Acta Acustica united with Acustica
Vol. 103 (2017) 812 – 829

DOI 10.3813/AAA.919110

# Investigation of the Relationships Between Perceived Qualities and Sound Parameters of Saxophone Reeds

Jean-François Petiot[1,2)], Pierric Kersaudy[1,2)], Gary Scavone[2)], Stephen McAdams[2)], Bruno Gazengel[3)]

[1)] Ecole Centrale de Nantes, IRCCyN, Nantes, France. Jean-Francois.Petiot@ec-nantes.fr, Pierric.kersaudy@gmail.com

[2)] CIRMMT, Schulich School of Music, McGill University, Montreal, Quebec, Canada. [gary, smc]@music.mcgill.ca

[3)] Université du Maine, LAUM, Le Mans, France. bruno.gazengel@univ-lemans.fr

**Summary**

The perceived quality of cane reeds used on saxophones or clarinets may be very different from one reed to another even though the reeds have the same shape and strength. The aim of this work is to better understand the differences in the perceived quality of reeds by making use of acoustical measurements. A perceptual study, involving a panel of 10 musicians, was first conducted on a set of 20 reeds of the same strength. Each musician assessed each of the 20 reeds according to three descriptors: *Brightness*, *Softness*, and *Global quality*. Second, signal recordings during saxophone playing (saxophone playing by a musician in the laboratory, called *in vivo* measurements) were made of the pressures in a player's mouth, in the mouthpiece, and at the bell of the instrument. These measurements enable us to deduce specific acoustical variables, such as the threshold pressure or the spectral centroid of the notes. After an analysis of the perceptual and acoustical data (assessment of the agreement among the assessors and the main consensual differences between the reeds), correlations between the perceptual and acoustical data were performed. A modeling of the descriptors *Brightness* and *Softness* according to the acoustical variables is proposed using multiple linear regression. Results show that the pressure in the mouth at the beginning of the permanent regime is an important variable to predict the softness of the reed. The performance of the models in the prediction of the perceptual dimensions provides important clues for a more objective assessment of perceived reed qualities.

**PACS no. 43.75.Ef, 43.75.Pq, 43.75.Yy**

## 1. Introduction

For a saxophone player, the quality of a reed (a piece of cane that the player clamps to the mouthpiece) is fundamental and has important consequences for the quality of the sound produced by the instrument. The experience of saxophone players shows that in a box of reeds, roughly 30% are of good quality, 40% are of medium quality, and 30% are of bad quality. Nevertheless, the only indicator a musician can see on a box of reeds is the strength, which is usually measured by the maker by applying a static force at a particular location near the tip. The reeds are then classified according to the strength measured. But this strength is not representative of the perceived quality of the reed. According to musicians, there are many differences among the reeds in a given box. And it is still difficult to understand which physical or chemical properties govern the perceived quality. The control of reed quality remains an important problem for reed makers, because of the high variability of this natural material (arundo donax) and the large number of influencing factors. To control their production, reed makers are interested in characterizing objectively the quality of reeds.

A thorough study of the perceived quality of reeds, and more generally of musical instruments, necessitates two categories of measurements on a set of products: subjective assessments given by musicians or listeners [1], and objective measurements (chemical or physical) made on a set of instruments [2]. The task is then to uncover (with statistical methods) a model for predicting subjective dimensions from the objective measurements. In [3] for example, the preferences of French horn players are correlated with geometrical and acoustical variables, in an attempt to understand what influences the quality of instruments. The main difficulty in the study of the perceived quality of musical instruments is to gather subjective assessments from musicians that are both reliable and suf-

Received 18 May 2016,
accepted 16 August 2017.

ficiently representative of the subtle interaction between the musician and the instrument. Many uncontrolled factors may influence this complex interaction. The subjective ratings of a "subject" may be non-reproducible and may be context-dependent, semantically ambiguous, and dependent on cultural background and musical training. A study of the reliability of violinists in assessing perceptual qualities of instruments is presented in [4], where the authors noticed large inter-individual differences in preference, but also in perceived qualities of the instruments. To get representative data, it is necessary to find an acceptable trade-off between realistic playing conditions and artificial assessments of stimuli that may be oversimplified and then become too caricatured [5]. And to trust the data, it is necessary to control the assessments with repetitions and with several independent assessors. In this context, experimental protocols and data analysis techniques developed in sensory analysis can be very useful [6]. Several statistical analysis methods are proposed to assess the evaluations of subjects and the panel's performance in descriptive analysis tasks [7].

With regard to reeds, the main investigations have focused on acoustical or mechanical measurements of the materials and subjective/objective experiments. In [8], optical measurements were used to assess the vibrational modes of clarinet reeds, which had been correlated with the quality as judged by musicians. The authors suggested different patterns of vibrations that should be representative of good reeds, results that must be confirmed given the small size of the reed sample used. A chemical analysis of the reed material was made in [9], but no significant differences could be identified between good and poor reeds. The influence of the relative humidity of a reed was studied in [10], where the authors noticed a great influence of water-soluble extracts on the frequency response of the material. The extraction of mechanical parameters of reeds was proposed in [11] with a validation using numerical models, but no correlation with the perceived quality was proposed. In [12], Gazengel and Dalmont proposed two categories of measurements to explain the behavior of a tenor saxophone reed. On the one hand, they performed *in vitro* measurements using a mechanical bench to characterize the mechanical response of the reed. The results showed that the repeatability of the measurements was low, and that the mechanical properties of the material may change significantly over time. Furthermore, apart from the stiffness, no variable extracted from the frequency response could explain the perceived differences among the reeds. On the other hand, they performed *in vivo* measurements during saxophone playing, by measuring the acoustic pressure at the bell of the saxophone and in the mouthpiece, as well as the pressure in the player's mouth. These studies showed that the perceived strength can be matched to the estimated threshold pressure in the musician's mouth, and that the perceived brightness correlates with the high-frequency content of the sounds and the spectral centroid [13, 14]. Of course, the spectral content of the sound perceived by the player (mainly by the ears but also by bone conduction inside the head) is different from the sound at the saxophone bell. Nevertheless, it is assumed that the same transformation applies to all the reeds. Therefore, relative ratings are unchanged, suggesting that the sound at the saxophone bell is a relevant measurement. It is also important to mention that the correlation between brightness and high- frequency content of the sound agrees with many studies on timbre [15]. These results were based on a small set of reeds (12), a single musician, and were limited to simple correlations between subjective variables and acoustical measurements. A study with larger sets of Bb clarinet reeds (50 and 150) was presented in [16]. Different perceptual descriptors (e.g. *ease of playing, brightness*) were assessed by a single expert, and correlated with mechanical parameters of the reeds, static or dynamic. The main results showed that the static and dynamic compliances (inverse of the stiffness) of the reeds were negatively correlated with the descriptor *ease of playing*. Again, the perceptual assessments were based on only one musician and on one-to-one correlations between perceptual and mechanical measurements. To understand the different dimensions of the perceived quality, and to be able to test their generalizability, a panel of musicians and multivariate modeling techniques are needed.

In a previous paper [17], we defined a predictive model of tenor saxophone reed quality with regression. This model was based on a set of 20 reeds and a panel of 10 musicians, each musician assessing all the reeds. This paper is the continuation of that work. It is centered specifically on the study of the performance of the panel of musicians and on the proposal of a model of the perceived qualities of reeds with data modeling techniques. The objective of the paper is first to assess the reliability of the perceptual assessments, and second to explain them with acoustical *in vivo* measurements.

The paper is organized as follows. Section 2 presents the details of the experiments carried out with a set of 20 reeds and a panel of 10 musicians. The acoustical *in vivo* measurements, obtained from performances with two different musicians, are described in detail. Section 3 is dedicated to the presentation of the results of the perceptual tests and the acoustical measurements. The agreement between the different assessments and the performance of the panel are presented. Section 4 presents different models of the *Softness* and *Brightness* of a reed using multiple linear regression. The last section draws general conclusions and discusses the contribution of this study.

## 2. Material and methods

### 2.1. Reed samples

A set of 20 tenor saxophone reeds of the same cut, strength (2.5), and brand (Classic *Vandoren*) was selected. Given that one of the objectives of the study is to understand the differences between reeds sold as similar, we did not make any selection of the reeds: they all came from 4 commercial boxes of 5 reeds each, bought in a music shop. This choice means that the differences between the reeds may be small, but they will be representative of what a saxophonist experiences in his/her everyday life when selecting reeds. An additional objective of the study is thus to assess the magnitude of the differences (perceived or measured) between 20 “similar” reeds.

## 2.2. Perceptual evaluations

### 2.2.1. Procedure

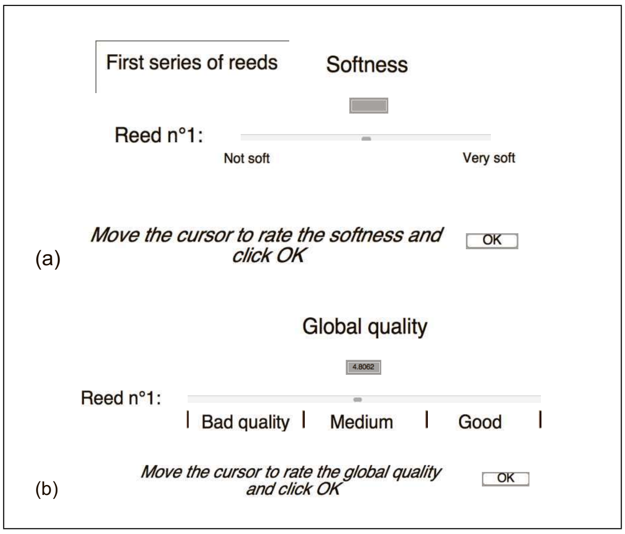
Ten musicians participated in the perceptual tests (9 males, average age = 20 years). They were all skilled saxophonists (students involved in a music curriculum at Schulich School of Music of McGill University), with more than 10 years of practice. For the sake of consistency, all musicians (denoted as “assessors” in the rest of the paper) used the same mouthpiece during the study (Vandoren V16 T7 Ebonite). However, they were asked to play on their own tenor saxophone. These tests took place at CIRMMT (Centre for Interdisciplinary Research in Music Media and Technology, McGill University in Montreal, Quebec, Canada) in the same room, the Performance and Recording Lab.

Different semantic dimensions are generally defined to assess perceptual differences between products. For saxophone reeds, interviews with saxophonists have shown that the most frequent dimensions relate to “ease of emission”, “quality of sound”, or “homogeneity” [18]. Inside these categories, a great diversity of terms is used by musicians to assess a reed (strength, projection, richness, centering, . . . ). Nevertheless, these terms come from different languages and no standard list of descriptors is available. On the basis of previous studies [8, 14], and from our experience with reed assessments, we proposed three perceptual descriptors to assess the reeds:

- The *Softness* of the reed, which corresponds to the ease of producing a sound. This dimension was assessed on a continuous scale from 0 (not soft) to 10 (very soft) (Figure 1a),
- The *Brightness* of the sound produced using the reed. This dimension was assessed on a continuous scale from 0 (not bright) to 10 (very bright),
- The *Global quality* of the reed. This dimension can also be related to the *preference* of the musician concerning the reed. It was assessed on an analogical-categorical scale [19], which was coded on a continuum from 0 to 10 with an indication of 3 categories on the scale: bad – medium – good (Figure 1b).

The test was divided into 3 phases: a training phase, an evaluation phase, and the filling out of a questionnaire concerning the mouthpiece, reed, saxophone, and musical style the musicians usually play, as well as their past experience.

A training phase was proposed to help the assessors understand the meaning of the two descriptors *Softness* and *Brightness* and to verify their use of the scale. The method is inspired from the training phase described in [20]. “Anchor reeds”, prepared in advance, and located at the extremes of the *Softness* scale, were proposed, and recorded sounds with different *Brightnesses* were played to the assessors. These anchor reeds were taken from boxes of reeds of lower and higher strength values (strength = 2.0 and 3.0). Finally, assessors were asked to participate in a short test to train themselves in the use of the scales and to verify their discrimination. Three quite different reeds (of different strength number 2.0, 2.5 and 3.0) were presented to the assessor, repeated once. A one-way analysis of variance with the factor “reed” was used to estimate whether the assessor could discriminate between the reeds on each scale. All assessors produced normally distributed data and discriminated the reeds (significant effect of the F-test for the reed factor with an individual one-way ANOVA), so they were all selected for the next evaluation phase.



Move the cursor to rate the softness and click OK

Move the cursor to rate the global quality and click OK

Figure 1. (a) Continuous unstructured scale for the assessment of *Softness*. (b) Continuous structured scale for the assessment of *Global Quality*.

For the evaluation phase, the musician was asked to play each of the 20 reeds in turn, and to rate them against the three descriptors on the graphical interface. Complete freedom was given to the musician both in terms of what they played and in the duration of the assessment. The reeds, disinfected first (hydro-alcoholic solution) and moistened with water and a sponge, were set on the mouthpiece by the experimenter. To reduce the effect of reed moistening on the evaluation, all the reeds were placed in water until saturated before playing. They were presented to the assessor in an order following a Williams Latin square, in order to control the order and carry-over effects [21]. The assessments were repeated two times in two independent blocks during the same day. Given that we had 20 reeds, 10 assessors and two repetitions, the presentation plan was perfectly balanced. Between the tests, the reeds were stored in their original boxes and plastic dispensers, in an air-conditioned room. For each of the 10 assessors, the perceptual data consisted of two arrays of quantitative values (one per repetition). The arrays had 20 rows (one per reed) and three columns (one per descriptor). The assessment of reed *i* by assessor *j* during session *k* according to *Softness* is denoted $y^1_{ijk}$, $y^2_{ijk}$ for *Brightness* and $y^3_{ijk}$, for *Global quality*. For a more generic notation, the assessment of reed $i$ by assessor $j$ during session $k$ according to any descriptor is denoted $y_{ijk}$.

2.2.2. Method for the analysis of the individual assessments

In sensory analysis, it is important to establish the performance of the assessors to ensure the quality of the data [7]. Three criteria are of prime importance in sensory evaluation: discrimination ability, reliability, and agreement among the panelists [22]. Our sensory panel consisted of J = 10 assessors who judged I = 20 products (reeds) during K = 2 repetitions (repetitions are called sessions in the following presentation) using M = 3 attributes. We use a particular notation for the representation of different mean values: considering the evaluation $y_{ijk}$, a dot in place of a subscript means average over that subscript [e.g., the notation $y_{\bullet j \bullet}$ indicates the mean of evaluations $y_{ijk}$ over the indices $i$ (product) and $k$ (repetition)].

We describe in this section the principles of the GRAPES method [23], which is a powerful tool for assessing the performance of a panel of experts in sensory analysis. It provides graphical representations of assessors' performance. The method focuses on the different uses of the scale, the reliability of the assessors, their repeatability, and their discrimination ability. We report is this section the six quantities that are defined in the GRAPES method to assess the individual performance of an assessor, and provide a brief explanation of their interest.

Two quantities are computed to compare the use of scales by assessors. LOCATION$_j$ (Equation 1) is the average of the scores given by assessor $j$ (in other words the mean rating):

$$\mathrm{LOCATION}_j = y_{\bullet j \bullet}, \tag{1}$$

and SPAN$_j$ (Equation 2) is the average across sessions of the standard deviation of the reed scores based on the mean session reed scores across reeds, $y_{\bullet jk}$.

$$\mathrm{SPAN}_j = \frac{1}{K} \sum_K \sqrt{\frac{\sum_i \left(y_{ijk} - y_{\bullet jk}\right)^2}{(I-1)}}. \tag{2}$$

SPAN$_j$ characterises the average variability in reeds across sessions according to assessor $j$, and represents the range of the assessments of this assessor.

Two coefficients are computed to assess the performance of the assessors in terms of their reliability and the influence of the different repetitions for each descriptor. The unreliability ratio, labeled UNRELIABILITY$_j$ (Equation 3), represents the repeatability error of the assessor, relative to the average variability in the ratings. The value is zero (perfectly reliable) if the assessor gives identical ratings of the products for the two sessions. It is given by

$$\mathrm{UNRELIABILITY}_j = \tag{3}$$
$$\frac{\sqrt{\frac{1}{(I-1)(K-1)} \sum_{i,k} \left(y_{ijk} - y_{ij\bullet} - y_{\bullet jk} + y_{\bullet j \bullet}\right)^2}}{\mathrm{SPAN}_j} =$$
$$\frac{\sqrt{\frac{1}{(I-1)(K-1)} \sum_{i,k} \left((y_{ijk} - y_{\bullet jk}) - (y_{ij\bullet} - y_{\bullet jk}) + y_{\bullet j \bullet}\right)^2}}{\mathrm{SPAN}_j}.$$

The DRIFT_MOOD$_j$ (Equation 4) is the between-sessions error relative to the average variability in the ratings (expressed in SPAN units). It represents the deviation of the ratings of the assessor across the sessions and is given by

$$\mathrm{DRIFT\_MOOD}_j = \frac{\sqrt{\frac{1}{(K-1)} \sum_k \left(y_{\bullet jk} - y_{\bullet j \bullet}\right)^2}}{\mathrm{SPAN}_j}. \tag{4}$$

Finally, two further quantities are proposed to assess the performance of an assessor (Equations 5 and 7)

$$\mathrm{DISCRIMINATION}_j = \tag{5}$$
$$\frac{\left[K \sum_i (y_{ij\bullet} - y_{\bullet j \bullet})^2\right] / (I-1)}{\sum_{i,k} (y_{ijk} - y_{ij\bullet} - y_{\bullet jk} + y_{\bullet j \bullet})^2 ] / (I-1)(K-1)}$$

is the classical F-ratio for testing the significance of a product-effect in an individual two-way ANOVA model (Equation 6)

$$Y_{ijk} = \text{grand mean} + \text{product}_i + \text{session}_k + \text{error}. \tag{6}$$

$$\mathrm{DISAGREEMENT}_j = \tag{7}$$
$$\frac{\left[KJ \sum_i (y_{ij\bullet} - y_{i\bullet\bullet} - y_{\bullet j \bullet})^2 + y_{\bullet\bullet\bullet})^2\right] / (I-1)(K-1)}{\sum_{i,j,k} (y_{ijk} - y_{ij\bullet} - y_{\bullet jk} + y_{\bullet j \bullet})^2 ] / (I-1)(K-1)}$$

measures the contribution of assessor $j$ to the product × assessor interaction F-ratio in the global ANOVA model presented in equation (8),

$$\begin{aligned} Y_{ijk} \;=\; & \text{grand mean} + \text{product}_i + \text{assessor}_j + \text{session}_k \\ & + \text{session} * \text{assessor} + \text{product} * \text{assessor} + \text{error}. \end{aligned} \tag{8}$$

## 2.3. Acoustical measurements

2.3.1. Procedure

The principle of *in vivo* measurements in the context of our experiment is to record acoustical variables when a musician is playing the reeds. The advantage is that we have a real playing situation, close to the perceptual assessment situation, but this method has the disadvantage of introducing variability, particularly because of the way the musician plays. Many factors can influence the tone quality (embouchure, amount of mouthpiece in the mouth, oral cavity manipulation, etc.). There are indeed different techniques that are taught for the embouchure of the saxophone ("loose" or "tight"), that may have an important influence on the sound produced. For example, musicians make a clear distinction between "classical" or "jazz" sound quality [24]. But no clear explanation of the influence of the player technique on tone quality is available and further studies are needed. Even if the variability in tone is important according to the musician, we consider that it is interesting to study how perceptual assessments of musicians concerning reeds correlate with playing parameters of the instrument, when it is played by a given musician.

We chose to measure the acoustic pressure $p_a(t)$ at the bell of the saxophone and the pressure in the musician's mouth $p_m(t)$. The mouth pressure was measured using an Endevco 8507-C1 differential pressure sensor attached to the front of the mouthpiece such that it was inside the mouth during normal playing. The small size of this microphone allows a minimally invasive pressure measurement, even if the musician needs some time and practice to become accustomed to its presence. The acoustic pressure was measured with a B&K 4190-L-001 microphone placed in front of the saxophone bell (at a constant distance equal to the diameter of the bell, 13 cm). The sampling frequency used was 44100 Hz. Two saxophonists (players A and B, not included in the assessors' panel) were responsible for the *in vivo* measurements, using the same mouthpiece and the 20 reeds as used in the perceptual test. The musicians performed two sessions of measurements two months apart, one session before the perceptual test and one session after. The pattern played by the saxophonists was a descending arpeggio of seven notes (C5, G4, Eb4, C4, G3, Eb3, C3-concert key, where C4 has a fundamental frequency of 261.6 Hz), played with a breath attack (no use of the tongue) and a mezzo-forte ($mf$) dynamic. This pattern was repeated five times for each reed and each saxophonist. An example of a measured signal is shown in Figure 2.

The playing of the seventh note (the lowest note: C3) was often imprecise, primarily due to the poor response of the lowest notes of the saxophone used. We chose to discard this note and to keep the data for only the first six notes. In summary, the acoustical measurements consist of the acoustic pressure and the mouth pressure measured on 20 reeds × 6 notes × 5 repetitions × 2 musicians × 2 sessions.

### 2.3.2. Playing variables estimation

From the acoustic pressure $p_a(t)$ at the bell of the saxophone and the pressure in the musician's mouth $p_m(t)$, several variables that characterize the interaction between the musician, the reed, and the saxophone were extracted. Each variable was computed for each note, each reed, and each of the five repetitions of the pattern. The variables were calculated by analyzing separately the transient and stationary parts of the signal. The general scheme used for this estimation is the following:

- Note detection using a threshold applied to the radiated pressure envelope,
- For each note:
  - Detection of the stationary part of the note,
  - Estimation of variables on the stationary part (mean mouth pressure, acoustic pressure parameters),
  - Estimation of variables on the transient part (threshold pressure, attack time),
  - Efficiency estimation.

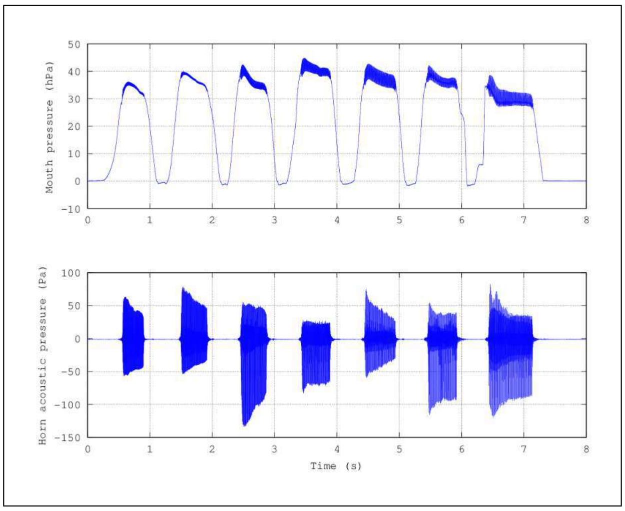


Figure 2. Example of signal measured when the musician played the 7-note arpeggio: (top) Mouth pressure; (bottom) Acoustic pressure at the saxophone horn output.

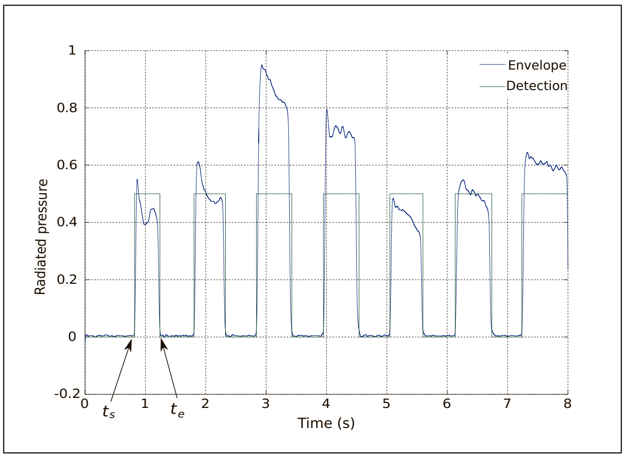


Figure 3. View of the note detection using the threshold applied to the normalized envelope of the radiated acoustic pressure. View of start time $t_s$ and end time $t_e$ of the first note.

The reader may refer to [12] and [25] for additional explanations concerning the definition of these variables. Each note is detected by using a threshold applied to the acoustic radiated pressure envelope. The envelope is estimated by convolving the absolute value of the acoustic pressure $p_a(t)$ with a Hann window $W[k] = (1 - \cos(2\pi k/N))/2$ of length $T_w = 1/F_c$, where $F_c$ is the cut-off frequency ($F_c = 20\,\text{Hz}$).

The comparison of the normalized envelope $E_a[n] = \text{abs}(p_a[n] * W[n])/\max(E_a[n])$ with a threshold enables us to deduce the start time $t_s$ and end time $t_e$ of each note as shown in Figure 3. The threshold value is chosen empirically by analysing different recorded signals.

For each note, the stationary part of the signal is estimated by calculating the energy of the signal as a function of $t$:

$$E(t) = \int_{t_s}^{t} p_a^2(\tau)\,\mathrm{d}\tau. \tag{9}$$

The stationary part of the signal is defined by $E(t) \in [0.05, 0.95] \times E_{\max}$, where $E_{\max}$ is the maximum energy obtained at the end of the note. The time at the beginning of the stationary part of the signal is $t_{\text{stat\_s}}$ ($E(t_{\text{stat\_s}}) = 0.05E_{\max}$), whereas the time at the end of the stationary part of the signal is $t_{\text{stat\_e}}$ ($E(t_{\text{stat\_e}}) = 0.95E_{\max}$).

A first category of variables concerns the acoustics of the sound, computed on the stationary part of the acoustic pressure $p_a(t)$. These variables were obtained from the frequencies $f_k$ and the amplitudes $A_k$ of the $k$ components of the sound, computed with a Discrete Fourier Transform. The first 40 harmonics of the spectral representation were considered (to respect the Shannon condition for all the notes including that of highest pitch).

The following variables were estimated:

- Spectral Centroid,

$$SC = \frac{1}{f_1}\frac{\sum_{k=1}^{40} A_k f_k}{\sum_{k=1}^{40} A_k}. \tag{10}$$

- Odd-harmonic Spectral Centroid,

$$OSC = \frac{1}{f_1}\frac{\sum_{h=0}^{19} A_{2h+1} f_{2h+1}}{\sum_{h=0}^{19} A_{2h+1}}. \tag{11}$$

- Even-harmonic Spectral Centroid,

$$ESC = \frac{1}{f_1}\frac{\sum_{k=1}^{20} A_{2h} f_{2h}}{\sum_{k=1}^{20} A_{2h}}. \tag{12}$$

- Ratio between Odd and Even harmonics,

$$OER = \frac{\sum_{h=0}^{19} A_{2h+1}^2}{\sum_{h=0}^{19} A_{2h}^2}. \tag{13}$$

- Amplitude of the harmonic signal,

$$Lv = \sqrt{\frac{\sum_{k=1}^{40} A_k^2}{2}}. \tag{14}$$

- 3 tristimuli ($TR1$, $TR2$, $TR3$) and an additional stimulus $TR4$ (ratio between the power of the harmonics above 4000 Hz and the total power of the harmonics),

$$TR1 = \frac{A_1^2}{\sum_{k=1}^{40} A_k^2}, \tag{15}$$

$$TR2 = \frac{A_2^2 + A_3^2 + A_4^2}{\sum_{k=1}^{40} A_k^2}, \tag{16}$$

$$TR3 = \frac{\sum_{k=5}^{40} A_k^2}{\sum_{k=1}^{40} A_k^2}, \tag{17}$$

$$TR4 = \frac{\sum_{k/f_k>4000}^{40} A_k^2}{\sum_{k=1}^{40} A_k^2}. \tag{18}$$

- the Attack Time ($AtT$); i.e., time to establish the permanent regime, defined by

$$AtT = t_{\text{stat\_s}} - t_s. \tag{19}$$

The “unitless” spectral centroid (also for odd and even) is used to be able to compare effects of notes with different fundamental frequencies.

A second category of variables is defined with respect to the pressure in the mouth $p_m(t)$. To detect the time at which the acoustic pressure measured at the saxophone bell shows a periodic component at the fundamental frequency of the played note (this frequency being a priori known by analyzing the whole signal over the note duration), a detection function is proposed, defined by

$$D(t) = \frac{\sqrt{U^2(t) + V^2(t)}}{\max\left[\sqrt{U^2(t) + V^2(t)}\right]}, \tag{20}$$

with

$$U(t) = \int_{t_s}^{t} p_a(\tau)\cos\left(2\pi f_1 \tau\right)\mathrm{d}\tau, \tag{21}$$

$$V(t) = \int_{t_s}^{t} p_a(\tau)\sin\left(2\pi f_1 \tau\right)\mathrm{d}\tau, \tag{22}$$

where $f_1$ is the estimated fundamental frequency on the stationary part. The comparison between indicator $D(t)$ and a threshold value (defined empirically) enables us to deduce the threshold pressure time $t_p$ of the note (beginning of the permanent regime with a fundamental frequency $f_1$). The threshold pressure ($PTh$) corresponds to the pressure in the mouth at the beginning of the permanent regime at frequency $f_1$,

$$PTh = p_m(t_p). \tag{23}$$

The mean Static Pressure ($StP$) is the mean of the pressure in the mouth during the stationary part of the signal,

$$StP = \frac{1}{t_{\text{stat\_e}} - t_{\text{stat\_s}}}\int_{\text{stat\_s}}^{\text{stat\_e}} p_m(t)\,\mathrm{d}t. \tag{24}$$

The efficiency ($Eff$) is defined as the ratio between the amplitude (RMS) of the harmonic pressure signal to the mean static pressure $StP$,

$$Eff = \frac{L_V}{StP}. \tag{25}$$

In conclusion, each reed is defined by 13 acoustical variables × 6 notes × 5 repetitions × 2 musicians × 2 sessions.

## 3. Results and discussion

### 3.1. Analysis of the perceptual assessments

#### 3.1.1. Individual assessor’s performance

This section focuses on the individual performance of the assessors, to determine whether the results of some participants should be discarded.

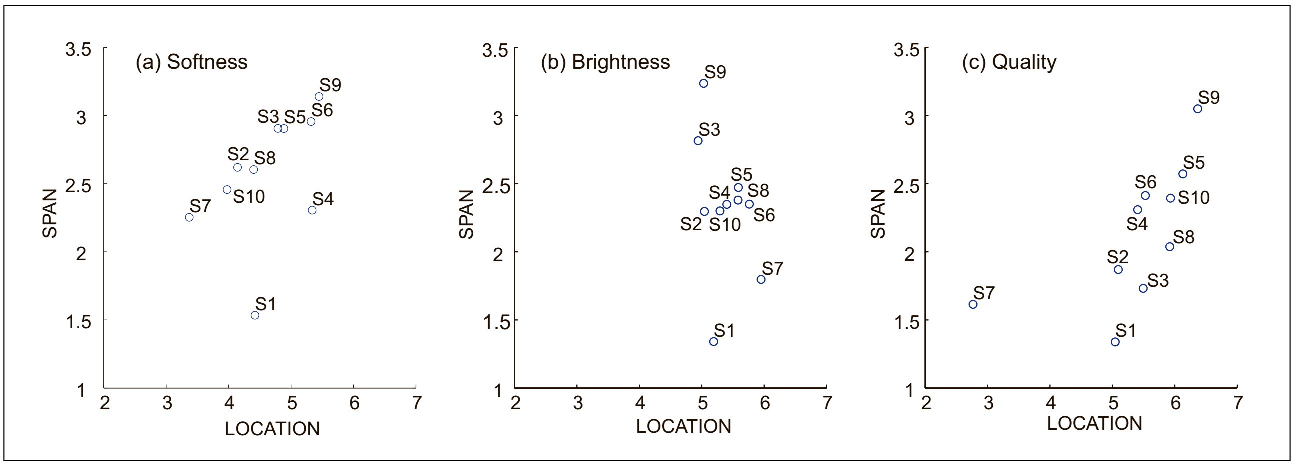


Figure 4. Plot of $\text{SPAN}_j$ vs. $\text{LOCATION}_j$ for each assessor $Sj$ and each descriptor.

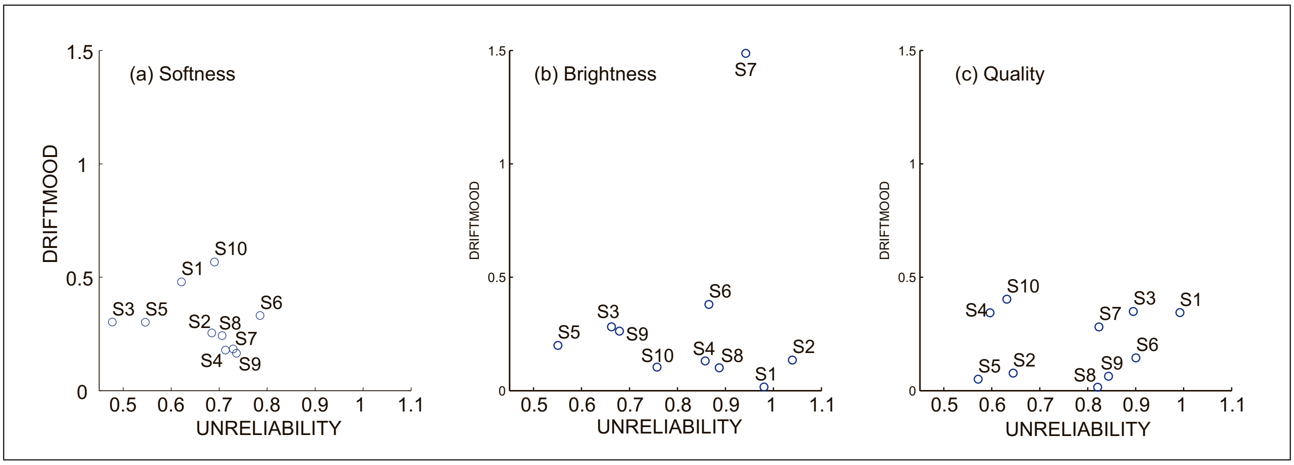


Figure 5. Plot of $\text{DRIFT\_MOOD}_j$ vs. $\text{UNRELIABILITY}_j$ for each assessor $Sj$ and each descriptor.

Figure 4 presents $\text{SPAN}_j$ vs $\text{LOCATION}_j$ for the assessors S1 to S10 and the three descriptors.

The results show that assessor S1 uses a small range for all the assessments (the SPAN is very small for all the descriptors), contrary to S9 who uses a wide range. Assessor S7 globally dislikes all the reeds (if we assume that the global quality of the reed can be an indicator of preference – LOCATION is low for this assessor for the quality descriptor – Figure 4c) and assesses them as not soft (Figure 4a).

Figure 5 represents, for each descriptor, the performance of the assessors according to DRIFT_MOOD and UNRELIABILITY.

For *Softness*, S6 is the least reliable, and S3 and S5 are the most reliable. S10 deviates the most between the two sessions (high DRIFT_MOOD). For *Brightness*, S2 is the least reliable, and S5 is the most reliable. S7 presents a very high deviation between the two sessions. For *Quality*, S1 is the least reliable, and S5 is the most reliable.

We can conclude that S5 is a particularly reliable assessor. We can also see that the worst value of unreliability for *Softness* (0.8 for assessor S6) is lower than most of the values for *Brightness* and *Quality*. This means that most assessors (S6, S4, S8, S1, S2, S7) are less reliable for *Brightness* than for *Softness*. This result is in accordance with the feedback from participants during the tests, who indicated having more difficulty assessing *Brightness* than *Softness*.

Figure 6 represents, for each descriptor, the performance of the assessors according to $\text{DISAGREEMENT}_j$ and $\text{DISCRIMINATION}_j$.

On these graphs, a vertical line is located at a value of DISCRIMINATION equal to a 5% significant Fisher variance ratio for reed-effect in the model of equation 6. Thus, the line allows a rapid interpretation of a statistical test on the reed effect: assessors located on the right side of the vertical line are significantly discriminant at the 5% level for the reed effect.

A horizontal line is located at a value of the average contribution of an assessor (for a panel of 10 assessors – assuming that all the 10 assessors have this same average contribution) corresponding to a 5% significant product × assessor interaction with the ANOVA model equation (8). In this case, this line is not equivalent to a statistical test. It is only an indication to evaluate whether an assessor contributes more than this average contribution (in this case it is located above the line) or less (below the line).

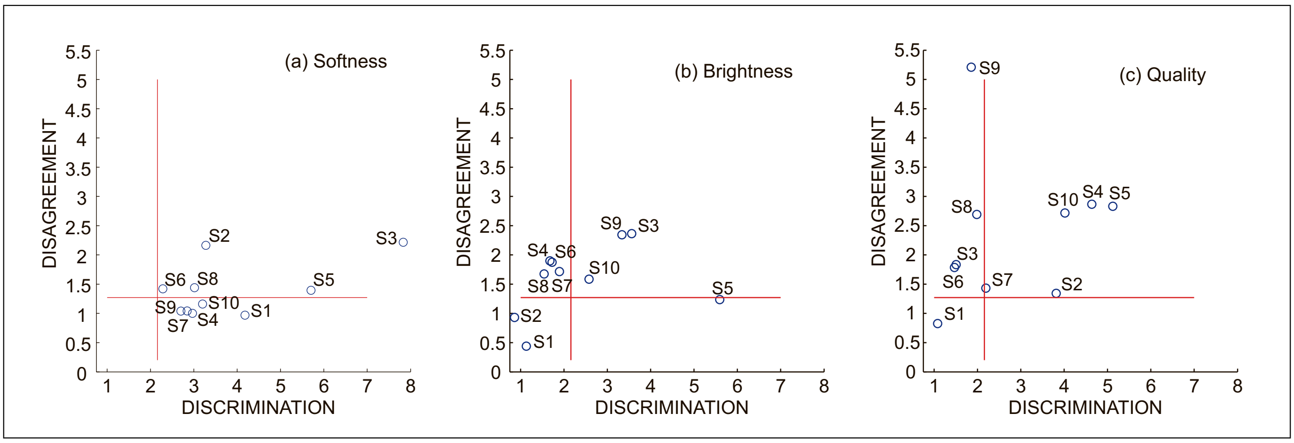


Figure 6. Plot of $\text{DISAGREEMENT}_j$ vs. $\text{DISCRIMINATION}_j$ for each assessor Sj and each descriptor.

For *Softness*, all the assessors are significantly discriminating. S2 and S3 disagree the most with the rest of the group. For *Brightness*, only S5, S3, S9 and S10 are discriminating, and S9 and S3 disagree the most with the group. For *Quality*, only S5, S4, S10, S2 and S7 are discriminating. S9 contributes a great deal to the disagreement. Furthermore, this disagreement is greater than for *Softness* and *Brightness*. This is not very surprising, given that *Quality* may express preferences of the musicians, which can be quite diverse.

These graphs are interesting to verify the quality of the individual assessments and to detect notable unreliability or misunderstanding in the ratings. In our panel, the assessors are much more reliable in the assessments of *Softness* than for *Brightness* and *Quality*. *Softness* is the most relevant for characterizing the reeds because all the assessors are discriminating and show the greatest agreement. Differences in *Brightness* are more difficult to assess by the panel (some assessors being non-discriminating), either because reeds are too similar or because assessors are not reliable enough. The disagreement between the assessors remains limited for *Brightness*, of the same order as *Softness*. This disagreement can be due to differences in the technique of the musicians (embouchure or amount of mouthpiece in the mouth, for instance).

*Quality* is also difficult to assess reliably, but a noticeable aspect is that the disagreement between the assessors for this descriptor is the highest. Important differences between the assessors in the quality of the reeds are reported, due to their individual preferences. Finally, given the results of the individual study, no assessor is discarded from the panel for *Softness* and *Brightness*. The high disagreement for *Quality* suggests that this descriptor should not be taken into consideration for the characterization of the reeds.

### 3.1.2. Global performance of the panel

*Agreement between the assessors* The agreement between the assessors in their evaluation of the reeds can be estimated by another method, consonance analysis; a method based on a principal component analysis (PCA) of the assessments. A description of this method can be found in [26]. Let us denote by $Y_k^m$ a matrix of size $(I \times J)$, of generic term $y_{ijk}^m$. To study the agreement between assessors for each descriptor $m$ (independent of the sessions), the two sessions are merged vertically to form the matrix $Y^m$ (Equation 26 – sessions are considered as different observations). A standardized PCA is performed on the matrix $Y^m$,

$$Y^m = \begin{bmatrix} Y_1^m \\ Y_2^m \end{bmatrix}. \tag{26}$$

The results of the PCA of the matrices $Y^m$ are given in Figure 7 for each descriptor. In this PCA, the variables are the assessors (S1 to S10), and the observations are the reeds. A perfectly consensual panel would consist of assessors who rate the reeds in the same way. In this case, the first component of the PCA would account for a very large variance. The more the panel is consensual, the more the arrows of the assessors point in the same direction. The percentage of the variance explained by the first principal component is considered as an indicator of the consonance of the panel (under the condition that the variable points are on the same side of the first component).

The highest agreement is obtained for the descriptor *Softness* (54.7% of the variance on the first component). The ratings of the assessors are the most convergent, and the agreement is the highest. For *Brightness* (29.3%), the agreement is weaker, even though no assessor is very discordant. For *Quality* (29.2%), assessors are even opposite on the first component, indicating that the agreement is the weakest. This is again not surprising, given that this descriptor may express the preferences of the saxophonist, which are in essence subjective and a function of the tastes of the musician. Assessors S1, S3, S9 are rather opposite to the rest of the panel, and assessor S8 is discordant with respect to the general trend of the group.

This analysis confirms the conclusions of the individual study obtained with the criterion DISAGREEMENT. For the descriptors *Softness* and *Brightness*, no particularly discordant assessor was identified (all the assessors are close according to DISAGREEMENT) and the descriptors are considered as consensual enough. For the de-

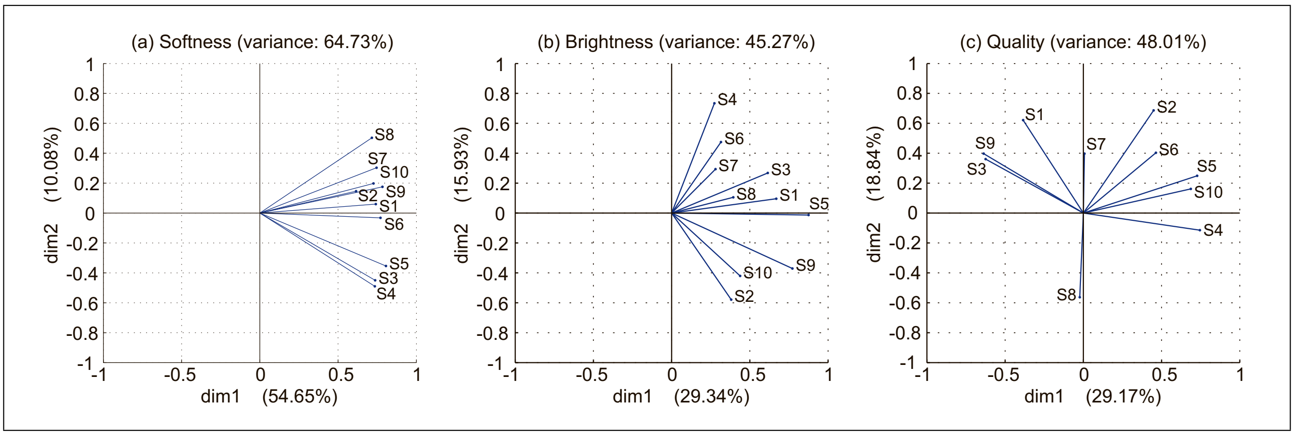


Figure 7. Consonance analysis for each descriptor: plot of the first two factors of the PCA (plane of the variables).

Table I. Results of two-way mixed model ANOVA for the three descriptors (Fisher test, eta-square $\eta^2$).

| Source | | *Softness* | *Brightness* | *Quality* |
|---|---|---|---|---|
| Reed (fixed) | F | F(19,171) = 15.64, p < 0.001 | F(19,171) = 4.74, p < 0.001 | F(19,171) = 1.48, p = 0.1 |
| | $\eta^2$ | 44.1% | 21.4% | 8.1% |
| Assessor (random) | F | F(9,171) = 4.23, p < 0.001 | F(9,171) = 0.78, p = 0.63 | F(9,171) = 6.23, p < 0.001 |
| | $\eta^2$ | 5.6% | 1.7% | 16.3% |
| Reed × assessor (random) | F | F(171,200) = 1.19, p = 0.11 | F(171,200) = 1.31, p = 0.032 | F(171,200) = 2.26, p < 0.001 |
| | $\eta^2$ | 25.3% | 40% | 49.8% |

scriptor *Quality*, the agreement is considered as not satisfying and a partitioning of the panel into more homogeneous subgroups should be made (see [27] for an analysis of the reeds according to the descriptor *Quality*). Additional analyses using another method, the eggshell plot [28] (not reported here), led to convergent conclusions.

*Performance of the panel* A general method to estimate the performance of a panel of assessors is Analysis of Variance (ANOVA). It is used in sensory analysis to study the differences between products and, more generally, to test the statistical significance of levels of qualitative factors [29]. The standard ANOVA model in sensory analysis is a two-way model, with product and assessor main effects together with a product × assessor interaction effect. To better generalize the results, the product effect is assumed to be fixed, whereas the assessor and interaction effects are random [29]. These random effects together with the fixed effect constitute the so-called mixed model ANOVA [30]. The assessment of product $i$ by assessor $j$ during session $k$ according to a descriptor being denoted $y_{ijk}$, the model (Equation 27) may be written as

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}, \qquad (27)$$

where $\mu$ is the intercept, $\alpha_i$ the product (reed) main effect (fixed) represents differences between the average score for the different reeds. A highly performing panel of assessors should get large product effects, if perceptual differences between products exist and the dependent variables characterize them well. $\beta_j$, the assessor main effect (random), represents differences in scoring levels between the assessors (use of scale). A trained and highly performing panel would lead to a non-significant assessor effect, but this condition is not imperative, because differences between assessors in the location on the scale are acceptable to get representative data. $\gamma_{ij}$, the assessor × product interaction (random), expresses differences between assessors in measuring differences between products. The interaction effect measures the lack of consensus, which can be the results of two effects: a scaling effect (differences between assessors in the magnitude of the differences between products) and a disagreement effect (disagreement in the ranking of the products) [31]. For the panel to be considered consensual the assessor × product interaction would have to be non-significant. This condition is important for a reliable interpretation of the assessments, because poor results can be obtained in interpreting the main effects when a high level of interaction is observed. $\epsilon_{ijk}$ is an error term, independent from observation to observation, $\epsilon_{ijk} \sim N(0, \sigma^2)$.

The results of the F-test with the model of equation (27) for the whole panel and each descriptor are given in Table I. Non-significant effects ($p > 0.05$) are depicted in grey. The effect size of each source of variation is assessed with the classical eta-square ($\eta^2$), the ratio between the variation (sum of square) attributable to the factor and the total variation.

The attributes *Softness* and *Brightness* show a significant reed effect ($p < 0.001$), whereas it is not significant for *Quality*. It signifies that the panel can discriminate the reeds for *Softness* and *Brightness* only. The average results for *Quality* are not adapted to discriminate the reeds.

The assessor effect is significant for *Softness* and *Quality* only, indicating differences in the location of the ratings on the scale by the assessors for these two descriptors. This result is confirmed by the plot of LOCATION in Figure 4, which shows the weakest differences among the assessors along the LOCATION axis for the descriptor *Brightness*. These differences represent level differences between assessors in the use of the scale and may be due to different calibrations of the assessors and their lack of training in the use of the scale. A training of the assessors (association of the magnitude of the sensation to the correct location on the scale) could solve this calibration problem. It is also important to mention that the size of these effects is small.

The interaction is significant for *Brightness* (p = 0.032) at the 5% level but not at the 1% level. A strong interaction is observed for *Quality* (p < 0.001), which confirms the lack of consensus in the panel for this descriptor. For *Softness*, the reed effect size dominates (44.1%), whereas the interaction effect size is the greatest for *Brightness* and *Quality*.

In conclusion, the assessments of the panel according to *Softness* are interesting to characterize the differences between the reeds: the assessments are considered as reliable, discriminating and consensual enough. For *Brightness*, the agreement between the assessors is weaker, but it has been considered as satisfying given that the reed effect is significant. For *Quality*, the assessments are not consensual enough to represent significant differences between the reeds. Individual analyses or clustering of assessors should be performed (see [27]). In the following sections of the study, only the *Softness* and *Brightness* descriptors will be considered to represent differences between the reeds (sensory profile).

### 3.1.3. Post Hoc analysis

After an overall assessment of the effect of the reeds with ANOVA, the following stage concerns the test of differences between pairs of reeds. For each reed, the mean value across the repetitions and the assessors are computed and represented in Figure 8 for *Softness* and in Figure 9 for *Brightness*, the reed being ranked in increasing order of value.

Significant differences between pairs of reeds are evaluated by a Duncan multiple comparison test. The Duncan groups (5% level) are represented in Figures 8 and 9 by horizontal lines connecting pairs of reeds: when pairs are connected by a line, the difference is not significant (e.g., for *Brightness* in Figure 9, R18 and R13 are not significantly different, whereas R18 and R11, not connected, are significantly different). Figures 8 and 9 show the differences between reeds that are significant for each attribute. The Duncan multiple comparison test enables discrimination between 9 (*Softness*) and 7 (*Brightness*) overlapping groups of reeds (Duncan groups).

The post-hoc test confirms that the discrimination between the reeds is better for *Softness* (9 groups) than for *Brightness* (7 groups). Although the reeds are very similar (same brand, strength, cut), the results show that the panel of musicians can significantly discriminate several groups of reeds, mainly for the *Softness* descriptor.

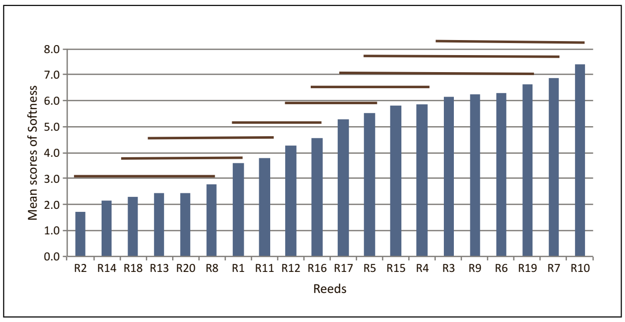


Figure 8. Mean value of *Softness* and Duncan groups (multiple comparison test, p = .05).

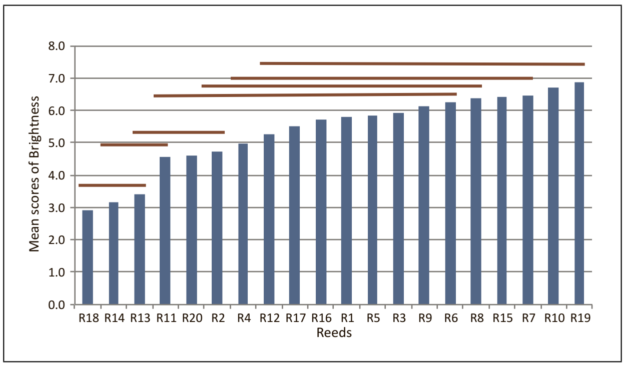


Figure 9. Mean value of *Brightness* and Duncan groups (multiple comparison test, p = .05).

### 3.1.4. Consensual configuration

The last stage of the perceptual study is to define a consensual configuration that describes the differences between the reeds and constitutes the sensory profiling. Several methods are proposed in sensory analysis to transform individual evaluations into an average multivariate description of products. The simplest method is to compute the average values of the ratings according to the sensory descriptors, for the ten assessors and the two sessions, denoted $y_{i\bullet\bullet}$. But this method must be used with care, the direct mean value of the assessments of all the assessors may lead to a poor description of the differences between products if the assessors are not in agreement (i.e., the mean value may be not representative). The sensory analyst is confronted with the dilemma of discarding dissonant assessors and losing information in this case, or leaving the data as such and getting a noisy assessment that is not representative.

In our experiment, the analysis of the performance of the panel showed that the agreement between the assessors was very weak for the descriptor *Quality*, with oppositions and dissident assessors. For this reason, this descriptor is excluded from the sensory profiling. The agreement for *Brightness* is better, with a significant reed × assessor interaction at the 5% but not at the 1% level. Furthermore, the results show that the disagreement is shared among all the assessors and not due to one or two outliers (Figures

Table II. Inter-session Spearman correlation coefficient and significance test for the 13 variables. Sessions A1–A2: $(x^m_{i11\bullet\bullet}–x^m_{i12\bullet\bullet})$, B1–B2: $(x^m_{i21\bullet\bullet}–x^m_{i22\bullet\bullet})$.

| Variable | *AtT* | *SC* | *OSC* | *ESC* | *OER* | *Lv* | *TR*1 | *TR*2 | *TR*3 | *TR*4 | *PTh* | *StP* | *Eff* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A1–A2 | −0.09 | 0.90 | 0.89 | 0.84 | 0.37 | 0.12 | 0.69 | 0.48 | 0.84 | 0.76 | 0.84 | 0.59 | 0.41 |
| B1–B2 | 0.53 | 0.20 | 0.25 | 0.37 | −0.34 | 0.23 | −0.12 | −0.19 | 0.30 | 0.10 | 0.75 | 0.09 | −0.15 |

6b and 7b). The assessments according to *Brightness* are considered satisfactory. For *Softness*, they are satisfactory given the significant reed effect and the non-significant reed × assessor interaction.

To characterize the reeds, the consensual configuration is simply the average value across the sessions and the assessors according to *Softness* and *Brightness*. To confirm the validity of this decision, we implemented three more sophisticated methods to compute consensual configurations: the STATIS method [32] and the GAMMA method [33], which weight the assessors according to their performance, and the Generalized Procrustean Analysis [34]. The results showed that the differences between the configurations obtained by these methods and the average configuration were weak (the average and maximum relative error was lower than 0.8% and 2%, respectively, given that the agreement between the musicians was high).

The sensory profile of the 20 reeds is finally a bidimensional representation, the average value of the assessments according to *Softness* and *Brightness*. The average position of the 20 reeds (R1 to R20) according to *Brightness* and *Softness* is given in Figure 10. The Kolmogorov-Smirnov normality tests showed that all the assessments followed a normal distribution for all the reeds. The 95% confidence intervals around the average position using the t-distribution are also given for information.

R10, R7, R19 are the most soft and bright reeds, R14, R18, R13 are the least soft and least bright reeds. There is also a correlation between the two descriptors *Brightness* and *Softness*: a bright reed is also generally soft (Pearson's correlation coefficient $r = 0.77$, $p < 0.01$). A noticeable result is that the brightness of hard reeds (low softness) has a greater variability (discrepancy with respect to the regression line) and larger confidence intervals than for soft reeds. The assessors disagree more on *Brightness* for "hard" reeds (softness under 5) than they do for soft reeds.

The average range of the assessments is larger for *Softness* (7.5 − 1.5 = 6) than for *Brightness* (6.8 − 2.8 = 4), showing that the average differences between the reeds are larger for *Softness* than for *Brightness*.

## 3.2. Analysis of the acoustical measurements

### 3.2.1. Individual results

The acoustical measurements consisted of J = 2 musicians (player A and B) who played I = 20 reeds during K = 2 sessions on L = 6 notes with N = 5 repetitions. A set of M = 13 variables was defined (described in section 2.3.2), the value of variable *m* of reed *i* by musician *j* during session *k*, note *l* and repetition *n* is denoted $x^m_{ijkln}$.

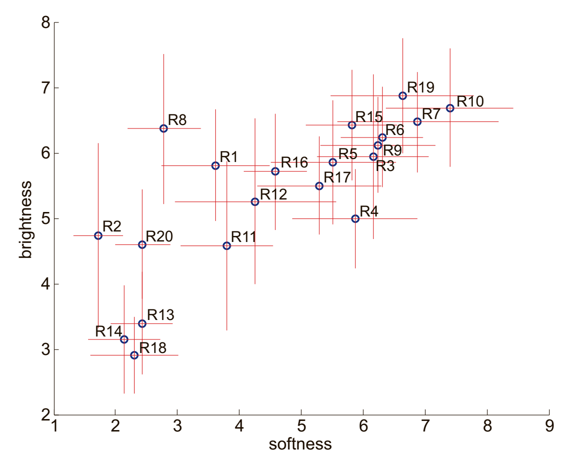


Figure 10. Position of the reeds according to *Softness* and *Brightness* (average configuration) and 95% confidence intervals around the average value using the t-distribution.

To assess the inter-session repeatability for each musician *j* and each variable *m*, the Spearman rank correlation coefficient between the average scores of session 1 $x^m_{ij1\bullet\bullet}$ and session 2 $x^m_{ij2\bullet\bullet}$ (averaged on note and repetition) was computed. The results are given in Table II. A test of the Spearman coefficient (with a Bonferroni correction for multiple comparisons) is carried out. Non-significant values of the coefficient (p-values higher than 0.05/13 = 0.0038) are depicted in grey.

The results show that the correlations between the variables for player B are generally low (even negative), except for the variable *PTh* (threshold pressure). These low correlations may be due to physical changes in the reeds' characteristics between the two sessions, the reeds having been played by all the participants of the perceptual study between the two sessions; but given that player A obtained higher correlations for several descriptors, we discarded this explanation and considered that the differences are due to a higher variability in the way of playing of player B between the two sessions: uncontrolled factors in musician B's playing may resulted in differences in the measurements between the two sessions. This explanation is strengthened by the fact that player A is a more skilled saxophonist than player B (considered as an amateur player), so we are more confident in the consistency of player A for a repeatable playing of the reeds. To avoid considering doubtful measurements, the data of player B are therefore discarded for the rest of the study. Only recordings from player A are used as the acoustical measurements to characterize the reeds.

Table III. F, $\eta^2$ and p-value of the Fisher test of the ANOVA (Equation 28).

| | | $AtT$ | $SC$ | $OSC$ | $ESC$ | $OER$ | $Lv$ | $TR1$ | $TR2$ | $TR3$ | $TR4$ | $PTh$ | $StP$ | $Eff$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Reed | F(19, 1174) | 3.2 | 63.2 | 53.0 | 59.0 | 4.5 | 13.8 | 7.4 | 3.7 | 19.5 | 18.4 | 63.3 | 48.2 | 32.9 |
| | p-value | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 |
| | $\eta^2$ | 3.9% | 11.0% | 8.1% | 15.8% | 1.4% | 1.8% | 1.6% | 1.1% | 5.7% | 10.1% | 31.9% | 14.2% | 7.8% |
| Session | F(1, 1174) | 264.4 | 548.3 | 540.9 | 333.4 | 0.0 | 2844.6 | 390.7 | 99.8 | 316.5 | 136.8 | 251.1 | 169.5 | 6247.5 |
| | p-value | <0.0001 | <0.0001 | <0.0001 | <0.0001 | 0.836 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 |
| | $\eta^2$ | 17.3% | 5.0% | 4.4% | 4.7% | 0.0% | 19.0% | 4.5% | 1.5% | 4.8% | 4.0% | 6.7% | 2.6% | 77.6% |
| Note | F(5, 1174) | 6.2 | 1596.7 | 1936.3 | 891.0 | 1006.0 | 2134.2 | 1383.7 | 1061.6 | 935.8 | 359.1 | 227.6 | 842.2 | 0.7 |
| | p-value | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | 0.639 |
| | $\eta^2$ | 2.0% | 73.2% | 78.1% | 62.9% | 80.0% | 71.4% | 80.2% | 79.8% | 71.5% | 52.0% | 30.2% | 65.1% | 0.0% |

To study the performance of player A for the acoustical measurements, we choose to fit an individual ANOVA model to the data for each variable $m$ labeled $x^m_{ikln}$ in equation (28) (the subscript "1" of musician A is dropped for clarity). This model takes into account the reed, the session and the note effect. $x^m_{ikln}$ is a generic notation that represents the value of the acoustical variable $m$ for the $i$th reed, the 1st musician (A), the $k$th session, the note $l$ and the $n$th repetition,

$$x_{ikln} = \mu + a_i + b_k + c_l + \epsilon_{ikln}, \qquad (28)$$

where $\mu$ is the intercept, $a_i$ the main effect of reed $i$, $b_k$ the main effect of session $k$, $c_l$ the main effect of note $l$, and $\epsilon_{ikln}$ is an error term, independent from observation to observation, $\epsilon_{ikln} \sim N(0, \sigma^2)$.

The results of the F-test of the ANOVAs are given in Table III. A Bonferroni correction for multiple comparisons is carried out. The effect size of each source of variation is assessed with the eta-squared ($\eta^2$), Non-significant values of the coefficient (p-values higher than 0.05/13 = 0.0038) are depicted in grey.

The results show that all the effects are significant, except "session" for $OER$ and "note" for $Eff$. The most interesting information concerns the effect sizes that are by far the most important for the factor "note" (around 70% for almost all the variables). This signifies that important differences between the played notes are observed, for all the variables except $AtT$ and $Eff$. The magnitude of the variables changes according to the played note. The "reed" effect is generally weak, except for the pressure threshold $PTh$. The session effect, even if significant, is not dominant except for $Eff$. Further investigations should be conducted to explain this important "session" effect of the variable $Eff$. Concerning $AtT$, the percentage of variance accounted for by the model (around 23% – our experiment being balanced, the sum of the eta-squareds for the three factors is equal to the determination coefficient $R^2$ of the model) is weak and interaction effects should be introduced. For the other variables, the percentage of variance is quite high, and it is unnecessary to introduce interaction effects. To summarize, it is therefore likely that the two variables $AtT$ and $Eff$ are useless in an explanatory model of the perceptual descriptors *Brightness* and *Softness*.

The session effect is due to three potential uncontrolled factors: variability of the musician in the way of playing, modification of the measurement chain, and changes of the reeds over time.

To investigate the differences between the sessions, a graphical representation of the reeds using Principal Component Analysis is provided. Let us denote by $X_k$ the matrix of size 20 × 13 of generic term $x^m_{ik\bullet\bullet}$ that represents the average scores (averaged on note and repetition) of reed $i$ and session $k$ for variable *m*. The two sessions are merged vertically to form the matrix $X$ (40 × 13) (Equation 29 – sessions are considered as different observations). A standardized PCA is performed on the matrix $X$,

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}. \qquad (29)$$

The results of the PCA of the matrix $X$ are given in Figure 11. The first two factors F1 and F2 account for more than 81% of variance: the 13 variables are therefore highly correlated. The positions of the reeds for the two sessions are noticeably separated in the plane, which illustrates the session effect noticed in the previous ANOVA. It is also interesting to mention that the relative position of the reeds inside the two sessions is rather similar. Additional studies are needed to investigate the cause of the offset in the measurements, which can be mainly due to modifications in the measurement conditions between the two sessions or changes in the reeds over time.

Figure 12 shows the plane of the variables of the PCA, with *Softness* and *Brightness* as supplementary variables. As expected, the variables $SC$, $OSC$, $ESC$, $TR3$ and $TR4$ are highly correlated, and opposite to $PTh$.

Given the large effect size of the note in the ANOVAs, two sets of acoustical variables are considered for the models: the values $x^m_{i\bullet\bullet\bullet}$ averaged over the sessions and the notes, and the values $x^m_{i\bullet l\bullet}$ averaged over the sessions only.

### 3.2.2. Choice of the acoustical variables for the model

The choice of the variables to include in an explanatory model between sensory data and instrumental data is not an easy task. A method based on a brute force search would be to test all the possible combinations of variables among the 13 candidates [35]. We consider that this strategy is beyond the scope of this paper. For the selection of the variables, an appropriate tradeoff between goodness of fit, generalizability, and stability of the results must be considered. Different strategies can be considered.

The first strategy is to consider only the variables that are similar enough between the two sessions of player A (significant correlation between the two sessions ($r \geq .6$, Table II). We exclude also $AtT$ (the $R^2$ of the ANOVA model Equation 28 is weak) and $Eff$ due to the very large session effect. Seven variables can be considered: $SC$, $OSC$, $ESC$, $TR1$, $TR3$, $TR4$, $PTh$.

The second strategy is to study the correlations between these variables and exclude the highly correlated variables. A Hierarchical Agglomerative Clustering (HAC) of the variables is made using the Pearson's similarity and the complete linkage aggregation rule [36]. Figure 13 shows the dendrograms for the case of the 7 variables $x^m_{i\bullet\bullet\bullet}$, and Figure 14 for the case of the 42 variables $x^m_{i\bullet l\bullet}$.

The results show that the variables $SC$, $OSC$, $ESC$ are highly correlated. It is thus unnecessary to include them in a model. From the dendrogram, four variables are finally considered, with a similarity threshold of 0.6: $PTh$, $TR1$, $SC$ and $TR3$.

The results show that for all the notes $l$, the variables $PTh\_l$, $TR1\_l$, $TR4\_l$ are highly correlated. $OSC$, $ESC$ and $SC$ are also highly correlated, except for note 6 ($SC\_6$, $OSC\_6$ and $ESC\_6$ group together later in the dendrogram).

From the dendrogram, five groups can be considered, with a similarity threshold of 0.4. The choice of the variables inside a group is somewhat arbitrary. Note 4 (in the middle of the tessitura of the saxophone) has been favored. Five variables are retained: $PTh\_4$, $TR1\_1$, $TR1\_6$, $SC\_4$, $TR3\_3$.

Two cases are finally considered to form the explanatory variables for the modeling:

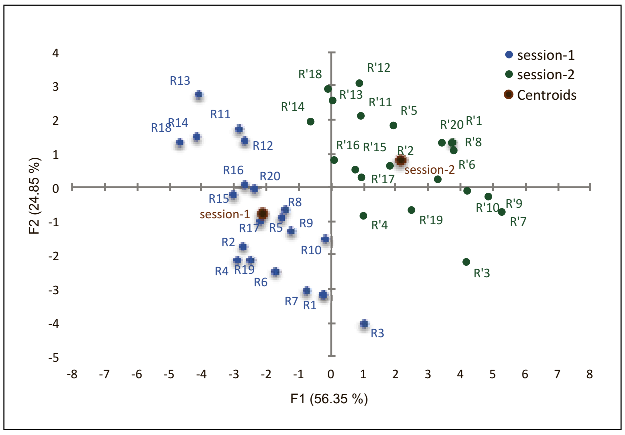


Figure 11. PCA of the reeds for the two sessions according to the 13 acoustical variables: plot of the first two factors of the PCA (plane of the observations).

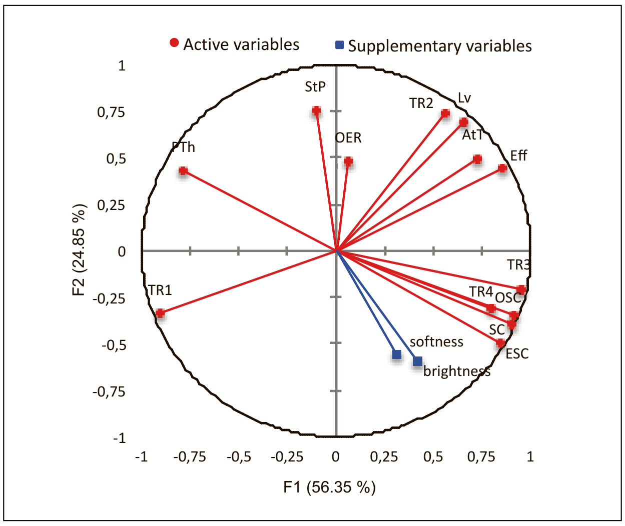


Figure 12. PCA of the reeds for the two sessions according to the 13 acoustical variables: plot of the first two factors of the PCA (plane of the variables).

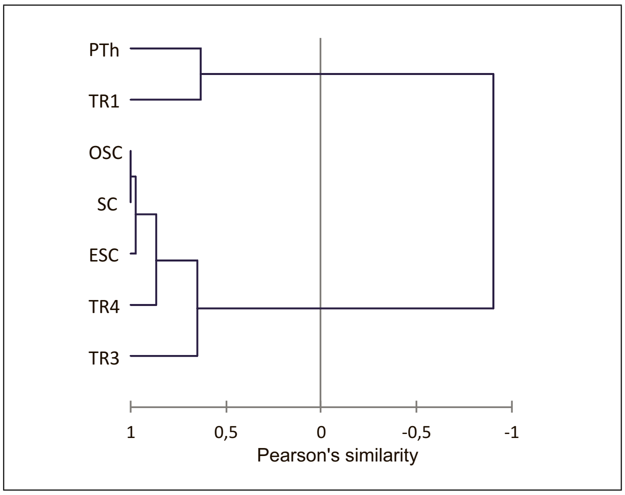


Figure 13. HAC of the 7 acoustical variables (averaged over the sessions and the notes) according to the Pearson's similarity.

Table IV. Correlation coefficients between the perceptual descriptors and the acoustical variables (musician A) and significance test.

| Variable | $AtT$ | $SC$ | $OSC$ | $ESC$ | $OER$ | $Lv$ | $TR1$ | $TR2$ | $TR3$ | $TR4$ | $PTh$ | $StP$ | $Eff$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Softness* | −0.03 | 0.54 | 0.50 | 0.58 | −0.44 | −0.70 | −0.27 | −0.29 | 0.57 | 0.33 | −0.73 | −0.73 | 0.28 |
| *Brightness* | −0.08 | 0.72 | 0.71 | 0.71 | −0.21 | −0.65 | −0.48 | −0.21 | 0.76 | 0.41 | −0.78 | −0.81 | 0.44 |

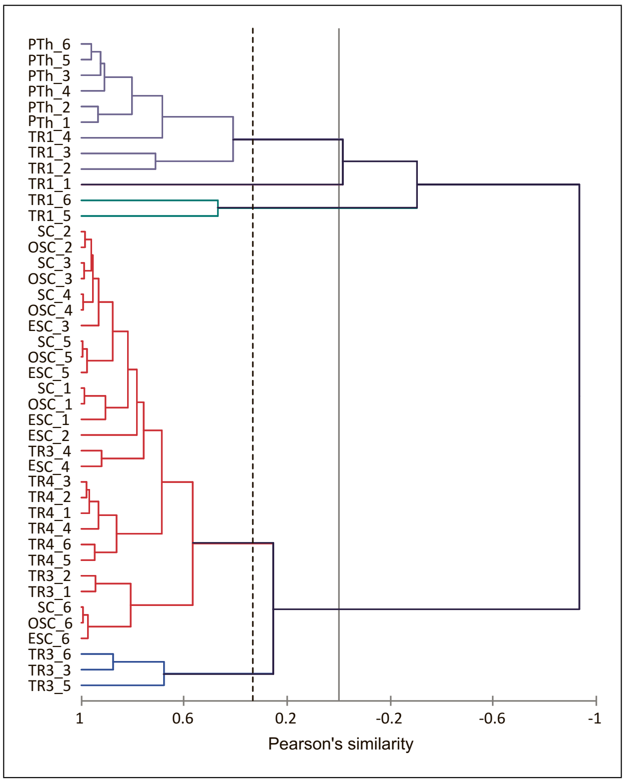


Figure 14. HAC of the 42 acoustical variables (averaged over the sessions) according to the Pearson's similarity.

- Data averaged across the sessions and the repetitions only, with the five variables $PTh$_4, $TR1$_1, $TR1$_6, $SC$_4, $TR3$_3. This approach is interesting to show whether particular notes have an important contribution in the model.
- Data averaged across sessions, notes and repetitions, with the four variables $PTh$, $TR1$, $SC$ and $TR3$.

## 4. Predictive Models of *Softness* and *Brightness*

### 4.1. One-to-One Correlation

A simple way to study the relationships between perceptual and acoustical variables is to compute the linear Pearson coefficient of correlation. In Table IV are presented the Pearson's correlation coefficients between the average value of the acoustical variable $x^{m}_{i\bullet\bullet}$ (averaged over sessions, notes, and repetitions) on the one hand, and the average values of the perceptual assessments of the reeds $y_{i\bullet\bullet}$ according to *Softness* and *Brightness* on the other hand. The Kolmogorov-Smirnov normality test showed that all the variables followed a normal distribution ($p >$ 0.05). Values of the Pearson coefficient of correlation non-significant at the $p = 0.05/13 = 0.0038$ level (Bonferroni correction for the multiple comparison problem) are depicted in grey.

The variable that are most correlated with *Softness* are the threshold pressure $PTh$ (−0.73) and the mean static pressure StP (−0.73). These negative correlations make sense from a physical point of view: a "soft" reed necessitates a low pressure and a "hard" reed a high pressure. The softer the reed, the lower the pressure in the mouth to trigger and maintain a note.

*Brightness* also has a strong correlation with the mean Static Pressure $StP$ (−0.81) and the threshold pressure $PTh$ (−0.78). This is reliable given that *Softness* and *Brightness* are correlated ($r = 0.77$, p<0.01). *Brightness* also presents strong correlations with timbral descriptors: the Tristimulus 3 $TR3$ (0.76), the Spectral Centroid $SC$ (0.72), the Odd Spectral Centroid $OSC$ (0.71) and the Even Spectral Centroid $ESC$ (0.71). These correlations make sense from a physical point of view: a reed with a high "brightness" score will produce a sound with a higher Spectral Centroid than a reed with a low "brightness" score, which is in agreement with the literature [15, 25].

### 4.2. Multiple Linear Regression Models

Linear regressions are classical techniques to explain the behavior of a dependent variable (here *Softness* or *Brightness*) based on the behaviors of a set of explanatory variables (here the different acoustical variables).

Two multiple linear regressions (MLR) are fitted to the data for each descriptor *Softness* and *Brightness*, using the two sets of explanatory variables described in section 3.2.2. An optimization of the model (choice of the variables in the set) according to the adjusted $R^2$ is carried out.

In addition to the MLRs, two simple linear regressions (LR) were considered to allow a comparison of the results: for *Softness*, the chosen regressor was the threshold pressure ($PTh$) (due to its highest correlation with *Softness*); for *Brightness*, the regressor was the spectral centroid ($SC$), given the ability of this descriptor to explain the brightness in the literature [15].

To assess the quality of the models, and define the optimal one, five classical criteria were used:

- the root mean squared error RMSE between the predictions by the model and the observations, estimating the goodness of fit of the model,
- the Root Mean PRESS (square root of the mean of the predicted residual error sum of squares). This metric estimates the generalizability of the models, by comput-

Table V. Values of RMSE, Root Mean PRESS, AIC, BIC and predicted $R^2$ for the different predictive models, for each descriptor.

| Descriptor | Model | Variables | RMSE | Root Mean PRESS | AIC | BIC | Predicted $R^2$ |
|---|---|---|---|---|---|---|---|
| *Softness* | $MLR\ so_1$ | 5 variables | 1.00 | 1.48 | 12.16 | 18.30 | 0.31 |
| | $MLR\ so_2$ | 4 variables | 1.09 | 1.35 | 11.45 | 15.43 | 0.43 |
| | $LR\ so_3$ | *PTh* | 1.21 | 1.30 | 11.62 | 13.62 | 0.47 |
| *Brightness* | $MLR\ b_1$ | 5 variables | 0.70 | 0.92 | −6.27 | −2.28 | 0.35 |
| | $MLR\ b_2$ | 4 variables | 0.69 | 0.85 | −6.84 | −2.86 | 0.45 |
| | $LR\ b_3$ | *SC* | 0.80 | 0.93 | −4.93 | −2.93 | 0.34 |

ing the RMSE with a cross validation (CV) procedure (LOOCV – Leave-one-out cross validation),

- the Akaike Information Criterion AIC, a predictive criterion based on a tradeoff between accuracy and parsimony,
- the Bayesian Information Criterion (BIC), an explicative criterion based on a tradeoff between accuracy and parsimony, but which also controls for the number of observations,
- The Predicted $R^2$ (based on PRESS), a measure that indicates how well the model predicts responses for new observations.

The best model (if it exists) should obtain a minimum value for the first four criteria and a maximum value for the last criterion.

### 4.3. Choice of the optimal model

The results of the different models (labeled $so_i$ for softness, $b_i$ for brightness, $i = 1$ to 2) are presented in Table V. They are compared with the results of a simple linear regression (LR) using only one acoustical variable: Pressure Threshold *PTh* for softness (model $so_3$), and the Spectral Centroid *SC* for *Brightness* (model $b_3$).

A model must be selected based on a tradeoff between goodness of fit and generalizability. For *Softness*, the LR $so_3$ model (simple linear regression with *PTh*) obtains the best performance on Root Mean PRESS, BIC and predicted $R^2$. For these reasons, the chosen model for *Softness* is therefore $so_3$.

For *Brightness*, the model MLR $b_2$ is the best according to all the criteria except BIC. The chosen model for *Brightness* is therefore $b_2$.

The two chosen models show a reasonable fit to the data: the RMSE is around 1 (1.21 for $so_3$, 0.69 for $b_2$, which gives an average relative error of around 10% given that the assessment of softness and brightness was specified on a scale from 0 to 10). The generalizability, given by the Root mean PRESS, shows that the average prediction error is on the order of 13.9% for softness and 9% for brightness.

Figures 15 (model $so_3$) and 16 (model $b_2$) show the magnitudes of the variables in the models (standardized coefficients).

For *Softness* (Figure 15), the pressure threshold *PTh* has a negative effect on *Softness* – the lower the pressure, the softer the reed – which conforms to the physical sense and the general opinion of musicians concerning soft reeds.

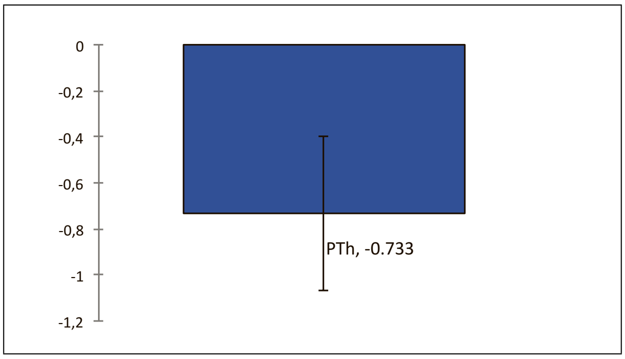


Figure 15. Standardized coefficient and confidence intervals (95%) of the variable *PTh* in model $so_3$ for *Softness*.

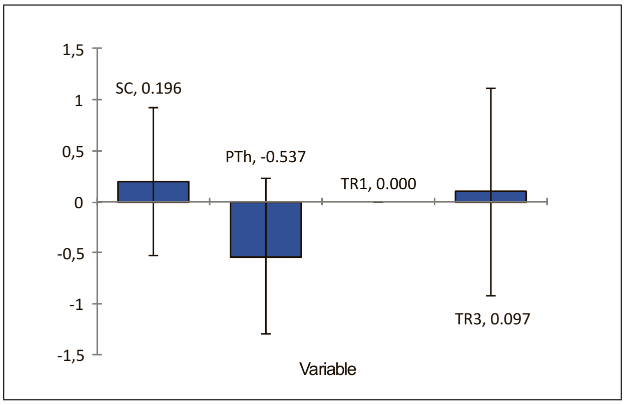


Figure 16. Standardized coefficients confidence intervals (95%) of the three variables for the model $b_2$ (*Brightness*).

For *Brightness* (Figure 16), the threshold pressure *PTh* and the spectral centroid are the most important variables: a bright reed has a high *SC* and low values for *PTh*. These conclusions need to be confirmed with additional reeds, the confidence intervals for the coefficients being large.

For both models, the importance of the variable related to the pressure controls (*PTh*) is higher than that of the variables related to the acoustic signal (*SC*, *TR*3). The variable directly controlled by the musician (threshold pressure *PTh*) has a greater effect on *Softness* and *Brightness* and is of prime importance in explaining them. Finally, the models make sense from a physical point of view. The higher *PTh* and the lower SC, the harder and less bright the reed, which conforms to the physical sense for saxophone playing. This could suggest which *in vitro*

Table VI. Partitioning of the reeds in three categories according to the average score of *Softness*.

| | hard | medium | soft |
|---|---|---|---|
| Reed label | R2, R14, R20, R13, R18, R8 | R1, R11, R12, R16, R17 | R5, R15, R4, R3, R9, R6, R19, R7, R10 |

measurements to use in a test bench in making an objective estimate of the perceived quality of reeds.

### 4.4. Results and discussion

The PRESS is interesting for comparing models, but it does not give a clear indication of the quality of the model from an operational point of view. To illustrate the results for reed makers and show how the models can predict the perceptual quality of a reed, a qualitative criterion was defined based on different categories of reeds.

We consider first only the descriptor *Softness*, which is the most discriminating. A Hierarchical Agglomerative Clustering (HAC) of the reeds was made according to softness, using the Euclidian distance and Ward's method as aggregation rule and an automatic truncation according to entropy [36]. Three classes of reeds were formed: hard, medium, soft (relative to reeds of strength 2.5). The partitioning of the reeds is given in Table VI.

For each reed, a score of softness is predicted with MLR model $so_3$, with a Leave-One-Out procedure (the model is trained on all the samples except one; then the model predicts the score of the withdrawn sample, this operation is performed N times for each sample). With this score, the reed is next assigned to the class whose center point is the closest (classification rule). The confusion matrix of the classification is given in Table VII. Note that the prediction error always occurs between adjacent categories.

Different performance measures of the classifier can be proposed to cover different aspects of a classification [37]. We consider, for each class, the *precision* (fraction of reeds correctly predicted in a class to the number of reeds of the class, Equation 30), the *recall* (fraction of reeds correctly predicted in a class to the number of reeds predicted in that class, Equation 31), the *F_measure* (harmonic mean of precision and recall, Equation 32). In addition, the global performance of the classifier is characterized by the average values over the three classes of *precision*, *recall*, and *F_measure*, and also the Correct Classification Rate (*CCR*, rate of reeds assigned to the correct classes by the classifier, Equation 33).

$$precision_{\text{class}} = \frac{\text{\# of reed correctly predicted in a class}}{\text{\# of reed in a class}}, \quad (30)$$

$$recall_{\text{class}} = \frac{\text{\# of reed correctly predicted in a class}}{\text{\# of reed predicted in that class}}, \quad (31)$$

$$F\_measure_{\text{class}} = \frac{2 \cdot precision_{\text{class}} \cdot recall_{\text{class}}}{precision_{\text{class}} + recall_{\text{class}}}, \quad (32)$$

$$CCR_{\text{soft}} = \frac{\text{\# of reed correctly predicted in the class}}{\text{total number of reed}}. \quad (33)$$

According to these measures, a perfect classifier would obtain a value of 1 for *precision*, *recall*, *F_measure* and *CCR*.

Table VII. Confusion matrix for the prediction of the "softness classes" by the MLR model $so_3$ and performance measures of the classifier.

| | | | Predicted | |
|---|---|---|---|---|
| | | Hard | Medium | Soft |
| | Hard | 3 | 3 | 0 |
| Observed | Medium | 0 | 5 | 0 |
| | Soft | 0 | 3 | 6 |
| | *precision* | 1 | 0.45 | 1 |
| | *recall* | 0.5 | 1 | 0.66 |
| | *F_measure* | 0.66 | 0.62 | 0.8 |
| | *Avg_precision* | | 0.81 | |
| | *Avg_recall* | | 0.72 | |
| | *Avg_F_measure* | | 0.69 | |
| | $CCR_{\text{soft}}$ | | 14/20 = 70% | |

The results show that the average *F_measure* of the classifier is 0.69, close to the $CCR_{\text{soft}}$, equal to 70%. It signifies that the model has 70 out of 100 chances to predict correctly the softness category. Performances in the classes are a little unbalanced, the *F_measure* in the "medium class" (0.62) being the weakest, compared to the performance in the "soft" class (0.8). The classifier performs better in predicting "extreme" reeds (soft or hard) than medium ones. The average performance of the model is far above a "random" *CCR* of 33%, corresponding to a random assignment of a reed to a category.

This classification of reeds from *in vivo* measurements with a rate of 70% is interesting for researchers working on reeds, who would like to rapidly obtain reed categories without conducting a complex and time consuming perceptual test with a panel of musicians. The study shows that with the models, the playing of the reeds from the same box by the player A can produce a typology of the reeds in three categories with a 70% correct classification rate.

This result is also an encouraging sign for the automatic classification of reeds for a reed manufacturer. It emphasizes the importance of the threshold pressure *PTh* in the perceived qualities (*Softness* or *Brightness*). The tester of the company could serve as the reference musician (as player A in our study) to develop the process. An automatic test bench could be developed by reed manufacturers to objectify the qualities of reeds, beyond the strength number based on the static stiffness.

The results of our study help define a test bench for a reed manufacturer. Previous studies using physical modeling of saxophone or clarinet playing have shown with linear stability analysis that the theoretical threshold pressure is proportional to the reed equivalent stiffness (using analytical models describing woodwind instruments) [38]. In our study, we showed that the *Softness* can be explained by the threshold pressure *PTh* in the mouth of the musician. Therefore, mechanical measurements of the stiffness (static or dynamic) should be investigated to understand softness differences, as perceived by musicians.

Our study agrees with the results presented in [16]: the ease of playing estimated by one expert clarinet player is correlated with the reed stiffness measured in a static and dynamic way for many reeds. The use of a panel of musicians allows a better generalization of the perceptual dimension of the reed.

## 5. Conclusions

This paper presented a combined perceptual and acoustical study of a set of 20 saxophone reeds. Three descriptors were assessed during the perceptual study by ten musicians: *Softness*, *Brightness* and *Global Quality*. Acoustical *in vivo* measurements were performed during saxophone playing and 13 acoustical variables were extracted from these measurements. Different models, based on multiple linear regression, were tested to explain the descriptors *Softness* and *Brightness* by the acoustical variables. For each descriptor, two optimal models were selected based on a tradeoff between goodness of fit and generalizability. These models were next used to predict the reed quality according to three categories (hard, medium, soft).

The results show first that even on a set of very similar reeds (from four boxes of the same strength), the panel of ten musicians was able, with our experimental protocol, to make a discrimination between the reeds, to provide reliable assessments, and to agree on their assessments for the descriptor *Softness*. For *Brightness*, the agreement between assessors was lower even though reeds were clearly discriminated. For *Global Quality*, the agreement was low, which may be due to differences in tastes and habits of the musicians.

Second, the results show that the multiple linear regression models have interesting prediction qualities and allow a determination of the most important variables in defining the perceived *Softness* and *Brightness*: the threshold pressure PTh and the spectral centroid SC. A Correct Classification Rate of 70% was obtained in cross validation.

The paper presented a rigorous experimental protocol for the perceptual assessment of reeds that can be used by researchers to set up different acoustical measurements (e.g., frequency response of reeds). A reed manufacturer could also implement a similar methodology to explain quality models depending on customers' preferences. After a study on a large number of saxophone players, different customer profiles could be defined and then characterized according to acoustical measurements. Future work will consist of developing *in vitro* measurements (for example by the use of artificial mouths), leading to an objectification of the perceived quality of reeds.

**Acknowledgements**

The authors would like to thank the ten musicians of the Schulich School of Music, McGill University, Montreal for their participation in the perceptual tests, to Yann Grenier for his contribution as well as to Philippe Courcoux from ONIRIS, Nantes, for his advice on statistics.

## References

[1] R. L. Pratt, J. M. Bowsher: The subjective assessment of trombone quality. Journal of Sound and Vibration **57** (1978) 425–435.

[2] R. L. Pratt, J. M. Bowsher: The objective assessment of trombone quality. Journal of Sound and Vibration **65** (1979) 521–547.

[3] G. R. Plitnik, B. A. Lawson: An investigation of correlations between geometry, acoustic variables, and psychoacoustic parameters for French horn mouthpieces. J. Acoust. Soc. Am. **106** (1999) 1111–1125.

[4] C. Saitis, B. L. Giordano, C. Fritz, G. P. Scavone: Perceptual evaluation of violins: A quantitative analysis of preference judgments by experienced players. The Journal of the Acoustical Society of America **132** (2012) 4002–4012.

[5] C. Fritz, D. Dubois: Perceptual evaluation of musical instruments: State of the art and methodology. Acta Acustica united with Acustica **101** (2015) 369–381.

[6] M. C. King, J. Hall, M. A. Cliff: A comparison of methods for evaluating the performance of a trained sensory panel. Journal of Sensory Studies **16** (2001) 567–582.

[7] N. Zacharov, G. Lorho: What are the requirements of a listening panel for evaluating spatial audio quality? Proceedings of the Spatial Audio and Sensory Evaluation, Techniques Workshop, University of Surrey, UK, 2006, iosr.surrey.ac.uk/projects/ias/papers/Zacharov_Lorho.pdf.

[8] F. Pinard, B. Laine, H. Vach: Musical quality assessment of clarinet reeds using optical holography. The Journal of the Acoustical Society of America **113** (2003) 1736.

[9] S. Glave, J. Pallon, C. Bornman, B. L. Olof, R. Wallen, J. Rastam, P. Kristiansson, M. Elfman, K. Malmqvist: Quality indicators for woodwind reed material. Nuclear Instruments and Methods in Physics Research B **150** (1999) 673–678.

[10] E. Obataya, M. Norimoto: Acoustic properties of a reed (Arundo donax L.) used for the vibrating plate of a clarinet. The Journal of the Acoustical Society of America **106** (1999) 1106–1110.

[11] P. A. Taillard, F. Laloë, M. Gross, J.-P. Dalmont, J. Kergomard: Statistical estimation of mechanical parameters of clarinet reeds using experimental and numerical approaches. Acta Acustica united with Acustica **100** (2014) 555–573.

[12] B. Gazengel, J.-P. Dalmont: Mechanical response characterization of saxophone reeds. Proceedings of Forum Acusticum, Aalborg, June-July 2011.

[13] B. Gazengel, J.-F. Petiot, E. Brasseur: Vers la définition d'indicateurs de qualité d'anches de saxophone. Proceedings of 10ème Congrès Français d'Acoustique, Lyon, April 2010.

[14] B. Gazengel, J.-F. Petiot, M. Soltes: Objective and subjective characterization of saxophone reeds. Proceedings of Acoustics 2012, Nantes, April 2012.

[15] S. McAdams, S. Winsberg, S. Donnadieu, D. G. J. Krimphoff: Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. Psychol. Res. **58** (1995) 177–192.

[16] B. Gazengel, J.-P. Dalmont, J.-F. Petiot: Link between objective and subjective characterizations of bb clarinet reeds. Applied Acoustics **106** (2016) 155–166.

[17] J.-F. Petiot, P. Kersaudy, G. Scavone, S. McAdams, B. Gazengel: Modeling of the subjective quality of saxophone reeds. Proceedings of ICA 2013, Montreal, Quebec, Canada, June 2013.

[18] A. Nykänen, O. Johansson, J. Lundberg, J. Berg: Modelling perceptual dimensions of saxophone sounds. Acta Acustica united with Acustica **95** (2009,) 539–549.

[19] R. Weber: The continuous loudness judgement of temporally variable sounds with an “analog” category procedure. – In: Fifth Oldenburg Symposium on Psychological Acoustics. A. Schick, J. Hellbrück, R. Weber (eds.). BIS, Oldenburg, 1991, 267–294.

[20] S. Droit-Volet, W. Meck, T. Penney: Sensory modality and time perception in children and adults. Behavioural Processes **74** (2007) 244–250.

[21] I. N. Wakeling, A. Hasted, D. Buck: Cyclic presentation order designs for consumer research. Food Quality and Preference **12** (2001) 39–46.

[22] ISO: Sensory analysis - General guidance for the selection, training and monitoring of assessors - part 1: Selected assessors 8586-1. AFNOR, La defense, Paris, 1993.

[23] P. Schlich: GRAPES: A method and a SAS program for graphical representation of assessor performances. Journal of sensory studies **9** (1994) 157–169.

[24] V. R. Hasbrook: Alto saxophone mouthpiece pitch and its relation to jazz and classical tone qualities. Doctoral project, University of Urbana-Champaign, 2005.

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

[26] G. Dijksterhuis: Assessing panel consonance. Food Quality and Preference **6** (1995) 7–14.

[27] J.-F. Petiot, P. Kersaudy, G. Scavone, S. MacAdams: Study of the perceived quality of saxophone reeds by a panel of musicians. Proceedings of SMAC 2013, Stockholm, August 2013.

[28] D. Hirst, T. Næs: A graphical technique for assessing differences among a set of rankings. Journal of Chemiometrics **8** (1994) 81–93.

[29] T. Næs, P. B. Brockhoff, O. Tomic: Statistics for sensory and consumer science. John Wiley and Sons Ltd., New York, 2010.

[30] A. I. Khuri, T. Mathew, B. K. Sinha: Statistical tests for mixed linear models. John Wiley, New York, 1998.

[31] P. B. Brockhoff: Statistical testing of individual differences in sensory profiling. Food Quality and Preference **14** (2003) 425–434.

[32] C. Lavit: Analyse conjointe de tableaux quantitatifs. Masson, Paris, 1976.

[33] S. Ledauphin, M. Hanafi, E. Qannari: Assessment of the agreement among the subjects in fixed vocabulary profiling. Food Quality and Preference **17** (2006) 277–280.

[34] G. Dijksterhuis, J. Gower: The interpretation of generalised procrustes analysis and allied methods. Food Quality and Preference **3** (1991/1992) 67–87.

[35] J. Francombe, R. Mason, M. Dewhirst, S. Bech: Modelling listener distraction resulting from audio-on-audio interference. Proceedings of ICA 2013, Montreal, Quebec, Canada, June 2013.

[36] L. Kaufman, P. J. Rousseeuw: Finding groups in data. An introduction to cluster analysis. Wiley, Hoboken, New Jersey, 1990, 2005.

[37] M. Sokolova, G. Lapalme: A systematic analysis of performance measures for classification tasks. Information Processing and Management **45** (2009) 427–437.

[38] J.-P. Dalmont, J. Gilbert, J. Kergomard: Reed instruments, from small to large amplitude periodic oscillations and the Helmholtz motion analogy. Acustica united with Acta Acustica **86** (2000) 671–684.