Acoustical Correlates of Timbre and Expressiveness in Clarinet Performance

Mathieu Barthet, Philippe Depalle,
Richard Kronland-Martinet, and Sølvi Ystad
CNRS Laboratoire de Mécanique et d'Acoustique, Marseille, France

This study deals with the acoustical factors liable to account for expressiveness in clarinet performances. Mechanical and expressive performances of excerpts from Bach's *Suite No. II* and Mozart's *Quintet for Clarinet and Strings* were recorded. Timbre, timing, dynamics, and pitch descriptors were extracted from the recorded performances. The data were processed using a two-way analysis of variance, where the musician's expressive intentions and the note factors were defined as the independent variables. In both musical excerpts, a strong effect of the expressive intention was observed on the timbre (attack time, spectral centroid, odd/even ratio), timing (intertone onset intervals) and dynamics (root mean square envelope) descriptors. The changes in the timbre descriptors were found to depend on the position of the notes in the musical phrases. These results suggest that timbre, as well as timing and dynamics variations, may mediate expressiveness in the musical messages transmitted from performers to listeners.

Received September 11, 2008, accepted March 14, 2010.

Key words: timbre, timing, dynamics, musical interpretation, expressive clarinet performance

Acoustical Correlates of Timbre and Expressiveness in Clarinet Performance

Since the beginning of the twentieth century, the authors of studies on musical performance have attempted to analyze, understand, and model the processes underlying the act of musical interpretation, namely "the act of performance with the implication that in this act the performer's judgment and personality necessarily have their share" (Scholes, 1960). Although many studies have focused on timing (e.g., note durations, tempo, chord asynchronization) and dynamics (see e.g., Kendall & Carterette, 1990; Repp, 1992; Todd, 1992), intonation, phrasing, and articulation (see e.g., Gabrielsson & Lindstrom, 1995), less attention has been paid so far to timbre (see Juslin & Laukka, 2003, for a review). Here we present the first part of a study on the role of timbre in the musical message transmitted from performers to listeners. For this purpose, mechanical and expressive clarinet performances were recorded and analyzed in order to determine which (if any) acoustical correlates of timbre change when a performer plays more expressively.

The Notion of Expressive Deviations

Early studies carried out around 1930 by C. E. Seashore's group at Iowa University led to the conjecture that "the artistic expression of feeling in music consists in aesthetic deviation from the regular—from pure tone, true pitch, even dynamics, metronomic time, rigid rhythms, etc." (Seashore, 1967). This suggests that expressiveness in music can be characterized by measuring acoustical features of the musical instrument's tones related to time, energy, frequency, and/or the performer's instrumental gestures. In line with these early studies, artistic expression has often been approached by measuring the deviations of time and frequency parameters with respect to fixed and regular values corresponding to a strictly "mechanical" rendering of the score (Gabrielsson, 1998; Seashore, 1967).

The Measurement of Musical Performance

Detailed reviews of studies taking a psychological approach to the measurement of musical performance have been published by Gabrielsson (1998) and Palmer (1997). Repp (1992) investigated the timing differences and commonalities between several famous pianists' interpretations of Schumann's *"Träumerei."* Statistical analyses of the intertone onset intervals (IOIs) showed the existence of recurrent patterns, such as the ritardandi observed at the end of musical phrases, corresponding to "how most pianists transmit musical structure and

Music Perception volume 28, issue 2, pp. 135–153, issn 0730-7829, electronic issn 1533-8312 DOI:10.1525/mp.2010.28.2.135

expression through timing variations," as well as other patterns reflecting the individuality and eccentricity of some performers, such as Vladimir Horrowitz and Alfred Cortot, in particular. In a study on the interpretation of a Mozart piano sonata, Palmer (1996) established that expert pianists consistently repeated the same prosodic timing and intensity patterns. These results tend to prove that timing and dynamic deviations are not random, but are linked to musicians' expressive intentions. Note that the observations from Repp (1992) and Palmer (1997) may be related to the fact that the piano is impoverished in terms of its ability to manipulate timbre and that performers naturally use the degrees of freedom at their disposal in a performance, namely timing and dynamics (however, due to the sympathetic resonances of the strings, the piano allows to perform subtle timbre modifications by varying the playing technique). In the case of timbre, based on their analysis of spectrograms of several interpretations of an aria by Verdi (*"Parmi veder le lagrime"* in the piece *"Rigoletto"*), Födermayr and Deutsch (1993) noted that one of the singers applied a subtle change of timbre to a vowel for expressive effect. The present study focuses on whether changes of timbre of this kind occur arbitrarily or whether, on the contrary, they are dictated by the performer's expressive intentions.

On the Definition of Timbre

Back in 1938, Seashore (1967) was already convinced that timbre contributes importantly to musical aesthetics, but no appropriate means of measurement were available for examining this parameter more closely: "We should here recognize 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." More than seventy years later, there still exists no widely accepted definition of timbre on which researchers can base general models. In the psychoacoustical context, timbre is defined as the attribute of the auditory sensation that allows to distinguish different sounds equal in pitch, loudness, and duration, depending on some of their temporal and spectral characteristics (ANSI, 1960). Timbre is hence closely related to the identity of the sound source. However, as remarked by Schaeffer (1966), this facet of timbre is paradoxical: how can we speak of an instrument's timbre when each of its tones also possesses a specific timbre? In his description of timbre, Schaeffer combines the causal invariants that may be partly responsible for the instrument's identity (e.g., the hammered strings in the case of the piano), with the sources of variations, some of which are linked to the instrument's register (e.g., the low registers are generally richer than the high registers), and others that are due to the performers' control gestures. A description of timbre cannot therefore be limited to the typological aspects mentioned above but should also include the morphological aspects. Timbre can therefore be regarded as an elementary perceptual property of sound that can vary in a single instrument with time. This two-fold nature of timbre (identity/quality) can be explained in terms of cognitive categorization's theories: musical sounds can be categorized either in terms of the sources from which they are generated, or simply as sounds, in terms of the properties that characterize them (Castellengo & Dubois, 2005; Handel, 1995).

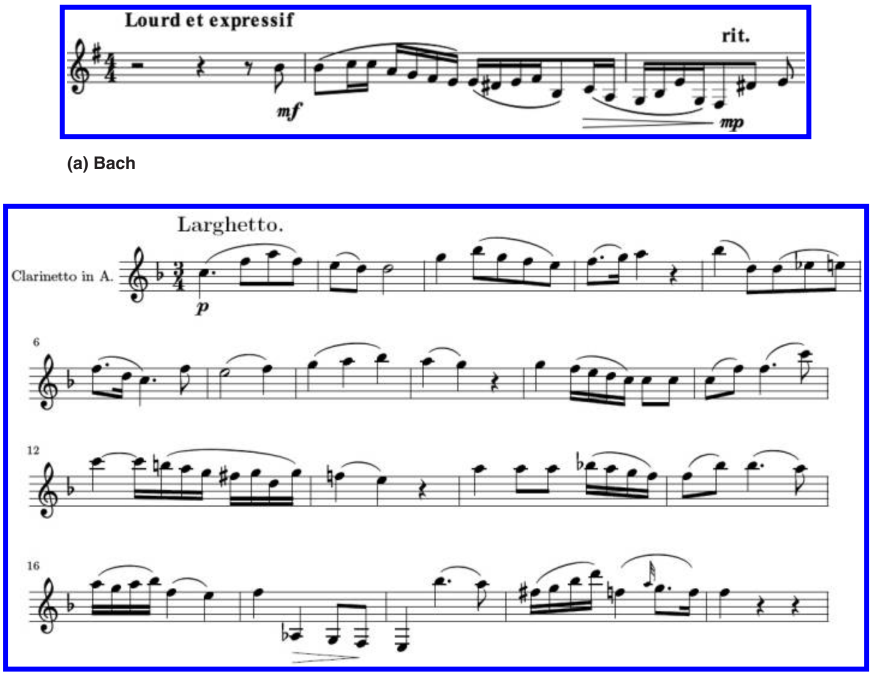
The Timbre Descriptors: The Acoustical Correlates of Timbre

Many studies on timbre have consisted of quantifying its acoustical correlates, which are also known as *timbre descriptors* (see Hajda, Kendall, Carterette, & Harshberger, 1997; McAdams, 1994, for detailed historical reviews). The methods used to address this issue are mostly based on multidimensional scaling (MDS) techniques, with which various timbres can be mapped in a low-dimensional space (the so-called *timbre space*), where the relative positions reflect the degree of perceived proximity. The structure of the perceptual representation of timbre is sensitive to the choice and number of stimuli used in these studies. However, the differences in timbre between orchestral instruments' tones are usually modeled in a three-dimensional perceptual space (see e.g., Grey, 1977; Kendall & Carterette, 1991; Krumhansl, 1989; McAdams, Winsberg, Donnadieu, De Soete, & Krimphoff, 1995; Wessel, 1979). The dimensions of this space are generally well correlated with descriptors based on the temporal (e.g., attack time), spectral (e.g., spectral centroid) and spectro-temporal (e.g., spectral flux) aspects of sounds (see Krimphoff, McAdams, & Winsberg, 1994, for the definitions of these descriptors).

Method

Procedure

We focused here on a monodic instrument, the clarinet, because its mechanical and acoustic properties make it possible for the player to control the timbre very closely while playing. Indeed, in clarinet-like systems, the self-sustained oscillations are due to the coupling between the exciter and the resonator, which is maintained practically throughout the duration of the sound. In order to test whether timbre plays an expressive role, we developed a method with which performances played with different expressive intentions can be compared. Contrary to what



(a) Bach

(b) Mozart

FIGURE 1. Scores of the Bach (a) and Mozart (b) musical excerpts.

occurs with pitch and rhythm, which can be compared with the composer's indications on the score, it is difficult to define a control level in the case of timbre. In this study, mechanical or inexpressive performances were used as a reference against which deviations in acoustical correlates of timbre are quantified.

Sound Corpus

Musical excerpts were recorded with the same professional clarinet player during two different recording sessions. The scores and a description of the musical excerpts are shown in Figure 1 and Table 1, respectively. Sound examples are available at: http://www.lma.cnrs-mrs.fr/~kronland/Interpretation_acoustical.

Bach's Allemande. The first musical excerpt was the opening phrase from the *Allemande* movement of Bach's *Suite No. II* (BWV 1008). Note that this Suite was written by Bach for the cello. An adaptation for the clarinet by U. Delécluse was used here by the performer. The musical phrase is in quadruple time and is played in the lowest register of the clarinet, the *chalumeau*. A ritardando is indicated at the end of the phrase. The clarinetist was asked to repeat the phrase 20 times with two different levels of expression. The first level corresponded to a mechanical or inexpressive rendering (keeping strictly to the indications on the score), whereas the second level corresponded to an expressive interpretation. A reference tempo of 48 bpm was chosen by the performer. During the mechanical performances, the metronome beats were delivered to the performer via earphones. During the expressive performances, the reference tempo was given only just before the recordings, and was then turned off.

TABLE 1. Description of the Sound Corpus.

Musical excerpt	*Allemande—Bach's Suite*	*Larghetto—Mozart's Quintet*
Duration (bars/notes)	1.8 bars (27 notes, N1 to N27)	17 bars (75 notes, N1 to N75)
No. of phrases	1	4
No. of mechanical performances	20 (P1 to P20)	2 (P1 to P2)
No. of expressive performances	20 (P21 to P40)	4 (P3 to P6)
Reference tempo (bpm)	48	44

Mozart's Larghetto. The second musical excerpt was the first 17 bars of the *Larghetto* movement of Mozart's *Quintet for Clarinet and Strings* (KV 581). This piece, written for the middle register of the clarinet, the *clarino*, is in triple time. The *Larghetto* movement was chosen because it seemed likely that a slow movement would probably give the performer more time to modulate the timbre while playing. The clarinetist was asked to give two performances in a mechanical way and four in an expressive way, at the self-selected reference tempo of 44 bpm.

Recordings. All the recordings were carried out in an anechoic chamber to prevent any room reflections from affecting the sound. As timbre is known to be influenced by the characteristics of the recording equipment and the settings, two different types of microphones and setting positions were used. The first was a system of microphones attached to the body and the bell of the instrument (SD Systems LCM 82 for Clarinet). The second was a Neumann KM 183 omnidirectional microphone, placed approximately 1.5 m from the instrument, at right angles to its body. The recordings of the Bach excerpt obtained with the SD Systems microphones were used for further analyses, as they were judged to reproduce more accurately the timbre of the instrument played in the lower register (*chalumeau*), in comparison to the recordings obtained with the Neumann microphone, which presented too much bass components. Conversely, the recordings obtained with the Neumann microphone were selected in the case of the Mozart sequence because they were judged to be more faithful to the timbre of the instrument in the *clarino* register than those obtained with the SD Systems microphones, which were very bright and short of bass components. Note that the use of different microphones for the Bach and Mozart excerpts did not affect the results since the analyses consisted of evaluating the differences between the mechanical and expressive performances within the same musical excerpt. All the performances were digitized at a sampling rate of 44.1 kHz.

Segmentation of the Performances

The performances were first segmented in order to analyze the timing of the notes (i.e., the onset and offset times). We previously developed a semi-automatic note segmentation procedure (some errors had to be corrected manually) based on the detection of the instabilities of the fundamental frequencies (F0) occurring at each transition from one note to another. The instabilities are due to the fact that the F0s are not clearly defined before the start of the self-sustained oscillations and after their end. Further details about this procedure are given in (Barthet, Kronland-Martinet, & Ystad, 2006).

Acoustical Analyses of the Performances

The next step consists of using acoustical analysis techniques to extract temporal, timbre, dynamics, and pitch descriptors from the recorded performances.

TEMPORAL DESCRIPTORS

Intertone onset interval deviation (ΔIOI). The duration of tones, which is classically quantified in terms of the intertone onset interval (IOI) (Repp, 1992), is used by performers as a means of expression (Palmer, 1996). In order to characterize the local changes of tempo during a performance, we computed the IOI deviation descriptor ΔIOI, defined as the difference between the measured IOIs (called the effective IOIs) and the IOIs obtained by directly transcribing the notations on the score (called the nominal IOIs):

$$\Delta IOI = IOI_{eff} - IOI_{nom} \quad (1)$$

where IOI_{eff} and IOI_{nom} are the effective and nominal IOIs, respectively.

Tempo (TEMPO). The mean tempo of a performance, denoted TEMPO, was defined as the ratio between the total number of beats in the musical excerpt and the sum of the effective IOIs.

TIMBRE DESCRIPTORS

In a previous study, we observed that the perceptual dissimilarities between synthesized clarinet tones could be represented accurately in a three-dimensional timbre space (Barthet, Guillemain, Kronland-Martinet, & Ystad, 2010), whose dimensions were well correlated with the attack time (AT), the spectral centroid (SC) and the odd/even ratio (OER). These three timbre descriptors were therefore used here to quantify clarinet timbre variations.

Attack time (AT). Attack time is correlated with the rate of energy increase in the onset of a sound. Results presented in (Barthet et al., 2010) have shown that the attack time of clarinet tones depends on two main control parameters, the player's blowing pressure and the force he imposes on the reed with the lower lip, which modulates the reed channel aperture.

There exists no computation methods so far that could be used to explain the perception of attack times in a large range of tone dynamics. As pointed out by Schaeffer (1966), the perception of attack is a complex phenomenon that paradoxically seems to depend not only on the physical attack transient but also on the shape of the dynamics during the successive phases of sounds. Gordon (1987) tested various models for attack times based on the amplitude envelope. In the model that best predicted perceptual measurements, the attack time was defined as the time the amplitude envelope takes to go beyond a certain threshold relative to its maximum value. Krimphoff et al. (1994) also used a threshold to account for the perception of the beginning of a sound. The expression used here for the attack time descriptor (AT) takes both thresholds into account:

$$AT = t_{e_{AT}} - t_{s_{AT}} \quad (2)$$

where $t_{s_{AT}}$ and $t_{e_{AT}}$ correspond to the times at which the amplitude envelope reaches 10% and 90% of the maximum amplitude, respectively (the thresholds were defined as in Peeters, 2004). This descriptor was found to be strongly correlated with the first dimension of the clarinet timbre space obtained in Barthet et al. (2010).

Spectral centroid (SC). Grey and Gordon (1978) proposed to characterize numerically the spectral energy distribution by its mean (centroid or "balance point"), in order to find an acoustical descriptor predicting the distribution of sounds along one of the dimensions of a timbre space obtained from dissimilarity judgments. This parameter, later referred as spectral centroid in the psychoacoustic literature, is a good index of our ability to distinguish broad differences in timbre between various musical instruments (Grey, 1977; Krumhansl, 1989; McAdams et al., 1995), as well as the finer differences in timbre produced by the same instrument (Barthet et al., 2010; Loureiro, Paula, & Yehia, 2004; Traube, 2004). Lichte (1941) probably made the first reference to the perceptual sensation associated with the spectral centroid (*brightness*) in a study on verbal descriptions of timbre.[1] Kendall and Carterette (1996) indeed observed that the spectral centroid accurately maps a perceptual quality that the authors called *nasality* (also referred as *brightness* or *sharpness*). Independent of the hearing level, the brightness may explain how the ear recognizes tones played *piano*, which are generally perceived as dull and mellow, or *forte*, which are generally perceived as bright and harsh (Risset, 1991).

In the context of musical performance, the spectral centroid might be a relevant means of accounting for the timbre variations produced by performers because it has often been found to be closely related to variations in the musician's control gestures. For example, Traube (2004) observed that the point at which a guitar string is plucked strongly affects the spectral centroid of the resulting sound: the closer to the middle of the string the pluck occurs, the lower the spectral centroid of the guitar tone, and the less bright the tone will be. In the case of the clarinet, the playing technique chosen by the performer, which is linked to choices about the instrument itself (such as the bending and length of the mouthpiece table and the strength of the reed) is known to affect the brightness of the tones. The "French" technique, where the clarinetist takes only a small part of the mouthpiece into his/her mouth, tends to brighten the tones, whereas the "German" technique, where a much larger part of the mouthpiece is held inside the mouth, generally yields less bright tones (Fritz, 2004). As shown in previous studies, even when a same playing technique is used, modulations of spectral centroids can be obtained by varying the control parameters of the instrument: a monotonous increase in the spectral centroid was observed as the mouth pressure and reed aperture increased, in the case of clarinet tones synthesized with a physical model (Barthet, 2008; Barthet, Guillemain, Kronland-Martinet, & Ystad, 2005; Helland, 2004). Listeners use changes in the spectral centroid to discriminate between different clarinet timbres, both in the cases of synthetic (Barthet et al., 2010) and natural (Loureiro et al., 2004) tones.

We have assumed the existence of a link between brightness and the perception of tension. As music usually consists of series of tensions and releases, brightness variations may provide performers with a means of communicating the musical structure to the listeners. In the context of music sequential integration, Wessel (1979) put forward the idea that differences in brightness could surprisingly induce melodic segregation in much the same way as differences in pitch (see also Bregman, 1994).

Different ways of defining the spectral centroid are given in the literature, depending on the amplitude and frequency scales adopted, and whether physiological auditory models are used. Although the spectral centroid descriptors based on physiological auditory data (e.g., the sharpness defined by (Zwicker & Fastl, 1990), or the descriptor proposed by (Marozeau, Cheveigné, McAdams, & Winsberg, 2003), based on the partial loudness measurements, and an equivalent rectangular band-rate scale) have increased the correlations between perceptual and acoustical data in timbre studies, these improvements are rather small in comparison to the predictions obtained using methods based on the traditional Fourier analysis (see Grey & Gordon, 1978; Marozeau et al., 2003, for comparisons), which lower

[1]An effort has been made in this article to distinguish between the spectral centroid as a measure of the spectral distribution and the brightness as a perceptual attribute of sound.

the cost of computation and are more suitable for synthesis applications (e.g., feature-based synthesis). As a means of analysis and synthesis, the spectral centroid has been efficiently used by Beauchamp (1982) to determine the parameters of a nonlinear/filter synthesis model via an automatic analysis procedure. The latter author's definition was used here to compute the short-term spectral centroid (SC):

$$SC(n) = \frac{\sum_{k=1}^{K} f(k) A_n(k)^2}{b_0 + \sum_{k=1}^{K} A_n(k)^2} \quad (3)$$

where $A_n(k)$ is the magnitude of the kth coefficient of the discrete Fourier transform (DFT) associated with the frame centered at time n, $f(k)$ is the frequency associated with the k^{th} spectral component, K denotes the last frequency bin to be processed, and b_0 is a positive amplitude threshold forcing the descriptor to decrease at very low amplitudes when noise predominates (e.g., during note transitions). As clarinet tones include both a deterministic part (corresponding to the harmonic signal resulting from the self-sustained oscillations) and a stochastic broadband part (resulting for instance from breath and key noises), the spectral centroid was calculated using a frequency scaling method that takes all the spectral components into account, and not only the harmonic partials. We also used a power amplitude scale, which assigns a greater weight to the dominant harmonics, as it has shown to increase the correlations with perceptual dissimilarity judgements compared to a linear amplitude scale (see Barthet et al., 2010). Note that due to the stabilization term b_0, the values of SC can be smaller than the fundamental frequency of the tone.

In order to characterize the spectral centroid variations at note level, we calculated the mean value and the range of variations of the spectral centroid within the duration of a tone, which were denoted the spectral centroid mean (SCM) and the spectral centroid range (SCR), respectively. These parameters are defined by the following equations:

$$\begin{cases} SCM = \dfrac{1}{IOI} \displaystyle\sum_{n=n_{on}}^{n=n_{off}} SC(n) \\ SCR = \underset{n_{on} \le n \le n_{off}}{Max} (SC(n)) - \underset{n_{non} \le t \le n_{off}}{Min} (SC(n)) \end{cases} \quad (4)$$

where n_{on} and n_{off} are the onset and offset times of the note, respectively, and IOI is its duration. Note that these descriptors are independent: notes can potentially have the same SCMs but different SCRs, and vice versa.

Odd/even ratio (OER). The odd/even ratio, which is used to analyze harmonic or quasi-harmonic sounds, accounts for the difference in the relative energy between odd and even harmonics (see e.g., Peeters, 2004). The odd/even ratio is particularly suitable for characterizing clarinet tones, since this instrument's "closed/open" cylindrical resonator is known to favor the odd harmonics at the expense of the even ones, which are very weak in this case (Benade & Kouzoupis, 1988). In Barthet et al. (2010), the odd/even ratio was found to be strongly correlated with one of the dimensions of the perceptual timbre space of synthetic clarinet tones.

We have defined the time-varying odd/even ratio (OER) by the following equation:

$$OER(t) = \frac{b_0 + \sum_{h=0}^{\frac{H}{2}-1} A_{2h+1}(t)^2}{b_0 + \sum_{h=1}^{\frac{H}{2}} A_{2h}(t)^2} \quad (5)$$

where $A_h(t)$ denotes the instantaneous amplitude of the h^{th} harmonic component and b_0 is defined as in equation 3. H is the total number of harmonics under consideration, which is assumed to be even in equation 5, so that an equal number of odd and even harmonics are compared. Note that OER is dimensionless. $OER < 1$ indicates that the even harmonics predominate, whereas $OER > 1$ indicates that the odd harmonics predominate. As with the spectral centroid, the following two note-level timbre descriptors were defined, based on the odd/even ratio: the odd/even ratio mean (OERM) and the odd/even ratio range (OERR). These parameters were computed in the same way as in equation 4.

DYNAMICS DESCRIPTOR

The root mean square (RMS) envelope was used to characterize the changes in the acoustical energy. This parameter has been classically defined as follows:

$$ENV(n) = \sqrt{\frac{\sum_{k=1}^{K} A_n(k)^2}{N}} \quad (6)$$

where the various quantities are defined as in equation 3 and N is the number of points used to calculate the discrete Fourier transform. As with SC and OER, we computed the mean value and the range of variation of the envelope during each of the tones, which were denoted ENVM and ENVR, respectively.

PITCH DESCRIPTOR

The pitch of complex harmonic tones is closely linked to the fundamental frequency (Terhardt, Stoll, & Seewann, 1982). The latter was used as a first approximation to characterize the pitch of clarinet tones. The instantaneous fundamental frequency F0 was obtained using the method developed in Jaillet (2005), which involves detecting spectral peaks in the time-frequency plane using a global optimization process. This method is implemented in the LEA software program produced by Genesis (2010). The mean value and the range of variation of the fundamental frequency during a tone will be denoted F0M and F0R, respectively.

COMPUTATION OF THE DESCRIPTORS

The short time discrete Fourier transform was computed using a 1024-point Hann window (approximately 20 ms at a sampling frequency of 44.1 kHz) with a 50% overlap; b_0 was set at a value giving a spectral dynamic of 60 dB. In order to compute the odd/even ratio, each tone was analyzed using a bank of bandpass filters, the frequencies of which matched the frequencies of the tone components (which correspond to a harmonic series in the case of sustained clarinet sounds). This provided us with short-band analytic signals associated with the frequency components of the tone. The instantaneous amplitude and phase of the tone components were then obtained from the short-band analytic signals (see, for example, Picinbono, 1997).

Synchronization of the Descriptors

As changes in IOI occurred between the various performances, a time synchronization procedure had to be performed at the note level, to be able to compare the descriptors SC, OER, ENV, and F0. For this purpose, a time-warping procedure was carried out on the descriptors. The temporal profiles of the descriptors associated with each tone were shortened or lengthened using cubic spline interpolation methods, so that the new durations corresponded to the mean IOI based on the repeated performances. Note that this time-warping procedure was used by Wanderley (2002), for instance, to examine the regularity of clarinetists' spatial movements.

Statistical Analyses

REPRODUCIBILITY OF THE EXPRESSIVE DEVIATIONS

In order to determine the level of similarity between the expressive deviations observed during performances played with the same expressive intentions (i.e., between all mechanical to mechanical and all expressive to expressive performances), Pearson product-moment correlations (r) were computed on the various time, frequency, and energy descriptors across the repeated performances.

COMPARISON BETWEEN MECHANICAL AND EXPRESSIVE PERFORMANCES

The analyses described above show the consistency of the acoustical parameters observed with a given musical intention, but they cannot be used to test whether any differences occur when the player's intentions change. In order to test whether the descriptors change depending on the performer's expressive intentions, two-way analyses of variance (ANOVA) were conducted with the player's expressive intentions and the note factors as independent variables. The dependent variables were the note level values of the descriptors (ΔIOI, AT, SCM, SCR, OERM, OERR, ENVM, ENVR, F0M, F0R). For all the descriptors, the one-way effects of the player's expressive intention and the note factors, and the two-way effect of interaction between these factors were analyzed. The magnitudes of the effects were estimated by using the partial eta squared (η^2) index of effect size. The definitions in (Cohen, 1977, p. 285) have been adopted to discuss the effect sizes (small effect size: $\eta^2 = .01$, medium effect size: $\eta^2 = .06$, large effect size: $\eta^2 = .14$). When interactions were observed, a multiple comparison procedure (MCP) based on the Holm-Sidak sequential procedure (Holm, 1979) was conducted to identify which tones in the musical sequence differed significantly between the mechanical and expressive performances. The Holm-Sidak procedure was used here as it is more powerful than non sequential multiple comparison tests, such as Bonferroni's, or Sidak's tests (Ludbrook, 1998).

An alpha level of .05 was used for all statistical tests.

Results and Discussion

Reproducibility of the Expressive Deviations

The mean correlations (r) within the mechanical and expressive performances were computed for each descriptor (ΔIOI, AT, SC, OER, ENV, and F0). For the Bach excerpt, these correlations were, on average, .93 for the mechanical performances (minimum: $r = .81$, $p < .001$) and .91 for the expressive performances (minimum: $r = .80$, $p < .001$). For the Mozart excerpt, the correlations were, on average, .93 for the mechanical performances (minimum: $r = .83$, $p < .001$) and .90 for the expressive performances (minimum: $r = .78$, $p < .001$). Hence, for both excerpts, ΔIOI, AT, SC, OER, ENV, and F0 were highly correlated across performances played with the same musical intention. These results show that the performer consistently repeated the patterns linked to time, frequency, and energy, whenever the same interpretative strategy was used.

Influence of the Note Factor

The results of the two-way analyses of variance conducted on the various note-level descriptors for the Bach and Mozart performances are presented in Table 2. It can be

TABLE 2. Two-Way Analyses of Variance Results for the Note-Level Descriptors (ΔIOI, AT, SCM, SCR, OERM, OERR, ENVM, ENVR, FOM, FOR) for the Bach and Mozart Performances.

	Bach			Mozart		
Source	*df*	*F*	η^2	*df*	*F*	η^2
ΔIOI						
Exp.	1	257.36***	.20	1	54.49***	.15
Note	26	663.22***	.94	74	70.28***	.95
Exp. x Note	26	16.80***	.30	74	2.68***	.40
Error	1026	(0.001)		300	(0.006)	
AT						
Exp.	1	33.47***	.03	1	28.23***	.09
Note	26	159.31***	.80	74	29.40***	.88
Exp. x Note	26	2.87***	.07	74	1.30	.24
Error	1026	(0.003)		300	(0.027)	
SCM						
Exp.	1	24.87***	.02	1	96.75***	.24
Note	26	914.67***	.96	74	128.43***	.97
Exp. x Note	26	6.33***	.14	74	3.45***	.46
Error	1026	(1450.7)		300	(2320.25)	
SCR						
Exp.	1	0.44	.00	1	24.15***	.08
Note	26	364.55***	.90	74	34.74***	.90
Exp. x Note	26	6.51***	.14	74	1.98***	.33
Error	1026	(2173.42)		300	(6932.37)	
OERM						
Exp.	1	54.86***	.05	1	190.84***	.39
Note	26	554.83***	.93	74	33.07***	.89
Exp. x Note	26	6.47***	.14	74	5.00***	.55
Error	1026	(0.41)		300	(0.48)	
OERR						
Exp.	1	2.84	.00	1	5.22*	.02
Note	26	197.68***	.83	74	30.96***	.88
Exp. x Note	26	3.97***	.09	74	3.21***	.44
Error	1026	(1.06)		300	(0.94)	
ENVM						
Exp.	1	125.02***	.11	1	308.16***	.51
Note	26	522.56***	.93	74	49.25***	.92
Exp. x Note	26	5.62***	.13	74	2.10***	.34
Error	1026	(.001)		300	(0.001)	
ENVR						
Exp.	1	41.38***	.04	1	0.31	.00
Note	26	702.57***	.95	74	27.23***	.87
Exp. x Note	26	1.64*	.04	74	1.85***	.31
Error	1026	(.001)		300	(0.002)	
FOM						
Exp.	1	5.22*	.01	1	50.53***	.14
Note	26	72978.64***	.99	74	192457.77***	1.00
Exp. x Note	26	2.43***	.06	74	0.59	.13
Error	1026	(4.73)		300	(0.973)	
FOR						
Exp.	1	1.16	.00	1	0.002	.00
Note	26	320.17***	.89	74	517.87***	.99
Exp. x Note	26	2.62***	.06	74	0.64	.14
Error	1026	(825.62)		300	(486.38)	

Note. The expressive intention and note factors are denoted Exp. and Note, respectively. Results enclosed in parentheses represent mean square errors. η^2 is the partial eta squared measure of effect size. $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$.

seen from this table that the note factor had a highly significant effect on all the descriptors, both for the Bach and Mozart performances. The effect sizes were found to be very large ($.80 \leq \eta^2 \leq 1.00$, $M = .92$). These results show that the values of the time, frequency, and energy descriptors varied according to the musical characteristics of the notes (such as pitch and duration) and/or their location in the musical structure. The influence of the note factor can be explained straightforwardly for the descriptors that are by definition correlated to the notes' characteristics, i.e., duration (ΔIOI) and pitch (F0M, SCM). The intrinsic mechanical and acoustical properties of the clarinet (for example, the increase of visco-thermal losses with frequency) also explain why the timbre descriptors depend on pitch. According to the model of musical expression proposed by Clarke (1988), the variability of the time, frequency, and energy descriptors as a function of the notes could also be related to the choice of controls made by the performer in order to communicate the musical structure to the listener.

Comparison Between Mechanical and Expressive Performances

TEMPORAL DESCRIPTORS

Δ*IOI*. The ANOVA showed a strong effect of the player's expressive intention on the IOI deviation descriptor both for the Bach and Mozart performances, $F(1, 1026) = 257.36$, $p < .001$, and $F(1, 300) = 54.49$, $p < .001$, respectively. The interaction between the expressive intention and the note factors was found to be highly significant for both excerpts (see Table 2), with large effect sizes ($\eta^2 = .30$ and .40 for the Bach and Mozart excerpts, respectively). As significant interactions were found between the main factors, multiple comparison procedures were performed. The results of the MCPs indicated that many tones (13 out of 27 for the Bach excerpt, and 13 out of 75 for the Mozart excerpt) lasted significantly longer than nominal in the expressive performances (see Figure 2, and Tables 3 and 4). Indeed, for both excerpts, the expressive performances lasted longer than the mechanical performances, on average

TABLE 3. Results of Multiple Comparisons on the Bach Excerpt.

Notes	F0M	F0R	Δ*IOI*	AT	SCM	SCR	OERM	OERR	ENVM	ENVR
N1	—	7.55**	—	3.51**	7.93**	3.90**	—	2.15*	6.73**	—
N2	—	—	—	—	—	—	—	—	2.85**	—
N3	—	—	—	—	—	—	—	—	—	—
N4	7.66**	—	—	—	—	—	—	—	—	—
N5	—	—	13.25**	6.27**	2.83**	—	—	—	—	2.65**
N6	—	—	7.23**	—	—	7.93**	5.08**	—	6.27**	2.22*
N7	—	—	—	—	—	4.91**	3.31**	—	3.52**	3.50**
N8	—	—	—	—	—	2.61**	—	2.60**	4.41**	—
N9	—	—	—	—	2.44*	—	—	—	4.49**	—
N10	—	—	—	—	2.79**	—	—	—	3.78**	—
N11	—	—	3.09**	—	3.16**	—	4.66**	2.82**	3.40**	—
N12	—	—	2.25*	—	2.41*	3.44**	2.30*	6.46**	2.52**	2.40*
N13	—	—	2.63**	—	4.52**	—	6.65**	2.23*	3.47**	—
N14	—	—	—	—	4.94**	3.37**	7.83**	2.87**	3.70**	2.19*
N15	—	—	—	4.05**	3.94**	2.80**	3.96**	3.34**	3.57**	—
N16	—	—	4.11**	—	—	—	—	—	2.42*	—
N17	—	—	2.44*	3.19**	—	—	—	—	—	—
N18	—	—	2.49*	—	—	—	—	—	—	—
N19	—	—	3.31*	2.87**	—	—	—	—	—	—
N20	—	—	—	—	—	—	—	—	1.98*	—
N21	—	—	—	2.33*	2.81**	3.30**	2.38*	—	2.36*	2.08*
N22	—	—	2.88**	—	—	2.05*	—	—	—	—
N23	—	—	5.26**	—	—	—	—	—	—	—
N24	—	2.03*	13.53**	—	—	—	3.72**	—	—	2.11*
N25	—	—	—	—	—	—	—	—	—	3.67**
N26	—	—	—	—	—	—	—	—	4.70**	—
N27	—	—	13.15**	—	—	—	2.32*	—	—	2.11*

Note. Comparisons were made on each of the 27 tones in the excerpt between the 20 mechanical performances (control group) and the 20 expressive performances. *T*-tests corrected for multiple comparisons (Holm-Sidak correction); $^*p < .05$, $^{**}p < .01$. Non-significant values are not shown.

TABLE 4. Results of Multiple Comparisons on the Mozart excerpt.

Notes	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12	N13
ΔIOI	3.06**	—	—	—	—	—	7.40**	—	—	—	—	—	—
SCM	3.97**	—	—	—	—	—	—	—	3.05**	—	—	—	—
SCR	—	2.49*	—	2.17*	—	—	—	2.73**	—	—	—	2.14*	—
OERM	3.40**	2.57*	5.00**	5.24**	2.18*	—	2.13*	—	—	—	2.72**	3.17**	3.53**
OERR	—	—	2.23*	—	—	—	—	—	—	—	2.70**	—	—
ENVM	2.09*	2.83**	2.92**	2.53*	—	—	—	2.33*	2.02*	—	3.28**	4.30**	3.40**
ENVR	—	—	—	—	—	—	—	4.35**	—	—	—	—	—

Notes	N14	N15	N16	N17	N18	N19	N20	N21	N22	N23	N24	N25	N26
ΔIOI	—	4.76**	3.44**	—	—	—	—	—	—	—	2.43*	—	—
SCM	—	3.39**	—	—	4.04**	5.43**	—	—	—	—	—	2.78**	—
SCR	—	—	—	—	—	2.85**	2.54*	—	—	—	—	3.20**	2.86**
OERM	2.00*	—	—	2.64**	2.89**	4.73**	—	2.43*	4.13**	—	2.33*	2.99**	—
OERR	—	—	—	—	2.87**	3.04**	—	1.99*	—	—	—	—	—
ENVM	2.90**	2.12*	—	—	2.85**	4.42**	3.19**	3.44**	2.12*	—	—	—	—
ENVR	—	—	—	—	—	—	—	—	—	—	—	2.83**	—

Notes	N27	N28	N29	N30	N31	N32	N33	N34	N35	N36	N37	N38	N39
ΔIOI	—	—	—	2.27*	2.71**	4.45**	2.83**	—	—	—	—	—	—
SCM	—	—	—	6.50**	4.43**	2.74**	—	—	—	—	—	2.73**	5.71**
SCR	—	—	—	—	—	2.40*	—	—	—	—	—	—	2.52*
OERM	—	—	—	5.11**	—	—	—	—	4.18**	—	—	—	5.95**
OERR	—	—	3.74**	3.44**	—	2.28*	—	—	2.95**	—	—	—	2.34*
ENVM	—	—	—	5.04**	5.17**	2.01*	—	—	3.56**	2.71**	3.25**	3.06**	4.35**
ENVR	—	—	—	3.40**	—	3.27**	—	2.01*	—	—	—	—	—

Notes	N40	N41	N42	N43	N44	N45	N46	N47	N48	N49	N50	N51	N52
ΔIOI	—	—	—	—	—	—	—	—	—	—	—	—	—
SCM	4.45**	—	—	—	—	2.62**	—	—	—	—	—	2.61*	3.90**
SCR	3.19**	3.42**	—	2.59*	2.13*	—	2.05*	—	—	—	—	—	—
OERM	4.33**	4.94**	2.90**	3.86**	3.04**	5.89**	4.54**	2.12*	—	3.02**	—	3.12**	—
OERR	2.60**	3.28**	—	—	—	—	—	—	—	—	—	—	—
ENVM	3.08**	2.31*	2.63**	2.35*	—	—	2.17*	—	—	1.98*	2.84**	3.90**	5.09**
ENVR	—	—	—	—	—	—	2.87**	—	—	—	—	—	—

Notes	N53	N54	N55	N56	N57	N58	N59	N60	N61	N62	N63	N64	N65
ΔIOI	—	3.84**	—	—	—	—	—	—	—	—	—	—	—
SCM	—	2.85**	—	3.75**	2.18*	—	—	—	—	—	—	—	2.39*
SCR	—	—	—	2.25*	2.09*	—	2.62**	—	—	—	—	—	—
OERM	2.02*	2.34*	—	—	—	4.81**	—	2.84**	—	2.69**	2.12*	—	2.04*
OERR	—	—	—	2.34**	2.58*	—	2.21*	—	—	—	—	—	—
ENVM	3.61**	2.34*	—	3.31**	—	—	—	—	—	—	—	—	—
ENVR	—	—	—	3.41**	—	—	2.08*	—	—	—	—	—	—

Notes	N66	N67	N68	N69	N70	N71	N72	N73	N74	N75
ΔIOI	—	—	—	—	—	—	—	2.61**	3.44**	5.22**
SCM	—	—	—	—	2.70*	—	—	—	2.67**	4.24**
SCR	—	—	—	—	—	—	—	—	—	—
OERM	—	3.09**	—	2.64**	—	4.45**	2.21*	—	—	2.89**
OERR	—	—	4.45**	—	—	—	—	—	—	—
ENVM	—	—	—	—	—	2.95**	3.03**	—	3.00**	3.57**
ENVR	—	—	—	—	—	—	—	—	—	—

Note. Comparisons were made on each of the 75 tones in the excerpt between the 2 mechanical performances (control group) and the 4 expressive performances. *T*-tests corrected for multiple comparisons (Holm-Sidak correction); $^*p < .05$, $^{**}p < .01$. Non-significant values are not shown.

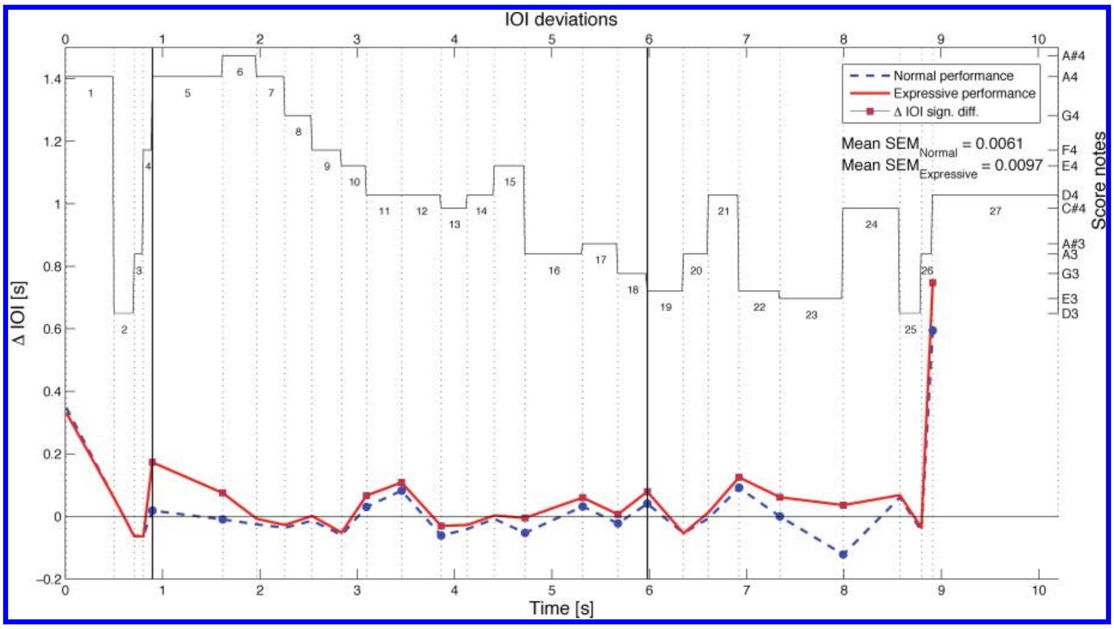


FIGURE 2. Comparison between the mean intertone onset interval deviations (ΔIOI) measured in the mechanical (dashed line) and expressive (solid line) performances of the excerpt from Bach's *Suite no. II*. The notes with which the multiple comparisons showed the existence of significant differences are indicated by circular and square markers. The dashed vertical lines correspond to the onsets and offsets of the tones. The thin vertical lines indicate the bars. The thick vertical lines indicate the beginnings and ends of the various musical phrases. The notes on the score are displayed at the top of the figure, along with their ranks in the musical sequence. The averages of the standard errors of the mean (SEM) computed across the data points are reported.

(Bach: $M = 10.75$ s, $SD = 0.21$ and $M = 9.79$ s, $SD = 0.13$, respectively; Mozart: $M = 72.46$ s, $SD = 0.89$ and $M = 67.80$ s, $SD = 0.64$, respectively). Consequently, the average tempi of the expressive performances were slower than the ones of the mechanical performances (Bach: $M = 40.47$ bpm, $SD = 0.78$ and $M = 44.45$ bpm, $SD = 0.57$, respectively; Mozart: $M = 42.23$ bpm, $SD = 0.43$ and $M = 45.13$ bpm, $SD = 0.43$, respectively), which were closer to the reference tempo (Bach: 48 bpm; Mozart: 44 bpm). This is not surprising as the mechanical performances were played in keeping with the metronome. In the case of the Bach excerpt, the shape of the IOI deviation pattern (ΔIOI) was very similar between the two interpretations (Figure 2). This shows that in both mechanical and expressive interpretations, the durations of the notes were lengthened or shortened with respect to the theoretical score indications, but the changes were more pronounced when the piece was played expressively. This pattern has often been reported to occur in studies on timing variations in musical performance (see e.g., Penel & Drake, 2004). For instance, the final ritardando in the Bach excerpt (N23 to N27) occurred in both interpretations, but it was more pronounced in the expressive performances.

PITCH DESCRIPTORS

F0M and F0R. The player's expressive intention significantly affected the mean fundamental frequency F0M both for the Bach and Mozart excerpts, $F(1, 1026) = 5.22$, $p = .02$ and $F(1, 300) = 50.53$, $p < .001$, respectively, but not the fundamental frequency range F0R (Table 2). The effect of interaction between the expressive intention and the note factors was found to be significant only in the case of the Bach performances with a medium effect size both for F0M and F0R ($\eta^2 = .06$). However, the results of the MCP on the Bach excerpt (see Table 3) showed that only one tone in the case of F0M (N4) and two tones in that of F0R (N1 and N24) showed significant differences when the player's intentions changed. These differences in F0M and F0R were due either to instabilities in the descriptor F0 at the onset and/or offset of the tones (a phenomenon induced by the method of analysis), or were very small. The maximum F0 difference in the sustained part of the tones was approximately 1 Hz (note N1), which was close to the frequency discrimination threshold measured in the case of pure tones (1 Hz in the case of a pure 200-Hz tone presented at 40 dB according to Wier, Jesteadtt, & Green, 1977). However, informal listening made by the authors did not reveal noticeable changes of pitch between the sequences. Based on these results, it appears that the contribution of pitch to the expression of the playing intention was weak at best in the case of the Bach excerpt, and non significant in the case of the Mozart excerpt.

TIMBRE DESCRIPTORS

AT. Highly significant effects of the player's expressive intentions on the tones' attack time (AT) were found for both the Bach and Mozart excerpts, $F(1, 1026) = 33.47$, $p < .001$ and $F(1, 300) = 28.23$, $p < .001$, respectively. However, the interaction between the expressive intention and the note factors was only significant for the Bach performances, $F(26, 1026) = 2.87$, $p < .001$, with a medium effect size ($\eta^2 = .07$). The posthoc analyses conducted for the Bach excerpt showed that 6 tones had significantly higher ATs in the expressive performances than in the mechanical ones (cf. Figure 3 and Table 3).

SCM and SCR. The ANOVA showed that the effect of the player's expressive intention on the spectral centroid mean SCM was highly significant for both the Bach and Mozart excerpts, $F(1, 1026) = 24.87$, $p < .001$ and $F(1, 300) = 96.75$, $p < .001$, respectively. For the spectral centroid range (SCR), the one-way effect of the expressive intention was only highly significant for the Mozart excerpt, $F(1, 300) = 24.15$, $p < .001$. However, for both excerpts, strong interactions between the expressive intention and the note factors occurred for SCM and SCR (Table 3). The effect sizes of the interaction between the expressive intention and the note factors were found to be large for both the Bach and the Mozart excerpts, although larger for the latter (Bach: $\eta^2 = .14$ for both SCM and SCR; Mozart: $\eta^2 = .46$ for SCM and $\eta^2 = .33$ for SCR). The results of the MCPs show that the mean and/or the range of the spectral centroid values were significantly different between the expressive and the mechanical performances in a large number of notes for both excerpts (Bach: 14 out of 27, as shown in Figure 4 in color plate section and Table 3, and Mozart: 33 out of 75, as shown in Figure 4 and Table 4). In order to evaluate if such changes would be noticeable from the perceptual point view, we used as a reference the mean difference threshold (just noticeable difference) in spectral centroid reported in Kendall and Carterette (1996), since it was obtained from perceptual experiments with human listeners. To address this issue, a F0-normalized spectral centroid ($\overline{SC}$) was computed as in (Kendall & Carterette, 1996). This was done by using a linear amplitude scale, $b_0 = 0$ and dividing by F0 in equation 3. For the Bach excerpt, the $\overline{SC}$ differences were higher than the JND threshold (0.117) for 13 out of the 14 tones for which significant differences of SCM and SCR were reported

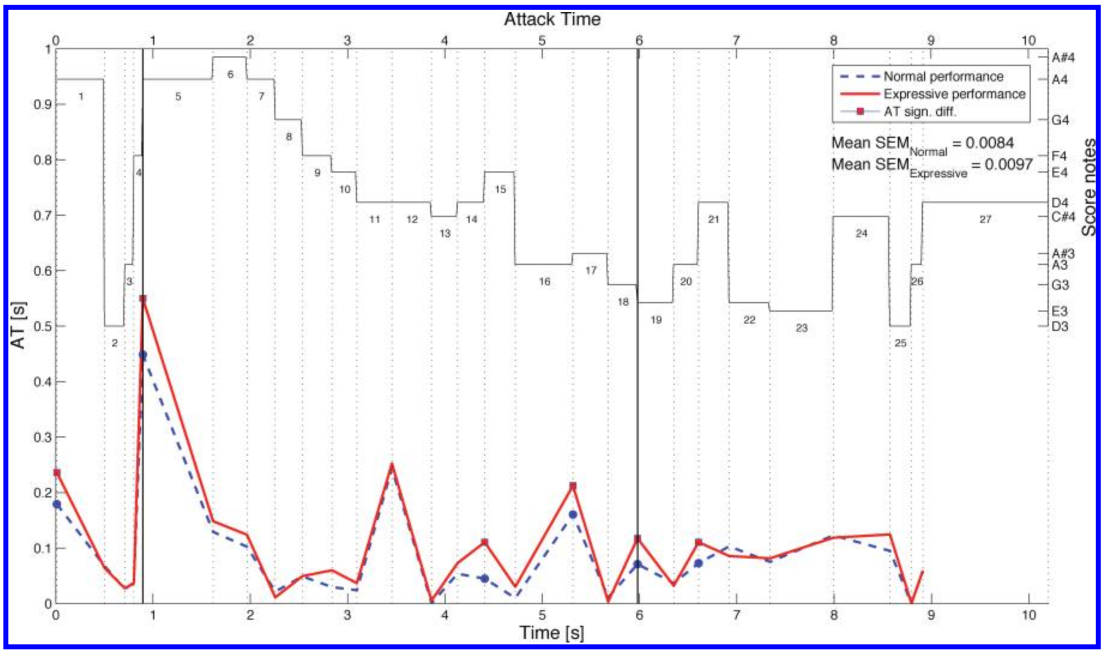


FIGURE 3. Average attack times in the mechanical (dashed line) and expressive (solid line) interpretations of the Bach excerpt. The notes and groups of notes with which the statistical analyses showed the existence of significant effects are indicated by circular and square markers. For other explanations, see the legend to Figure 2.

($0.04 \leq \Delta\overline{SC} \leq 2.9$). For the Mozart excerpt, the *SC* differences were higher than the JND threshold for 27 out of the 33 tones for which significant differences of SCM and SCR were reported ($0.01 \leq \Delta\overline{SC} \leq 1.95$).

In both excerpts, the changes observed in the spectral centroid with the expressive performances depended on the musical structure. In some parts of the musical phrases, the SCM was higher in the expressive than in the mechanical performance: for instance, at the beginning of the phrase, from notes N1 to N5, for the Bach excerpt, or in the third musical phrase, at the 12th bar, for the Mozart excerpt. In other parts, the opposite pattern occurred, i.e., the SCM was lower in the expressive than in the mechanical performance: for instance, in the middle of the phrase, from notes N9 to N15, for the Bach excerpt, and from the 7th to 9th bar, for the Mozart excerpt. Upon listening informally to the Bach performances, the authors noted that the sequence from notes N9 to N15 sounded mellower in the expressive performances, where low SCM values had been observed, than in the mechanical ones (see sound examples *1a* and *1b*).

It is worth noting that the significant changes in spectral centroid detected in the expressive rendering were not due to changes in F0, except possibly for the first note N1 in the Bach excerpt, since the interaction effect between the expressive intention and the note factors for F0 was either non-significant or weakly significant, but in this case caused by changes occurring on distinct notes than for SC (see Tables 2 and 3, and the discussion on the pitch descriptors, above). Furthermore, within a musical phrase, some tones with the same F0 had very different spectral centroid values. For instance, in the mechanical performances of the Bach excerpt, notes N11, N12, and N27, which all corresponded to a D4 ($\overline{F0} \approx 293.67$ Hz), had significantly different SCM values (448.93, 355.72, and 82.07 Hz[2], respectively): N11 vs. N12, $t(351) = 21.45$, $p < .001$, N11 vs. N27, $t(351) = 30.46$, $p < .001$, and N12 vs. N27, $t(351) = 21.45$, $p < .001$. These findings confirm that the spectral centroid variations depend on the position of the notes in the musical structure.

OERM and OERR. As for the other timbre descriptors, the results of the ANOVA showed that the effect of expressive intention on the odd/even ratio mean (OERM) were highly significant for both the Bach and the Mozart excerpts, $F(1, 1026) = 54.86$, $p < .001$, and $F(1, 300) = 190.84$, $p < .001$, respectively. For the odd/even ratio range (OERR), the one-way effect of expressive intention was non-significant for the Bach excerpt, $F(1, 1026) = 2.84$, $p = .09$, and weakly significant for the Mozart excerpt, $F(1, 300) = 5.22$, $p = .02$ ($\eta^2 = .02$), which is probably due to the strong one-way effect of the note factor (Table 2). Indeed, the interactions between the expressive intention and the note factors were highly significant both for OERM and OERR, in both excerpts. Again, the effect size of this interaction was larger for the Mozart excerpt ($\eta^2 = .55$ for OERM and $\eta^2 = .44$ for OERR) than for the Bach excerpt ($\eta^2 = .14$ for OERM and $\eta^2 = .09$ for OERR). The results of the MCPs showed that significant differences in OERM and/or OERR were observed with 12 notes for the Bach excerpt (see Figure 5 in color plate section and Table 3), and with more than half of the notes (48 notes out of 75) for the Mozart excerpt (see Figure 5 and Table 4). Note that the odd/even ratio was mostly greater than one, except during the attack parts of a few notes (see Figure 5), which points out, as was to be expected, the dominance of odd harmonics compared to even harmonics for clarinet tones.

DYNAMICS DESCRIPTOR

ENVM and ENVR. The effect of the performer's expressive intention on the mean value of the RMS envelope ENVM was found to be highly significant for both the Bach and the Mozart excerpts, $F(1, 1026) = 125.02$, $p < .001$, and $F(1, 300) = 308.16$, $p < .001$, respectively. Interactions between the expressive intention and the note factors were highly significant for ENVM in both the Bach and Mozart excerpts, with medium ($\eta^2 = .13$) and large effect sizes ($\eta^2 = .34$), respectively (Table 2). Regarding the range of variations of the RMS envelope ENVR, only a weakly significant effect of the expressive intention was found for the Bach excerpt, $F(1, 1026) = 41.38$, $p < .05$, with a small effect size ($\eta^2 = .04$). However, the interactions between the expressive intention and the note factors were significant for both the Bach and Mozart excerpts, with small ($\eta^2 = .04$) and large ($\eta^2 = .31$) effect sizes, respectively. The multiple comparison procedure (see Table 3) showed the existence of significant differences in the ENVM and ENVR values with many notes (20 notes in all for the Bach excerpt, and 43 notes for the Bach excerpt).

RELATIONSHIPS BETWEEN THE DESCRIPTORS

It is worth noting that for the Bach excerpt, the short notes (such as the grace notes N2, N3, N26) did not generally show any significant differences in ΔIOI, AT, SC, or OER, possibly due to the fact that when playing short notes, the performer did not have enough time to make expressive timing or timbre variations.

Some notes systematically showed disparities in terms of both timbre and temporal descriptors. For instance, for the Bach excerpt, the first note in the first bar (N5) after the anacrusis (N1) and the grace notes (N2 to N4)

[2]The fact that SCM can be smaller than F0 is due to the stabilization term b_0 in equation 3.

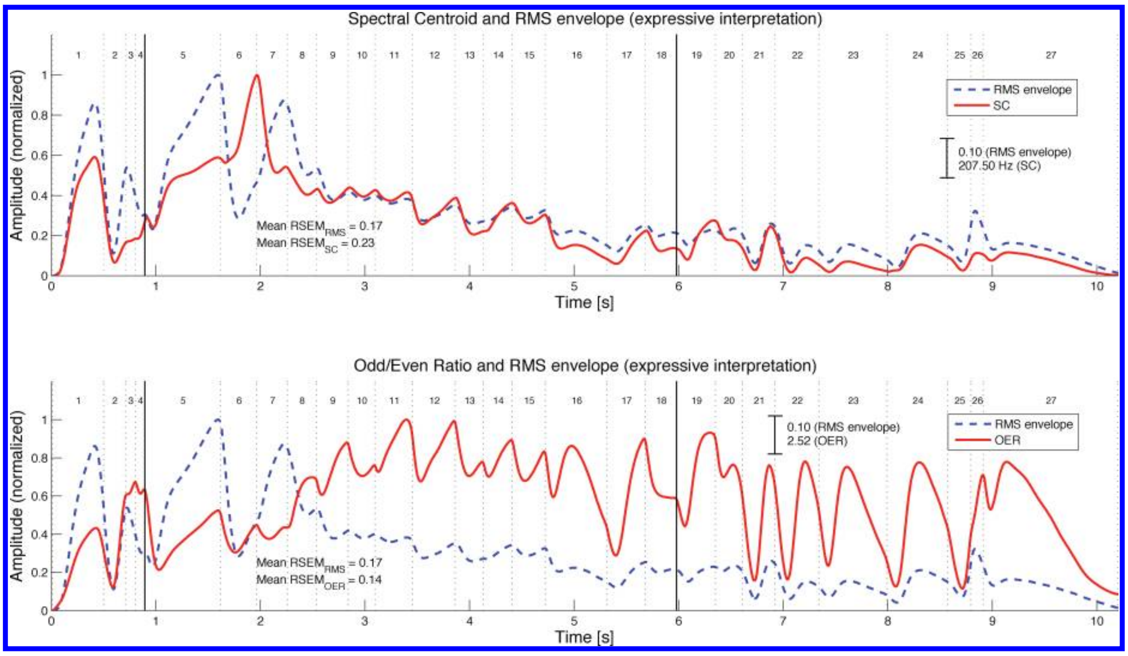


FIGURE 6. Comparison between the averaged spectral centroid (SC) and odd/even ratio (OER) variations and the root mean square envelope (ENV) in the case of the expressive interpretations of the Bach excerpt. The values of SC, OER, and ENV have been normalized with respect to their maximal values, for display purposes. The averages of the relative standard errors of the mean (RSEM) computed across the data points are reported.

lasted longer, and had a longer attack time, and a higher spectral centroid in the more expressive interpretation. It is worth noting that this note plays an important role in the musical structure because it is the first in the musical phrase. Timbre-related changes and timing variations of this kind may be used by the performers to emphasize the importance of a specific tone by increasing the musical tension locally.

The correlations (Pearson) between the spectral timbre descriptors and the acoustical energy were also analyzed, as some of the perceptual effects of these features have been studied in (Barthet, Depalle, Kronland-Martinet, & Ystad, in press). Figure 6 shows comparisons between the average spectral centroid, root mean square envelope and odd/even ratio patterns, computed from the expressive performances of the Bach excerpt. The spectral centroid was closely correlated in a linear way with the RMS envelope for both the Bach ($r = .82$, $p < .0001$) and Mozart ($r = .92$, $p < .0001$) excerpts. As explained by the Worman-Benade laws, the spectral richness of acoustical wind instrument tones is highly sensitive to the blowing pressure, which is correlated with the resulting acoustical energy output (Benade & Kouzoupis, 1988). Conversely, the relationship between the odd/even ratio and the RMS envelope was more complex. For some notes, the differences in the odd/even ratio depending on the levels of expression were in line with those in the spectral centroid (for instance, with OERM and SCM in the passage from N11 to N15 of the Bach excerpt, see Table 3). However, this was not systematic (for instance, notes N6, N7, N24, and N27 of the Bach excerpt showed significant differences in the OERM but not in the SCM values, see Table 3). The linear correlation between OER and ENV was very weak for the Bach ($r = .08$, $p < .05$) and Mozart ($r = .41$, $p < .0001$) excerpts. It can be seen from Figure 6 that OER and ENV were linearly correlated at the beginning of the phrase of the Bach excerpt, when the acoustical energy was high. Then, from note N8 onwards, a non-linear jump in the OER occurred towards higher values when the acoustical energy decreased. The overall increase in OER observed during the decrescendo phase was linked to the fact that when the playing level is weak, clarinet tones are nearly sinusoidal and contain very few even harmonics (this also explains why the spectral centroid decreased during the decrescendo).

Discussion

Several conclusions can be drawn from the results of these experiments. First, the changes in the timbre, temporal, dynamics, and pitch descriptors were found to be highly consistent when the performer's expressive intentions

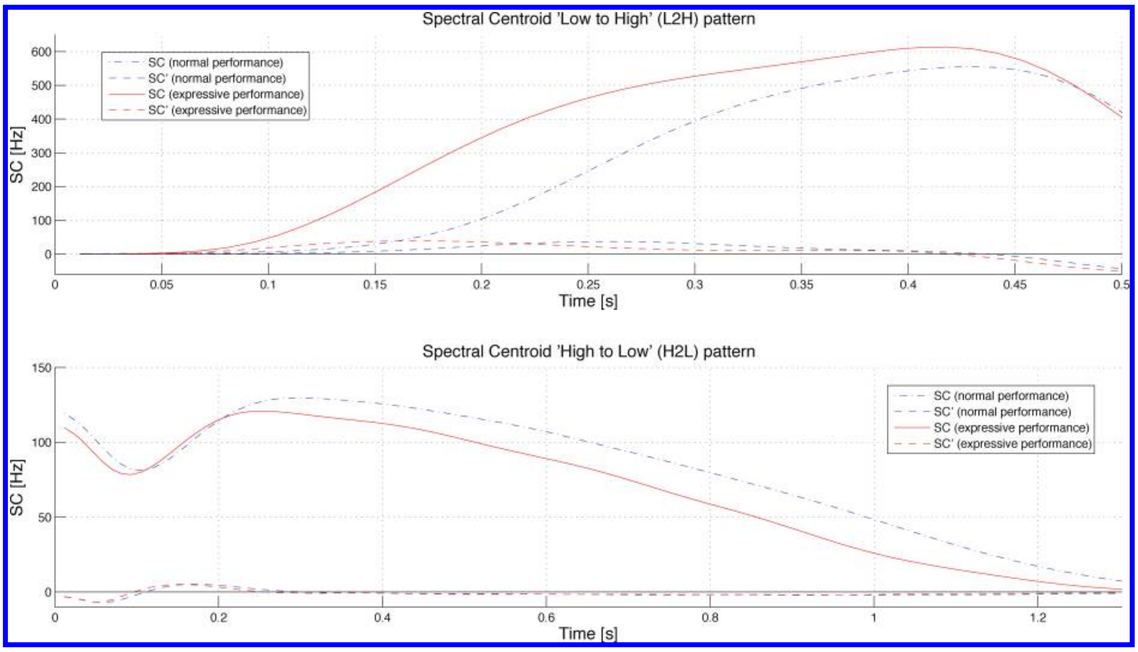


FIGURE 7. Examples of spectral centroid 'Low to High' (top) and 'High to Low' (bottom) patterns. The spectral centroid patterns and their derivatives SC′ correspond to tones in the mechanical and expressive performances of the Bach excerpt.

were the same. These findings confirm that expert performers are able to faithfully reproduce their expressive deviations at a given level of expressiveness. The behavior of the timbre descriptors is therefore not random but seems on the contrary to follow rules of interpretation. This suggests that the performer used the acoustical parameters associated with timbre to vary his interpretation. This would certainly be true if changes in the timbre descriptors occurred when the level of expression changed. The above statistical analyses show in fact that most of the time, strong interactions occured between the performer's expressive intentions and note factors, for the timbre descriptors. Although the timbre-related changes observed in the case of the Bach excerpt were quite subtle, they were found to be significant (medium to large effect sizes) in the case of 18 notes out of 27 (upon combining the AT, SCM, SCR, OERM, and OERR data). The changes in the timbre descriptors made for expressive purposes were higher in the performances of the Mozart excerpt (58 notes out of 75 showed significant differences with large effect sizes, when the SCM, SCR, OERM, and OERR data were combined). This piece contains many more long notes (half notes, quarter notes, dotted quarter notes) than in the Bach excerpt, which mainly contains semiquavers notes. In other words, a close control of the tone quality may be possible and worthwhile only when the duration of the notes is sufficiently long.

Several kinds of spectral centroid patterns were found to occur recurrently. For instance, notes N1, N6, N12, N14 in the Bach excerpt showed an increasing spectral centroid trend, which we have called the 'Low to High' (L2H) pattern, whereas notes N16 and N27 showed a decreasing spectral centroid trend, which we have called the 'High to Low' (H2L) pattern. During notes N22 to N24, the spectral centroid successively increased and decreased, corresponding to what we have called the 'Low to High to Low' (L2H2L) pattern. Figure 7 gives some examples of typical spectral centroid patterns at each of the two expressive levels in the case of the Bach excerpt. The L2H spectral centroid pattern corresponding to the first note in the sequence (N1) can be seen at the top of the figure, and the H2L spectral centroid pattern corresponding to the last note in the sequence (N27) is shown at the bottom. It is worth pointing out that the shape of the L2H pattern depended on the player's expressive intentions. The slope was higher at the beginning of the tone in the case of the more expressive performance (see the derivative SC′, which was higher during the first part of the tone in the expressive performance than in the mechanical one). The shape of the H2L pattern given in this example was roughly the same in both interpretations, as shown by the fact that the derivatives were almost the same (the curves differed in their means, but not in their range, see Table 3). In order to determine whether these variations in spectral centroid corresponded to audible changes of brightness, we used a transformation relying on an additive-based synthesis model for harmonic tones to freeze the spectral centroid of the tone at a given time. This was done by eliminating the change in shape of the spectral envelope over time as in the amplitude-envelope coherence transformation defined in (McAdams, Beauchamp, & Meneguzzi, 1999). The spectral centroids of the tones were frozen to their respective values at the beginning of the tone, after the attack part, and at the end of the tone, before the release part (cf. sound examples *3a* to *3c* corresponding to note N1, and *4a* to *4c* corresponding to note N27). These synthesized tones give the net changes of brightness between the beginning and the end of the tones. It is worth noting that the L2H pattern was concurrent with an increase in the overall amplitude, and the H2L pattern was concurrent with a decrease in the overall amplitude. The question arises as to whether the timbre- or dynamic-related variations are more important from the perceptual point of view as means of conveying expressiveness. This point is addressed in our companion article (Barthet et al., in press).

Summary and Conclusions

In this study, we investigated whether timbre, temporal, dynamics, and pitch descriptors showed any changes when performers played expressively. To answer this question, mechanical and expressive clarinet performances of musical excerpts from the Baroque and Classical repertoire were compared. When the clarinetist repeated the performances with the same musical intentions, the timbre descriptors showed a high level of consistency. The performer also reproduced the intertone onset interval deviations and root mean square envelope highly consistently, as previously reported to occur in studies on timing and dynamics variations (see e.g., Kendall & Carterette, 1990; Palmer, 1996; Repp, 1992). The temporal and spectral timbre descriptors studied here (attack time, spectral centroid, and odd/even ratio), as well as the timing and dynamics descriptors, differed significantly between mechanical and expressive performances. The timbre-related changes across expressive levels did not occur at every note, but were specific to some notes, or groups of notes in the musical phrases (such as the first note in a phrase, or specific passages). The most conspicuous changes were in the mean spectral centroid and odd/even ratio values and the range of variations in the duration of the tones. These changes seemed to be made more frequently in the case of long notes (such as half and quarter notes), possibly because a performer needs to have sufficient time to control the timbre while playing. According to the communication theory of musical expression proposed by Kendall and Carterette (1990) in the context of traditional Western art music, musical messages are first recoded from ideas to notations by the composer, then recoded from notations to acoustical signals by the performer, and finally recoded from acoustical signals to ideas by the listener. The findings reported in this study support this model of musical communication, since the distinct signatures across expressive levels showed by the timbre, temporal and dynamics descriptors afford a basis for perceptual discrimination between the performances.

In line with the "musical expression" hypothesis put forward by Clarke (1988) to explain the role of expressive timing, it seems likely that timbre as well as timing may be used by performers to communicate the musical structure to listeners, in addition to mediating expressiveness. For instance, by increasing the brightness of a tone while it is being held (as reflected in an increase in the spectral centroid), a performer may insist on the role of this specific tone in the musical structure. This might be so in the case of some notes in particular, which all showed specific spectral centroid patterns across the repetitions of mechanical and expressive performances, such as the 'Low to High' (L2H) and 'High to Low' (H2L) spectral centroid patterns.

These results suggest that, in the case of instruments involving self-sustained oscillations like the clarinet, there may exist a link between the process of musical interpretation and the way the acoustical correlates of timbre vary with time. Some timbre-related variations seem to be intrinsically linked to intensity variations due to the physics of the instrument (cf. the strong correlation between spectral centroid and root mean square envelope), whereas others are not linearly correlated to intensity variations (cf. the weak linear correlation between the odd/even ratio and the RMS envelope). Further research is now required to determine whether the relationships between spectral centroid, odd/even ratio and acoustical energy simply result from the intrinsic acoustical characteristics of the instrument, or whether they are deliberately induced by performers. The question of whether performers intended to produce intensity or timbre variations is hard to answer. These variations are likely to be sought by performers concomitantly, but a specific experimental procedure would have to be designed to test this hypothesis.

Therefore, variations in the temporal and spectral parameters contributing to timbre are likely factors that account for the disparities and commonalities between performers along with timing variations. The next step will consist of comparing the variations of the timbre descriptors observed between performances of the same musical excerpt by several performers using the same musical instrument. This issue could also be investigated with performers playing different instruments, in order to establish whether there exist any similarities in the variations of the timbre descriptors between the performances of a clarinetist and a cellist, for example. In the present study, which was the first step towards elucidating the role of timbre in musical interpretation, we focused on the timbre-related differences occurring at the single note level and in small groups of notes. As music usually involves larger phrases and movements, it might also be interesting to explore the variations in the timbre descriptors occurring at a higher level in the musical structure.

This study has shed light on some of the acoustical parameters that mediate a performer's expressive intentions, but the perceptual effects of these parameters were not investigated. In our companion article (Barthet et al., in press), we have compared the perceptual effects of spectral centroid, intertone onset interval, and acoustical energy variations on the musical appreciation of listeners. To address this issue, an analysis-by-synthesis approach was adopted in order to assess the effects of the expressive deviations detected in recorded clarinet performances.

Author Note

We would like to thank the clarinetist Claude Crousier for participating in this project and for many fruitful discussions. We are grateful to Mitsuko Aramaki and Henri Burnet from the Institut des Neurosciences Cognitives de la Méditerranée and Jean Pierre Durbec from the Centre d'Océanologie de Marseille for their precious help. We also wish to thank the reviewers for their helpful comments and corrections. This project was partly supported by the French National Research Agency (ANR JC05-41996 "senSons", http://www.sensons.cnrs-mrs.fr/).

Author Philippe Depalle is now affiliated with the Sound Processing and Control Laboratory, The Schulich School of Music, McGill University, Montreal, Canada.

Correspondence concerning this article should be addressed to Mathieu Barthet, CNRS Laboratoire de Mécanique et d'Acoustique, 31 chemin Joseph-Aiguier, 13402 Marseille Cedex 20, France. E-mail: barthet@lma.cnrs-mrs.fr

References

ANSI. (1960). *USA standard acoustical terminology.* New York: American National Standards Institute.

Barthet, M. (2008). *De l'interprète à l'auditeur: Une analyse acoustique et perceptive du timbre musical [From performer to listener: An acoustical and perceptual analysis of musical timbre]*. Unpublished doctoral dissertation, Université Aix-Marseille II.

Barthet, M., Depalle, P., Kronland-Martinet, R., & Ystad, S. (in press). Analysis-by-synthesis of timbre, timing, and dynamics in expressive clarinet performance. *Music Perception*.

Barthet, M., Guillemain, P., Kronland-Martinet, R., & Ystad, S. (2005). On the relative influence of even and odd harmonics in clarinet timbre. In *Proceedings of the International Computer Music Conference* (*ICMC'05*) (pp. 351–354). Barcelona, Spain: International Computer Music Association (ICMA).

Barthet, M., Guillemain, P., Kronland-Martinet, R., & Ystad, S. (2010). From clarinet control to timbre perception. *Acta Acustica united with Acustica*, *96*, *4*, 678-689.

Barthet, M., Kronland-Martinet, R., & Ystad, S. (2006). Consistency of timbre patterns in expressive music performance. In *Proceedings of the 9th International Conference on Digital Audio Effects (DAFx06)* (pp. 19–24). Montreal, Quebec, Canada.

Beauchamp, J. W. (1982). Synthesis by spectral amplitude and brightness matching of analyzed musical instrument tones. *Journal of the Audio Engineering Society, 30*, 396–406.

Benade, A. H., & Kouzoupis, S. N. (1988). The clarinet spectrum: Theory and experiment. *Journal of the Acoustical Society of America, 83*, 292–304.

Bregman, A. S. (1994). *Auditory scene analysis–The perceptual organization of sound.* Cambridge, MA: MIT Press.

Castellengo, M., & Dubois, D. (2005). Timbre ou timbres? Propriété du signal, de l'instrument, ou construction cognitive? [Timbre or timbres? Property of the signal, the instrument, or cognitive construction?]. In *Proceedings of the Conference on Interdisciplinary Musicology (CIM05)*. Montréal, Québec, Canada.

Clarke, E. F. (1988). Generative principles in music performance. In J. A. Sloboda (Ed.), *Generative processes in music. The psychology of performance, improvisation and composition* (pp. 1–26). Oxford, UK: Oxford University Press.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Födermayr, F., & Deutsch, W. A. (1993). Parmi veder le lagrime. One aria, three interpretations. In A. Friberg (Ed.), *Proceedings of the Stockholm Music Acoustics Conference* (pp. 96–107). Stockholm, Sweden: Royal Swedish Academy of Music.

Fritz, C. (2004). *La clarinette et le clarinettiste: Influence du conduit vocal sur la production du son [The clarinet and the clarinetist: Influence of the vocal tract on the sound production]*. Unpublished doctoral dissertation, Université Paris 6 and University of New South Wales.

Gabrielsson, A. (1998). The performance of music. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 501–602). San Diego, CA: Academic Press.

Gabrielsson, A., & Lindstrom, B. (1995). Emotional expression in synthesizer and sentograph performance. *Psychomusicology, 14*, 94–116.

Genesis. (2010). *Lea software*. http://www.genesis-acoustics.com/. Aix-en-Provence, France.

Gordon, J. W. (1987). The perceptual attack time of musical tones. *Journal of the Acoustical Society of America*, *82*, 88–105.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America*, *61*, 1270–1277.

Grey, J. M., & Gordon, J. W. (1978). Perception of spectral modifications on orchestral instrument tones. *Computer Music Journal*, *11*, 24–31.

Hajda, J. M., Kendall, R. A., Carterette, E. C., & Harshberger, M. L. (1997). Methodological issues in timbre research. In I. Deliége & J. A. Sloboda (Eds.), *Perception and cognition of music* (2nd ed., pp. 253–306). New York: Psychology Press.

Handel, S. (1995). Timbre perception and auditory object identification. In B. C. J. Moore (Ed.), *Handbook of perception and cognition* (2nd ed., pp. 425–461). San Diego, CA: Academic Press.

Helland, R. T. (2004). *Synthesis models as a tool for timbre studies.* Unpublished master's thesis, Norwegian University of Science and Technology, Trondheim, Norway.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65–70.

Jaillet, F. (2005). *Représentation et traitement temps-fréquence des signaux audionumériques pour des applications de design sonore [Audio signals representation and time-frequency processing for sound design applications]*. Unpublished doctoral dissertation, Université de la Méditerranée–Aix-Marseille II.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*, 770–814.

Kendall, R. A., & Carterette, E. C. (1990). The communication of musical expression. *Music Perception*, *8*, 129–164.

Kendall, R. A., & Carterette, E. C. (1991). Perceptual scaling of simultaneous wind instrument timbres. *Music Perception*, *8*, 369–404.

Kendall, R. A., & Carterette, E. C. (1996). Difference thresholds for timbre related to spectral centroid. In B. Pennycook & E. Costa-Giomi (Eds.), *Proceedings of the 4th International Conference on Music Perception and Cognition (ICMPC)* (pp. 91–95). Montreal, Canada: McGill University.

Krimphoff, J., McAdams, S., & Winsberg, S. (1994). Caractérisation du timbre des sons complexes, II Analyses acoustiques et quantification psychophysique [Characterization of complex sounds' timbre, II Acoustical analyses and psychophysical quantification]. *Journal de Physique IV, Colloque C5, 4*, 625–628.

Krumhansl, C. L. (1989). Why is musical timbre so hard to understand? In S. Nielzén & O. Olsson (Eds.), *Proceedings of the Marcus Wallenberg Symposium held in Lund, Sweden* (pp. 43–53). Amsterdam: Excerpta Medica.

Lichte, W. H. (1941). Attributes of complex tones. *Journal of Experimental Psychology*, *28*, 455–480.

Loureiro, M. A., Paula, H. B. de, & Yehia, H. C. (2004). Timbre classification of a single instrument. In *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR 2004)*. Barcelona, Spain.

Ludbrook, J. (1998). Multiple comparison procedures updated. *Clinical and Experimental Pharmacology and Physiology*, *25*, 1032–1037.

Marozeau, J., Cheveigné, A. de, McAdams, S., & Winsberg, S. (2003). The dependency of timbre on fundamental frequency. *Journal of the Acoustical Society of America*, *114*, 2946–2957.

McAdams, S. (1994). La reconnaissance de sources et d'événements sonores [The recognition of sources and sound events]. In S. McAdams & E. Bigand (Eds.), *Penser les sons [Think sounds]* (1st ed., pp. 157–213). Paris: Presses Universitaires de France.

McAdams, S., Beauchamp, J. W., & Meneguzzi, S. (1999). Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters. *Journal of the Acoustical Society of America*, *105,* 882–897.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, *58,* 177–192.

Palmer, C. (1996). Anatomy of a performance: Sources of musical expression. *Music Perception*, *13,* 433–454.

Palmer, C. (1997). Music performance. *Annual Review of Psychology*, *48,* 115–138.

Peeters, G. (2004). *A large set of audio features for sound description (similarity and description) in the Cuidado project.* Paris: I.R.C.A.M. (Unpublished technical report).

PENEL, A., & DRAKE, C. (2004). Timing variations in music performance: Musical communication, perceptual compensation, and/or motor control? *Perception and Psychophysics*, *66,* 545–562.

PICINBONO, B. (1997). On instantaneous amplitude and phase of signals. *IEEE Transactions on Signal Processing, 45*, 552–560.

REPP, B. H. (1992). Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's Träumerei. *Journal of the Acoustical Society of America*, *92,* 2546–2568.

RISSET, J.-C. (1991). Timbre et synthèse des sons [Timbre and sound synthesis]. In Christian Bourgois (Ed.), *Le timbre, métaphore pour la composition [Timbre, metaphor for composition]* (pp. 239–260). Paris: I.R.C.A.M.

SCHAEFFER, P. (1966). *Traité des objets musicaux [Treaty of musical objects]*. Paris, France: Seuil.

SCHOLES, P. A. (1960). *The Oxford companion to music (2nd ed.)*. Oxford, UK: Oxford University Press.

SEASHORE, C. E. (1967). *Psychology of music*. New York: Dover Publications.

TERHARDT, E., STOLL, G., & SEEWANN, M. (1982). Pitch of complex signals according to virtual-pitch theory: Tests, examples, and predictions. *Journal of the Acoustical Society of America*, *71,* 671–678.

TODD, N. P. M. (1992). The dynamics of dynamics: A model of musical expression. *Journal of the Acoustical Society of America, 91*, 3540–3550.

TRAUBE, C. (2004). *An interdisciplinary study of the timbre of the classical guitar*. Unpublished doctoral dissertation, McGill University, Montreal, Canada.

WANDERLEY, M. (2002). Quantitative analysis of non-obvious performer gestures. In I. Wachsmuth & T. Sowa (Eds.), *Gesture and sign language in human-computer interaction* (Vol. 2298, pp. 241–253). Heidelberg, Germany: Springer Berlin.

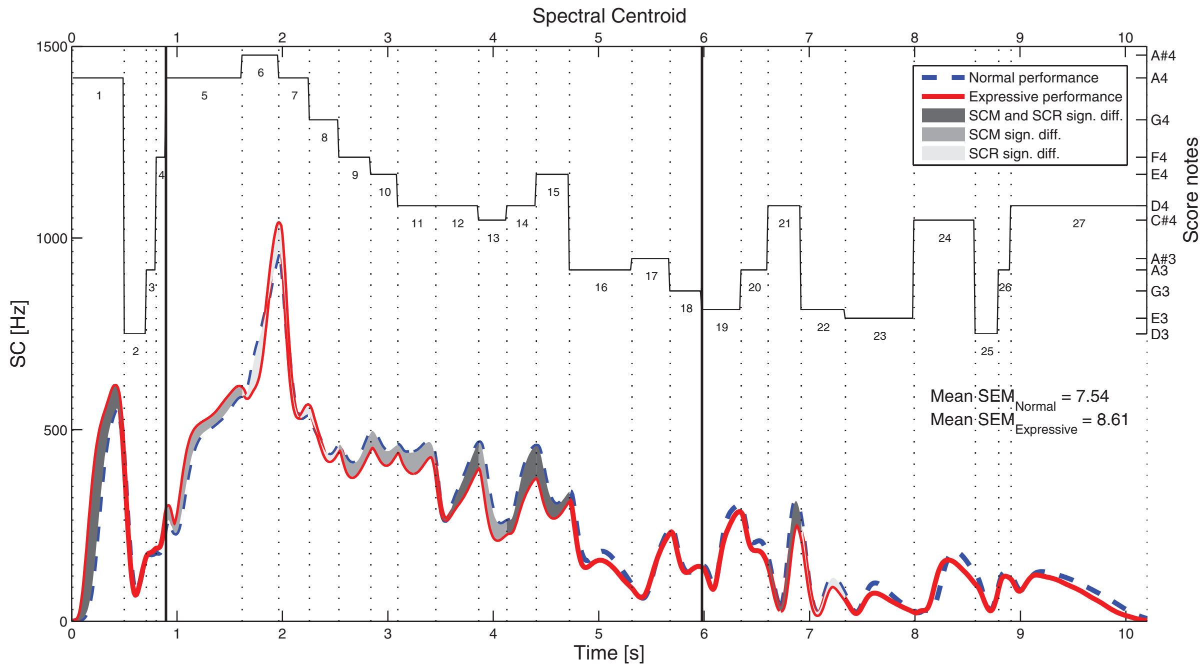
WESSEL, D. L. (1979). Timbre space as a musical control structure. *Computer Music Journal*, *3,* 45–52.

WIER, C. C., JESTEADTT, W., & GREEN, D. M. (1977). Frequency discrimination as a function of frequency and sensation level. *Journal of the Acoustical Society of America*, *6,* 178–184.

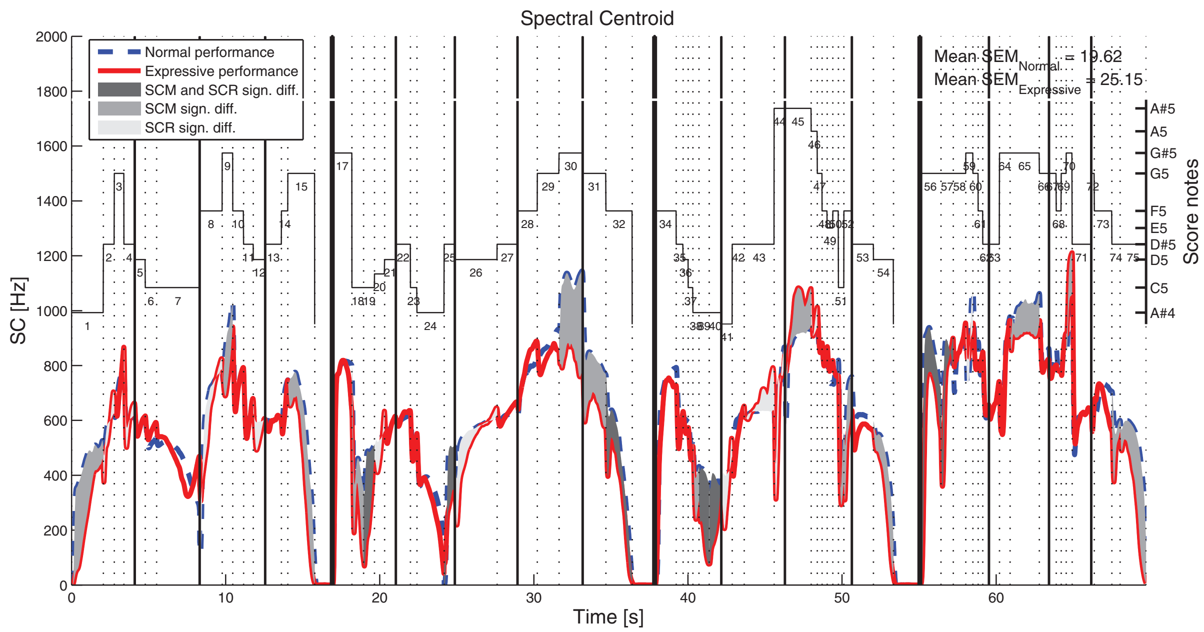
ZWICKER, E., & FASTL, H. (1990). *Psychoacoustics: Facts and models*. New York: Springer-Verlag.

color plates

Mathieu Barthet, Philippe Depalle, Richard Kronland-Martinet, & Sølvi Ystad, *Acoustical Correlates of Timbre and Expressiveness in Clarinet Performance.* Figures 4 and 5

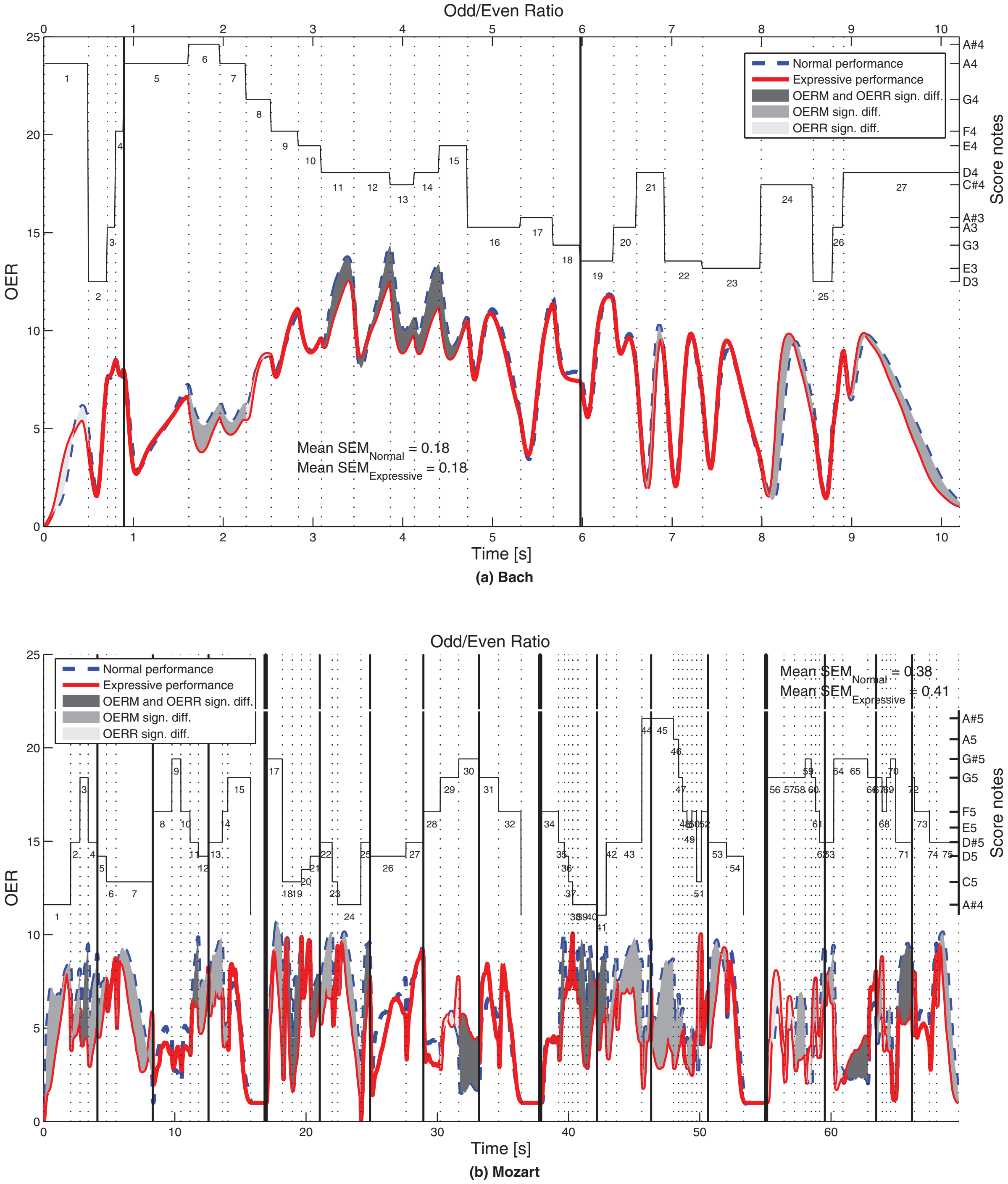


(a) Bach



(b) Mozart

MATHIEU BARTHET, PHILIPPE DEPALLE, RICHARD KRONLAND-MARTINET, & SØLVI YSTAD, **FIGURE 4.**
Average short-term spectral centroid in the mechanical (dashed line) and expressive (solid line) interpretations of the Bach (a) and Mozart (b) excerpts. The notes and groups of notes with which statistical analyses showed the existence of significant effects are shown in gray. The dark gray areas indicate significant differences in both the spectral centroid mean (SCM) and spectral centroid range (SCR). The pale gray areas indicate significant differences in SCM only. The light gray areas indicate significant differences in SCR only. For other explanations, see the legend to Figure 2.



MATHIEU BARTHET, PHILIPPE DEPALLE, RICHARD KRONLAND-MARTINET, & SØLVI YSTAD, **FIGURE 5.**
Average short-term odd/even ratio in the mechanical (dashed line) and expressive (solid line) interpretations of the Bach (a) and Mozart (b) excerpts. The notes and groups of notes with which statistical analyses showed the existence of significant effects are shown in gray. The dark gray areas indicate significant differences in both the odd/even ratio mean (OERM) and odd/even ratio range (OERR). The pale gray areas indicate significant differences in OERM only. The light gray areas indicate significant differences in OERR only. For other explanations, see the legend to Figure 2.