig. 1. Mean Spectral Centroid (a) and loudness (b) of isolated clarinet tones ($f0 \approx 170$ Hz) generated with a physics-based synthesis model for different control parameter values. The control parameters of the model are the dimensionless mouth pressure (γ) and reed aperture (ζ).

dynamics and timbre. Risset [31] showed that the signature of trumpet sounds resides in the correlation between the input force (pressure) and the spectral spread.

Natural instruments offer performers controls that are not yet taken into account by physical models and which can also act on the timbre of the generated sounds. In the case of woodwind instruments, performers can modify their embouchures to produce different effects. The consequences are that the vocal tract geometry and force applied by the lips to the reed change. Fritz [32] showed that different configurations of the vocal tract (e.g. "aw", "ee") induced different distributions of the harmonics energy in the sounds' spectra. Research is currently done to better understand the complex interaction between the vocal tract and the instrument bore [33]. We asked a professional clarinetist to play a same note with different timbres while trying not to alter the global sound level. Three tones with different sonorities were produced (cf. Sound Examples 2a to 2c [29]). The tones have approximately the same loudness (respectively 58, 61 and 62 phones). As mentioned by the player, the Sound Example 2a corresponds to a "normal" timbre and was obtained with a classical mouthpiece configuration. The Sound Example 2b was produced by taking much less mouthpiece, which consequently generated a "narrower", "tighter" timbre. The Sound Example 2c, which was obtained on the contrary by taking too much mouthpiece, induced the timbre to become "twisted". It hence seems possible to generate tones with different timbres but close sound levels on a natural clarinet. Reciprocally, modifications of the level without alterations of the timbre seem to be a harder task. The clarinetist performed a crescendo by trying to keep a stable timbre (cf. Sound Example 3 [29]). This effect can be sought in music but its accomplishment is only partially attainable. Although the color of the tone remains quite similar

during the first portion of the sound, it then brightens up. A sudden increase of brightness occurring close to the end of the tone is clearly audible. We showed that this phenomenon was linked to the changeover to the beating reed mode (situation where the reed beats against the mouthpiece) inducing an increase of energy in a frequency range close to the resonance frequency of the reed and a global increase of energy for high-order harmonics [34].

Although physics-based models can help to uncover interesting links between the control of an instrument and the generated sound, they are generally still too simplified to reproduce the subtle possibilities offered by natural instruments. We therefore preferred to focus on natural instruments to study the influence of timbre variations on musical interpretation.

2.2 Timbre Temporal Shape

As seen, timbre can be considered as a global characteristic of a sound (helping us for instance to discriminate amongst different clarinet tones), but also as a time-varying variable. Audible timbre variations can be heard within the duration of a tone (report for instance to the crescendo of Sound Example 3). In order to point out such differences, we compared the Instantaneous Timbre Qualities at the beginning and at the end of a clarinet tone. This was done by freezing in time the signals close to the tone's onset and offset (cf. Sound Example 4a to 4c [29]). The timbre variations - to which we may not have paid attention at first glance due to the speed of the change - could then be highlighted: the signal corresponding to the end of the original tone is perceived much brighter than the one corresponding to the beginning of the original tone.

These considerations on a simple clarinet tone raise many interesting questions. From the performer's point of view: is the process of modifying timbre in time used by performers ? If it is the case, is there a link between the Timbre Temporal Shape produced by the performer and his/her musical intention ? From the listener's point of view: are the corresponding variations perceived ? What are their impacts on the musical likings ?

2.3 Consistency of Timbre Temporal Shape

In a former experiment we showed that when a performer reproduces a musical excerpt several times while keeping the same musical intention, the Timbre Temporal Shape can also be reproduced in a very faithful way [35] [26]. The results were obtained by analyzing 20 clarinet performances of the first two bars of an *Allemande* of Bach. Figure 2 presents the Spectral Centroid temporal variations corresponding to five of the recorded performances. A time warping process similar to the one used by Wanderley [36] for the analysis of gestural data was first performed in order to synchronize the Spectral Centroid variations across the repetitions.

The nature of the expressive deviations of timing (subtle changes of the tones durations compared to the values indicated by the scores) and dynamics (nuance) have shown to explain some of the commonalities and differences between
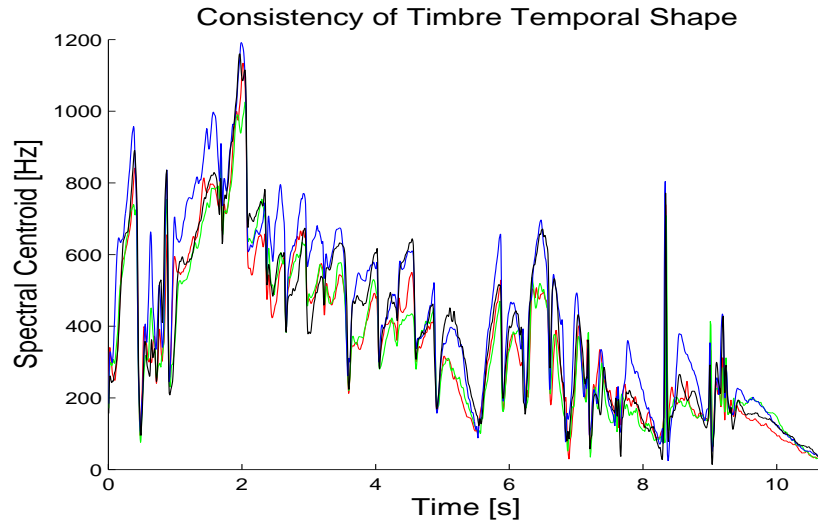
Fig. 2. Spectral Centroid temporal variations across five repetitions of the beginning of a Bach's *Allemande* played by a professional clarinetist. The Spectral Centroids were first synchronized according to the mean tones' durations thanks to a time-warping process.

performers [3] [1]. The strong correlation of the Spectral Centroid variations between the repetitions tends to indicate that the Timbre Temporal Shape (here limited to brightness aspects) could also be a signature of a performer's interpretation.

2.4 Perceptual Influences of Brightness, Timing and Dynamics Variations

There is good evidence that expressive deviations have a great effect on how listeners perceive the musicality of a performance [3] [5]. So-called "neutral", "inexpressive" or "mechanical" performances are those for which the expressive deviations are less pronounced. In a former experiment [26], the effects on music perception induced by a removal of the performer's deviations of brightness, timing, and dynamics were investigated. For that purpose, three transformations acting on an original clarinet performance were designed: (1) tones' brightness temporal variations freezing, (2) expressive timing removal, (3) dynamics flattening. The combinations of these transformations allowed us to generate seven stimuli for which the various expressive deviations (timbre, timing, dynamics) were either kept or suppressed. These musical sequences and the original expressive performance were all compared in pairs by 20 skilled musicians who were asked to select which of the two performances they prefer.

The participants' answers were then derived into marks assigned to each performance by counting the number of times a performance has been preferred. The medians of these marks are shown in figure 3.
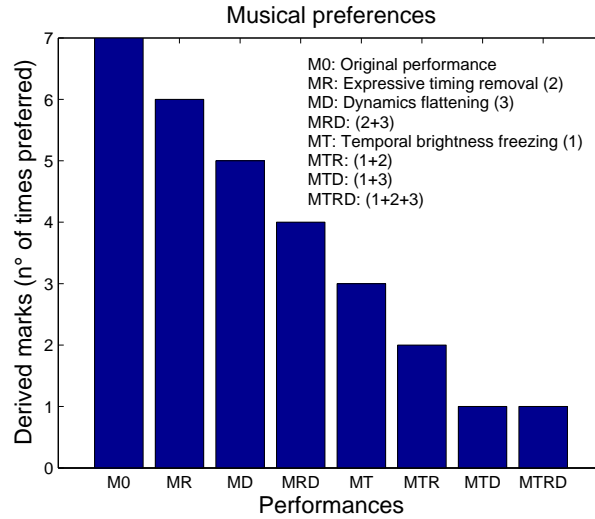

Fig. 3. Bar plot of the medians of the derived marks assigned to each performances according to the preferences of the participants. The performances are sorted according to the descending order of their marks. Given that 8 different performances were compared two by two, the marks range from 0 to 7 times preferred.

As expected, the clarinet performance with no modifications (M0) is the most expressive as it has been systematically preferred to the others. Surprisingly, the removal of the expressive deviations of timing did not induce a major loss in musical quality (see MR), but this is most likely due to the fact that Bach's music lends itself well with an interpretation with minor variations of tempo. Neither did the dynamics flattening (see MD) although the appreciation of the performances decreases significantly when both expressive timing removal and dynamics flattening transformations are combined (see MRD). Conversely, all the performances for which the brightness variations were removed (see MT, MTR, MTD, MTRD) were on average the least preferred. Similar results were obtained with another musical excerpt from a Mozart piece. This tends to prove that the tones' temporal variations of brightness are an important factor of expressiveness in clarinet interpretation.

Since removing the brightness deviations decreases the quality of an expressive performance, one might reciprocally expect that adding appropriate deviations of brightness to a "neutral" performance could improve the interpretation. We therefore conducted a second experiment aiming at comparing performances

presenting "neutral" brightness temporal variations and performances presenting brightness temporal variations issued from an expressive playing.

3 Method

3.1 Stimuli

An easy way to generate plausible but inexpressive performances in a reproducible way is to use samplers that control natural recorded sounds. These can be driven by MIDI (Musical Instrument Digital Interface) data which main parameters are note durations, pitch, and velocity. The velocity is closely linked to sound level and timbre. Hence, by using a constant velocity, sound level and timbre variations can be minimized. A MIDI-score was derived from the expressive clarinet performance of Bach's *Allemande* used in the previous experiment. This was done by extracting the tones' timing information from the acoustical signal and by using the pitch information given by the original score (see [35] for details on the procedure). This midi-score was then used with six different sampler-based instruments as resumed in figure 4.
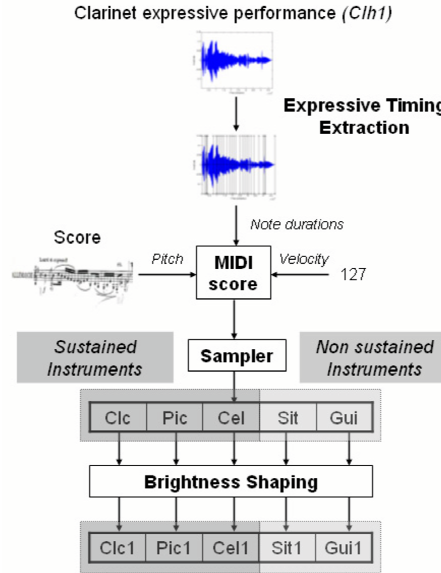

Fig. 4. Design of the stimuli

Three sustained instruments, the clarinet (*Clc*), the piccolo flute (*Pic*), the cello (*Cel*), and two non-sustained instruments, the guitar (*Gui*), and the sitar

(Sit) were selected. These generated sequences hence present no musically mean-ingful intensity or timbre variations (fixed velocity) and are all time-synchronized on the expressive clarinet performance used as a reference ($Clh1$).

A feature-based synthesis technique was developed to adjust the brightness temporal variations of each input performance according to those of the reference. It relies on the use of a dynamic low-pass filter aiming at sculpting harmonically rich signals so that their brightness match the expected value. This procedure appears to be easier than adding energy to harmonically poor signals to increase their brightness (in such case a high-pass filtering would induce an amplification of the high-frequency noise which is not desired). The input signals should therefore be brighter than the reference signal. This is in practice done by setting the velocity of the sampler to the highest value (127). The adjustment method is based on the optimization of the cutoff frequency F_c of a dynamic low-pass filter. Dynamic filter optimization has for instance been used in Aramaki and Kronland-Martinet [37] to determine the parameters of an impact sounds synthesis model. Figure 5 describes the different steps of the time-varying brightness shaping transformation.
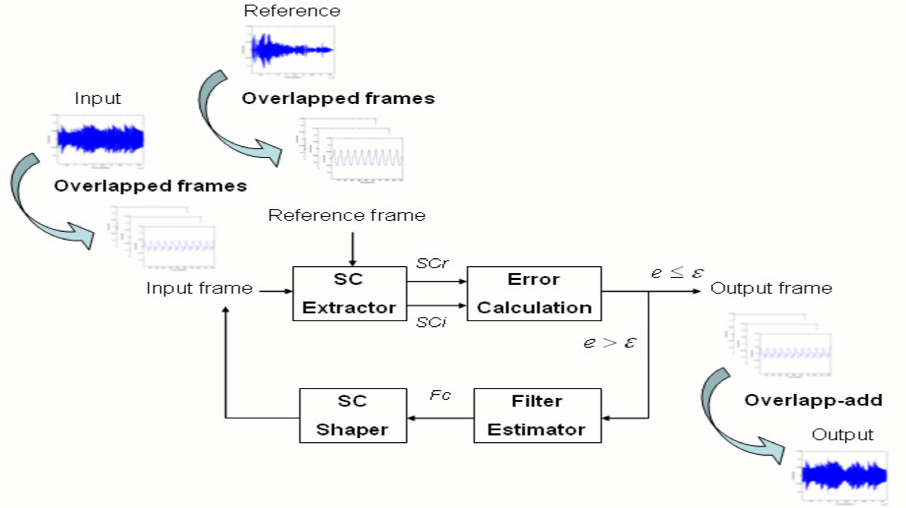

Fig. 5. Time-varying brightness shaping transformation of an input signal according to a reference signal. For each frame, the cutoff frequency F_c of a low-pass filter is optimized so as to minimize the error function e based on the distance between the Spectral Centroid values corresponding to the input SCi and reference SCr.

The reference and input signals are first segmented into overlapped frames (using 4096 samples Hann window with 50% overlap). The window size is long enough to detect the lowest fundamental frequency of the sequence (D2 - 146.83 Hz, sample frequency $f_s = 44.1$ kHz). At each time step t, the instantaneous

Spectral Centroids of the reference $SC_r(t)$ and input $SC_i(t)$ signals are derived from the Discrete Fourier Transforms of the current frame. A 2nd order lowpass digital Butterworth filter is used for brightness shaping. The unknown cutoff frequency of the filter is estimated by minimizing the error criterion $e(t)$ defined by the following expression:

$$e(t) = |SC_i(t) - SC_r(t)| \qquad (1)$$

A bound constraint is added in the minimization process so that the cutoff frequency F_c of the filter satisfies $0 < F_c < \frac{f_s}{2}$. As the gradient of the objective function e cannot be computed analytically, a finite-difference approximation to the Hessian matrix of the Lagrangian associated to e is used. The BFGS (Broyden, Fletcher, Goldfarb, Shanno) Quasi-Newton optimization method was chosen to update the Hessian [38]. The optimization loop ends up when the error criterion is inferior or equal to a threshold set up to $\epsilon = 1e^{-15}$. The output signal is finally obtained by performing an overlap-add on the optimized input frames.

The results of the time-varying brightness shaping transformation of the sampler-based clarinet performance (Clc) according to the reference performance $(Clh1)$ are shown in figure 6. The Spectral Centroid adjustment process works well as the error between the time-varying Spectral Centroid of the output SCo and the reference SCr is almost null. The process has only a minor effect on the dynamics of the sequence as the differences between the Root Mean Square envelopes of the input and output performances are negligible.
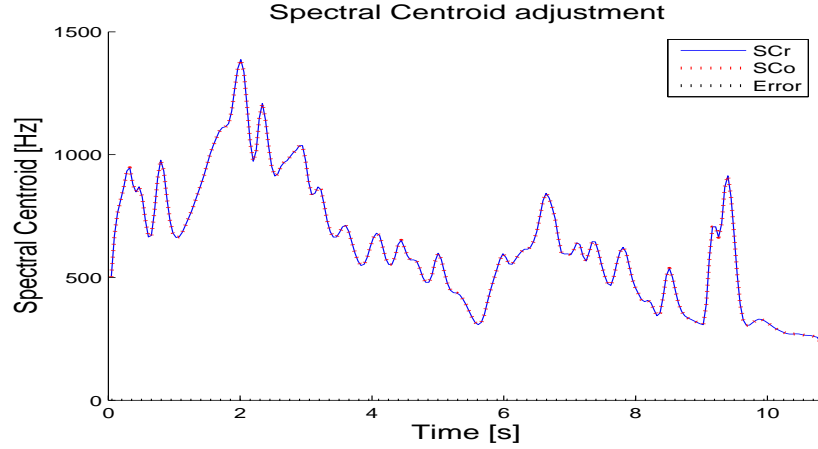

Fig. 6. Results of the optimization loop used in the time-varying brightness shaping procedure applied on the sampler-based clarinet performance (Clc). The time-varying Spectral Centroid of the output performance (SCo) matches well the one of the reference (SCr).

The generated MIDI sequences Clc, Pic, Cel, Gui, Sit are very bright since the highest velocity value is used as explained previously. Their corresponding sculpted versions $Clc1$, $Pic1$, $Cel1$, $Gui1$, $Sit1$, each of which follow the same Spectral Centroid variations extracted from the natural clarinet performance $Clh1$, have a much lower global brightness. In order to form an homogeneous group of stimuli, the mean Spectral Centroids of the input sequences (mean value of the time-varying Spectral Centroid) were adjusted on the one of the reference. This was also done by optimizing the cutoff frequency of a digital lowpass filter. The resulting sequences were all equalized in loudness. The set of stimuli used in the experiment comprise 11 sequences: the original expressive clarinet performance ($Clh1$), the five "neutral" performances ($Clc0$, $Pic0$, $Cel0$, $Gui0$, $Sit0$) and their five brightness shaped versions ($Clc1$, $Pic1$, $Cel1$, $Gui1$, $Sit1$). The index "0" hence refers to the "neutral" timbre temporal patterns whereas the index "1" refers to the expressive timbre temporal pattern. Sound Examples 5a to 5c respectively correspond to the expressive clarinet performance ($Clh1$), the "neutral" piccolo flute performance ($Pic0$) and its brightness shaped version ($Pic1$) [29].

3.2 Participants

Twenty participants (aged 23-50 years old, 6 female) took part to the experiment. Most of them came from our acoustic laboratory. 15 had received a musical training and 5 were non musicians. The musicians had received on average 12.7 years of musical practice. All had normal audition.

3.3 Apparatus

The experimental user interface was developed in the Matlab environment and was run on a Apple iMAC G4 workstation. The experiment took place in a soundproof audiometric cabin. The participants listened to the sounds with a *STAX SRM-310* headphones system including a dedicated stereo amplifier.

3.4 Procedure

As asking individuals to rank a group of performances from first to last according to their preference may not correspond to a natural "behavior", a method of paired comparisons was chosen. As a matter of fact, preferences may not be transitive (if there are three objects to be compared, A, B, and C, a participant may prefer A to B, B to C but prefers C to A).

The experiment comprised three different stages (listening, training and test). During the listening stage, the participants were asked to listen to some of the 11 stimuli (distributed randomly on the computer screen) in order to get familiarized with the range of variation occurring between the performances. After four practice trials, the listeners attended the test stage based on paired comparisons. 55 pairs resulting from the combinations of the 11 stimuli were used.

All pairs of stimuli were presented in a random order. Within a pair, the first and second stimuli were also chosen randomly. This was done to further minimize possible order effects of presentation between and within the pairs. For each pair, listeners were first asked to rate the dissimilarities in terms of musical interpretations between two performances. They were then asked to select which of the two performances they preferred. The dissimilarity ratings were made by locating a cursor on a scale with the mouse. The scale boundaries which were given in french corresponded to "very similar interpretation" and "very dissimilar interpretation". The ratings were transcribed on a scale from 0 to 1. The participants were allowed to listen to the pairs as many times as they wished. Care was taken to explain that they had to judge the differences in musical interpretation and not solely the differences between the instruments. They were told that the different performances were all synchronized in time. At the end of the test, they had to fill in a questionnaire concerning the strategy(ies) they used to differentiate the interpretations and their preferences.

3.5 Statistical Analyses

Notations In the following, the total number of participants and stimuli will respectively be noted N_p and N_s.

Analysis of the Performance Dissimilarities A hierarchical clustering (average linkage technique) was first performed in order to detect if the ratings of the musicians and the ratings of the non musicians would systematically differ from each other. In such a case, it would be preferable to conduct further analyses in separate ways for the two groups for the sake of clarity of the results. Dissimilarity measures between the participants were based on the Pearson correlation coefficient (by complementation to 1).

The dissimilarity ratings of the selected participants were then averaged and processed with a non metric multidimensional scaling analysis (MDSCAL algorithm [39]). As the perceptual ratings do not necessary respect ratio or interval scale assumptions (the differences between any two numbers of the scale may not have a meaning), a non metric procedure was chosen. For such procedures, an ordinal scale assumption is sufficient. The algorithm is built to keep a monotonous relation between the perceptual distances and their corresponding distances in the MDS space so that the rank orders are best respected. Kruskal's stress was selected as the goodness-of-fit minimization criterion. The initial configuration of points was found using the classical multidimensional scaling CMDSCAL solution. The rate of decline of the stress as dimensionality increases and Shepard diagrams helped us define the reliable number of dimensions retained for the MDS space.

Parallel to the MDS analysis, a hierarchical clustering (complete linkage technique) was also performed on the averaged dissimilarity ratings to investigate if the performances could be separated into distinct classes according to their perceived specificities. A confrontation of the results obtained with the MDS

and the cluster analysis was further done. It is important to specify at this point that the purpose of the MDS analysis is not here to verify if the dimensions of the MDS space describe a perceptive continuum. As a matter of fact, we do not expect the stimuli to be continuously distributed along the dimensions of the perceptual space since the stimuli were not designed by continuous evolutions along some predefined acoustical dimensions. Due to the design of the stimuli, the nature of the sound corpus tends to be of the categorical type: the performances present two kinds of brightness temporal variations, the "neutral" ones (indexed by 0) and the expressive ones (indexed by 1). The purpose of the MDS analysis is here to get a spatial representation of the stimuli helping to better visualize the perceptual distances between them (which appear less clearly on a dendrogram representation).

Analysis of the Musical Preferences The preference measurements consist in binary nominal data (the preference of a performance A over another one B equals 1 if A has been preferred and 0 otherwise). For this reason and in order to have an estimation of the degree of agreement in the preferences of the participants, a nonparametric measure of association, the Kendall coefficient of agreement (u) for paired comparisons [40], has been used. It is computed from the sum of the individual preferences in the sample. It is worth noticing that, independently of the number of raters, u can attain the maximum value of 1 when there is complete agreement. Alternatively, when there are more than two raters, the minimum value of u cannot be -1 because they cannot all be in disagreement. In order to get a measure of association ranging from 0 to 1 independently from the number of raters, we also computed the Wt coefficient defined by Siegel (see [40]) and which derives from u.

The test of significance of u determines the probability that random samples presenting no association would yield a correlation as large as or larger than the one observed. The chosen level of significance was $\alpha = 0.01$. As the number of participants is large ($N_p > 6$) and the number of stimuli being ranked is large ($N_s > 8$), a large-sample approximation to the sampling distribution is used. In this case, the test statistic (X^2) derived from u, is asymptotically distributed as a chi-square distribution whose number of degrees of freedom dof is the total number of pairs of stimuli ($dof = 55$).

The step described above consists in analyzing whether or not the musical preferences of the participants are related. The way the transformations defined in section 3.1 influence the musical preferences of the participants was then investigated. For that purpose, for each participant, the performances were attributed a mark derived from their frequencies of preference when compared to the other ones. The distributions of these marks across the sample hence characterize the ranking of the performances by the participants.

4 Results and Discussion

4.1 Performance Dissimilarities

The hierarchical clustering analysis made on the dissimilarities between the participants did not reveal specific distinctions between ratings of musicians and non musicians. Although we will not infer general consequences of such results since the number of participants representing each category is disproportionate in this experiment (15 musicians versus 5 non musicians), we have chosen to pursue further statistical analyses by considering the data from all the samples (20 participants).

The MDS method yields a 3-dimensional space ($stress \approx 0.07$). The projection of the MDS solution on the first two dimensions is shown in figure 7. The figure also shows the two main levels of categorization from the cluster analysis made on the dissimilarities between the performances. The resulting dendrogram is given in the appendix (see figure 9).
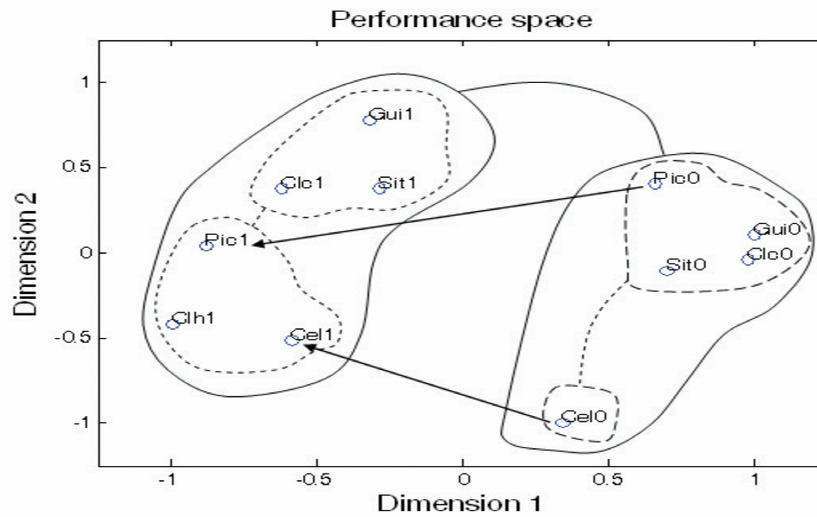

Fig. 7. Two-dimensional projection of the MDS configuration corresponding to the 11 performances' dissimilarity ratings. The first two levels of grouping resulting from the cluster analysis are superimposed on the MDS solution. The first two main clusters are represented in solid lines and the two subordinate clusters in dashed lines. The shape of the clusters has no other meaning than to represent the groups formed by the cluster analysis. The arrows put forward the change of positions of the performances played by the piccolo flute (Pic) and cello (Cel) when their brightness temporal variations are adjusted to imitate that of the expressive clarinet performance ($Clh1$).

The first dimension of the MDS performance space clearly opposes the performances with "neutral" brightness temporal variations (symbolized by "0")

from the ones with expressive brightness temporal variations (symbolized by "1"). This distinction made by the listeners according to the type of brightness pattern is confirmed by the cluster analysis. The two main clusters (solid lines) located on both sides of the first dimension indicates that the listeners judged performances possessing the same brightness temporal patterns as similar. It is worth noticing that when evaluating the similarity of two musical interpretations, the type of timbre pattern prevailed on the type of instrument that played the excerpt. Although the clarinet and the guitar belong to two different families of instruments and would be far from each other on a hierarchical structure of timbre, the performances played on these two instruments can be judged as very close as long as their brightness temporal variations are similar (see for instance the couples $(Clc1, Gui1)$ and $(Clc0, Gui0)$. The arrows which have been superimposed on the figure are intended to point out the perceptual effects of the time-varying brightness shaping transformation according to the reference clarinet performance $Clh1$. The reference clarinet performance $Clh1$ seems to act as a magnet on the other performances when their brightness temporal variations are adjusted: see for instance the positions of the performances played by the piccolo flute and the cello before and after the transformations $(Pic0 \rightarrow Pic1$ and $Cel0 \rightarrow Cel1)$. Although this effect of perceptual closeness to the reference appears globally for all the instruments, it is less pronounced for the plucked string instruments, the sitar and the guitar. For such non-sustained instruments, transforming the brightness variations according to the ones of a sustained instrument like the clarinet, may be inappropriate.

As figure 9 reveals, the subordinate clusters seem more related to the specificities of each instrument: the plucked string instruments (guitar and sitar) having the same timbre temporal pattern remain grouped; $Gui0$ and $Sit0$ are classified in the same sub-cluster, as well as $Gui1$ and $Sit1$. The same happens for the two sustained wind instruments; the piccolo flute and the clarinet ($Clc0$ and $Pic0$ belong to a same cluster as well as $Clc1$ and $Pic1$). The fact that instruments of a same family remain grouped appears legitimate. The perception of a musical interpretation is not just a function of the expressive intentions of a performer (either they are issued from a natural playing or modeled in an artificial way), but remains linked to some global aspects of the sounds related to the acoustic properties of the source.

Hence, the modification of the brightness variations of a sequence according to those of an expressive performance has an effect on the resulting musicality of the performance. The following study of the musical preferences of the raters will help to uncover how this effect affects the listeners' musical likings.

4.2 Musical Preferences

The value of the Kendall coefficient of agreement based on the preferences of the 20 participants is $u = 0.21$. The corresponding coefficient of concordance value is $Wt = 0.25$. The derived statistic $X^2 = 277$ ($p < 0.0005$) distributed as a chi-square distribution with 55 degrees of freedom indicates that the hypothesis of no association among the participants may be rejected with the $\alpha = 0.0005$

level. We conclude that there is a strong agreement among the participants in their preferences. The transformations on the brightness hence seem to produce a common effect on the likings of the listeners. The distributions of the marks assigned to each performance across the participants are described by figure 8.
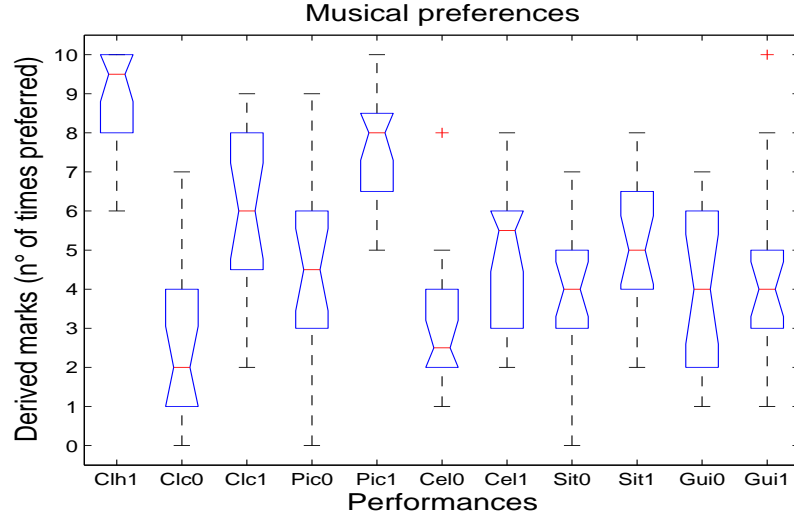

Fig. 8. Box and whisker plots of the marks assigned to each performance across the participants. The marks range from 0 to 10. The box has lines at the lower quartile, median, and upper quartile values. The length of the whiskers is 1.5 the interquartile range. Outliers are represented by crosses. The box notches represent an estimate of the uncertainty about the medians.

Given that there are 11 stimuli which have been compared in a pairwise way, the marks corresponding to the frequencies of preference of a performance to the others range from 0 to 10. The brightness transformation has a striking influence on the preferences of the participants. The performances which went through the time-varying brightness shaping according to the reference $Clh1$ tend to be preferred to their corresponding "neutral" versions. However, this seems to depend on the type of instrument. In order to compare the different rates of preference increase between the instruments, we computed the rate of change of the performance marks before and after the brightness transformations (cf. respectively to the sequences indexed by 0 and by 1). The rates were calculated from the medians of the performance marks and expressed as a deviation percentage from the initial mark. They are presented in table 1.

The increase of preferences after transformation is strong for all the sustained instruments (clarinet, piccolo flute, and cello). The change of brightness temporal variations is most efficient for the clarinet Clc (200% of increase). This is most likely linked to the fact that reference brightness temporal variations were

Table 1. Rate of increase of the preferences due to the time-varying brightness shaping. The rates are given for each instrument as a deviation percentage between the median marks of the performances before and after the transformation.

Instruments	Clc	Pic	Cel	Sit	Gui
Rate of preference increase (%)	200	78	120	25	0

extracted from the same type of instrument, a clarinet ($Clh1$), and that the listeners judged the effect as consistent. The increase of preferences is much less pronouced or even null for the non-sustained instruments (25% for the sitar and 0% for the guitar). Although the participants did judge the interpretations of $Sit1$ and $Gui1$ closer to the reference than their "neutral" versions, they may have found the consequence of the transformation incoherent with the acoustic behavior of the instruments. Effectively for plucked instruments, the coupling between the excitor and the resonator appears just at the beginning of the sound and this prevents the performer from acting on the timbre of the sound during its production.

4.3 Analysis of the Questionnaire

Several points that arise from the analysis of the questionnaire of the participants are worth to be mentioned. When questioned on the stimuli, the participants often referred to a specific performance which was often preferred to the others as it appeared more expressive and more natural (notably due to the "breath noise" and the "key noise"). These distinctions clearly refer to the reference performance played by the clarinet player $Clh1$. The fact that it was recorded in an anechoic chamber here explains the very precise rendering of the natural clarinet playing (there is no reverberation that could diminish the presence of accessory noises). These remarks coincide with the statistical analysis: this sequence is the one that is the leftmost along the first dimension of the MDS performance space (see figure 7) and which obtained the highest mark (the mark's median is 9.5/10 as seen in figure 8). Even after application of the brightness transformation the other performances were judged as less expressive than this one. It is worth noticing that the accessory noises of the clarinet playing appear in this context as a positive attribute of the sound and contribute to its pleasantness. Conversely, the cello performance $Cel0$ was often described as sounding unnatural. This sequence effectively appears in a distinctive subordinate cluster within the cluster of "neutral" performances (see figure 9) and has been poorly marked (see figure 8). These considerations may be linked to the low-pass filtering performed during the design stage of the stimuli applied to adjust the global brightness of the different performances and making the timbre of the cello too dull. Nevertheless, the interesting point is that the brightness transformation brings the cello performance closer to the reference and that the modified version is considered as a much better performance (average increase of the preferences of 120%).

Three main factors explaining how the participants rated the performances emerged from their remarks. The first one is linked to the dynamics aspects of the performances (definition of dynamics from the Freedictionary [41]: "variation in force or intensity, especially in musical sound"), the second one is related to the articulation, and the third one refers to timbre (the order of importance is linked to the frequencies of verbalizations associated to each factor). The following examples have been translated from French.

Dynamics Factor Some interesting examples of verbalizations linked with the dynamics factor are: "nuance", "amplitude variations", "sound intensity", "accents, location in time", "force fluctuation of the tone", "shape of the tone (development in time, round, flat, etc.)". It is worth noticing that some analyses come from the musical domain ("nuance", "accents", "shape"), whereas others come from the acoustical domain ("amplitude", "intensity", "force") which might be linked to the fact that most participants work in the acoustic field. It is striking to see the strong presence of words expressing the notions of variation or shape of the acoustical signal with time. Nevertheless, although the transformations of the brightness temporal variations only slightly affected the acoustical level (characterized by the Root Mean Square envelopes) of the performances, they undeniably seem to alter their perceived level. Experiments made by Zwicker showed that the annoyance produced by sounds of equivalent loudness increases when the acuity (auditive equivalent of the physical Spectral Centroid) of the sounds increases [28]. What prevail for the perception when modifying the Spectral Centroid variations of a musical performance hence seems to be mainly related to loudness variations. As no accurate models of loudness yet exist for non-stationary sounds (as are the sounds of a musical performance), we did not have means to verify by measurements the considerations evoked above. It is however interesting to see that the correlations observed between force, dynamics and timbre in the production process of the sounds seem to have a counterpart in the perception process.

Articulation Factor The use of expressions such as "fluidity", "flow", "legato", "attack" lead to think that modifying the brightness temporal variations has an effect on the articulation of the performance. Adding brightness variations issued from an expressive clarinet performance to the sequences generated by the sampler hence seems to enhance the coherence or the links between the tones. Effectively, sampler-based performances provide poor connections between the tones since they constitute a patchwork of isolated tones recorded separately.

Timbre Factor References to the third timbre factor were "brightness of certain tones", "sound quality", "heat of the sound", "attack". Faure [42] showed that the verbal descriptions of timbre by the use of a vocabulary coming from other sensorial modalities ("brightness", "heat") are common. Note that the term "attack" has been placed in the articulation and timbre factors, as it appears to be connected to both factors. The fact that few participants directly referred to timbre may be linked to the fact that it is still a misunderstood notion and

that most people only consider the typological facet of timbre by associating the timbre to the instrument, the notion of timbre variation then having no meaning.

The free verbalizations of the participants hence give interesting cues to understand the strategies they used to judge the stimuli. It brings complementary elements to the cluster and the MDS analysis that the latter cannot reveal. The modification of the brightness variations seems to affect the perception of dynamics, articulation and timbre aspects in a combined manner. It (indirectly) induced the listeners to divide the performances into two different categories: a category gathering the expressive and preferred performances with variations providing them coherence and liveliness (cluster on the left along the first MDS dimension), and a category of performances which are less preferred relatively to the performances of the first category, since they are less homogeneous and present less variations (cluster on the right along the first MDS dimension).

Independently of their musical experience, several participants reported that they found the test difficult. The evaluation of an intricate process such as music performance has already been reported as problematic; regarding the criteria used by raters in such tasks, the author of [43] states that "the measures employed are typically subjective judgments based on irregular and uncontrolled observations" even when the raters are professional music teachers. For that matter, research on the design of specific music performance rating scales based on the use of a common set of evaluative dimensions was conduced in [44]. In our case, at the expense of the difficulty of the task, we did not use such common sets of music descriptors or semantic differential scales not to bias the evaluation process by an a priori knowledge. Difficulties in rating the dissimilarities between the interpretations may also have been due to difficulties in using the rating scale to its whole extent: the two main groups obtained by the cluster analysis tend to show that the interpretations were judged either very close, either very different. The complexity of the task may also be correlated to the relatively long duration of the musical excerpt (approximatively *10 s*). However, long sequences have already been employed successfully in paired-comparison tasks. For example, in [45], the authors used sequences of *15 s* duration in a study on the perceived quality of sound-reproducing systems (after having judged that 30 s long sequences were inappropriate for paired comparisons). The comparison or assignment of rates to sounds of long durations may be hard tasks for the perceptive and cognitive systems. Supposing that we would know how the ear integrates the information it gets across time, the final judgment still can be biased by a preponderant influence of the last heard events due to a memorization process [46]. A so-called recency effect has been reported by one of the participant who was more sensible to the three last tones of the performances. We hope this effect was minimized across the sample thanks to the possibility the participants had to hear the stimuli as many times as they wished.

5 Conclusions and Outlooks

The exploration of the role of timbre in music performance is a large problem. We decided to focus our study on one aspect of timbre, the brightness, which is quantified by the Spectral Centroid derived from the Fourier analysis of the spectrum. Brightness has caught our attention for several reasons. In the case of the clarinet as for most instruments, a variation of the input force induces a variation of brightness of the resulting sound. The control of the instrument hence generally offers performers sufficient degrees of freedom to act on this dimension of timbre. From the perceptual point of view, brightness changes seem to well explain how the auditory system discriminates different musical sounds [20]. Regarding music perception, Wessel demonstrated that brightness could control stream segregation in a similar way as pitches: when a same ascending triplet of tones of different brightness is repeated in time, the ear tends to follow tones with the same brightness [47].

Even at a tone level, audible brightness variations can occur. We highlighted such variations by time-stretching small portions of sounds close to the onset and the offset of a clarinet tone (cf. Sound Examples 4a to 4c [29]). The study of various repetitions of an excerpt from a Bach piece by a professional clarinet player have shown that Spectral Centroid variations are highly consistent across the repetitions [26]. There hence seems to be a link between the expressive intention of the performer and the way he manipulates timbre. A first perceptual experiment showed that clarinet performances generated with tones having no brightness variations were judged as inexpressive [26]. Conversely, in this study we investigated the effect of applying the brightness temporal variations issued from an expressive clarinet performance to sequences generated by samplers set up with a fixed velocity. The performances appear to be discriminated by listeners according to their temporal brightness pattern. This shows that modifications carried out on brightness temporal variations while keeping other musical modalities constant (timing, dynamics) change the perception of the musical interpretation. A performer might then change the significance of the musical message he conveys to the listener by solely acting on the timbre dimension. Secondly, the performances that were considered as the most expressive are those with brightness temporal variations issued from an expressive clarinet playing. Improvements of listeners' musical preferences occurred especially for sustained instruments (clarinet, piccolo flute, cello) for which timbre variations are known to be possible during the course of the tones. For non sustained instruments (guitar, sitar), the much less pronounced benefits of the transformation on the affect judgments of the participants tend to show its incoherence for this type of instruments. This might be related to the impossibility for listeners to recognize familiar sounds which could have been produced with these instruments. As mentioned in [8], sounds that are identifiable tend to be more appreciated than sounds that we are not accustomed to. The incapacity to alter the brightness during the course of the tones in a standard guitar playing may not be extraneous to the success met by the famous "wah-wah" effect pedal which acts on the sound spectrum by means of a sweeping bandpass filter.

Being inspired by the demonstration of Sundberg which shows that reversing the tones' subtle timing deviations from the score occurring in an expressive performance (a lengthening becoming a shortening and vice versa) induced absurd musical performances, we generated excerpts for which the brightness temporal variations were reversed during the course of the tones (an increase of brightness becoming a decrease and vice versa). The result clearly sounds unmusical and seems to go against a natural playing (cf. Sound Example 6a and 6b, respectively generated with normal and reversed tones' brightness variations [29]). Deviations of timing and dynamics participating to the musicality of a performance are designed by "expressive deviations" in the literature (see for instance [1]). It appears legitimate to use the qualification "expressive deviations of timbre" when referring to timbre temporal variations that procur the sensation of expressiveness.

The notions of timbre variations and shape, emerged from the various statistical and semantic considerations, have shown to be preponderant for the listeners' affect judgments. The prospects of these results are interesting regarding the modeling of expressive performances. Models aiming at efficiently reproduce an expressive play on instruments allowing temporal modifications of timbre should then take into account the temporal variations of brightness in addition to timing, dynamics, intonation, etc. A more systematic description of the different timbre temporal variation profiles (shapes, role of the range of variation, derivative of the brightness profile, etc.) could improve the understanding of musical expressiveness perception. The importance of timbre in music has already been sensed by contemporary composers. When analyzing the temporal structure of the piece *Chronochromie* from Olivier Messiaen, Lerdhal [8] fairly remarks that: "Timbre is no more a simple "color", it becomes an essential element holding shape." (translation). In his preface to the score of *Chronochromie* Messiaen effectively remarks that one of the functions of timbre is to bring out the framework of durations (which is no more handled in by the structure of pitches).

We introduced the concept of Timbre Temporal Shape to include that aspect of timbre in music analysis and perception which refers to timbre temporal variations and which may be put in the shade by the classical psychoacoustic definition of timbre, the latter tending to exclude timbre from time. We do not however pretend to have defined the notion of Timbre Temporal Shape in a rigorous way yet and hope, on the contrary, that it will give rise to further investigations.

References

1. Gabrielsson, A.: The Performance of Music. In: Psychology of Music, Academic Press, 2nd ed. (1999)
2. Scholes, P. A.: The Oxford Companion to Music, p. 521, Oxford University Press, 2nd ed. (1960)
3. Repp, B. H.: Diversity and Commonality in Music Performance: an Analysis of Timing Microstructure in Schumann's Träumerei. J. Acoust. Soc. Am., vol. 92, no. 5, pp. 2546–2568 (1992)
4. Palmer, C.: Music Performance. Annu. Rev. Psychol., vol. 48, pp. 115–138 (1997)
5. Sundberg, J., Friberg, A., Frydén, L.: Rules for Automated Performance of Ensemble Music. Contemporary Music Review, vol. 3, pp. 89–109 (1989)
6. Widmer, G., Goebl, W.: Computational Models of Expressive Music Performance. J. New Music Research, vol. 33, no. 3, pp. 203–216 (2004)
7. De Poli, G.: Expressiveness in Music Performance. In: Algorithms for Sound and Music Computing, Creative Commons (2006)
8. Lerdahl, F.: Les Hiérarchies de Timbre. In: Le Timbre, Métaphore pour la Composition, pp. 182–203, I.R.C.A.M. (1991)
9. Deutsch, D.: Grouping Mechanisms in Music. In: Psychology of Music, pp. 299–348, Cognition and Perception, Academic Press, 2nd ed., New York (1999)
10. Seashore, C.E.: Psychology of Music, McGraw-Hill - Reprinted 1967 by Dover Publications, New York (1938)
11. Marozeau, J.: L'Effet de la Fréquence Fondamentale sur le Timbre. PhD thesis, Université Pierre et Marie Curie, Paris VI (2004)
12. Cadoz, C.: Timbre et Causalité. In: Le Timbre, Métaphore pour la Composition, pp. 17–46, I.R.C.A.M. (1991)
13. ANSI: USA Standard Acoustical Terminology (1960)
14. AFNOR: Recueil des Normes Françaises de l'Acoustique, Tome 1 (Vocabulaire), NF S 30-107 (1977)
15. Bregman, A.: Sequential Integration. In: Auditory Scene Analysis, p. 97 (1990)
16. Schaeffer, P.: Traité des Objets Musicaux. Editions du Seuil ed. (1966)
17. Grey, J.M.: Multidimensional Perceptual Scaling of Musical Timbres. J. Acoust. Soc. Am., vol. 61, pp. 1270–1277 (1977)
18. Krumhansl, C.L.: Why Is Musical Timbre so Hard to Understand ?. In: Nielzén, S., Olsson, O. (eds.) Structure and Perception of Electroacoustic Sound and Music: Proc. of the Marcus Wallenberg Symposium Held in Lund, Sweden, pp. 43–53, Excerpta Medica, Amsterdam (1988)
19. McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., Krimphoff, J.: Perceptual Scaling of Synthesized Musical Timbres: Common Dimensions Specificities, and Latent Subject Classes. Psychological Research, vol. 58, pp. 177–192 (1995)
20. Caclin, A., McAdams, S., Smith, B.K., Winsberg, S.: Acoustic Correlates of Timbre Space Dimensions: A Confirmatory Study Using Synthetic Tones. J. Acoust. Soc. Am., vol. 118, no. 1, pp. 471–482 (2005)
21. Grey, J.W., Gordon, J.W.: Perception of Spectral Modifications on Orchestral Instrument Tones. Computer Music Journal, vol. 11, no. 1, pp. 24–31 (1978)
22. Beauchamp, J.W.: Synthesis by Spectral Amplitude and Brightness Matching of Analyzed Musical Instrument Tones. J. Audio Eng. Soc., vol. 30, no. 6, pp. 396–406 (1982)
23. Jensen, K.: The Timbre Model - Discrimination and Expression., In: Proceedings of the Mosart Midterm Meeting, Esbjerg, Denmark (2002)

24. Hoffman, M., Cook, P.: The Featsynth Framework for Feature-Based Synthesis: Design and Applications. In: Proc. Int. Comp. Music Conf. (ICMC'07), Copenhaguen, Denmark, vol. 2, pp. 184–187 (2007)

25. Risset, J.-C., Wessel, D.L.: Exploration of Timbre by Analysis and Synthesis. In: Psychology of Music, Academic Press, 2nd ed. (1999)

26. Barthet, M., Depalle, P., Kronland-Martinet, R., Ystad, S.: From Performer to Listener: an Analysis of Timbre Variations. J. Acoust. Soc. Am. (under revision)

27. Guillemain, P.: A Digital Synthesis Model of Double-reed Wind Instruments. Eurasip Journal on Applied Signal Processing, Special Issue on Model-based Sound Synthesis, vol. 7, pp. 990–1000 (2004)

28. Zwicker, E., Fastl, H.: Psychoacoustics, Facts and Models. Springer Verlag (1990)

29. Barthet, M., Kronland-Martinet, R., Ystad, S.: Improving Musical Expressiveness by Time-Varying Brightness Shaping, `http://www.lma.cnrs-mrs.fr/~kronland/cmmr2007/cmmr2007.html`

30. Loureiro, M.A., de Paula, H.B., Yehia, H.C.: Timbre Classification of a Single Instrument. In: ISMIR 2004 5th International Conference on Music Information Retrieval, (Barcelona, Spain), Audiovisual Institute, Universitat Pompeu Fabra (2004)

31. Risset, J-C.: Computer Study of Trumpet Tones. J. Acoust. Soc. Am., vol. 38, no. 912 (1965)

32. Fritz, C., Wolfe, J.: Acoustic Impedance Measurement of the Clarinet Players' Airway. In: CFA/DAGA'04, (Strasbourg), pp. 101–102 (2004)

33. Guillemain, P.: Some Roles of the Vocal Tract in Clarinet Breath Attacks: Natural Sounds Analysis and Model-Based Synthesis. J. Acoust. Soc. Am., vol. 121, no. 4, pp. 2396-2406 (2007)

34. Barthet, M., Guillemain, P., Kronland-Martinet, R., Ystad, S.: On the Relative Influence of Even and Odd Harmonics in Clarinet Timbre. In: Proc. Int. Comp. Music Conf. (ICMC'05), Barcelona, Spain, pp. 351–354 (2005)

35. Barthet, M., Kronland-Martinet, R., Ystad, S.: Consistency of Timbre Patterns in Expressive Music Performance. In: Proc. 9th Int. Conf. on Digital Audio Effects (DAFx06), (Montreal, Quebec, Canada), pp. 19–24 (2006)

36. Wanderley, M.: Quantitative Analysis of Non-Obvious Performer Gestures. In: Gesture and Sign Language in Human-Computer Interaction: International Gesture Workshop, p. 241, Springer, Berlin (2002)

37. Aramaki, M., Kronland-Martinet, R.: Analysis-Synthesis of Impact Sounds by Real-Time Dynamic Filtering. In: IEEE Trans. on Acoust., Speech, and Sig. Proc., vol. 14, no. 2, pp. 695–705 (2006)

38. Goldfarb, D.: A Family of Variable Metric Updates Derived by Variational Means. Mathematics of Computing, vol. 24, pp. 23–26 (1970)

39. Dillon, W. R., Goldstein, M.: Multivariate Analysis. In: Wiley series in probability and mathematical statistics, John Wiley & Sons, New York (1984)

40. Siegel, S., John Castellan Jr., N.: Non Parametric Statistics for the Behavioral Sciences, ch. Measures of Association and their Tests of Significance, p. 272, McGraw-Hill International Editions, 2nd ed. (1988)

41. Farlex: The free dictionary, `http://www.thefreedictionary.com`

42. Faure, A.: Des sons aux mots, comment parle-t-on du timbre musical ? PhD thesis, École des Hautes Etudes en Sciences Sociales (2000)

43. Whybrew, W. K.: Measurement and Evaluation in Music, p. 63, William C. Brown Company Publishers, Dubuque (1962)

44. Abeles, H. F.: Development and Validation of a Clarinet Performance Adjudication Scale. Journal of Research in Music Education, vol. 21, no. 3, pp. 246–255 (1973)

45. Gabrielsson, A., Lindstrom, B.: Perceived Sound Quality of High-Fidelity Loud-speakers. J. Audio Eng. Soc., vol. 33, no. 1, pp. 33–53 (1985)
46. Susini, P., McAdams, S.: Effet de Récence dans une Tâche de Jugement de la Sonie. In: 5ème Congrès français d'acoustique, Lausanne, Suisse (2000)
47. Wessel, D.L.: Timbre Space as a Musical Control Structure. Computer Music Journal, vol. 3, no. 2, pp. 45–52 (1979)

Appendix

The results of the hierarchical cluster analysis made on the performance dissimilarities are presented in figure 9.
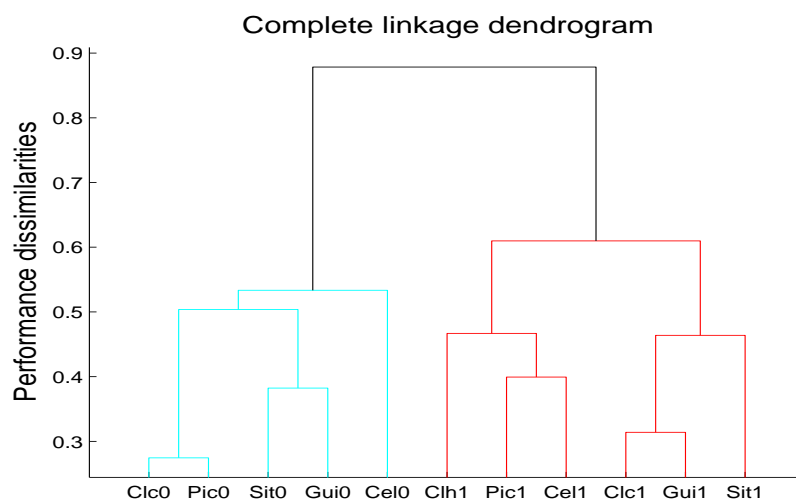

Fig. 9. Complete linkage dendrogram from the hierarchical cluster analysis made on the performance dissimilarities