The effect of timbre in clarinet interpretation

Mathieu Barthet, Richard Kronland-Martinet, Solvi Ystad, Philippe Depalle

THE EFFECT OF TIMBRE IN CLARINET INTERPRETATION

Mathieu Barthet[1], *Richard Kronland-Martinet*[1], *Sølvi Ystad*[1], *Philippe Depalle*[1,2]
[1] CNRS - Laboratoire de Mécanique et d'Acoustique, Marseille, France
[2] McGill University, Montréal, Qc, Canada

ABSTRACT

The perceptual importance of timbre variations was investigated in clarinet expressive music performance. Three basic transformations acting on timbre, rhythm and dynamics and four combinations of them were applied to solo clarinet recordings in order to remove or flatten some of the expressive variations of the performer. Twenty skilled musicians were asked to choose the interpretations they preferred in a pair comparison task. The rankings of the performances were strongly consistent and this, for two different musical excerpts coming from Bach and Mozart pieces. Multidimensional scaling showed that the most prominent factor used by listeners was linked to the timbre of the tones.

1. INTRODUCTION

Musical interpretation is an intricate concept which is hard to describe from a scientific point of view. Many authors from various domains such as musicology, psychology and acoustics worked on the description and understanding of the phenomena linked to expressive music performance either focusing on the control of the instrument or the sound that it produces [1]. Differences between performers can be characterized by small or large deviations of parameters related to musical expression including timing, dynamics, timbre and pitch [2]. A performer can reproduce expressive deviations of timing and dynamics in a very faithful way when repeating performances of a given piece of music with the same musical intention [1] [2]. Timbre expressive patterns can also present a high within-individual consistency during repeated performances [3] [4]. These results tend to prove that expressive deviations are not random but due to the interpretive intentions of the performer. Among expression parameters, timing (e.g. tempo, note durations, asychrony) and dynamics have shown to be essential when evaluating the quality of a performance [5] [6]. Fewer studies have been effectuated on the influence of timbre variations in music performance, probably due to the difficulties in defining and analyzing this parameter. In this paper, we investigate the relative importance of timbre, rhythm, and dynamics expressive patterns in performance evaluation. To assess this issue our approach was based on analysis-synthesis giving us the possibility to alter just one parameter or several simultaneously. Perceptual tests were carried out on 20 skilled musicians to analyze the effect of the transformations by means of preference jugments. We chose to focus on the classical music repertoire and on a specific instrument, the clarinet, which acoustic properties allow a great control of the timbre during the sound production.

2. METHODOLOGY

2.1. Sound corpus

We asked a professional clarinettist to perform two excerpts of standard classical musical pieces several times with the possibility of slightly modifying his interpretation. The first excerpt comes from Bach's Suite II for Cello and the second one from Mozart Clarinet Quintet in A major (K. 581). Both excerpts were played at a rather slow tempo (respectively 48 and 44 bpm) giving the performer a great scope to play expressively. The musical sequences were recorded in an anechoic chamber.

2.2. Analysis-Synthesis model

An additive synthesis approach has been adopted to resynthetize the original recorded sequences with or without alterations. It is particularly suitable for such a task since sounds are reconstituted as a superposition of a number of frequency components each of which can be controlled separately in amplitude and frequency. The tones of the original sequences are first segmented (see [4] for more details) and then the analysis synthesis process is applied to each of the separated tones. The resynthetized sequences are obtained by the juxtaposition of the resynthetized tones. Our sound model decomposes a sound into a deterministic part, constituted by a sum of quasi-sinusoidal components, plus a noise part. Thus, the original tone $s(t)$ can be written as:

$$\begin{aligned} s(t) &= \sum_{h=1}^{H} A_h(t) cos[\Phi_h(t)] + b(t) \\ \Phi_h(t) &= 2\pi \int_0^t f_h(t)dt + \Phi_h(0) \end{aligned} \tag{1}$$

where $A_h(t)$, $\Phi_h(t)$, $\Phi_h(0)$ and $f_h(t)$ are respectively the instantaneous amplitude, phase, initial phase and frequency of the h^{th} among H sinusoids, and $b(t)$ is the time-varying noise component.

A bandpass filter bank analysis in the frequency domain let us compute the analytic signals associated to the tone's frequency components. The frequencies of the filters are adapted to match the frequencies of the tones' components (which follow an harmonic series in the case of the clarinet). The instantaneous amplitude and phase of the tones' components are then derived from the analytic signals. As the deterministic part preserves the phases of the original sound, the noise (or residual) part is obtained by a time-domain substraction between the original and the synthesis of the deterministic part. Hence, when no transformations are carried out on the control parameters of the model (amplitude, phase) the original and resynthetized sounds are identical.

2.3. Transformations

Three basic transformations have been defined to modify independently the rhythm, dynamics or timbre expressive variations. Hence, we are able to cancel the performer's rhythmic deviations from the score, soften the original dynamics variations or cancel some timbre variations. In order to avoid the influence of the intonation as an interpretation criterion, we decided to fix the fundamental frequencies of the tones to their mean value $\overline{f0}$ during the sustained part. The instantaneous frequencies of the h tones' components were fixed to $h * \overline{f0}$. To avoid non-ecologic changes in the signal, we kept the original attack and release of the tones.

The segmentation process mentioned above let us obtain the interonset interval namely the duration from the beginning of a tone to the beginning of the next tone. The rhythm transformation T_R cancels the performer's rhythmic deviations from the score, by fixing the durations of the tones to durations corresponding to a mathematically accurate rendering of the musical score. This transformation is obtained from time-scale modifications on the instantaneous frequencies and amplitudes of the tone's components.

A compressor/limiter has been implemented in order to flatten the dynamics variations of the original musical sequences without altering the timbre (T_D). The compression was followed by a loudness equalization.

As the Spectral Centroid of a tone proved to be one of the most important perceptual dimension of timbre, we wanted to test the impact of its alteration on the perception of musical interpretation. To eliminate the Spectral Centroid variations during the course of a tone we used the transformation defined in [7] in the case of isolated tones which consists in eliminating the spectral flux without changing the RMS envelope. As a matter of fact eliminating the changes in shape of the spectral envelope over time induces the Spectral Centroid to be fixed to a constant value. The Timbre Transformation (T_T) can be written as follows:

Stimuli	*Transformation description*
M_0	No transformations
M_R	"Mechanical" rhythm
M_D	Dynamics flattening
M_T	Spectral Centroid freezing
M_{RD}	Combination of T_R and T_D
M_{TR}	Combination of T_T and T_R
M_{TD}	Combination of T_T and T_D
M_{TRD}	Combination of T_T, T_R and T_D

Table 1. Description of the stimuli

$$T_T : A_h(t) \mapsto A'_h(t) = \frac{\widetilde{A_h} E_{rms}(t)}{\sqrt{\sum_{h=1}^{H} \widetilde{A_h}^2}} \quad (2)$$

where $A'_h(t)$ is the new instantaneous amplitude of the h^{th} tone's component, $\widetilde{A_h}$ is the mean Spectral Centroid calculated within the sustained part of the tone, $E_{rms}(t)$ is the RMS envelope of $s(t)$.

2.4. Perceptual test

From the Bach and Mozart recording series, we retained the most expressive performances. We applied the three basic transformations and their four combinations to produce the performances listed in table 1.

A preference judgment test by paired comparisons [8] was carried out by $k = 20$ participants (judges). Rather than asking a group of judges to rank the set of musical sequences, we decided to present them by pairs and asked them which musical interpretation they preferred. The different pairs were the $\frac{N(N-1)}{2} = 28$ combinations of the $N = 8$ stimuli. In this way, each stimulus was forced to be compared to the others which could not have been the case in a ranking test. Due to the nature of the task and the sometimes subtle differences between the stimuli, the participants were all skilled musicians and represented a large variety of instrumentalists (clarinetist, guitarist, pianist, violinist, etc.). In order to assess the influence of the musical excerpt, each participant passed two sessions one using the sound corpus derived from the Bach sequences, and the other the one derived from the Mozart sequences. The order of the sessions was determined randomly. A training stage was first carried out so that participants got used to the task and the computer interface. The simuli used for the training were not the same as the ones used in the experience. All pairs of different stimuli were presented in a random order. The designation of the first and second stimuli within a pair was also random. Participants could choose to listen to the sequences of each pairs as many times as they wished. At the end of the test, participants had to answer a questionnaire asking them which strategies they used to make their choices.

Test session	*Nonparametric statistics*		
	u	X^2	dof
Bach	0.58	338.20	28
Mozart	0.52	303.80	28

Table 2. Agreement among raters

3. RESULTS AND DISCUSSION

3.1. Perceptual data

The perceptual data can be represented by k individual preference matrices P_s defined for each judge s. The elements of P_s, noted $a_s(i,j)$, design whether the stimulus i has been preferred to the stimulus j :

$$\forall (i,j) \in [1;N]^2_{\mathbb{N}}\,, i \neq j\,, \tag{3}$$

$$a_s(i,j) = \begin{cases} 1 & \text{if the judge } s \text{ preferred the stimuli } i \\ 0 & \text{if the judge } s \text{ preferred the stimuli } j \end{cases} \tag{4}$$

The diagonal of the matrices P_s were set up to 0. Considering a sample of k judges, we defined the sample preference matrix M by the sum of the individual preferences matrices P_s :

$$M = \sum_{s=1}^{k} P_s \tag{5}$$

For each different pairs, the matrix M reveals the number of judges in the sample who preferred the first stimulus compared to the second one. Its elements will be noted $m(i,j)$.

3.2. Sample homogeneity

In order to calculate the degree of agreement among individuals in their preferences we used a nonparametric measure of association, the Kendall coefficient of agreement u for paired comparisons. Values of u have been calculated for the Bach and Mozart session and can be found in table 2. We can note that in both sessions the agreement coefficient is very high. The next step consists in checking its validity.

We tested the null hypothesis that there is no agreement among the raters against the alternative that the degree of agreement is greater than what one would expect had the paired comparisons been done at random. As the total number of judges is large ($k > 6$), a large-sample approximation to the sampling distribtution, X^2, is used. It is asymptotically distributed as a χ^2 distribution with 28 degrees of freedom. Results can be found in table 2. In both cases (Bach and Mozart), we may reject the null hypothesis with a risk $\alpha < 0.001$. We concluded that there is strong agreement among the participants in their preferences both when the sequences were excerpts from Bach or Mozart pieces.

3.3. Musical performance ranking

As the ratings of the judges were not random but on the contrary strongly correlated, one would observe the global ranking of the musical performances. Individual rankings were identified by adding up the number of times each of the eight sequences was chosen as the preferred one over the 28 trials. This number ranges from 0 to 7 (times preferred). The median and quartiles of these individual rankings are shown on figure 1, which hence gives a picture of the judges mean ranking during the Bach session. Results for the Mozart session are quite similar.

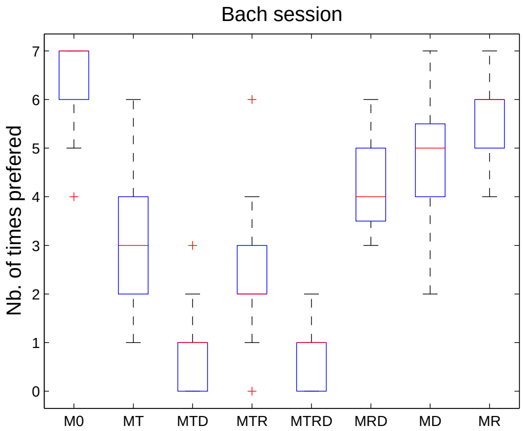

Figure 1. Box and whisker plots of the judges' rankings for each sequences of the Bach session. The box has lines at the lower quartile, median, and upper quartile values. The whiskers show the extent of the rest of the data. Outliers are represented by crosses. Note that there are border effects as the number of times each sequence has been preferred is neccessarily between 0 and 7.

The sequence which has been on average the best rated is the one which preserves the original variations of rhythm, dynamics and timbre M_0. Note that independently of the subjective nature of the judgments, all the sequences which went through the timbre transformation (M_T, M_{TD}, M_{TR}, M_{TRD}) were on average the least preferred. The cancellation of the Spectral Centroid variations during the sustained part of the tones induces a loss of the tone quality which is perceived as a loss of musical expressivity. Hence, timbre variations appear to be one major factor of music performance.

3.4. Multidimensional scaling analysis (MDS)

We carried out muldimensional scaling in order to map the musical sequences in a muldimensional space so that their relative positions in the space reflect the degree of perceived proximity between them. Therefore, the first step has been to transform the $(N * N)$ sample preference matrix defined in section 3.1 into a $(N * N)$ dissimilarity matrix.

The notion of perceptual distance between the musical performances can be reflected by the preferences of the judges. Let us consider the number of judges who preferred the interpretation in a sequence A compared to that in a sequence B. The higher (or the lower) the number of judges, the more dissimilar were considered A and B. A number of judges close to $\frac{k}{2}$ indicated that the judges did not agree either for subjective reasons, or because A and B were perceptually similar leading to almost random choices. We retained the latter hypothesis thanks to the remarks of the raters who found some stimuli hard to discriminate due to their perceptual closeness. Following these considerations, we built a class of functions f_n which transform the sample preference matrix M into a dissimilarity matrix D :

$$\forall x \in [0;k]_{\mathbb{N}} \text{ , } f_n(x) = (2x-k)^n \qquad (6)$$

where x designs the $m(i,j)$ and n is a non-null even integer. The following results have been obtained by choosing $n = 2$.

Both metric (assumption of interval measurements) and nonmetric (assumption of ordinal measurements) MDS were performed. The nonmetric procedure (MDSCAL) was retained as it produced a better fit (Kruskal's stress) to the data than the metric one. This procedure yields solutions such that the distances in the derived space are in the same rank order as the original data. The initial configuration of points was found using the classical multidimensional scaling solution. The rate of decline of the stress as dimensionality increases and Shepard diagrams helped us to define a reliable number of dimensions. The method indicated a 3-dimensional space for both Bach ($stress = 9.7e^{-5}$) and Mozart ($stress = 1.1e^{-3}$) sessions. Figure 2 shows the projection of the MDS solutions on the first two axis for both the Bach and Mozart sessions.

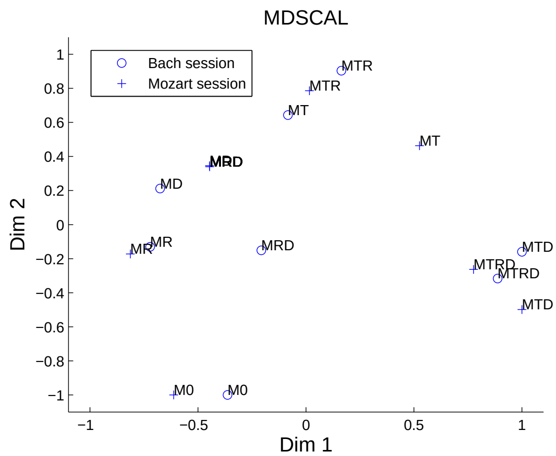

Figure 2. Two-dimensional projections (Dim 1 vs Dim 2) of the MDS configurations for the Bach (circles) and Mozart (crosses) sessions

The locations of the eight different types of performances along the two principal dimensions are very similar for the Bach and Mozart musical excerpts, which proves that the participants used the same factors to rank the performances in each session. The horizontal axis seem to refer to a tone factor as it opposes the sequences with fixed Spectral Centroid variations (M_{TRD},M_{TD},M_T) from the one with the orginal spectral centroid variations (M_0,M_R, etc). This opposition also appeared in the remarks of the participants who were many to mention that they preferred the performances with ”lively” tones than the one with ”static” tones. We do not yet have a reliable interpretation for the second axis.

4. CONCLUSION

This paper investigated the importance of timbre variations compared to rhythmic and dynamics variations in clarinet music performance. For this purpose, an analysis-synthesis procedure helped us to generate, from Bach and Mozart solo clarinet performances, two sets of performances for which timbre, rhythmic and dynamics variations were progressively removed or flattened. There was a strong consistency of the preferences between participants for both sets of performances. The average ranking of the participants showed that the least preferred sequences were those for which the timbre variations (represented by the Spectral Centroid) were removed. Multidimensional scaling underlined that the predominant factor used to rate the performances was the tone quality which is linked to the timbre.

5. REFERENCES

[1] Gabrielsson, A., The Performance of Music, *Psychology of Music*, Academic Press, 2nd ed., 1999.

[2] Palmer, C., “Music Performance”, *Annu. Rev. Psychol. 48: 115-138*, 1997.

[3] Farner, S., Kronland-Martinet, R., Voinier, T., and Ystad, S., “Timbre variations as an attribute of naturalness in clarinet play”, *CMMR05, pp. 45-53*, Springer, Pisa, Italy, 2005.

[4] Barthet, M., Kronland, R., and Ystad, S., “Consistency of timbre patterns in expressive music performance”, *DAFx06, pp. 19-24*, McGill University, Montreal, Canada, 2006.

[5] Repp, B. H., “Diversity and commonality in music performance: An analysis of timing microstructure in Schumann’s Trumerei”, *J. Acoust. Soc. Am., 92(5):2546-2568*, 1992.

[6] Sundberg, J., Firberg, A., and Frydn, L., “Rules for automated performance of ensemble music”, *Contemporary Music Review, 3: 89-109*, 1989.

[7] McAdams, S., Beauchamp, J. W., and Meneguzzi, S., “Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters”, *J. Acoust. Soc. Am., 105(2): 882-897*, 1999.

[8] Susini, P., McAdams, S., and Winsberg, S., “A Multidimensional Technique for Sound Quality Assessment”, *Acta Acustica, 85: 650-656*, 1999.