ACTA ACUSTICA UNITED WITH ACUSTICA
Vol. 96 (2010) 678 – 689

DOI 10.3813/AAA.918322

From Clarinet Control to Timbre Perception

Mathieu Barthet, Philippe Guillemain, Richard Kronland-Martinet, Sølvi Ystad
CNRS Laboratoire de Mécanique et d'Acoustique, 31 chemin Joseph-Aiguier, 13402 Marseille Cedex 20, France.
barthet@lma.cnrs-mrs.fr

Summary
This study investigates the relationships between the control gestures of the clarinet, the generated timbres and their perceptual representation. The understanding of such relationships can provide great interest in several research contexts: synthesis and control (e.g., to improve the quality of current synthesis models), music analysis and perception (e.g., to study music performance), and music information retrieval (e.g., to find relevant acoustical descriptors for automatic instrument and/or performer identification). A physics-based model was used to generate synthetic clarinet tones by varying the main control parameters of the model (related to the blowing pressure and lip pressure on the reed). 16 participants had to rate the dissimilarities between pairs of different tones and describe the factors on which they based their judgments in a questionnaire. The collected data were subjected to various statistical analyses (multidimensional scaling and hierarchical clustering) in order to obtain a low-dimensional spatial configuration (timbre space) which best represents the dissimilarity ratings. The structure of the clarinet timbre space was interpreted both in terms of control parameters and acoustical descriptors. The analyses revealed a 3-dimensional timbre space, whose dimensions were well correlated to the Attack Time, the Spectral Centroid, and the Odd/Even Ratio. Comparisons of natural and synthetic clarinet tones showed that the Odd/Even Ratio appears to be a good predictor of the beating reed situation, specific to single-reed instruments.

PACS no. 43.66.Jh, 43.75.Pq, 43.75.Yy, 43.75.Zz

1. Introduction

Since its conception in the 17^{th} century, the clarinet, a single-reed instrument from the woodwind family, has aroused the interest of composers as a result of its unique timbre [1]. From the acoustical point of view, the clarinet can be described as the association of an exciter (the reed) and a resonator (the air column contained in the bore of the instrument). Under certain control conditions imposed by the performer, the coupling between the exciter and the resonator, can lead to the establishment of self-sustained oscillations which are responsible for the sounds generated by the instrument. By varying their control gestures, performers can produce different timbres (see e.g., [2]). This study aims to increase understanding of the relationships between the control of the instrument, the generated sounds, and how they are perceived. The understanding of such relationships can provide great interest in several research contexts: synthesis and control (e.g., to improve the quality of current synthesis models), music analysis and perception (e.g., to study music performance), and music information retrieval (e.g., to find relevant acoustical descriptors for automatic instrument and/or performer identification). For synthesis purposes, perceptual clues can be used to calibrate the parameters of a synthesis model according to the behavior of an acoustic instrument, or to propose control devices based on a high-level perceptual description (e.g., "play a bright sound"). For music analysis purposes, the study of the links between the control of musical instruments and the perception of generated tones allow to determine the mechanical/acoustical parameters used by performers to alter the timbre of the tones, and find the relevant acoustical descriptors that account for such timbre variations. Such investigations are of importance when studying the role of timbre in musical interpretation (see e.g., [3, 4]) and can give new pedagogical insights. In the music information retrieval context, the results of such studies may be useful to achieve automatic instrument or performer identification based on timbre descriptors.

By varying the two main control parameters (related to the blowing pressure and reed aperture) of a physics-based synthesis model of clarinet [5], isolated clarinet tones were generated. Dissimilarity judgments were then collected and subjected to various statistical analyses to uncover the psychomechanical and psychoacoustical factors best predicting the perceptual representation of the various clarinet tones.

Timbre is the attribute of the auditory sensation that allows tones of equal pitch, loudness and duration to be distinguished [6]. From the cognitive point of view, timbre refers both to the identity of the sound sources (timbre is the attribute that allows two different musical instruments playing the same note, with identical sound lev-

Received 30 July 2009,
accepted 4 March 2010.

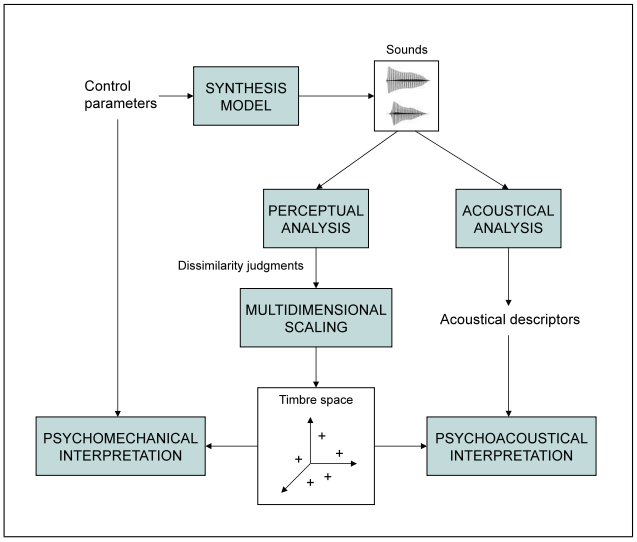

Figure 1. Multidimensional analysis of timbre.

els and durations, to be distinguished) and to the sound quality (e.g., two different guitar tones played with different plucking positions have different timbral qualities) [7]. The seminal works of Grey [8], Wessel [9], Krumhansl [10], Kendall and Carterette [11], and McAdams *et al.* [12] developed a methodology to represent and interpret the perceptual relationships between sounds differing in timbre. This methodology, which we rely upon in the present study, involves three main steps (see Figure 1): the collection of dissimilarity judgments among sounds varying in timbre, the construction of a geometrical representation of the dissimilarity judgments with a multidimensional scaling (MDS) procedure, and the interpretation of the representation with acoustical factors (timbre descriptors) and/or mechanical factors (control parameters). MDS techniques yield timbre spaces where the distances between the sounds represent the perceived dissimilarities statistically well [13]. Such techniques benefit from the fact that no assumptions are needed regarding the underlying acoustical descriptors that may predict the structure of the perceptual representation. MDS studies based on natural sounds from musical instruments, or synthetic tones generated to imitate the instruments of the orchestra, generally reported two- to four-dimensional timbre spaces. Caclin *et al.* [14] tested the perceptual relevance of some of the acoustical correlates of timbre space dimensions proposed in the psychoacoustic literature. This study confirmed that the Attack Time (linked to the rate of energy increase during the transient regime), the Spectral Centroid (frequency of the centroid of the spectrum), and the spectrum fine structure were major determinants of timbre. The results of MDS studies depend on the set of stimuli used in the experiment; the acoustical correlates found to characterize the broad differences of timbre between different musical instruments are not necessarily the same as the ones used to characterize the fine differences in timbre produced by the same instrument. One purpose of the present study is to find timbre descriptors well adapted to the range of timbres producible on the clarinet and which are potentially used by performers in a musical context (see [3]).

In section 2, the physics-based synthesis model used to generate clarinet tones will be presented. Some comparisons between natural and synthetic tones will be given in order to validate the use of the sounds produced by such a model in the study of the clarinet timbre. In section 3, the methods used in the perceptual experiment will be described. The statistical analyses conducted on the dissimilarity ratings will be presented and discussed in section 4. Finally, section 5 will present the conclusions and perspectives of the study.

2. Comparison of synthetic and natural clarinet tones

2.1. Description of the physics-based synthesis model

The physical model that was used to generate clarinet tones which formed the stimuli of the perceptual experiment is made of three coupled parts. The first part is linear and represents the bore of the instrument. The second part expresses the reed channel opening, linked to the reed displacement, by modeling the reed as a pressure driven mass-spring oscillator. The third part couples the previous ones in a nonlinear way. In what follows, we use dimensionless variables for the pressure, flow, and reed displacement, according to [15]. The internal variables of the model $p_e(t)$, $u_e(t)$ and $x(t)$ denote the dimensionless acoustic pressure, flow, and reed displacement, respectively. Capital letters denote their Fourier transforms. The dimensionless control variables $\gamma(t)$ and $\zeta(t)$ are related to the blowing pressure and the lip pressure on the reed, respectively, and will be detailed later.

2.1.1. Bore model

The bore is considered as a perfect cylinder of length L. Under classical hypothesis, its input impedance Z_e which links the acoustic pressure P_e and flow U_e in the mouthpiece, in the Fourier domain, is classically written as:

$$Z_e(\omega) = \frac{P_e(\omega)}{U_e(\omega)} = \mathrm{i}\tan\left(k(\omega)L\right). \quad (1)$$

The wavenumber $k(\omega)$ corresponds to the classical approximation (the bore radius is large with respect to the thicknesses of the boundary layers). It is worth noting that for any flow signal, the acoustic pressure mostly contains odd harmonics since the input impedance corresponds to that of a quarter-wave resonator. At high frequencies, the increase of losses taken into account in the wavenumber induces non-zero values of the impedance for even harmonic frequencies. Hence, if the flow contains high frequency even harmonics, these will also appear in the pressure.

2.1.2. Reed model

The classical single mode reed model used here describes the dimensionless displacement $x(t)$ of the reed with respect to its equilibrium point when it is submitted to the dimensionless acoustic pressure $p_e(t)$:

$$\frac{1}{\omega_r^2}\frac{\mathrm{d}^2x(t)}{\mathrm{d}t^2} + \frac{q_r}{\omega_r}\frac{\mathrm{d}x(t)}{\mathrm{d}t} + x(t) = p_e(t), \qquad (2)$$

where $\omega_r = 2\pi f_r$ and $1/q_r$ are the angular frequency and the quality factor of the reed resonance, respectively. The reed displacement behaves like the pressure below the reed resonance frequency, as a pressure amplifier around the reed resonance frequency, and as a low-pass filter at higher frequencies.

2.1.3. Nonlinear characteristics

The classical nonlinear characteristics used here are based on the steady Bernoulli equation and link the acoustic flow (the product of the opening of the reed channel and the acoustic velocity) to the pressure difference between the bore and the mouth of the player. The effective reed channel opening $S(t)$ is expressed in terms of the reed displacement by

$$S(t) = \zeta(t)\Theta\big(1-\gamma(t)+x(t)\big)\big(1-\gamma(t)+x(t)\big), \qquad (3)$$

where Θ denotes the Heaviside function, the role of which is to keep the opening of the reed channel positive by cancelling it when $1-\gamma(t)+x(t) < 0$. The parameter $\zeta(t)$ characterizes the whole embouchure and takes into account both the lip position and the section ratio between the mouthpiece opening and the resonator. It is proportional to the square root of the reed position at equilibrium and inversely proportional to the reed resonance frequency. The parameter $\gamma(t)$ is the ratio between the pressure $p_m(t)$ inside the player's mouth (assumed to be slowly varying on a sound period) and the static beating reed pressure P_M, i.e. the pressure needed to close the reed channel when there are no oscillations. In a lossless bore and a massless reed model, $\gamma(t)$ evolves from $\frac{1}{3}$ which is the oscillation threshold, to 1, which corresponds to the extinction threshold. The value $\frac{1}{2}$ corresponds to the so-called beating reed threshold, from which the reed touches (beats) the mouthpiece table during the course of each period of the oscillations.

Since the reed displacement corresponds to a linear filtering of the acoustic pressure, the reed opening mostly contains odd harmonics. Nevertheless, the function Θ introduces a singularity in $S(t)$ for playing conditions (given by $\zeta(t)$ and $\gamma(t)$) yielding a complete closing of the reed channel (dynamic beating reed case). This leads to a rise of even harmonics in $S(t)$ (saturating nonlinearity) and the generation of high frequencies.

The acoustic flow is finally given by

$$u_e(t) = S(t)\mathrm{sign}\big(\gamma(t)-p_e(t)\big)\sqrt{|\gamma(t)-p_e(t)|}. \qquad (4)$$

This nonlinear relation between pressure and opening of the reed channel explains why the flow spectrum contains all the harmonics.

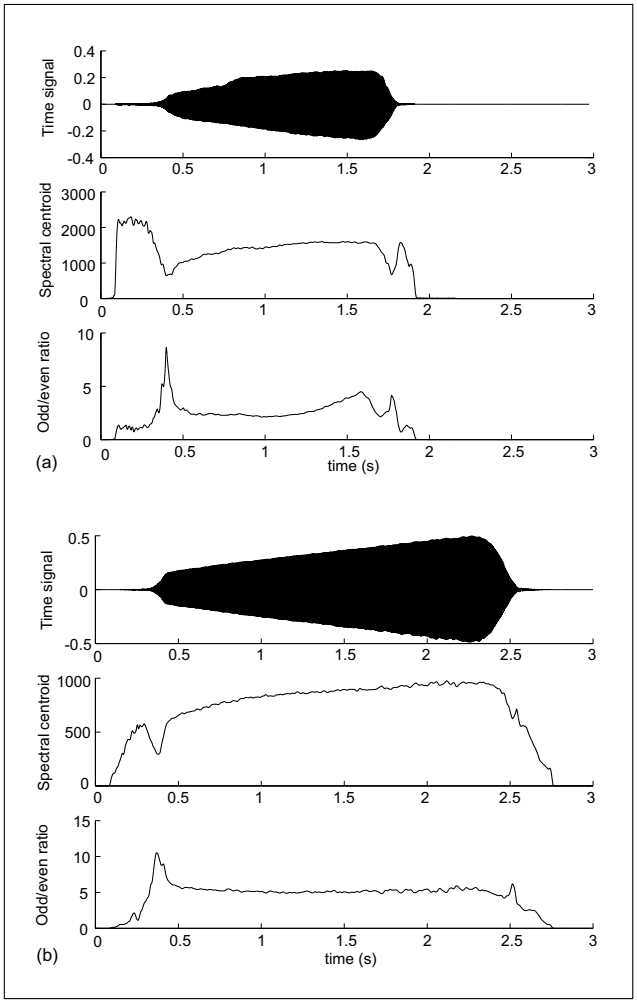

Figure 2. External pressure and timbre descriptors (Spectral Centroid and Odd/Even Ratio) in the cases of a natural (a) and a synthetic (b) clarinet tone.

2.1.4. Coupling of the reed and the resonator

Combining the impedance relation, the reed displacement and the nonlinear characteristics, the acoustic pressure, acoustic flow and reed displacement in the mouthpiece are solutions of the coupled equations (1), (2) and (3). The digital transcription of these equations and the computation scheme that explicitly solves this coupled system are achieved according to the method described in [5].

2.1.5. External pressure

From the pressure and the flow inside the resonator at the position of the mouthpiece, the external pressure is calculated by the relationship: $p_{ext}(t) = \frac{d}{dt}(p_e(t) + u_e(t))$, which corresponds to the simplest approximation of a monopolar radiation. This expression shows that in the clarinet spectrum, the odd harmonics are generated from both the flow and the pressure, while the even harmonics mostly come from the flow. Therefore, the ratio between odd and even harmonics can be considered as a signature of the "strength" of the nonlinearity in a non beating-reed situation.

2.2. Validation of the model

To investigate the relationships between the control of the instrument and the perception of the resulting sounds, we have chosen to use a synthesis model, since the control parameters $\gamma(t)$ and $\zeta(t)$ of such a model can be altered in a systematic and constrained manner. It is therefore necessary, as a first step, to check that the perceptual features considered here behave similarly for natural and synthetic sounds.

For that purpose, both an acoustic clarinet and a synthetic one were blown with the same time-varying pressure (parameter related to $\gamma(t)$). For the acoustic clarinet, an artificial mouth piloted in pressure by a PID (Proportional Integral Derivative) controller using a repetitive command was used. The role of the PID is to make an accurate pressure regulation to minimize the difference between the actual pressure in the mouth and the command pressure. Although the lip pressure on the reed can be measured on a human player, it varies significantly during transients. Furthermore, since the mechanical properties of the reed and the lip are unknown, the lip pressure cannot be linked to the reed channel opening at rest and therefore to the control parameter of the model $\zeta(t)$. Hence, the artificial mouth appeared to us as the only possibility to maintain this parameter as constant as possible in order to facilitate the comparison between the acoustic instrument and the model. The target pressure comprised a transient onset, followed by a linear pressure increase and then a decay transient. This specific pattern allows comparison of the behaviors of the timbre descriptors during transients and of the consistency of their variations with respect to a given steady-state pressure. The linear pressure increase has been chosen to test several constant blowing pressures within a single experiment. The actual mouth pressure (the result of the command) in the artificial mouth was recorded and used to feed the real-time synthesis model. In both experiments, the lip pressure on the reed was kept constant. The associated sound examples A1 and A2 are available at
www.lma.cnrs-mrs.fr/~kronland/ClarinetTimbre/.

Figure 2a shows respectively, from top to bottom, the variations in the external pressure, the Spectral Centroid (SC), and the Odd/Even Ratio (OER), over the duration of the note produced by the acoustic clarinet (the definitions of these timbre descriptors are given in section 3.5). At the oscillation threshold, around $t = 0.4$ s, the Spectral Centroid exhibits a minimum while the Odd/Even Ratio exhibits a maximum. This corresponds to the birth of the sound during which mostly the first harmonics are present. Then, the Spectral Centroid keeps on increasing steadily until $t = 1.8$ s, which corresponds to the release of the blowing pressure. A short increase followed by a short decrease of the centroid can be observed. After the attack, the Odd/Even Ratio decreases very quickly and then remains nearly constant with a slight decrease until $t = 1$ s. It then increases until $t = 1.6$ s. At the release of the blowing pressure, the OER exhibits an overshoot.

Non real-time simulations, for which all the physical variables of the problem are known, show that the amplitude of the flow increases steadily, and the balance between its even and odd harmonics remains constant, until the reed starts beating. This explains why the OER remains nearly constant with a slight decrease until $t = 1$ s. After that time, since the reed is beating, the amplitude of the flow oscillations are limited while those of the pressure are not. Moreover, the singularity in the flow due to its cessation over half a period introduces high frequency odd harmonics. This explains why the OER increases after $t = 1$ s.

The same features can be observed in Figure 2b for both the SC and the OER in the case of a synthetic tone. It can be observed that the duration of the sound is different to that produced by the acoustic clarinet. This difference can be attributed to the fact that the oscillation threshold is different. Indeed, it is important to point out that, although the lip pressure on the reed (related to $\zeta(t)$) is constant in both cases, they are different as the reed opening at rest cannot be measured on the artificial mouth. This probably also explains why the range of variations of the Spectral Centroid is similar in both cases while the average values are different, and why the average value of the Odd/Even Ratio is higher and its range of variations is smaller for the synthetic sound.

It is important to mention that the purpose of this example is not to make a model inversion using timbre descriptors, i.e to find the commands of the synthesis model from the analysis of an acoustic sound, but to demonstrate that the perceptual features under consideration evolve similarly with respect to a given transient or constant pressure command and that their variations can be related to the behavior of the physical variables of the problem.

3. Methods

3.1. Stimuli

A musician was asked to play a short sustained tone (E3, $f_0 \approx 164.81$ Hz) with the real-time synthesis model driven by a MIDI (Musical Instrument Digital Interface) controller adapted to wind instruments (Yamaha-WX5). The latter namely allows the control of the mouth pressure and the lip pressure on the reed. The blowing pressure was measured in the player's mouth cavity using a pressure probe (a very thin plastic tube, that does not disturb the player during tone production) linked to a pressure sensor (reference: Honeywell ASCX05DN, 0-5 psi differential piezzo-resistive pressure sensor). Figure 3 shows the variations in the measured blowing pressure over the duration of the note. The measured blowing pressure was normalized by dividing by the static beating reed pressure. This normalized blowing pressure, denoted $\gamma^{ref}(t)$, served as a reference when generating the stimuli by providing a maximum value, denoted γ_m.

A set of 150 tones ($f_0 \approx 164.81$ Hz, duration ≈ 2 s) was generated using the synthesis model by varying γ_m

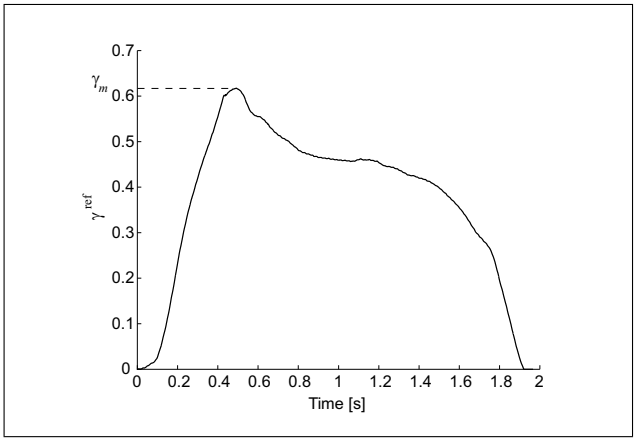

Figure 3. Blowing pressure of reference $\gamma^{ref}(t)$ (dimensionless). The maximal value of the pressure profile is denoted γ_m.

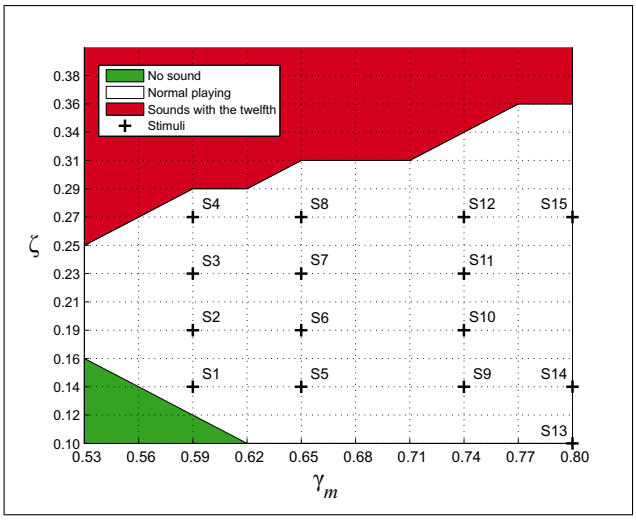

Figure 4. Control parameter space: the couples (γ_m, ζ) corresponding to the stimuli are indicated by crosses.

in the range $0.53 \leq \gamma_m \leq 0.80$ and ζ in the range $0.10 \leq \zeta \leq 0.40$. From this set, 15 tones were selected (see sound examples B1 to B15) which represented the palette of timbres producible with the model in traditional playing conditions (no squeaks, nor tones with the twelfth). Figure 4 represents the values of the control parameters associated with the stimuli. The sounds were equalized for loudness based on the informal loudness scaling of one listener. This was done by adjusting the loudness of the various sounds according to a reference signal (a 1kHz pure tone presented with an intensity of 70 dB SPL as measured at the position of the listener). Three other listeners validated that the amplitude-adjusted sounds were equal in loudness. This procedure was chosen since an automatic loudness equalization process based on the model defined by Zwicker and Fastl [16] for stationary sounds did not produce satisfying results. This may be due to the fact that clarinet tones are non stationary as they both contain a transient and a steady-state part. Furthermore, the synthesized tones present large differences in spectral richness, a factor which acts on the loudness [16] and which is poorly taken into account by the loudness model.

3.2. Participants

The experiment was conducted by a group of 16 participants (10 males, 6 females; age=23-47 years). 10 of them had received musical training and none reported any hearing loss.

3.3. Apparatus

The experiment took place in an audiometric cabin and the sounds were stored on the hard drive of an Apple iMac G4. The user interface was implemented in the Matlab environment[1]. The sounds were played to the participants using a STAX SRM-310 headphones system.

3.4. Procedure

The participants were first asked to listen to the 15 stimuli presented in random order to get used to the range of variation. After carrying out a training stage with randomly chosen practice trials, the participants had to rate the dissimilarity for all 105 possible pairs of non-identical stimuli. The within-pair order and the order of the pairs were randomized. The dissimilarity judgments were made by adjusting the position of a slider on a scale whose end points were labelled “Very similar" and “Very dissimilar" (in French) on the computer screen. The perceptual dissimilarities were digitized on a 0 to 1 scale with a discretization step of 0.01. The participants could listen to the stimuli as many times as they wished and were requested to keep the same rating strategy during the experiment. At the end of the test, they had to answer a questionnaire which was designed to establish the criteria they used to discriminate the various stimuli.

3.5. Presentation of the acoustical correlates of timbre

A large set of timbre descriptors was investigated in order to find some acoustical predictors of the perceptual structure underlying the discrimination of the various clarinet tones (see Table I). A precise description and formulation of these parameters can be found in [17]. The only definitions given here are those deemed most important in light of the results obtained in this study.

The **Attack Time (AT)** is given by

$$AT = t_{e_{AT}} - t_{s_{AT}}, \tag{5}$$

where $t_{s_{AT}}$ and $t_{e_{AT}}$ are the start and end of the attack times, respectively. They are defined as in [18], as the times at which the Root Mean Square (RMS) envelope attains 10% and 90% of its maximum value, respectively.

The short-term **Spectral Centroid (SC)** is defined as

$$SC(n) = \frac{\sum_{k=1}^{K} f(k) A_n(k)^2}{b_0 + \sum_{k=1}^{K} A_n(k)^2}, \tag{6}$$

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

Table I. List of the timbre descriptors.

Timbre descriptor	Description
Attack Time (AT)	Linked to the rate of energy increase during the transient regime
Log. of the Attack Time (LAT)	Logarithm (decimal base) of AT
Release Time (RT)	Linked to the rate of energy decrease during the release part
Temporal Centroid (TC)	Time of the centroid of the acoustical energy
Spectral Centroid (SC)	Frequency of the centroid of the spectrum
Spectral Spread (SS)	Variance of the spectrum around its mean
Spectral Skewness (SSK)	Measure of the asymmetry of the spectrum around its mean
Spectral Kurtosis (SKU)	Measure of the flatness of the spectrum around its mean
Spectral Roll-off (SRO)	Cutoff frequency of the spectrum so that 95% of the energy is below
Spectral Flux (SF)	Measure of the fluctuation of the spectrum over time
Spectral Flux Attack (SFAT)	Spectral Flux calculated during the attack part
Harmonic Spectral Centroid (HSC)	Frequency of the centroid of the harmonic spectrum
Odd Spectral Centroid (OSC)	Frequency of the centroid of the spectrum of the odd harmonics
Even Spectral Centroid (ESC)	Frequency of the centroid of the spectrum of the even harmonics
Odd/Even Ratio (OER)	Ratio between the odd harmonics and the even harmonics energy
Spectral Irregularity (IRRKRI)	Measure of the irregularity of the spectral envelope (Krimphoff)
Spectral Irregularity (IRRKEN)	Measure of the irregularity of the spectral envelope (Kendall)
Spectral Irregularity (IRRJEN)	Measure of the irregularity of the spectral envelope (Jensen)
Tristimulus, 1^{st} coefficient (TR1)	Ratio between the fundamental component energy and the total energy
Tristimulus, 2^{nd} coefficient (TR2)	Ratio between the energy of harmonics 2, 3, and 4 and the total energy
Tristimulus, 3^{rd} coefficient (TR3)	Ratio between the energy of higher-order harmonics and the total energy

where $A_n(k)$ is the magnitude of the k^{th} coefficient of the Discrete Fourier Transform (DFT) associated with the frame centred at time n, $f(k)$ is the frequency associated with the k^{th} spectral component, and K denotes the last frequency bin to be considered. We used a threshold value b_0 as in the formulation given by Beauchamp [19], in order to force the descriptor to decrease at very low amplitudes when noise predominates.

The time-varying **Odd/Even Ratio (OER)** is obtained from

$$OER(t) = \frac{b_0 + \sum_{h=0}^{\frac{H}{2}-1} A_{2h+1}(t)^2}{b_0 + \sum_{h=1}^{\frac{H}{2}} A_{2h}(t)^2}, \quad (7)$$

where $A_h(t)$ is the instantaneous amplitude of the h^{th} harmonic component, H is the total number of considered harmonics (assumed here to be even so that an equal number of odd and even harmonics are compared). The b_0 threshold is used, as in equation 6, to prevent the descriptor from tending to infinity when noise predominates. OER is dimensionless; $OER < 1$, indicates that even harmonics are dominant, whereas $OER > 1$ indicates that odd harmonics are dominant.

The spectral and spectro-temporal descriptors were computed from the Short Term Discrete Fourier Transform (STDFT) by using a 1024-point Hann window (approximately 20 ms at the sampling frequency of 44.1 kHz) with a 50%-overlap. The DFT was calculated over 8192 points (the discretization step of the frequency scale is approximately 5 Hz). The computation of the spectral and spectro-temporal descriptors was made with two different amplitude scales (linear and power) in order to find the one which best fits the perceptual measures (see [20, 21] for a discussion on amplitude scales). b_0 was set at a value giving a spectral dynamic of 60 dB. The harmonic descriptors were calculated from the components' instantaneous amplitudes and frequencies. The latter were derived from short-band analytic signals associated with each component of the tone (see [22]). To obtain a global measure characterizing the whole signal, the instantaneous descriptors were averaged between the start of the attack part and the end of the release part.

4. Results and discussion

4.1. Multidimensional scaling (MDS)

Before analyzing the dissimilarity ratings using an MDS procedure, tests were first carried out to determine if the answers of the participants depended on whether they were musicians or non-musicians. To address this issue, correlations (Pearson) were computed among participants' dissimilarity ratings. A Hierarchical Cluster Analysis (complete linkage) was then performed on distances derived from the correlation measures (one minus Pearson's correlation coefficient) to detect whether certain participants systematically answered differently from the others. As the HCA did not reveal systematic differences between musicians and non-musicians, the dissimilarity ratings of all the participants were averaged for the MDS analysis.

A nonmetric MDS procedure (MDSCAL) was adopted (see e.g., [23]) as the dissimilarities possess ordinal scale properties. The procedure yields solutions such that the distances in the MDS space are in the same rank order as the original data. The goodness-of-fit criterion that it was necessary to minimize was Kruskal's stress. The number of dimensions of the MDS configuration was found by

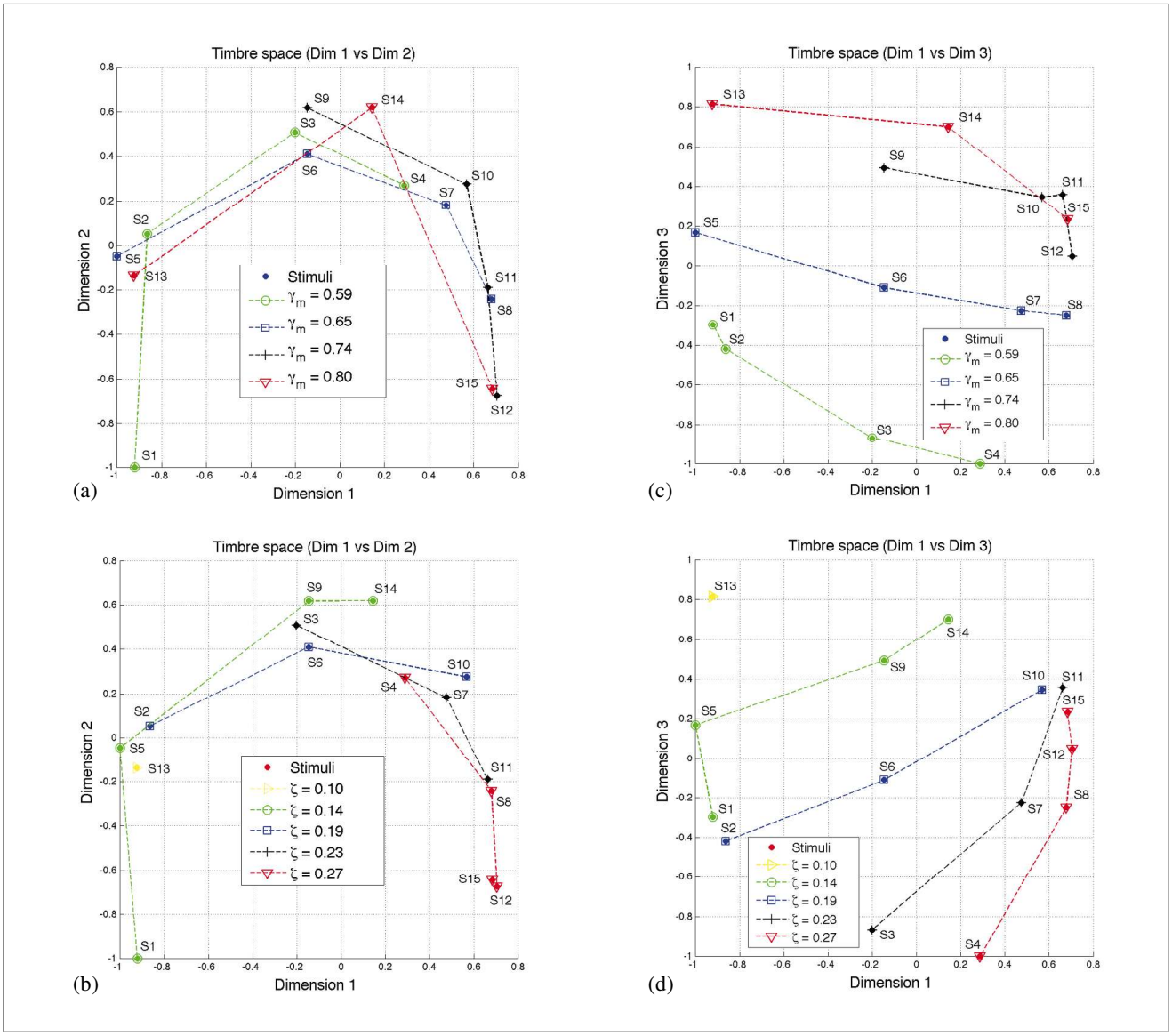

Figure 5. Two-dimensional projections of the MDS space. In Figures (a) and (c), the stimuli with same γ_m (related to the blowing pressure) are connected, in figures (b) and (d), the stimuli with same ζ (related to the lip pressure on the reed) are connected.

using several empirical criteria: the rate of decline of the stress as dimensionality increases (scree-elbow criterion) and the analysis of the Shepard diagram (representation of the distances in the MDS space as a function of the dissimilarities). The method indicated a 3-dimensional (3D) perceptual space (stress ≈ 0.05).

4.2. Interpretation of the clarinet timbre space

4.2.1. Mechanical correlates of the timbre space

Figure 5 shows the two-dimensional projections of the timbre space. To better visualize the influence of the control parameters on the positions of the stimuli in the perceptual space, the stimuli with identical γ_m values (Figures 5a and 5c) and ζ values (Figures 5b and 5d) have been linked together. The linear correlations (Pearson) between the positions of the stimuli in the space and the values of the control parameters are reported in Table II. It can be seen that the first dimension is correlated with ζ [r(13)=0.76, p<0.001] (the higher ζ, the higher the stimuli coordinate along the first dimension). Meanwhile, the third dimension is correlated with γ_m [r(13)=0.88, p<0.001] (the higher γ_m, the higher the stimuli coordinate along the third dimension). None of the control parameters were significantly correlated with the second dimension. Indeed, Figures 5a and 5b indicate a non-linear behavior of the stimuli coordinates along the second dimension as a

Table II. Pearson's correlation coefficient between the stimuli coordinates in the perceptual space and the control parameters γ_m and ζ (degrees of freedom = 13). Values that are not significant are not reported. The probabilities that the measures are independent are indicated as follows: *p<0.05, **p<0.01,***p<0.001.

	Dimension 1	Dimension 2	Dimension 3
γ_m	-	-	0.88***
ζ	0.76***	-	-0.54*

Table III. Pearson's correlation coefficient between the stimuli coordinates in the perceptual space and the timbre descriptors (d.o.f. = 13). When both linear and power amplitude scales were used in the computation of the descriptors, the corresponding correlations are reported. Values that are not significant are not reported. The probabilities that the measures are independent are indicated as follows: $^{*}p<0.05$, $^{**}p<0.01$,$^{***}p<0.001$. For each dimension, the two strongest correlations are indicated in bold.

Descriptors	Dimension 1		Dimension 2		Dimension 3	
Scale	*linear*	*power*	*linear*	*power*	*linear*	*power*
AT	-0.93^{***}		-		-	
LAT	$\mathbf{-0.95}^{***}$		-		-	
RT	-		-		-0.78^{***}	
TC	-		-		$\mathbf{0.80}^{***}$	
SC	0.90^{***}	$\mathbf{0.94}^{***}$	-	-	-	-
SS	0.63^{*}	0.88^{*}	$\mathbf{-0.63}^{*}$	-	-	-
SSK	-0.93^{***}	-0.72^{**}	-	-0.54^{*}	-	-
SKU	-0.88^{***}	-0.77^{***}	-	-	-	-
SRO	0.92^{***}		-	-	-	-
SF	-	-0.58^{*}	-	-	-	-
SFAT	$\mathbf{-0.95}^{***}$		-	-	-	-
HSC	0.91^{***}	$\mathbf{0.94}^{***}$	-	-	-	-
OSC	0.91^{***}	$\mathbf{0.94}^{***}$	-	-	-	-
ESC	0.90^{***}	0.81^{***}	-	-	-	-
OER	-	0.84^{***}	-	-	$\mathbf{0.87}^{***}$	-
IRRKRI	-0.79^{***}		-		-	
IRRKEN	-		0.53^{*}		0.69^{**}	
IRRJEN	0.79^{***}		-		-	
TR1	-0.92^{***}	-0.88^{***}	-	-	-	-
TR2	-0.71^{**}	0.62^{*}	$\mathbf{0.67}^{**}$	0.62^{*}	-	-
TR3	$\mathbf{0.95}^{***}$	$\mathbf{0.95}^{***}$	-	-	-	-

function of γ_m and ζ (see the bell-like shape when one of the parameters is fixed and the other varies). The second dimension might be correlated with a non-linear combination of the parameters γ_m and ζ, but we did not look for such a combination as it would not be easily interpretable from the physical point of view.

4.2.2. Acoustical correlates of the timbre space

Table III shows the correlations between the positions of the stimuli along the various dimensions of the perceptual space and the timbre descriptors previously presented in Table I.

It can be seen that the first dimension is correlated with the Logarithm of the Attack Time (LAT) [r(13)=-0.95, p<0.001], with various Spectral Centroid descriptors (SC, HSC, OSC [r(13)=0.94, p<0.001]), and with the third Tristimulus coefficient TR3 [r(13)=0.95, p<0.001]. Indeed, this dimension separates tones having long Attack Times and low Spectral Centroids, generated with small values of reed opening and blowing pressure (S1, S2, S5), or very small reed opening and high blowing pressure (S13), from tones having brief Attack Times and high Spectral Centroids, generated with high values of reed opening and mouth pressure (S8, S11, S12, S15). The fact that the Attack Time, the Spectral Centroid, and the third coefficient of Tristimulus are intercorrelated (see Table IV) is consistent with the physics of the instrument, as the greater the reed opening and mouth pressure, the faster the start of the self-sustained oscillations (small AT), and the greater the spectral richness (high SC and TR3) (see the Worman-Benade laws in [24]). Note that AT, SC and TR3 are not correlated with the other dimensions of the timbre space. It is also worth pointing out that the correlations with the Spectral Centroid increased when the computation of the descriptor was made using a power amplitude scale, which assigns a greater weight to the dominant harmonics.

Referring back to Table III, the second dimension is only weakly correlated with the second Tristimulus coefficient TR2 [r(13)=0.67, p<0.01], and the Spectral Spread [r(13)=-0.63, p<0.05] computed with a linear amplitude scale. Furthermore, these descriptors are more significantly correlated with the first dimension (see Table III), as they covary with AT, SC and TR3 (see Table IV). The descriptor TR2, based on two even harmonics (2 and 4) and one odd harmonic (5), is not adapted to characterize the difference in behavior between odd and even harmonics.

The third dimension is well correlated with the Odd/Even Ratio [r(13)=0.87, p<0.001] computed on a linear amplitude scale, and with the Temporal Centroid [r(13)=0.80, p<0.001]. Figure 6 displays the evolution of the Odd/Even Ratio as the mouth pressure and the reed aperture increases. The Odd/Even Ratio globally increases as the mouth pressure and reed opening increase (the odd harmonics grow faster than the even ones), attains a maximal value and then decreases past a certain threshold of ζ. We showed in a previous study that the decrease of OER is due to the beating reed situation, occurring when the reed

Table IV. Intercorrelations between the major timbre descriptors and the control parameters. The spectral descriptors have been computed using a power amplitude scale. ** and * indicate that the correlation is significant at the 0.01 and 0.05 levels, respectively.

	AT	SC	TR3	TR2	SS	OER	TC	γ_m	ζ
AT	1	−0.94**	−0.95**	−0.77**	−0.83**	−0.80**	0.66**	−0.20	−0.80**
SC		1	0.99**	0.69**	0.96**	0.89**	−0.56*	0.35	0.83**
TR3			1	0.69**	0.94**	0.87**	−0.58*	0.31	0.84**
TR2				1	0.52*	0.74**	−0.28	0.42	0.36
SS					1	0.89**	−0.42	0.46	0.77**
OER						1	−0.14	0.71**	0.49
TC							1	0.56*	−0.87**
γ_m								1	−0.18
ζ									1

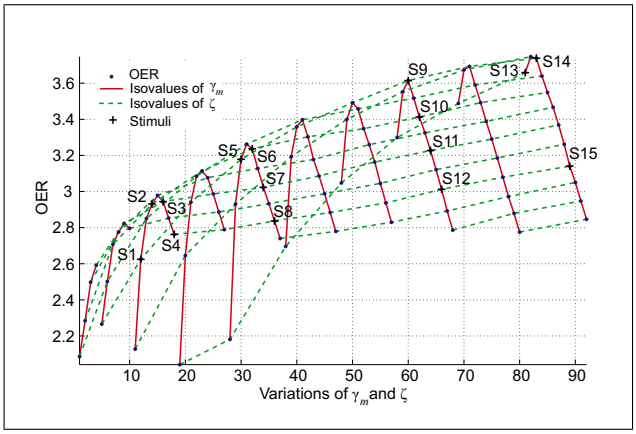

Figure 6. Evolution of the Odd/Even Ratio (OER) as a function of the control parameters γ_m (related to the blowing pressure) and ζ (related to the lip pressure on the reed). OER has been computed on a linear amplitude scale. The parameters used to generate the sounds are given in Figure 4 for normal playing conditions. The solid lines link the sounds with isovalues of γ_m. The dashed lines link the sounds with isovalues of ζ.

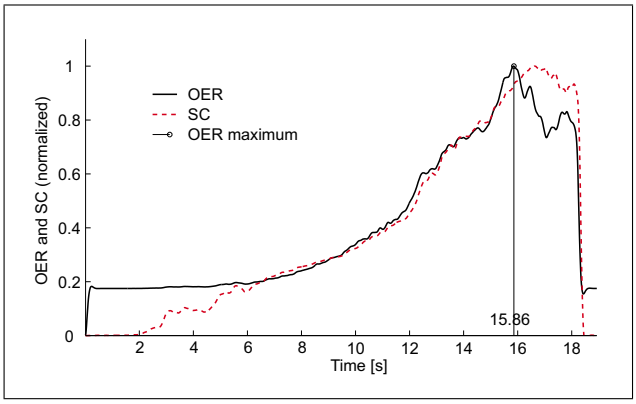

Figure 7. Evolution of the Odd/Even Ratio (OER) and Spectral Centroid (SC) for a natural clarinet crescendo.

beats the mouthpiece table and closes the reed channel aperture during a portion of its oscillations [25]. The closure of the reed channel induces the appearance of high-order even harmonics in the external pressure which grow faster than their relative odd harmonics. We verified that OER showed a similar behavior on a natural clarinet sound (see Figure 7 and sound example C). It is worth pointing out that a net timbre change is audible when OER attains its maximal value (t ≈ 15.9 s) and then decreases (due to the appearance of high-order even harmonics which give a metallic color to the sound). OER hence brings complementary information on timbre to that provided by the Spectral Centroid, as this timbre change at about 15.9 s is not revealed by SC, which keeps on increasing during the whole crescendo. These findings on OER are consistnt with the fact that the stimuli coordinates along the third dimension of the perceptual space globally increase as γ_m increases, and decrease as ζ increases (see Figure 5c).

4.3. Hierarchical Cluster Analysis (HCA)

The dissimilarity measures were also subjected to a Hierarchical Cluster Analysis (complete linkage) in order to test if certain groups of sounds presented systematic differences. Three main clusters of sounds were obtained. The first one associates tones generated using small (γ_m, ζ) values with the ones generated using high γ_m and very small ζ values (S1, S2, S5, S13). This cluster corresponds to tones which are not bright (small Spectral Centroid) and have a long attack part (high Attack Time). These results show that from the control point of view such timbral qualities (small brightness, long attack) can be obtained with two different techniques, either by applying a weak blowing pressure (small γ_m), while keeping a tight embouchure (small ζ), or by applying a strong blowing pressure (high γ_m), while keeping an even tighter embouchure (very small ζ). The second cluster gathers sounds generated using moderate values of γ_m and ζ (S3, S6, S9, S10, S14). The corresponding tones have moderate values of SC and AT. The third cluster associates tones generated using high (γ_m, ζ) values with ones generated using a high ζ value and a very small γ_m value (S4, S7, S8, S11, S12, S15). The tones gathered in this cluster are very bright (high SC) and present a brief attack part (small AT). Again, the results indicate that two combinations of mouth pressure and reed aperture can be used by the performer to obtain bright tones with a brief attack part.

Figure 8 represents the three-dimensional clarinet timbre space obtained with the MDS analysis and shows the

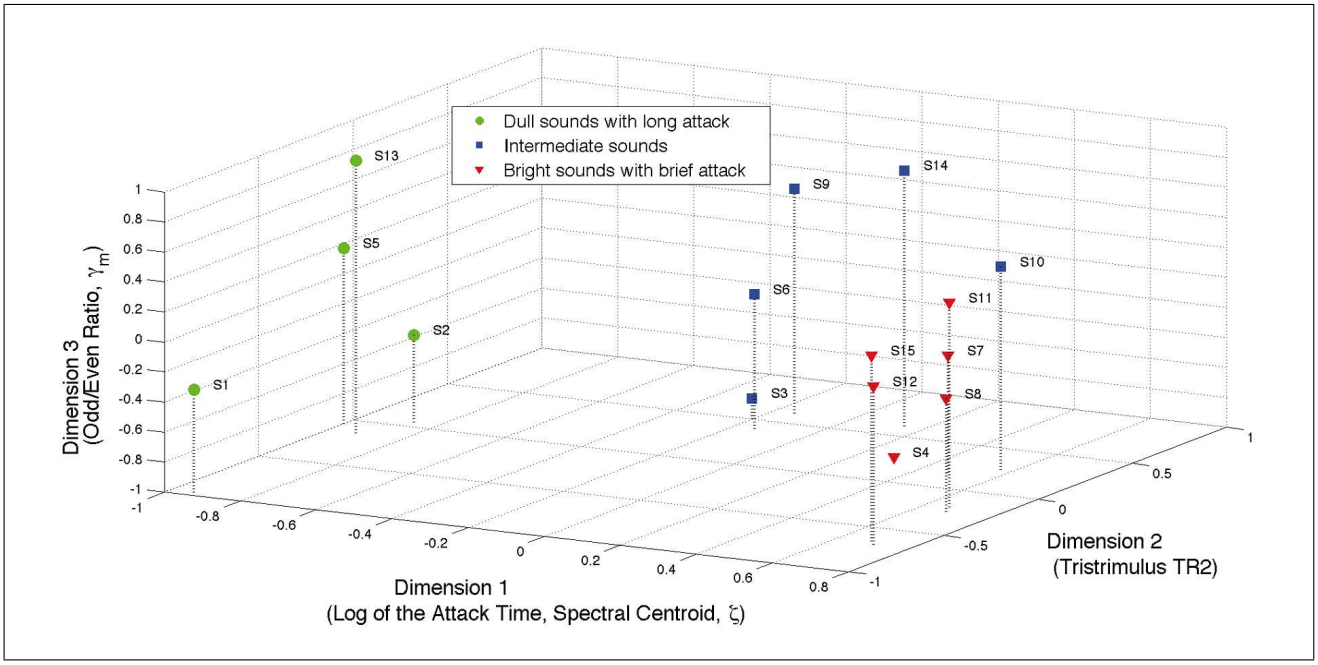

Figure 8. Three-dimensional clarinet timbre space and its mechanical and acoustical correlates. The three main clusters yielded by the HCA are represented with distinct markers.

Table V. Expressions used by the participants (P1 to P16) to describe their sound discrimination strategy (translated from French). They are gathered in three main categories: brightness, attack, and tension. Note that these categories are not necessarily independent. Certain expressions can belong to several categories at the same time.

	Brightness	**Attack**	**Tension**
P1	“nasal"		“soft"
P2	“rich and nasal"	“attack more or less soft"	
P3	“power of high-order harmonics"		
P4	“dull vs bright sounds"	“softness of the attack"	“softness of the attack"
P5		“attack"	
P6	“nasal"	“attack"	
P7	“richness of the sound in the trebles"	“dynamic of the transient"	
		“duration of the transient"	
P8			“aggressive vs soft sounds"
P9	“sharp sound, bright sound"	“attack"	“sharp sound"
P10	“round notes" / “nasal notes"		
P11	“brightness of the sound after the attack"	“attack time", “attack intensity"	
P12	“sizzling"		
P13	“richness of the sounds (more high harmonics)"		
P14	“intensity of the sizzling"		
P15	“round sound, full versus narrow sound"	“attack (soft or brief and intense)"	
	“brightness"		
P16	“brightness"	“attack (or more generally transient)"	

timbre descriptors and control parameters which best predict the distribution of the tones in the space. The three main clusters of tones yielded by the HCA are also reported in the figure (dull sounds with long attack, intermediate sounds, and bright sounds with brief attack).

4.4. Qualitative analysis of the verbal descriptions

The analysis of the questionnaire revealed three main factors of sound discrimination: the brightness of the sounds, the nature of their attack, and the sensation of tension (see Table V).

15 out of 16 participants employed expressions which refer to the percepts of *brightness* (also called *nasality*), either directly (e.g., “bright", “nasal"), or indirectly (“rich", “power of high-order harmonics"). 10 out of 16 participants also mentioned having relied on the nature of the attack (e.g., “attack time", “attack intensity") to discriminate the sounds. These qualitative observations are consistent with the preceding quantitative analyses of the timbre space. The first dimension indeed appeared to be well correlated with descriptors characterizing the nature of the attack (e.g., the Attack Time, characterizing its rapidity, the

Spectral Flux computed within the attack part and characterizing the non-synchronicity of the harmonic components), or with descriptors characterizing the spectral richness (e.g., the Spectral Centroid, the third Tristimulus coefficient TR3). It is worth pointing out that the Spectral Centroid strongly maps with the sensation of brightness [26], and has recurrently proved to be a good acoustical correlate of timbre space dimensions (see [8, 10, 12, 14]).

In [27], the Logarithm of the Attack Time and the Spectral Centroid are the best predictors of the first two dimensions of the 3D-timbre space associated with the discrimination of orchestral instruments' tones. As pointed out in section 4.2.2, the clarinet tones used in the perceptual experiment present covariant Attack Times and Spectral Centroids. It is then logical that AT and SC are both correlated with the first dimension. However, this does not prove that AT and SC would also be covariant with other configurations of the control parameters (e.g., different blowing pressure profiles). Hence, both descriptors seem to be needed to build a timbre model of clarinet tones.

Some expressions seem to refer to the timbral quality associated with the beating reed situation (e.g., "sizzling", "intensity of the attack", "richness of the sounds") which, as shown above, is highlighted by the shape of the Odd/Even Ratio. The timbre changes associated with Spectral Irregularity variations are well discriminated by listeners [26]. These results support the fact that the Odd/Even Ratio showed itself to be a good acoustical correlate of the third dimension of the timbre space.

Certain participants have discriminated the sounds according to the sensation of tension that they procured (e.g., "soft", "aggressive"). The soft tones may correspond to those which are dull and have a long attack, whereas the aggressive ones may be associated with those which are very bright and have a fast attack. Indeed, Zwicker and Fastl showed that the sensation of annoyance procured by sounds equal in loudness increased as the acuity (Spectral Centroid formulation based on an auditory model) increased [16].

5. Conclusions

This study shows that the perceptual relationships among different clarinet timbres generated with a synthesis model can be modelled in a 3-dimensional space. The underlying attributes used by participants when rating the dissimilarities between pairs of stimuli were investigated both from quantitative (mechanical and acoustical correlates of the timbre space dimensions) and qualitative (verbal descriptions) analyses. These analyses revealed that the participants were sensitive to changes of timbre induced by variations of the control parameters of the model (correlated to the blowing pressure and the lip pressure on the reed). The perceptual representation of clarinet timbres was best explained by timbre descriptors characterizing the attack of the tones (Attack Time), the spectral richness (Spectral Centroid, third Tristimulus coefficient), and the irregularity of the spectrum (Odd/Even Ratio). It is worth pointing out that the results from a study on the perceptual categorization of natural clarinet tones also showed that the Spectral Centroid is a major determinant of clarinet timbre [28].

Similar behaviors of the perceptual features (Spectral Centroid, Odd/Even Ratio) were found between natural and synthetic clarinet tones which validate the use of the physics-based synthesis model to explore the relationships between the control gestures of the instrument, the generated timbres and their perception. In particular, these analyses indicated that the Odd/Even Ratio is a good predictor of the change of timbral quality occurring during the beating reed situation for which high-order even harmonics are amplified.

The verbal descriptions given by the participants highlighted the influence of certain timbre changes on the sensation of tension. These results are interesting in the musical context, as music usually involves a succession of tensions and releases. According to Vines *et al.* [29], the sensation of tension is indeed correlated to the emotional responses of listeners. In other work, we have shown that musicians use timbre variations to vary their expression, and that such timbre variations induce changes in the emotional responses of listeners [3, 4].

6. Acknowledgments

This project was partly supported by the French National Research Agency (ANR JC05-41996, "senSons" http://www.sensons.cnrs-mrs.fr/, and ANR "Consonnes" http://www.consonnes.cnrs-mrs.fr/).
The authors would like to thank the reviewers for their helpful comments and corrections.

References

[1] J. Brymer: Clarinette. Hatier, Paris, 1979, (Collection Yehudi Menuhin).

[2] J. Kergomard: Le timbre des instruments à anche. – In: Le timbre, métaphore pour la composition. C. Bourgeois (ed.). I.R.C.A.M., 1991, 224–235.

[3] M. Barthet, P. Depalle, R. Kronland-Martinet, S. Ystad: Acoustical correlates of timbre and expressiveness in clarinet performance. Music Perception (in press) (2010).

[4] M. Barthet, P. Depalle, R. Kronland-Martinet, S. Ystad: Analysis-by-synthesis of timbre, timing, and dynamics in expressive clarinet performance. Music Perception (in press) (2010).

[5] P. Guillemain, J. Kergomard, T. Voinier: Real-time synthesis of clarinet-like instruments using digital impedance models. J. Acoust. Soc. Am. **118** (2005) 483–494.

[6] ANSI: USA Standard Acoustical Terminology. American National Standards Institute, New York, 1960.

[7] J. M. Hajda, R. A. Kendall, E. C. Carterette, M. L. Harshberger: Methodological issues in timbre research. 2nd ed. Psychology Press, New York, 1997, 253–306.

[8] J. M. Grey: Multidimensional perceptual scaling of musical timbres. J. Acoust. Soc. Am. **61** (1977) 1270–1277.

[9] D. L. Wessel: Timbre space as a musical control structure. Computer Music Journal **3** (1979) 45–52.

[10] C. L. Krumhansl: Why is musical timbre so hard to understand ? Proc. of the Marcus Wallenberg Symposium held in Lund, Sweden, Amsterdam, 1989, S. Nielzén, O. Olsson (eds.), Excerpta Medica, 43–53.

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

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

[13] R. Plomp: Timbre as a multidimensional attribute of complex tones. – In: Frequency Analysis and Periodicity Detection in Hearing. R. Plomp, G. F. Smoorenburg (eds.). A. W. Sijthoff, Leiden, 1970.

[14] A. Caclin, S. McAdams, B. K. Smith, S. Winsberg: Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. J. Acoust. Soc. Am. **118** (2005) 471–482.

[15] J. Kergomard: Elementary considerations on reed-instrument oscillations. – In: Mechanics of Musical Instruments. A. Hirschberg et al. (eds.). Springer-Verlag, New York, 1995.

[16] E. Zwicker, H. Fastl: Psychoacoustics: Facts and models. Springer-Verlag, New York, 1990.

[17] M. Barthet: De l'interprète à l'auditeur: une analyse acoustique et perceptive du timbre musical. Dissertation. Université Aix-Marseille II, 2008.

[18] G. Peeters: A large set of audio features for sound description (similarity and description) in the cuidado project. Tech. Rept. version 1.0, I.R.C.A.M., Paris, 2004.

[19] J. W. Beauchamp: Synthesis by spectral amplitude and brightness matching of analyzed musical instrument tones. J. Audio Eng. Soc. **30** (1982) 396–406.

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

[21] J. Marozeau, A. de Cheveigné, S. McAdams, S. Winsberg: The dependency of timbre on fundamental frequency. J. Acoust. Soc. Am. **114** (November 2003) 2946–2957.

[22] B. Picinbono: On instantaneous amplitude and phase of signals. IEEE Transactions on Signal Processing **45** (1997) 552–560.

[23] W. R. Dillon, M. Goldstein: Multivariate analysis. John Wiley & Sons, New York, 1984, (Wiley series in probability and mathematical statistics).

[24] A. H. Benade, S. N. Kouzoupis: The clarinet spectrum: Theory and experiment. J. Acoust. Soc. Am. **83** (January 1988) 292–304.

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

[26] R. A. Kendall, E. C. Carterette: Verbal attributes of simultaneous wind instrument timbres: I. von bismarck's adjectives. II. adjective induced from piston's orchestration. Music Perception **10** (1993) 445–468; 469–502.

[27] S. McAdams, J. W. Beauchamp, S. Meneguzzi: Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters. J. Acoust. Soc. Am. **105** (1999) 882–897.

[28] M. A. Loureiro, H. B. de Paula, H. C. Yehia: Timbre classification of a single instrument. ISMIR 2004 5th International Conference on Music Information Retrieval, Barcelona, Spain, 2004, Audiovisual Institute, Universitat Pompeu Fabra.

[29] B. W. Vines, C. L. Krumhansl, M. M. Wanderley, D. J. Levitin: Cross-modal interactions in the perception of musical performance. Cognition **101** (2006) 80–113.