# Analysis by synthesis of the expressive intentions in musical performance

Sergio Canazza, Giovanni De Poli , Antonio Rodà and Alvise Vidolin
canazza, depoli, vidolin@dei.unipd.it
CSC-DEI, University of Padova, Via Gradenigo 6a, 35131 Padova Italy

**Abstract**

Musical performance introduces some deviations from nominal values specified in the score. Music reproduced without such variations is usually perceived as mechanical. Most investigations explore how the musical structure influences the performance. There are a few studies on how the musician's expressive intentions are reflected in the performance. The purpose of this work is to analyze, both from a perceptual and sonological point of view, how the performance of a score differs when the musicians are requested to play it with different expressive characters. Professional musicians were asked to play short pieces of music in different versions, namely, so that it sounds light, heavy, soft, hard, bright, and dark. For comparison a normal, standard performance was recorded.
By analyzing the results of perceptual analysis, two quite distinct expressive directions were observed, one tending towards brightness and the other, towards softness of the pieces. The two dimensional space so obtained represents a model of expressivity, by which the subjects arranged the pieces in their own minds.
Sonological analysis allowed us to identify the variations of the main sonological parameters on varying the expressive intentions. It was possible to identify two distinct expressive sources. The first refers the musical structure of the period, its division into phrases and the continuous alternation of tension and relaxation points. The second depends on the expressive intentions that the musician wants to convey to the listeners.

## 1 Introduction

It is known that musical performance introduces some deviations from nominal values specified in the score. Music reproduced without such variations is usually perceived as mechanical.
Theoretical studies on musical performance and interpretation carried out according to rigorous scientific principles are a relatively recent development, and one of the biggest difficulties lies in the player's inability to explain his/her interpretation clearly, in theoretical terms. At the same time, there are also some difficulties in gathering musical data by the physical measurement of the overall sonorous result of a performance. Moreover, the study of musical performance and interpretation raise problems which go beyond the specific area of taking measurements, running into the wider field of non-verbal expressive communication that involves various aspects of human perception.
Musical performance, in the western tradition, is based on the score. This is a document of graphic symbols by means of which the composer's musical idea is communicated to the listener, even when they belong to different historical periods. The player usually takes a certain amount of freedom in his/her interpretation of the score in order to better express the sense of a musical phrase. A stylistically correct performance which is played paying the greatest respect to the style of the score, is here termed as normal. Furthermore, a musician may play the same piece differently on different occasions (changing each time his interpretation of the piece) and different musicians may differ much among themselves in the performance of the same piece. The (micro)structure of musical performance can never be fully understood without taking the performer's intentions, and the listner's experience, into account.
This paper presents a perceptual and sonological analysis of expressive deviations in the interpretation of a musical phrase performed with different expressive intentions. In this context, *expressive intention* is taken to mean how a musician's inspiration varied according to certain adjectives that had been given before each performance [5]. This aspect of a performance has not been widely studied, in that it has always been considered a pre-eminently artistic one and could not, therefore, be analyzed. With the advent of computers, however, new methods of analysis have been developed and, therefore, the possibility of verifying, by synthesis, the validity of the analytical data or performance model formulated on

the theoretical plane. Consequently, more attention has been given to this argument and some interesting studies have been carried out: for an extensive review see Gabrielsson [6]. This field of investigation is stimulating ever greater interest not only from the scientific and cognitive point of view, but also from the applicative one both in terms of composing music and, more generally, in multimedia systems.

This study is based on the analysis of musical fragments of the classical period played on different instruments. Professional musicians were asked to performe a musical phrase, inspired by the following (italian) adjectives: *light* (leggero), *heavy* (pesante), *soft* (morbido), *hard* (duro), *bright* (brillante), and *dark* (duro). The scholastic performance, described by the adjective *normal* (normale), was also added to them and used as a standard measure of comparison in the sonological analysis of the various interpretations. Unconventional adjectives, in the musical field, were deliberately chosen in order to give the performer the greatest possible liberty of expression. The recordings were carried out in three cycles, each cycle consisting of the seven different interpretations. The musician chose, then, the ones that, in his opinion, best corresponded to the proposed adjectives, trying to minimise the influence that the order of execution might have had on the performer. The recordings were carried out at the CSC of Padua University in the monophonic digital form at 16 bits and 44100 Hz.

We made sonologic and perceptual analysis from these performances. Some of these analysis are shown in this article. The combination of both kinds of analysis allowed us to develop a model for rendering the different expressive intentions.

## 2 Perceptual analysis of expressive intentions

There are a few studies on how expressive intentions are perceived by listeners [7,9]. We conducted some perceptual analyses in order to characterize the listening impressions on performances with own different expressive intentions [2]. We will report the most complete one.

Seven different interpretations of a fragment of Mozart's K622 Concert for Clarinet (see fig. 1), were performed by a professional clarinet player and recorded as described above.

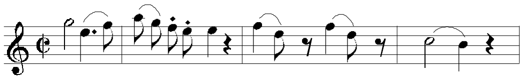

Fig. 1: Mozart's K622 Concert for Clarinet

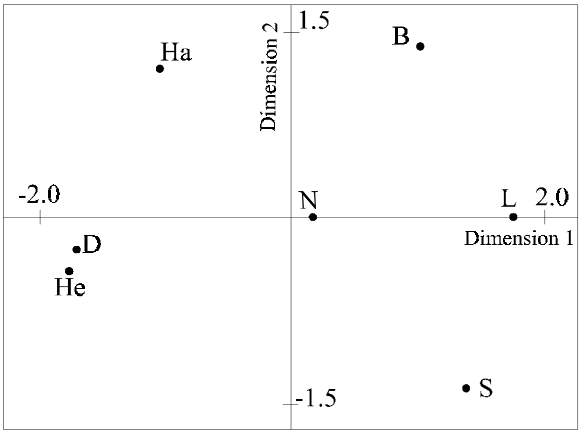


Fig. 2: MultiDimensional Scaling space (Stress=0.012). B=Bright, D=Dark, Ha=Hard, L=Light, S=Soft, N=Normal, He=Heavy.

We made an experiment to determine the judgement categories used by subjects called in to listen to the various interpretations of the same musical piece. The test was carried out on a group of 24 subjects, 12 musicians graduated at the Padua Conservatory, and 12 subjects without specific musical preparation.

A first experiment of simple identification has been made [2]. Here it is documented a second one, where the subjects were asked to describe the performances along 17 scales of evaluation adjectives of sensorial nature: **bla**ck (nero), **opp**ressive (greve), **ser**ious (grave), **dis**mal (tetro), **mas**sive (massiccio), **rig**id (rigido), **mel**low (soffice), **ten**der (tenero), **swe**et (dolce), **lim**pid (limpido), **air**y (aereo), **gen**tle (lieve), **eff**ervescent (spumeggiante), **vap**orous (vaporoso), **fre**sh (fresco), **abr**upt (brusco), **sha**rp (netto).

This list of adjectives did not contain those used in the performances and did not include their opposites . They were chosen so as to offer the subjects a exhaustive sampling of a semantic space.

From MultiDimensional Scaling (MDS) analysis of judgements it was seen that the listeners arranged the pieces separately in their own minds and tended to place them at an equal distance from the normal performance (see fig. 2).

We made two different factor analysis. Factor analysis on adjectives allowed us to determine a semantic space defined by the adjectives proposed to the listeners. The performances could be placing placed on it, according to factor scores [8], in order to observe in which semantic sector they map. Two significant components (i.e. with eigenvalues greater than 1) accounted for 87.2% of the variance. Varimax rotation was used in order to simplify the factors' interpretation.

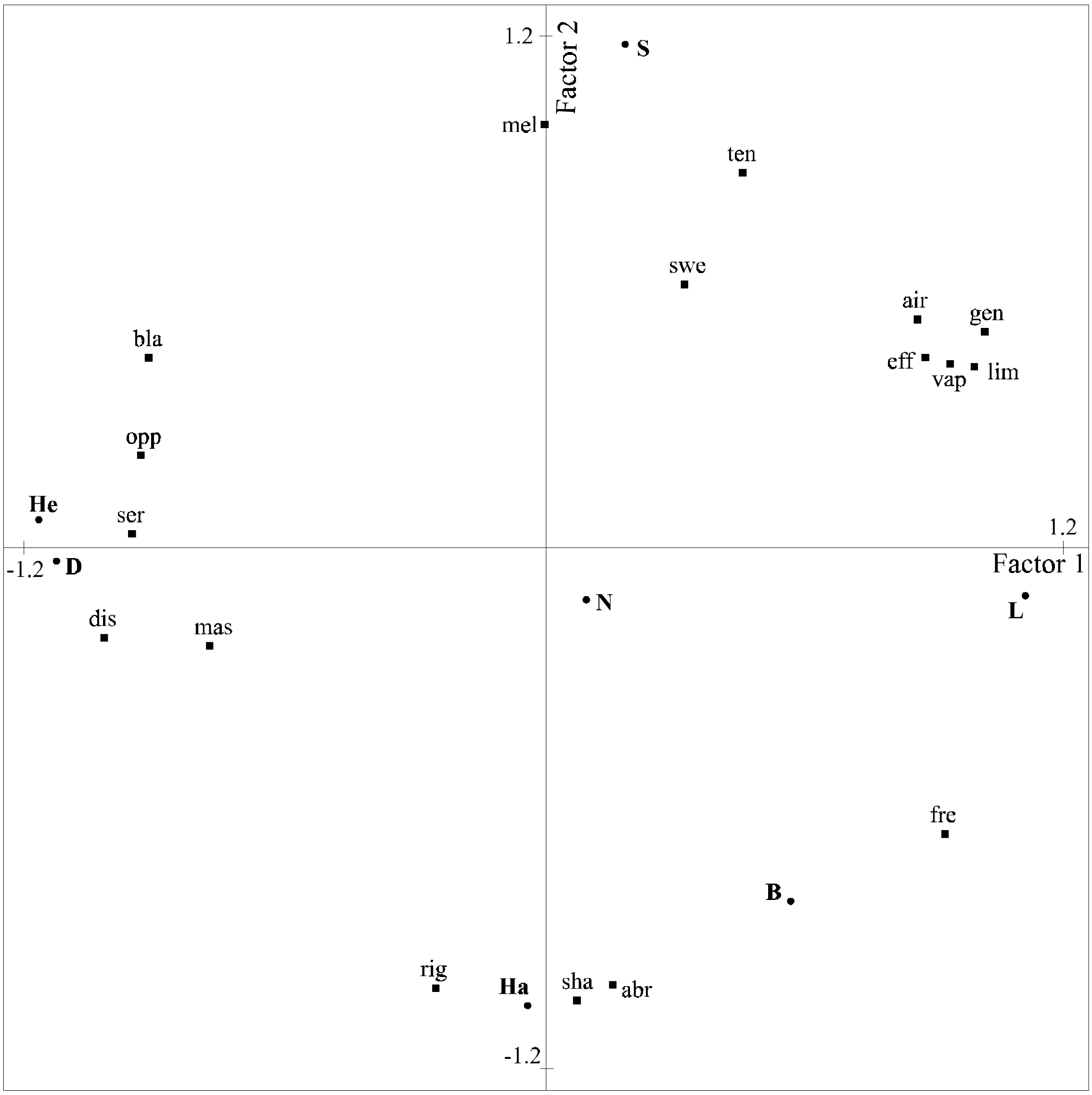


Fig. 3: factor analysis on the adjectives. The first factor explains 60% of the total variance, the second 27.2%.

Fig. 3 shows the position of the evaluation adjectives in the resulting space. As can be seen, adjectives associated with the extremes of these factors were for factor 1 dismal vs. limpid, for factor 2, mellow vs. rigid. By means of factor scores, it was possible to insert the performances into this space. The comparison of the positions of performances with that of the evaluation adjectives demonstrated a good recognition of the performers intentions by the subjects. For instance soft (S) performance is placed near mellow, tender and sweet adjectives. Moreover it can be noticed that normal performance is placed near the center of the space, far from all the adjectives.

The second factor analysis used performances as variables. It showed that the subjects had placed the performances along only two axes. The first two factors, in fact, explained 75.2% of the total variance and were the only ones having eigenvalues greater than 1.

The two dimensional space so obtained represents a model of expressivity, by which the subjects arranged the pieces in their own minds. This space is in agreement with the found one by Namba and co-workers [7]. Fig. 4 shows the arrangement of the pieces along these axes. The first factor (bright vs. dark) seemed to be closely correlated to the tempo, while the second factor (soft vs. hard) was connected to the amplitude envelope of the notes, and particularly to the attack time.

Notice that *normal* performance tend to be placed toward *bright* performance, as should be expected in a stylistically correct performance of a Mozart's piece.

The judgement categories that emerged from both analysis were similar and pointed out a factor bright-dark indepedent from the soft-hard one. These results were

similar both for musically trained and untrained subjects. However cluster analysis of answers showed a behaviour distinction between the group of trained musicians which was highly cohesive and the group of non musicians which showed a greater variance in the judgements given.
An important conclusion of perceptual analysis is that normal performance tend to be placed near the origin of the semantic space, thus suggesting the hypothesis of two sources of expressive deviations, as it will be discussed later on.

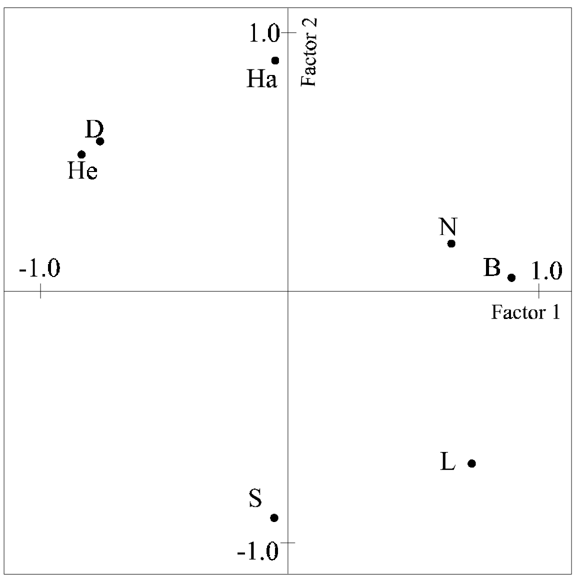


Fig. 4: Graph of factor loadings deduced from factor analysis. The first factor explains 49.6% of the total variance, the second 28.5%.

# 3 Sonological analysis

The principle aim was to identify which physical parameters, and how many of them, were subject to modifications when the expressive intention of the performer was varied.
Every musical instrument has its own expressive resources (vibrato in strings, the tongue in wind instruments, etc.), which are used by the musician to communicate his expressive intention. It is inevitable, therefore, that the results of any sonological measure depend, not only on the score, but also on the characteristics of the instrument used and the choices effected by the musician. Consequently, it is necessary to compare the data relative to different scores, musicians and instruments, in order to identify the expressive rules that can be considered valid in a general way and which are specific cases.
Some sonological analyses have been carried out on various musical pieces using different instruments and performers. Up to now, performances involving the clarinet [3], the violin and the piano have been analyzed. In this paper, we will illustrate one of these analyses, based on seven performances of a short excerpt from Arcangelo Corelli's Violin Sonata in A Major, V Op. (Fig. 5). The violin, in fact, is the instrument that it offers to performer the greater expressive resources. The results relating to the most representative parameters in conveying the expressive intentions will be presented.

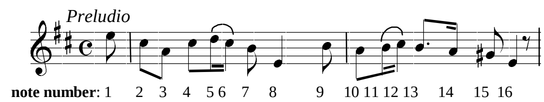


Fig. 5: First Movement of the violin Sonata in A Major op. V by A. Corelli.

## 3.1 Definition of the parameters

Time analysis starts, especially for instruments such as the violin, by measuring and modeling the amplitude envelopes of the various notes.
In the performances the amplitude envelopes have shapes and characteristics that vary considerably, both in terms of expressive intentions as well as among different notes within the same interpretation. Some notes have an extremely rapid onset and the amplitude reaches a peak value almost immediately. Other notes, on the other hand, have a slow, gradually rising onset. Moreover, an amplitude vibrato was present in almost all tones. There is no standard procedure for measuring the envelopes. Fig. 6 shows the parameters taken into consideration in relation to the envelopes.

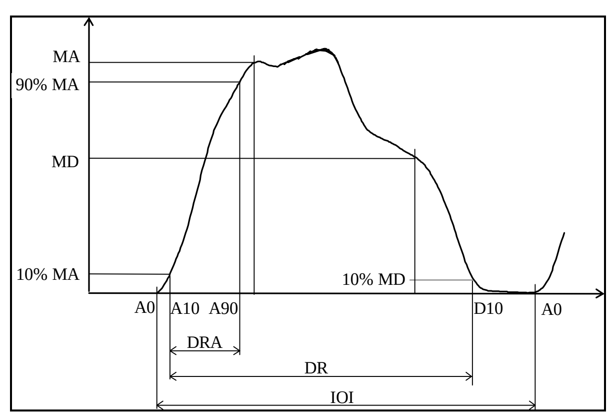


Fig. 6: the characteristic parameters of a note's amplitude envelope.

As a result of the great variability in the shape, it is very difficult to distinguish between the times of attack, sustain and decay. It was necessary, therefore, to analyze the envelope of each single note in order to identify the MA (end of attack) and MD (start of decay) points. The choice of which parameters to measure and how to measure them is a critical question in every sonological analysis. Often the characteristics of the sounds produced by different musical instruments lead to choices which change on each different occasion.

Unfortunately, differences in measuring techniques and definitions sometimes make it difficult to compare the results of different investigations. The tempo, for example, is a fictitious variable, as it can not be measured directly. What is, in fact, measured is the A0 onset-time of a tone, so that the tempo may only be estimated numerically. The average number of beats per minute throughout the whole piece can be regarded as a sufficient estimate of the tempo. This value in some cases, however, underestimates the underlying tempo somewhat because it includes retards, lengthenings and pauses at phrase endings. In order to obtain more accurate estimates it would better to measure the tempo in the measures, which do not have any lengthenings or pauses. Choosing which procedure to follow depends, above all, on the score to be analyzed. In the case of short performances, such as those analyzed here, the various methods tend to give very similar results. There are, in fact, no conspicuous lenthenings or pauses. After having compared a variety of methods, a least square interpolation technique was chosen to estimate the tempo of the seven interpretations [10].

Another problem involves the definition of the beginning and ending of a tone. It can be variously defined as the point in time when a change in the amplitude or waveform can be seen (A0), or when the amplitude of the tone has reached a certain level (A10 and D10). With reference to duration, a distinction must be made between the duration from the beginning of a tone to the beginning of the next one (Inter-Onset Interval), and the beginning of a tone to its end (note duration). The *Inter-Onset Interval* (of *n*-th note) has been defined as

$$IOI(n) = A0(n + 1) - A0(n).$$

The choice of A0 allows for a more simple comparison with the measurements carried out on the piano, where A0 corresponds to the point in time when a key is pressed. The *note duration* has, instead, been defined by

$$DR(n) = D10(n) - A10(n),$$

given that the use of A10 and D10 is often assumed to better reflect the perceptual onset and offset of a tone.

On the basis of these choices, articulation has been expressed by means of the legato parameter defined by

$$L(n) = DR(n) / IOI(n).$$

Lastly, the *attack duration* has been defined as the time taken by the amplitude envelope to pass from 10 to 90% of the MA value, that is

$$DRA(n) = A90(n) - A10(n).$$

## 3.2 Results of the analysis

Fig. 7 shows the performance tempo of the seven interpretations and it can be seen that those inspired by the adjectives *bright* and *light* were played very fast, while the *heavy* piece was the slowest. The violinist made great use of this expressive resource in order to differentiate his performances. It can be seen that there was, in fact, a wide range of tempo represented, with the fastest (*bright*) performance being played more than 40% faster than the slowest (*heavy*). Fig. 8 shows the values measured for the legato parameter of each note. For each variation observed, we have worked backwards from the performances to uncover, or more accurately, to recover, the source of the performance expression.

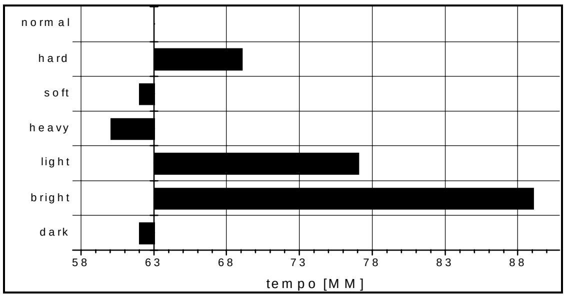


Fig. 7: value of tempo MM (quarter/min) in the seven performances.

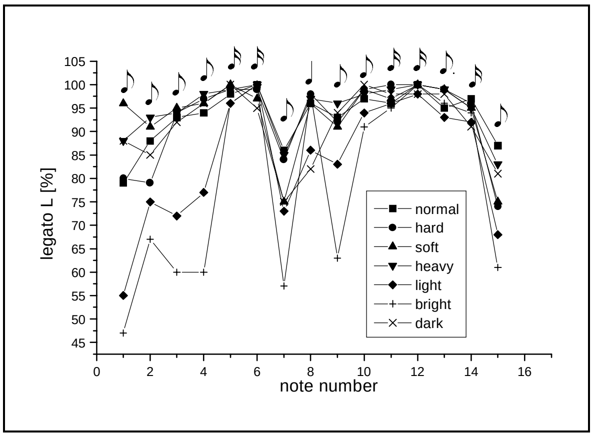


Fig. 8: value of legato in the seven performances.

Fig. 8 shows how all the profiles tend to be of the double curve type. This trend reflects the musical structure of the phrase, or rather its division into two half-phrases (the first going from note 1 to note 8 and the second going from note 9 to note 16). Other than this common trend, quite considerable differences can be found between one performance and another. In particular, the *bright* piece has legato values, which are inferior to those of the other pieces. This might be explained by the fact that the variations depend on the expressive intentions of the musician.

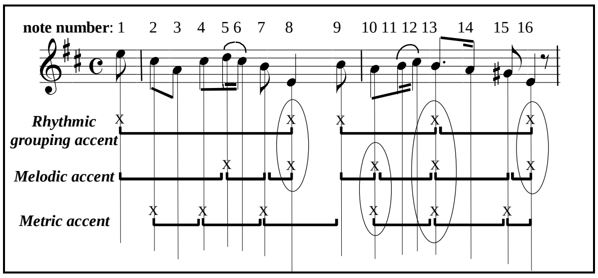


Fig. 9: accent structures in Corelli's score.

By observing the data it is possible to identify two distinct expressive sources. The first refers the musical structure of the period, its division into phrases, half-phrases and the continuous alternation of tension and relaxation points. The second depends on the expressive intention that the musician wants to convey to the listener. In this latter case, the choice was induced by the seven adjectives proposed earlier in the experiment.

An analysis of variance (ANOVA) on the legato values by intentions (light, heavy, etc.), showed that the adjective had a considerable effect on how the piece was played ($F(6,98) = 4.5$, $p < 0.001$). The average legato value was, therefore, calculated for each performance. On the basis of these values, it was possible to arrange the pieces in the following order: *bright* (L = 0.79), *light* (L = 0.84), *dark* ( = 0.91), *hard* (L = 0.92), *normal* (L = 0.93), *soft* (L = 0.94), *heavy* (L = 0.95). It is interesting to note that, especially in the performance of the *bright* and *light* pieces, the legato value of some notes draw away quite sensibly from the calculated average value. It is, therefore, necessary to analyze the legato values of each single note when the expressive intention varies. Fig. 8 shows that the 5, 6, 10, 11, 12, 13, 14 notes and, to a lesser extent note 8, have legato values which are very close to each other in all the performances. It would, then, be rather interesting to try to determine whether the way these notes are played are due to technical reasons or whether to expressive ones. On looking at the score, illustrated in Fig. 5, it can be seen that the 5, 6, 11, 12 and 14 notes are all sixteenth with a IOI duration of about 200 ms. It might be hypothesized that there is a technical limit to these notes. In fact, the need to play them staccato (with a low legato value), would require to resort to particular bowstroke (ricochet), which is not well suited to the piece. Furthermore, the score itself indicates that the 5 and 11 notes should be tied to the successive ones.

A different question arises, however, with reference to the 10, 13 and, in part to note 8. It might be said that these notes have an important role within the musical structure of the phrase.

In order to confirm this hypothesis, the definition of accents after Drake & Palmer [4], was applied to Corelli's score

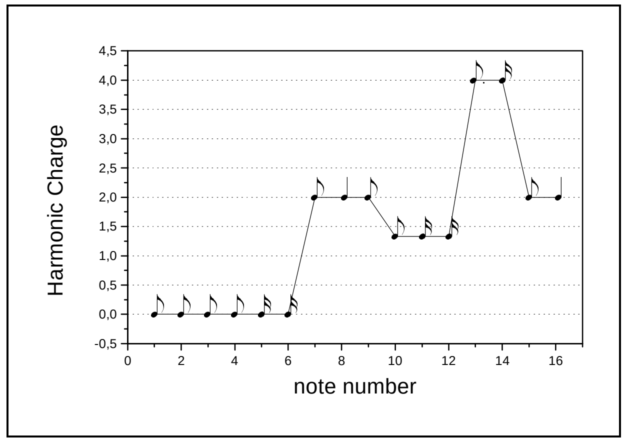


Fig. 10: Harmonic Charge calculated in Corelli's score.

Accents, in this sense, mean those rhythmic-melodic events, which attract the listener's attention most. These are very closely linked to the score and play an important role in the segmentation of the period and in its structural hierarchy. Fig. 9 shows the structure of the accents that have been identified. The greatest number of accents fall on the 8, 10, 13 and 16 notes and they are also those which change least on the varying in the expressive intentions.

A comparison with the harmonic charge as defined in [1], would seem to confirm the hypothesis that there is a constrain to deviations induced by the player's expressive intention. Given the harmonic context, the harmonic charge measures the singularity of a chord. Fig. 10 shows that note 13 has a higher value. The KTH expressive rules, which depend on the structure of the score, indicate that notes with a high harmonic charge should be underlined with greater intensity and by a longer duration. It might be said, then, that according to this expressive rule, the violinist has stressed note 13, playing it legato in all the various interpretations. It was therefore seen that the 8, 10 and 13 notes play an important role within the phrase, a role that is indispensable for the transmission of the musical structure. If these were played staccato as, for example, the *bright* piece would call for then, probably, the understanding of the musical phrase would be compromised.

It could be surmised that when expressive intentions are varied, there are rigid notes and elastic ones. In order to transmit the expressive intention induced by the adjectives, the musician could, therefore, only vary those notes that were not particularly important in terms of the rhythmic-melodic-harmonic structure of the musical phrase. The others, which were more important and therefore more rigid, could, then, be less varied by the musician's expressive intentions.

Fig. 11 shows the attack time values measured for each note. All the performances had a double curve trend, which reflected the musical structure of the phrase that could be divided into two half-phrases. As well as this common trend, sensible differences between the single performances could be identified. The attack times of the *bright* piece were generally shorter. It could be supposed that these variations depend on the expressive intentions of the musician. The analysis of variance (ANOVA) was made on the attack time values by intentions (light, heavy, etc.). It showed that these intentions had a significant effect on performance ($F(6,91) = 2.51$, $p < 0.03$).

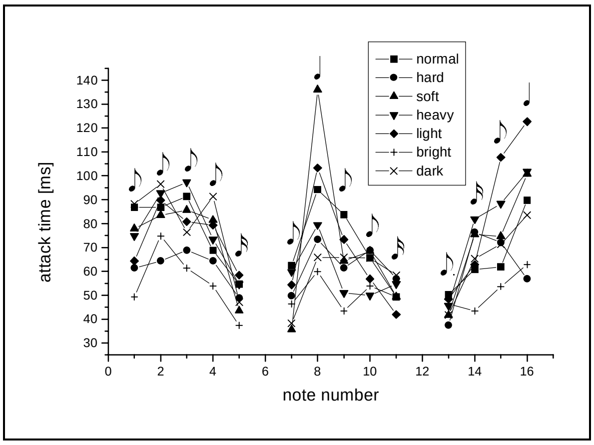


Fig. 11: attack duration values for the seven performances.

The average attack duration for each performance was then calculated. On the basis of these values it was possible to put the pieces into the following order: *bright* (DRA = 53 ms), *hard* (DRA = 61 ms), *dark* (DRA = 68 ms), *heavy* (DRA = 71 ms), *normal* (DRA = 72 ms), *soft* (DRA = 73 ms) and *light* (DRA = 75 ms).
From the observation of fig. 11, it should be noted that on varying the expressive intention, some notes have been performed with very similar attack durations. Similar considerations can also be made in this case, that is, those regarding the rigidity of some notes in the legato parameter. The 13 note is very rigid, given that in all the performances it was played with rather short attack times. Short attack time was used to emphasize the note. From a structural analysis of the musical period it is possible to identify - in that particular note - the point where the greater expressive stress converges. The musician, in order to communicate such structure clearly, has been induced to play this note with an ever diminishing attack time, notwithstanding the soft nature that would have been obtained, for example, had the attack been slower.

### 3.3 Synthesis

Others parameters were measured: dynamic range, ratio between up-beat and down-beat duration, ratio between amplitude MA and MD (fig. 6) and timbre brightness. Table 1 summarizes the behavior of the main parameters on varying the expressive intentions.
A model of the expressive intentions has been elaborated on the basis of these measurements and this model was then used to generate synthetic performances. The analysis-by-synthesis method was used to check the consistency of the resulting approximation. To this end, physical model synthesis was used, so that it was possible to apply the results of the sonological analysis in order to obtain performances that were much more similar to those of a real violin. The use of a physical model allows for controlling the performance, not only note-by-note but also at the level of the phrase. Moreover, the parameters controlled (speed of the bow, pressure on strings, etc.), allow for making an immediate reference to the expressive gestures of the musician.
The Yamaha VL1 physical model synthesizer, was used to realize the synthesis. Seven different performances were effected which, notwithstanding the timbre poverty of the VL1 violin, maintained the expressive color (hard, soft, etc.), of the originals.
It is important to notice that the measured parameters are refered to acoustic signal. It is therefore necessary to proceed backwards in order to derive the control parameters, expressed in terms of bow speed. In VL1, the bow speed has been controlled by three straight lines (attack, sustain, and decay). The function that controls the bow speed has been set to zero only in correspondence of the direction changes of the bow. In this way, it was possible to play more notes on the same bowing, reproducing the phrasing of a real violin.

| | Ha | S | He | L | B | D |
|---|---|---|---|---|---|---|
| Tempo | + | | - - | + + | + + + | |
| legato | | + | + | - | - - | |
| DRA | - | + | | + | - - | - |
| Dynamic | | + | | - | - | + |
| Up-beat / Down-beat | | - | | + | | |
| MA/MD | + | - - | + | | | - |
| Brightness | + + | - - | + | - | + + | - - |

Tab. 1: observed variations in the main sonological parameters on varying the expressive intentions.

# 4 Conclusion

Various studies on performance have led to suggesting models that could render synthesized music less monotonous and mechanical. Most investigations explore how the musical structure influences the performance.

There are a few studies on how the musician’s expressive intentions are reflected in the performance. We analyzed both from a perceptual and sonological point of view, how the performance of a score differs when the musicians are requested to play it with different expressive characters. Sensorial adjectives (light, heavy, soft, hard, bright, dark) were chosen in order to have a compact and coherent semantic space. For comparison a normal, standard performance was added.

By analyzing the results of perceptual analysis, two quite distinct expressive directions were observed, one tending towards brightness and the other, towards softness of the pieces. The two dimensional space so obtained represents a model of expressivity, by which the subjects arranged the pieces in their own minds.

Sonological analysis allowed us to identify the variations of the main sonological parameters on varying the expressive intentions. It was possible to identify two distinct expressive sources. The first refers the musical structure of the period, its division into phrases and the continuous alternation of tension and relaxation points. The second depends on the expressive intentions that the musician wants to convey to the listeners.

The results of the sonological analysis seem to confirm the hypothesis that various performances of the same musical piece have their own sonological characteristics that can, in reality, be measured. The measurements agreed with the intentions that had been verbally described by the musician in an interview after the performances.

The combination of both kinds of analysis allowed us to develop a model for rendering the different expressive intentions. The model computes note timing and a small set of acoustic parameters, important for characterizing the performances. It takes into account the adjectives and the structure of the score. The use of the model in synthesizing performances allowed us to refine the analysis and to compare the relevant parameters.

# References


[1] Friberg A. 1991. “Generative rules for musical performance: a formal description of a rule system,” Computer Music Journal, Volume 15, Number 2, pp. 56-71.

[2] Canazza S., De Poli G., and Vidolin A. 1997. “Perceptual analysis of the musical expressive intentions in clarinet performance,” in M. Leman (ed) *Music, gestalt, and computing. Studies in cognitive and systematic musicology*, Berlin, Heidelberg: Springer-Verlag

[3] Canazza S., De Poli G., Rinaldin S., and Vidolin A. 1997. “Sonological analysis of clarinet expressivity,” in M. Leman (ed) *Music, gestalt, and computing. Studies in cognitive and systematic musicology*, Berlin, Heidelberg: Springer-Verlag

[4] Drake C., and Palmer C. 1993. “Accent structures in music performance,” Music Perception, Volume 10, Number 3, pp. 343-378.

[5] Gabrielsson A. 1995. “Expressive intention and performance,” in *Music and mind machine*, 37-47, Berlin, Heidelberg, New York: Springer-Verlag.

[6] Gabrielsson A. 1997. “Music Performance,” in D. Deutsch (ed) *The psychology of Music*,.2nd edn. Academic, New York.

[7] Namba S., Kuwano S., Hatoh T., and Kato M. 1991 “Assessment of musical performance by using the method of continous judgment by selected description,” Music Perception, Volume 8, Number 3, pp. 251-276.

[8] Repp B.H. 1990. “Patterns of expressive timing in performances of a Beethoven minuet by nineteen pianists,” Journal of Acoustical Society of America, Volume 88, Number 2, pp. 622-641.

[9] Senju M., and Ohgushi K. 1987. “How are the player's ideas conveyed to the audience?,” Music perception, Volume 4, Number 5, pp. 311-324.

[10] Todd N. 1995. “The kinematics of musical expression,” Journal of Acoustical Society of America, Volume 97, Number 3, pp. 1940-1949.