Analysis-By-Synthesis of Timbre, Timing, and Dynamics in Expressive Clarinet Performance

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

In a previous study, mechanical and expressive clarinet performances of Bach's *Suite No. II* and Mozart's *Quintet for Clarinet and Strings* were analyzed to determine whether some acoustical correlates of timbre (e.g., spectral centroid), timing (intertone onset interval), and dynamics (root mean square envelope) showed significant differences depending on the expressive intention of the performer. In the present companion study, we investigate the effects of these acoustical parameters on listeners' preferences. An analysis-by-synthesis approach was used to transform previously recorded clarinet performances by reducing the expressive deviations from the spectral centroid, the intertone onset interval and the acoustical energy. Twenty skilled musicians were asked to select which version they preferred in a paired-comparison task. The results of statistical analyses showed that the removal of the spectral centroid variations resulted in the greatest loss of musical preference.

Received September 11, 2008, accepted May 4, 2010.

Key words: timbre, timing, dynamics, expressive clarinet performance, preference

Analysis-By-Synthesis of Timbre, Timing, and Dynamics in Expressive Clarinet Performance

The model for musical communication proposed by Kendall and Carterette (1990) in the context of traditional Western art music describes a musical performance as an act during which a performer transforms the composer's notational signals into acoustical signals that must be decoded by a listener. Whether this act of performance can be said to constitute an (aesthetic) interpretation depends on how the performer translates the musical notations into sounds. As a matter of fact, the notion of interpretation is defined as "the act of performance with the implication that in this act, the performer's judgment and personality necessarily have their share" (Scholes, 1960). Previous studies on musical performance have shown that performers use timing and intensity variations to play expressively (see Gabrielsson, 1999, for a review). However, less attention has been paid so far to timbre (see e.g., Juslin & Laukka, 2003, for a review), the perceptual attribute of sound that was described by Seashore (1938/1967) as "the most important aspect of tone," and which "introduces the largest number of problems and variables." This article is the second part of a study designed to test whether timbre variations are linked to the expressive intentions of performers and to the musical preferences of listeners. Previous analyses of recorded clarinet performances helped to determine the acoustical correlates of a performer's expressive variations in timbre, timing and dynamics (Barthet, Depalle, Kronland-Martinet, & Ystad, 2010). The perceptual effects of these acoustical parameters were investigated in the present study in terms of musical preferences, using an analysis-by-synthesis approach (Risset & Wessel, 1999).

Modeling Musical Interpretation

In addition to measuring musical performances, many authors have focused on modeling musical expressiveness. Detailed historical reviews of models for musical interpretation have been published by Widmer and Goebl (2004) and DePoli (2006). The strategies most commonly used for this purpose are analysis-by-measurement (see for example, Todd, 1992, 1995; Windsor, Desain, Penel, & Borkent, 2006), analysis-by-synthesis, and music theory knowledge (see for example, Friberg, 1995), machine learning (see for example, Goebl, Pampalk, & Widmer, 2004; Tobudic & Widmer, 2003), and combinations of these methods. Some of these models are predictive and account for the act of performance at its source, based on the assumption that interpretation can be described as a set of generative rules, while others are based on descriptive parameters and account for the effects of interpretation, i.e., the expressive deviations. All of the above

Music Perception volume 28, issue 3, pp. 265–278, issn 0730-7829, electronic issn 1533-8312 DOI:10.1525/mp.2011.28.3.265

models focus mainly on timing, dynamics, articulation, and intonation, whereas timbre has often been neglected. It seems necessary to model changes of timbre in order to efficiently simulate expressive performances, especially in the case of self-sustained instruments such as the clarinet or the violin, with which the sound continues to be controlled after the onset of a note (Kergomard, 1991). In their model, Canazza, Rodá, and Orio (1999) and Canazza, De Poli, Drioli, Rodá, and Vidolin (2004) took timbre variations into account in addition to the rhythmic and dynamic aspects of musical performance.

Timbre, A Multidimensional Perceptual Attribute of Complex Tones

Timbre is by definition the perceptual attribute that allows one to distinguish tones of equal pitch, loudness, and duration (ANSI, 1960). According to Handel (1995), the identification of timbre depends on on our ability to recognize various physical factors that determine the acoustic signal produced by musical instruments (called "source" mode of timbre perception in Hajda, Kendall, Carterette, & Harshberger, 1997), as well as to analyze the acoustic properties of sound objects perceived by the ear, which has traditionally been modelled as a time-evolving frequency analyser (called "interpretative" mode of timbre perception in Hajda et al., 1997). Timbre therefore involves two complementary facets, as it relates to both the identity and the quality of sound sources. No general models for timbre have been developed so far. However, the seminal research by Grey (1977), Wessel (1979), Krumhansl (1989), Kendall and Carterette (1991) and McAdams, Winsberg, Donnadieu, De Soete, and Krimphoff (1995) developed geometrical models for timbre that represent the organization of perceptual distances (the so-called timbre space), measured on the basis of dissimilarity judgments between tones with equal pitch, loudness, and perceived durations. Two- to four-dimensional timbre spaces have often been found in dissimilarity studies between various natural or synthetic tones corresponding to orchestral instruments (Caclin, McAdams, Smith, & Winsberg, 2005). The main acoustic correlates of timbre-space dimensions are the attack time (correlated with the rate of energy increase in a sound, see Krimphoff, McAdams, & Winsberg, 1994), and the spectral centroid (the mean of the spectral energy distribution, see Grey & Gordon, 1978). The spectral flux (measure of the fluctuation of the spectrum over time, see McAdams et al., 1995), and the spectral irregularity (an index to the disparities between the harmonic components, see Krimphoff et al., 1994) are examples of other timbre descriptors that have often been proposed as correlates of timbre space dimensions.

On the Role of Timbre in Musical Performance

In Western traditional music—which is the type of music with which most studies on musical performance have dealt—the role of timbre seems to be understimated compared to those of rhythm, pitch, and dynamics. This is notably revealed by the Western traditional notation system that almost omits timbre except in the references to the instrument (typological aspect of timbre), or in certain musical terms (e.g., *dolce*, *duramente*, *con brio*, *legato*), which can refer directly or indirectly to morphological aspects of timbre (Risset, 1994). It is worth noting that some highly complex timbre notation systems including more than one hundred symbols have been developed for some traditional Chinese and Japanese instruments such as the Chin, an ancient Chinese seven-string lute (see Traube, 2004), and the Shakuhachi, a Japanese bamboo flute. In the music played with these instruments, timbre therefore seems to be given as much importance as rhythm or pitch. It was not until the contemporary period and the new possibilities provided by digital sounds that timbre was placed at the foreground of Occidental music by composers such as Varèse, who proposed not only to compose with sounds but to compose the sounds themselves (Risset, 1994). The lack of notational information relating to timbre in traditional Western tonal music certainly does not mean that performers do not use timbre as a parameter to express feelings and emotions, however.

In a previous experiment (Barthet et al., 2010), we investigated the acoustical parameters accounting for expressiveness in clarinet playing. To address this issue, mechanical and expressive performances of excerpts from the Classical and Baroque repertoire were recorded and analyzed. Several temporal and spectral parameters were computed to characterize the acoustical features of the performer's interpretations. Statistical analysis of the data showed significant effects of the expressive intention on three timbre descriptors adapted to the clarinet (the attack time [AT], the spectral centroid [SC], and the odd/even ratio [OER], see Barthet, 2008), the intertone onset interval (IOI), quantifying the durations of the tones (see Repp, 1992), and the root mean square envelope (ENV), characterizing the variations of the acoustical energy. These results showed that changes in the timbre descriptors (AT, SC, OER), timing descriptor (IOI), and dynamics descriptor (ENV) occurred when the performer played in a more expressive way. Among the three timbre descriptors of clarinet tones (attack time, spectral centroid, odd/even ratio) that were analyzed in (Barthet et al., 2010), the spectral centroid was found to be the main predictor of the performer's expressive intentions (among all descriptors, SC was the one that most frequently differentiated between mechanical and expressive performances).

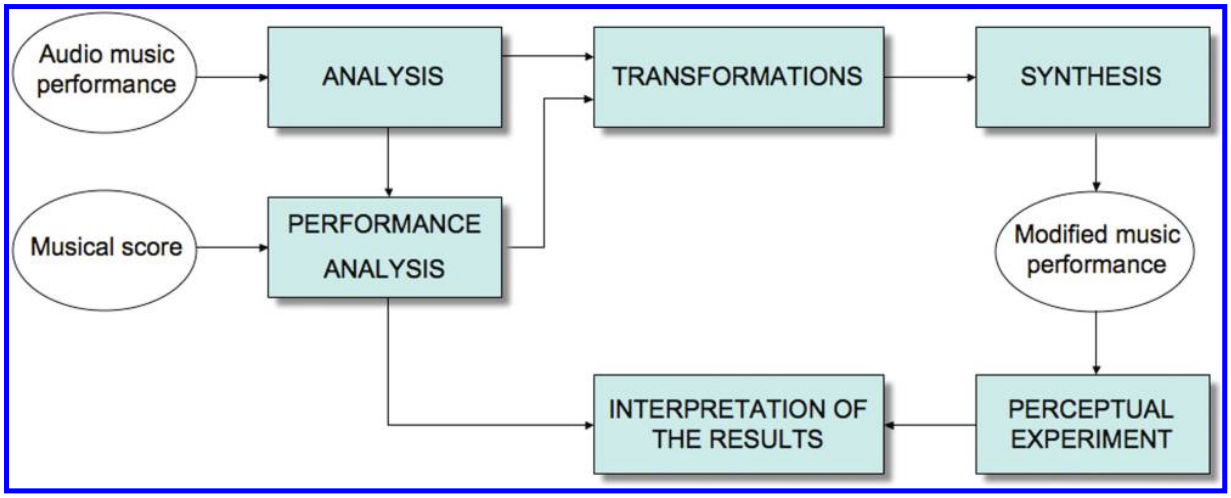


FIGURE 1. Exploration of the acoustical correlates of musical expressiveness based on the analysis-by-synthesis approach.

We therefore decided to focus on this timbre descriptor in the present experiments. The variations of acoustical features characterizing recorded clarinet performances were modified using signal processing techniques in order to further assess the effects of these changes on the musical preferences of listeners.

Method

General Methodology

We developed a general methodology to explore the role of timbre in musical performance, based on the analysis-by-synthesis paradigm (Risset & Wessel, 1999). The method, which is summarized in Figure 1, comprises four steps: the extraction of a representation from the signal (analysis), the transformation of this representation in the frequency domain, the conversion of the modified representation back to the time domain (synthesis), and the analysis of the perceptual effects induced by the transformation.

In order to identify the acoustical parameters that contribute most to musical aesthetic judgments, we investigated the effects of reducing the performer's original expressive deviations on preferences expressed by listeners in a paired-comparison task.

Stimuli

SOUND CORPUS

Expressive clarinet performances of the *Allemande* movement of Bach's *Suite No. II* (BWV 1008) and the *Larghetto* movement from Mozart's *Quintet for Clarinet and Strings* (KV 581) were recorded in an anechoic chamber (see Barthet et al., 2010, for further details). The scores of the excerpts and the sound examples related to the study are available at: http://www.lma.cnrs-mrs.fr/~kronland/Interpretation_perceptual. The first musical phrases in these pieces were selected to generate the stimuli. These excerpts were chosen so that they would be sufficiently long to have a musical meaning (a musical phrase), but sufficiently short for a paired-comparison task (10 to 15 s excerpts; see (Gabrielsson & Lindstrom, 1985) for a discussion on the stimuli durations in paired comparisons).

ANALYSIS-SYNTHESIS MODEL

The musical sequences were resynthesized using an additive-based synthesis model. The sound model decomposes an audio signal in to a deterministic part, consisting of a sum of quasi-sinusoidal components plus a residual part. A tone $s(t)$ can thus be written as follows:

$$s(t)=\sum_{h=1}^{H} A_h(t)\cos[\phi_h(t)]+r(t)$$
$$\phi_h(t)=2\pi\int_0^t f_h(\tau)d\tau+\phi_h(0) \quad (1)$$

where $A_h(t)$, $\phi_h(t)$, and $f_h(t)$ are the instantaneous amplitude, phase, and frequency, respectively, of the hth among H sinusoids, $\phi_h(0)$ is the initial phase, and $r(t)$ is the residual. This method is particularly suitable for changing the characteristics of the tones related to timbre and timing, for example, since sounds are reconstructed as a superposition of partial components, the frequency and amplitude of which can be individually controlled.

In addition to their harmonic structure, clarinet tones contain ancillary noises (for instance breath and key noises) that also contribute to the identity of the instrument. The latter noises are partly contained in the residual. The residual was obtained by performing a time-domain subtraction between the original sequence and the resynthesized deterministic part with no transformations. Resynthesized sequences were then obtained by juxtaposing the resynthesized tones. This procedure gives a high-quality resynthesis of the clarinet performances (cf. sound examples 1 to 3).

TRANSFORMATIONS

Recordings of a players' original performances can be transformed by appropriately manipulating the control parameters in the model. In order to assess the perceptual effects of spectral centroid (SC) variations, intertone onset interval (IOI) deviations, and acoustical energy (ENV) variations, three transformations were carried out to independently modify these parameters.

Spectral centroid freezing (T_T). A transformation acting on the spectral centroid was designed to control the shape of the tones' spectral centroids without affecting their acoustical energy. The method used for this purpose was based on the transformation described by McAdams, Beauchamp, and Meneguzzi (1999), which consists in eliminating the spectral flux (measure of the fluctuation of the spectrum over time), while leaving the RMS envelope intact. As a matter of fact, the removal of the spectral flux corresponds to canceling the time-dependent variations in the spectral envelope's shape, thus cancelling the spectral centroid variations. As the drastic elimination of all spectral centroid variations may cause tones to sound too artificial, the microfluctuations in the instantaneous amplitudes were preserved by separating the slow variations in the instantaneous amplitudes $L_h(t)$, defined as the amplitude variations of frequencies below 10 Hz, from the fast variations $H_h(t)$, defined as the amplitude variations of frequencies above 10 Hz. The modified instantaneous amplitude $\tilde{A}h(t)$ of the hth component of a given tone is obtained as follows:

$$\tilde{A}_h(t) = \beta_h(t) ENV(t)$$
$$\beta_h(t) = \frac{\bar{A}_h + H_h(t)}{\sqrt{\sum_{h=1}^{H} [\bar{A}_h + H_h(t)]^2}} \quad (2)$$

where $ENV(t)$ is the RMS envelope of the tone, and $\beta_h(t)$ is a term computed to fluctuate around the time-averaged amplitude $\bar{A}_h$ of the hth harmonic. $\bar{A}_h$ was determined during the sustained part of the tone. Since the high-frequency fluctuations $H_h(t)$ are small in comparison with $\bar{A}_h$, the term $\beta_h(t)$ is almost constant with time. The shape of the tones' new instantaneous amplitudes $\tilde{A}_h(t)$ is there fore very similar to that of the RMS envelope of the tone $ENV(t)$ with various scale factors. Figure 2 shows the instantaneous amplitudes of a clarinet tone before and after the transformation. The modified spectral centroid is almost frozen over time, although it varies quickly around the mean value of the initial spectral centroid calculated during the sustained part of the tone.

The application of this transformation to one of the clarinet performances is given, as example, in Figure 3. The original spectral centroid variations are presented in Figure 3(a), and the modified ones in Figure 3(b).

Intertone onset interval deviation cancellation (T_R). We designed a transformation for replacing the effective intertone onset intervals (as played by the performer) with the nominal IOIs given by the transcription of the-score notations. This was done by applying time-scale modifications to the instantaneous frequencies and amplitudes of the tone's components. These changes were applied only to the sustained and release portions of the tones, sparing the original attack, which is known to be an important attribute of timbre. The time dilation/contraction coefficient α was computed for each tone as follows:

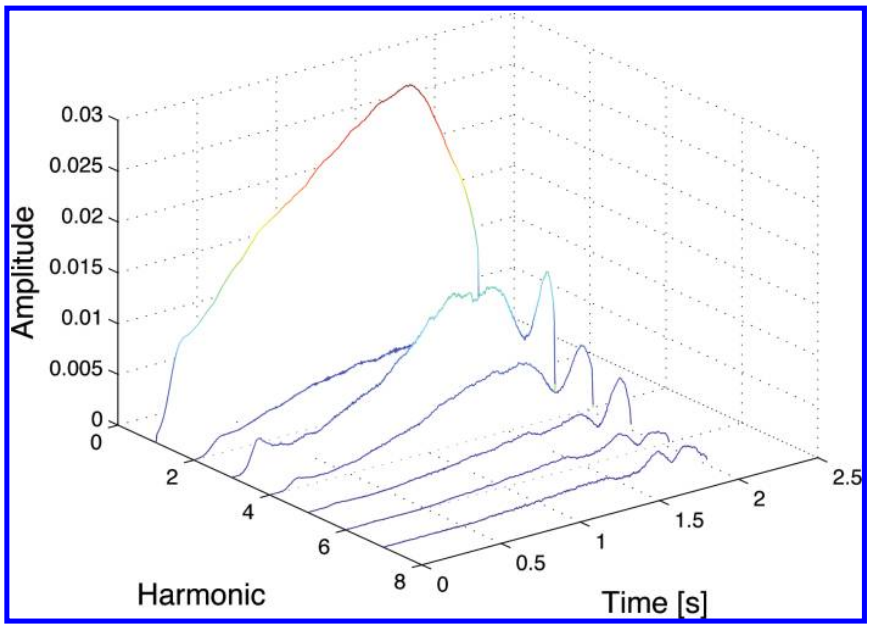
$$\alpha = \frac{IOI_{eff} - AT}{IOI_{nom} - AT} \quad (3)$$

where IOI_{eff} and IOI_{nom} denote the tone's effective and nominal IOIs, and AT is the attack time. For each tone, after the attack, the instantaneous frequencies $\tilde{f}_h(t)$ and amplitudes $\tilde{A}_h(t)$ were transformed as follows:

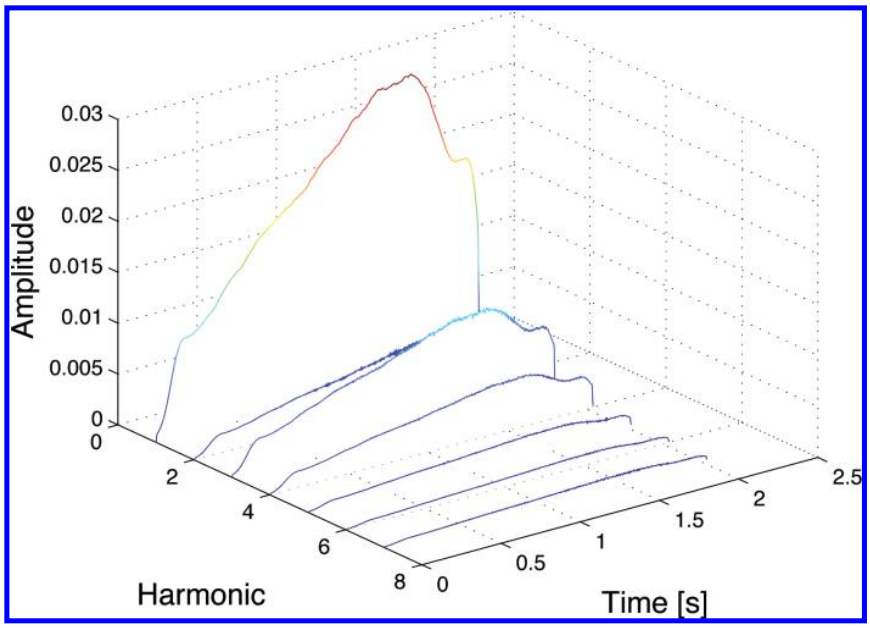
$$\tilde{f}_h(t) = f_h(\alpha t)$$
$$\tilde{A}_h(t) = A_h(\alpha t) \quad (4)$$

Figures 3(c) and 3(d) show the effect of the transformation on the IOI deviation descriptor ΔIOI. This transformation yields "mechanical" performances that are exactly in line with the timing indications given on the score.

Compression of the dynamics (T_D). In order to reduce the variations in the acoustical energy, we used a dynamic range controller serving as a compressor and limiter (see Zölzer, 1997). The signal's input level was determined via an envelope follower based on peak measurements. A gain factor was then used to adjust the amplitude of the input signals. This compressor/limiter was controlled



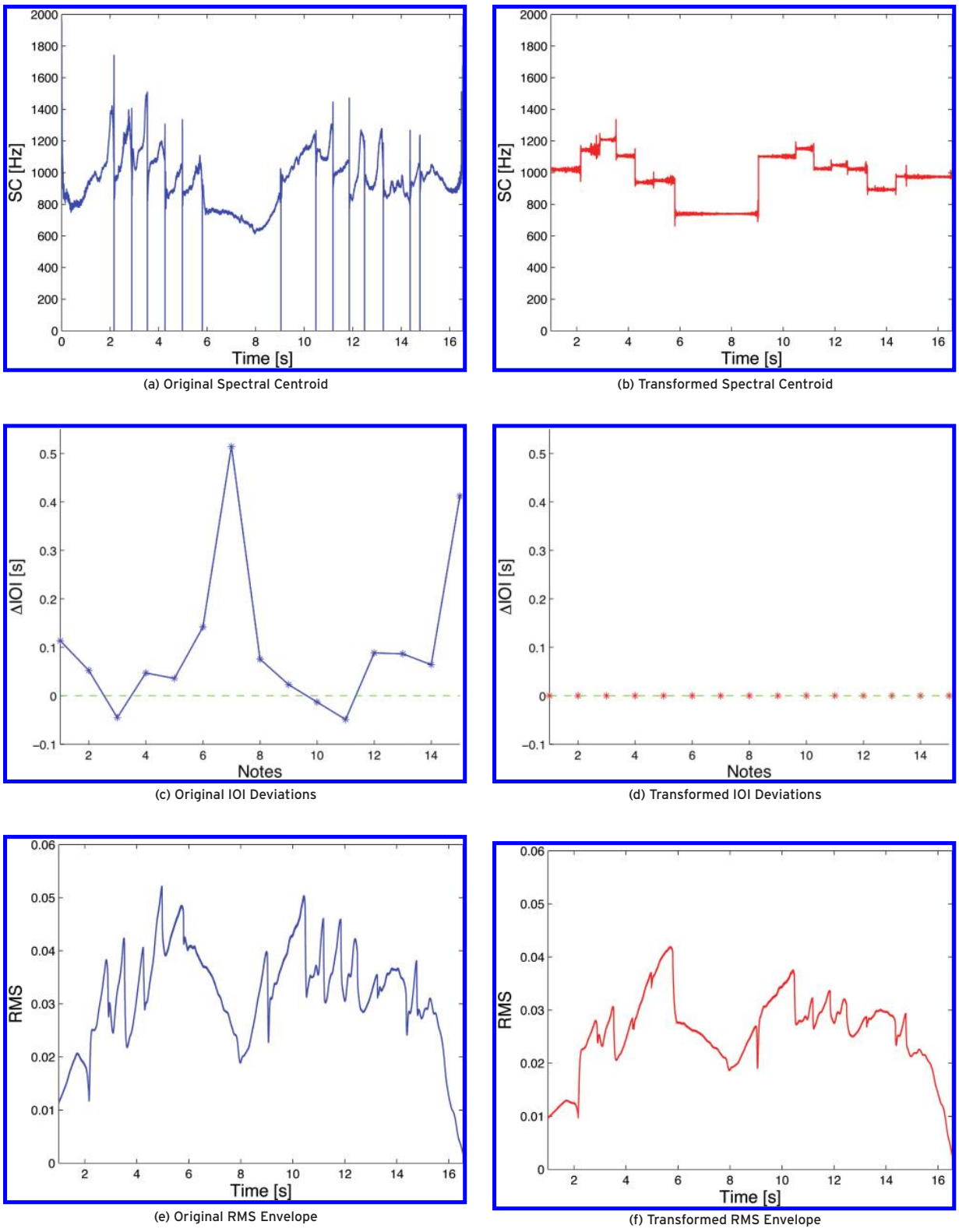
(a) Original Instantaneous Amplitudes



(b). Transformed Instantaneous Amplitudes

FIGURE 2. Transformations of the instantaneous amplitudes (harmonics 1 to 7): (a) original clarinet tone - (b) after the spectral centroid freezing.

by the parameters generally used with devices of this kind, i.e., by the thresholds and slope of the limiter, the thresholds and slope of the compressor, and the attack and release times. Although this method is based on nonlinear processing procedures liable to cause harmonic distortions, the control parameters of the dynamic range controller were carefully selected to prevent the occurrence of any audible changes in the timbre. The attack and release times were both set at 10 ms. As compression and limiting procedures lead to



(a) Original Spectral Centroid

(b) Transformed Spectral Centroid

(c) Original IOI Deviations

(d) Transformed IOI Deviations

(e) Original RMS Envelope

(f) Transformed RMS Envelope

FIGURE 3. Original (left) and transformed (right) expressive patterns for a clarinet performance of the Mozart excerpt.

changes in the signal levels, the loudness of each modified sequence was equalized to a constant value.

Examples of original and transformed RMS envelopes are given in Figures 3(e) and 3(f), respectively. As was to be expected, the range of variation of the acoustical energy is much smaller in the modified version.

To avoid effects of pitch on the preferences of the listeners, the fundamental frequencies of the tones were set at their mean values computed in the sustained parts, for each transformation. The instantaneous frequencies of the h tones' components were therefore set at $h\overline{f_0}$, where $\overline{f_0}$ denotes the mean fundamental frequency of the tone. We checked that the frequency changes were only weakly perceptible in the musical excerpts selected (cf. sound example 4).

DESIGN OF THE STIMULI

The three basic transformations (T_T, T_R, T_D) and their four combinations (T_{TR}, T_{TD}, T_{RD}, T_{TRD}) were applied to the expressive clarinet performances. As the original performances were recorded in an anechoic chamber, a slight reverberation was added to the resynthesized versions to make them sound more natural. The eight stimuli listed in Table 1 therefore were generated for each musical excerpt. The stimuli associated to the Bach and Mozart excerpts correspond to the sound examples 5 to 12, and 13 to 20, respectively.

Participants

Given the relatively demanding requirements of the auditory discrimination task, in which the participants had to rate the interpretations with various levels of expression, the experiment was carried out with skilled musicians (14 males, 6 females; age = 19–50 years) as listeners. Most of the participants were students in musicology practicing various musical instruments such as clarinet, guitar, piano, violin, etc., who were participating in improvisation workshops at the GRIM (Groupe de Recherche et d'Improvisations Musicales, Marseille).

TABLE 1. Description of the Stimuli.

Stimuli	Transformation description
M_0	No transformation
M_T	Freezing of the spectral centroid (T_T)
M_R	Canceling of the IOI deviations (T_R)
M_D	Compression of the dynamics (T_D)
M_{TR}	Combination of T_T and T_R
M_{TD}	Combination of T_T and T_D
M_{RD}	Combination of T_R and T_D
M_{TRD}	Combination of T_T, T_R and T_D

Apparatus

The experiment was carried out in an audiometric cabin. The user interface was implemented in the Matlab environment.[1] The sound files were stored on the hard drive of an Apple iMac G4 computer and delivered to the participants via a STAX SRM-310 headphone system.

Procedure

Participants were asked to select which stimuli they preferred in a paired comparisons task. They first underwent a training phase in order to become familiar with the task. In order to assess the influence of the musical excerpt being used, each participant attended two sessions, one with the Bach sequences as stimuli, and the other with the Mozart sequences as stimuli. The order of the two sessions (denoted 'Bach' and 'Mozart') was counterbalanced across participants. At the end of the test, the participants were asked to complete a questionnaire specifying what criteria they had used to assess the recordings.

During the experiment, participants listened to several successive pairs of clarinet performances separated by a 1 s interval. They could listen to each performance as many times as they wished. At each trial, participants were asked to indicate which version they preferred. All the possible combinations (28) of the eight stimuli (the three basic transformations, their four combinations and the original performance) were presented to each participant. The within-pair order and the order of the pairs were randomized. Each session lasted approximately 20 minutes.

Results and Discussion

Presentation of the Perceptual Data

The responses of each participant m are presented in the form of a preference matrix P_m. The elements of the preference matrix, denoted $P_m(i,j)$, designate whether stimulus i was preferred to stimulus j. The global preference matrix P of the sample is defined as the sum of the individual preference matrices P_m.

With each participant, the various performances were given preference scores $S_m(i)$, depending on the number of times they were preferred to the others; with each performance i, this corresponds to summing the preference matrix elements $P_m(i,j)$ across the columns. The scores range from 0 to 7 (times preferred). The mean

[1] http://www.mathworks.com/products/matlab/

preference scores, denoted $S(i)$, were computed on the basis of the preference scores $S_m(i)$, associated with the ratings of the participants.

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

Sample Homogeneity

The degree of agreement among the participants was computed using the Kendall coefficient of agreement u for paired comparisons, as described by Siegel and John Castellan Jr. (1988). This nonparametric measure of association can be written as follows:

$$u = \frac{2\sum_{i=1}^{N_s}\sum_{j=1}^{N_s} C_2^{P(i,j)}}{C_2^{N_m} C_2^{N_s}} - 1 \qquad (5)$$

where C_n^k denotes the binomial coefficient, N_s is the total number of stimuli, and N_m is the total number of participants. Here, $N_s = 8$ and $N_m = 20$. When N_m is even, u can range from $\frac{-1}{N_m - 1}$ to 1, the latter meaning that there is complete agreement among the participants. Siegel and John Castellan Jr. (1988) defined an index W_T based on u, which can range from 0 to 1. The statistic u can be taken to be an estimate for a population parameter ν, which stands for the true degree of agreement in the population. We tested the null hypothesis (H_0: $\nu = 0$) that there was no agreement among the participants against the alternative (H_1: $\nu \neq 0$) that the degree of agreement was greater than what one would have expected had the paired comparisons been done at random. As the total number of participants was large ($N_m > 6$), we used a large-sample approximation of the sampling distribution, asymptotically distributed as a χ^2 distribution ($df = 28$). The values of u and W_T calculated with both the Bach and Mozart excerpts can be found in Table 2. These results show that the agreement among the participants' preferences was significantly higher than chance both with the Bach ($u = 0.58$, $p < .001$) and Mozart ($u = 0.52$, $p < .001$) sequences.

Analyses of Variance (ANOVA)

In order to assess the influences of the musical excerpts (two modalities) and the transformations (eight modalities) on the participants' preferences, the preference scores $S_m(i)$ were subjected to a two-way repeated measures analysis of variance (ANOVA). The results of the ANOVA, which are presented in Table 3, show that the differences of preference scores between the two excerpts were not sufficiently large to exclude the possibility that they might be due to chance. Indeed, the effect of the musical excerpt on the preference scores was not found to be significant (the mean values of the preference scores were identical for both excerpts). The interaction between the musical excerpt and the transformations was also not significant, $F(7, 133) = 1.05$, $p = .40$. Conversely, the effect of the transformations on the preference scores was highly significant, $F(7, 133) = 118.99$, $p < .001$. In order to compare the between-transformation effects, multiple comparison tests (Tukey Honestly Significant Difference tests) were conducted for each excerpt. This procedure determined the significant differences existing between the mean preference scores in each 2 by 2 combination between the various transformations. The results of the multiple comparison procedure are presented in Tables 4 and 5. The preference scores associated with the various performances are described by the box-and-whisker diagrams shown in Figure 4.

For both the Bach and Mozart excerpts, the version M_0 was the most frequently preferred rendering (see Figure 4). This is not surprising, since the expressive deviations associated with timbre, timing, and dynamics had not been removed from this version. It is also not surprising that the performance M_{TRD} to which the three basic transformations were applied was the least preferred on average.

The removal of the IOI deviations (T_R) was the transformation that resulted on average in the least loss of musical preference in the case of both excerpts (see Figure 4). The fact that the IOI deviations' removal had minor effects on the musical preferences does not mean that the IOI deviations are not an important feature of preference, but means that their effect was weak compared to that of the acoustical energy and the spectral centroid variations for these excerpts.

TABLE 2. Coefficients of Agreement Among the Participants at the 'Bach' and 'Mozart' Sessions.

Session	*df*		*u*	W_t	X^2
Bach	28	0.58	($p < .001$)	.60	338.20
Mozart	28	0.52	($p < .001$)	.54	303.80

TABLE 3. Results of the Two-Way Repeated Measures Analysis of Variance of the Preference Scores.

Source	*df*	*F*	*p*
Excerpt	1	0	1.00
Transformation	7	118.99***	< .001
Excerpt × Transformation	7	1.05	.40
Error	133	(1.11)	

Note. The result enclosed in parentheses represent the mean square error.
*$p < .05$, **$p < .01$, ***$p < .001$

TABLE 4. Results of the Multiple Comparison Procedure Within the 'Bach' Session.

	M_0	M_T	M_R	M_D	M_{TR}	M_{TD}	M_{RD}	M_{TRD}
M_0	—	13.21***	3.60 ($p = .18$)	6.61***	15.82***	22.62***	8.81***	22.22***
M_T		—	9.61***	6.61***	2.60 ($p = .59$)	9.41***	4.41*	9.01***
M_R			—	3.00 ($p = .40$)	12.21***	19.02***	5.21**	18.62***
M_D				—	9.21***	16.02***	2.20 ($p = .78$)	15.62***
M_{TR}					—	6.81***	7.01***	6.41***
M_{TD}						—	13.82***	0.40 ($p = 1$)
M_{RD}							—	13.41***

Note. The table presents the results of the pairwise multiple comparison procedure associated to the eight transformations (Tukey HSD tests). The values of the studentized range statistic q are reported; *$p < .05$, **$p < .01$, ***$p < .001$.

However, it is worth noting that the differences of preference between the sequence M_R, without the IOI deviations, and the expressive sequence M_0 were not significant for both excerpts (see Tables 4 and 5), although significant effects of the performer's expressive intentions on the IOI deviations' descriptor were found for both excerpts in our companion study (Barthet et al., 2010). Hence, even if the IOI deviations showed significant differences from the acoustical point of view, the differences might have been too subtle to significantly alter the preferences of the listeners (cf. sound examples 7 and 15). This finding proves the interest of an analysis-by-synthesis approach to investigate the perceptual effects of specific acoustical features. In the case of the Bach excerpt, the fact that the IOI deviations had minor effects on preference is probably due to the style of the musical piece, which is an instrumental dance. Listeners might therefore expect the excerpt to be played with smaller IOI deviations from nominal durations (i.e., in a "mechanical" way). In the case of the Mozart excerpt, the fact that the removal of the IOI deviations had little effect on the listeners' preferences is more surprising, as timing deviations are more likely to occur in slow tempo movements (*Larghetto* in this case). However, the variance of the preference scores attributed to M_R was greater with the Mozart excerpt than with the Bach excerpt.

After the removal of the IOI deviations, the compression of the dynamics was the transformation that had the least effect on the musical preference. However, the differences between M_D and the reference M_0 were significant with both excerpts (see Tables 4 and 5).

At both sessions, the performances that were processed by freezing the spectral centroid (M_T, M_{TD}, M_{TR}, M_{TRD}) were consistently the least preferred (see Figure 4). As shown in Tables 4 and 5, these versions showed the most significant differences with the reference version M_0. The spectral centroid freezing procedure therefore resulted in a greater loss of musical preference than the removal of the IOI deviations, or the compression of the dynamics. It is worth noting that removal of the spectral centroid variations had more degrading effects than the removal of the IOI deviation and the dynamic transformations combined (cf. score of M_T versus the one of M_{RD}).

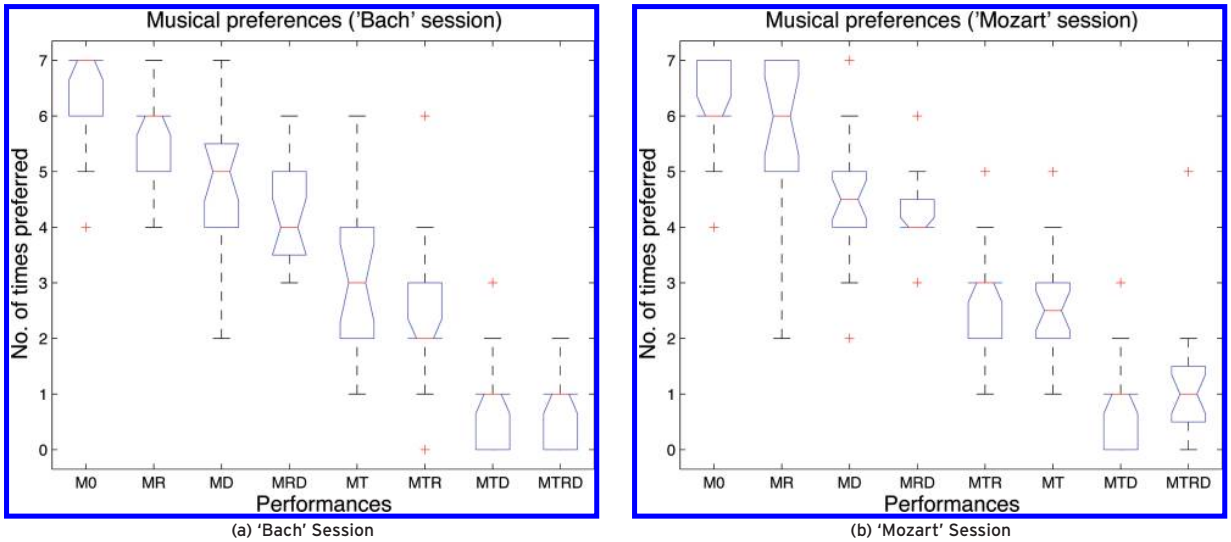
Hierarchical Cluster Analysis (HCA)

In order to determine which performances showed systematic similarities or differences in terms of musical

TABLE 5. Results of the Multiple Comparison Procedure Within the 'Mozart' Session.

	M_0	M_T	M_R	M_D	M_{TR}	M_{TD}	M_{RD}	M_{TRD}
M_0	—	14.62***	1.40 ($p = .98$)	6.81***	13.61***	21.02***	7.21***	20.02***
M_T		—	13.21***	7.81***	1.00 ($p = 1$)	6.41***	5.37**	5.41**
M_R			—	5.41**	12.21***	19.62***	6.01***	18.62***
M_D				—	6.81***	14.22***	0.60 ($p = 1$)	13.21***
M_{TR}					—	7.41***	6.21***	6.41***
M_{TD}						—	13.61***	1.00 ($p = 1$)
M_{RD}							—	12.61***

Note. Legend: see Table 4.



(a) 'Bach' Session (b) 'Mozart' Session

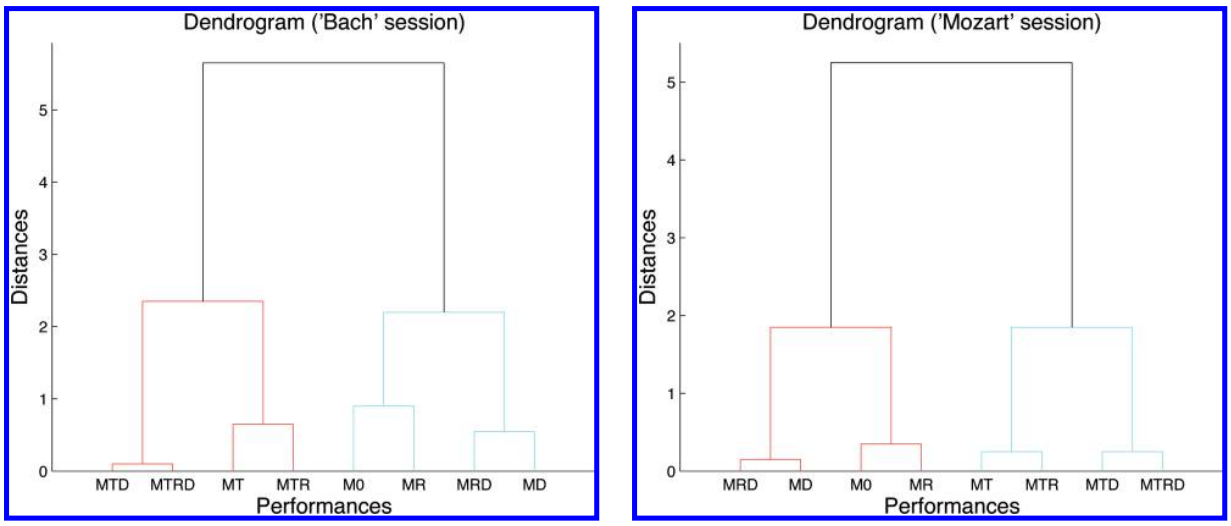
FIGURE 4. Box and whisker plots of the listeners' preference scores at the 'Bach' session (a) and the 'Mozart' session (b). The various sequences are arranged from left to right in decreasing order, based on the median values of the scores. The lines in the box are at the lower quartile, median, and upper quartile values. The whiskers show the range of the rest of the data. Their lengths was 1.5 times the interquartile range. Outliers are represented by crosses. The notches on the box give a robust estimate of the uncertainty of the medians. Note that the number of times each sequence was preferred ranged necessarily between 0 and 7.

preference, the mean preference scores associated to the performances were analyzed using a hierarchical cluster analysis (HCA). Between-performance distances were obtained by computing the Euclidean distances between the mean preference scores $S(i)$. Two different hierarchical clustering methods of the performances were tested: (1) the complete linkage, which is based on the furthest distance between the elements of each cluster, and (2) the Ward linkage, which is based on the increase in variance for the clusters being merged (see e.g., Dillon & Goldstein, 1984). Both methods returned similar hierachical cluster trees (dendrograms). As shown in Figure 5, which presents the dendrograms obtained with the complete linkage method, the two main clusters associated to the 'Bach' and 'Mozart' sessions were identical. One of them contains all the performances that underwent the spectral centroid freezing transformation (M_T, M_{TD}, M_{TR}, M_{TRD}), and the other contains the remaining performances (M_0, M_R, M_D, M_{RD}). These results show that the spectral centroid freezing transformation induced a drastic change of musical preference relative to the intertone onset interval deviation cancellation and/or the compression of the dynamics. They also show that the preference scores of the performances M_R, M_D, and M_{RD} were systematically closer to the score of the reference M_0, in comparison to the scores of the performances that underwent the timbre transformation.

General Discussion

The results showed that the preferences of the participants depended on which acoustical parameters had been modified (spectral centroid and/or intertone onset interval and/or RMS envelope). The preference scores of sequences submitted to multiple transformations were lower or, at best, equal to the score of the basic constituent transformation that was the least preferred. For instance, the score obtained by the sequence M_{RD}, for which both the IOI deviations and the dynamics were modified, was lower to that obtained by the sequence M_D, which was less preferred than the sequence M_R.

No significant effect of the musical excerpt on the preferences was observed. Among the seven transformations, the greatest effects observed were those caused by the freezing of the spectral centroid. This transformation results in a greater loss of preference than the one caused by removal of the IOI deviations or dynamic compression or these two transformations combined. McAdams et al. (1999) investigated subjects' ability to discriminate between various isolated instrumental tones in which spectrotemporal simplifications had been made. Among these simplifications, the freezing of the spectral centroid (induced by the spectral flux freezing process) was found to be most easily discriminated by the listeners. This effect was less marked,



(a) 'Bach' Session (b) 'Mozart' Session

FIGURE 5. Dendrogram representations of the Hierarchical Cluster Analysis of the between-performance distances (complete linkage method), for the excerpts from Bach (a) and Mozart (b).

however, with clarinet and oboe tones since their spectra generally undergo much less spectral flux than most other instruments, namely the brass instruments (Grey, 1977). The strong influence of the removal of spectral centroid variations may be due to the fact that this descriptor is correlated to one of the main timbre dimensions (Grey, 1977; Krumhansl, 1989; McAdams et al., 1995). Although analysis/synthesis techniques allow the control of the spectral centroid independently from other sound dimensions (such as the acoustical energy, in particular), this process is not necessarily acceptable regarding the timbral identity of instrumental tones. However, the participants never preferred the versions in which both the spectral centroid and the acoustical energy variations had been transformed. The spectral centroid freezing probably alters the original timbre of the instrument so that it becomes unnatural, or at least different. In this case, the transformation would do more than simply remove the expressive deviations of timbre, since it would change the nature of the instrument. It should be noted that the participants did not report that they had relied on timbre identity changes. Most of them showed a preference for performances with "lively" rather than "static" tones, which refers directly to the presence or absence of variations in the intensity and/or timbre during the tones. The clarinetist who had played the performances perceived that the transformed sequences were different from his original performances but still thought they were played with a clarinet. Among the possible causal explanations for the differences, he suggested, for instance, that the instrument might have been poorly controlled by the player (e.g., students who move their mouthpiece), or that it might have been in poor condition (e.g., an old reed), resulting in a general lack of homogeneity in the interpretation. Abeles (1973) developed a clarinet performance adjudication scale and observed that the timbre of the tones was one of the most important factors used by music teachers to rate clarinet performances. It is therefore not so surprising that the present participants based their assessments mainly on the timbre of the clarinet tones.

In these experiments, the intricate process of interpretation was reduced to a simple linear, additive model (SC variations +/– IOI deviations +/– energy variations). This is a first step towards reaching a better understanding of the influence of the temporal and spectral parameters related to timbre, timing, and intensity. However, these parameters may interact at the level of both performers (in the sound production process) and listeners (in the decoding of the musical signals). For instance, when listening to his own performance of the excerpt from Mozart's *Larghetto*, the clarinetist was convinced that he had deliberately lengthened one of the tones in the sequence, causing the following one to be late. However, the analysis showed that the onset of the second tone was perfectly on the beat. The decrescendo in the first of these two tones, which was associated with a concomitant decrease in brightness, may have induced this feeling of ritardando.

Summary and Conclusions

This study focused on the perceptual effects of variations in acoustical correlates of timbre, timing, and dynamics on musical preference. To address this issue, an experimental method based on the analysis-by-synthesis approach was developed, which consisted of transforming the expressive content of recorded clarinet performances and assessing the effects of these changes on listeners' musical preferences. Seven transformations were designed to remove or compress the variations in the spectral centroid (a timbre parameter), intertone onset interval (a timing parameter), and acoustical energy, either separately or in various combinations. The statistical analyses carried out on the listeners' aesthetic judgments showed that the spectral centroid freezing transformation most significantly decreased the musical preference of the performances. This finding seems to be due to the fact that this transformation altered the original timbre of the clarinet tones (the identity of the instrument), as well as drastically affecting the time-evolving spectral shapes, causing the tones to be static and unlively (the quality of the sound).

These results confirmed that the performer's choices of timbre, timing, and dynamics variables (see the companion article by Barthet et al., 2010) affect the listener's perception of the musical preference. The variations in the spectral centroid during tone production seem to be an important feature of preference in the musical message transmitted from performers to listeners. Indeed, in another study, we established that controlling the time-evolving spectral centroid of tones improved the preference of sampler-based generated sequences (Barthet, Kronland-Martinet, & Ystad, 2008).

These findings suggest that it might be worth developing a general set of rules related to timbre, as previous authors have done in the case of temporal and intensity deviations (see e.g., De Poli, 2006; Mathews, Friberg, Bennett, Sapp, & Sundberg, 2003; Widmer & Goebl, 2004). By taking the variations in the acoustical parameters associated with timbre into account in computational models of music performance, it might then be possible to improve the automatic rendering of musical pieces by computers.

Author Note

We would like to thank the clarinetist Claude Crousier for participating in this project and 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 would also like to thank the reviewers for their helpful comments and advices. This project was partly supported by the French National Research Agency (ANRJC05-41996, "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

Abeles, H. F. (1973). Development and validation of a clarinet performance adjudication scale. *Journal of Research in Music Education, 21*, 246–255.

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, Marseille, France.

Barthet, M., Depalle, P., Kronland-Martinet, R., & Ystad, S. (2010). Acoustical correlates of timbre and expressiveness in clarinet performance. *Music Perception, 28*, 135–153.

Barthet, M., Kronland-Martinet, R., & Ystad, S. (2008). Improving musical expressiveness by time-varying brightness shaping. In R. Kronland-Martinet, S. Ystad, & K. Jensen (Eds.), *Sense of sounds* (pp. 313–336). Berlin/Heidelberg: Springer-Verlag.

Caclin, A., McAdams, S., Smith, B. K., & Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. *Journal of the Acoustical Society of America, 118*, 471–482.

Canazza, S., De Poli, G., Drioli, C., Rodá, A., & Vidolin, A. (2004). Modeling and control of expressiveness in music performance. *Proceedings of the IEEE, 92*, 686–701.

Canazza, S., Rodá, A., & Orio, N. (1999). A parametric model of expressiveness in musical performance based on perceptual

and acoustical analyses. In *Proceedings of the International Computer Music Conference* (pp. 379–382). Beijing, China: International Computer Music Association (ICMA).

De Poli, G. (2006). *Algorithms for sound and music computing.* Creative Commons Attribution-NonCommercial-ShareAlike licence. Retrieved from http://www.dei.unipd.it/~musica/IM06/Dispense06/7_espressiveness.pdf

Dillon, W. R., & Goldstein, M. (1984). *Multivariate analysis.* New York: John Wiley & Sons.

Friberg, A. (1995). *A quantative rule system for musical performance.* Unpublished doctoral dissertation, Royal Institute of Technology, Stockholm, Sweden.

Gabrielsson, A. (1999). 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. (1985). Perceived sound quality of high-fidelity loudspeakers. *Journal of the Audio Engineering Society, 33*, 33–53.

Goebl, W., Pampalk, E., & Widmer, G. (2004). Exploring expressive performance trajectories: Six famous pianists play six Chopin pieces. In S. D. Lipscomb, R. Ashley, R. O. Gjerdingen, & P. Webster (Eds.), *Proceedings of the 8th International Conference on Music Perception and Cognition* (pp. 505–509). Evanston, IL: Causal Productions.

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. Deliège & 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.

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.

Kergomard, J. (1991). Le timbre des instruments à anche [The timbre of reed instruments]. In C. Bourgois (Ed.), *Le timbre, métaphore pour la composition [The timbre, metaphor for the composition]* (pp. 224–235). Paris: I.R.C.A.M.

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.

Mathews, M., Friberg, A., Bennett, G., Sapp, C., & Sundberg, J. (2003). A marriage of the director musices program and the conductor program. In R. Bresin (Ed.), *Proceedings of the Stockholm Music Acoustics Conference (SMAC 03)* (Vol. 1). Stockholm, Sweden.

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.

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. (1994). Quelques aspects du timbre dans la musique contemporaine [A few aspects of timbre in contemporary music]. In A. Zenatti (Ed.), *Psychologie de la musique [Psychology of music]* (1st ed., pp. 87–114). Paris: Presses Universitaires de France.

Risset, J.-C., & Wessel, D. L. (1999). Exploration of timbre by analysis and synthesis. In D. Deutsch (Ed.), *Psychology of music* (2nd ed., pp. 113–169). San Diego, CA: Academic Press.

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

Seashore, C. E. (1938/1967). *Psychology of music.* New York: McGraw-Hill. (Reprinted 1967, New York: Dover Publications).

Siegel, S., & John Castellan Jr., N. (1988). *Non parametric statistics for the behavioral sciences* (2nd ed., pp. 272). New York: McGraw-Hill International Editions.

Tobudic, A., & Widmer, G. (2003). Playing Mozart phrase by phrase. In K. D. Ashley & D. G. Bridge (Eds.), *Lecture Notes in Computer Science (LNCS), 5th International Conference on Case-Based Reasoning (ICCBR)* (Vol. 2689, pp. 552–566). Berlin Heidelberg: Springer-Verlag.

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

TODD, N. P. M. (1995). The kinematics of musical expression. *Journal of the Acoustical Society of America*, *97*, 1940–1949.

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

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

WIDMER, G., & GOEBL, W. (2004). Computational models of expressive music performance. *Journal of New Music Research*, *33*, 203–216.

WINDSOR, W. L., DESAIN, P., PENEL, A., & BORKENT, M. (2006). A structurally guided method for the decomposition of expression in music performance. *Journal of the Acoustical Society of America*, *119*, 1182–1193.

ZÖLZER, U. (1997). Dynamic range control. In *Digital Audio Signal Processing* (pp. 207–219). Chichester: John Wiley & Sons.