ADDING EXPRESSIVENESS TO AUTOMATIC MUSICAL PERFORMANCE

Canazza Sergio, De Poli Giovanni, Di Sanzo Gianni, Vidolin Alvise
Dipartimento di Elettronica e Informatica
Università di Padova - Via Gradenigo 6a – 35100 Padova – Italy
{canazza, depoli, vidolin}@dei.unipd.it.

Abstract

In this paper we will present an overview of a study about expressivity in music. Acoustic and perceptual analyses made on different performances, played it with different expressive intentions, suggested by a set of sensorial and affective adjectives, we will present. The acoustic analyses shown the acoustic parameters which separate the different performances of the same score. Perceptual analyses confirmed that listener's experience and performer's intention were basically agreed.

A model of expressiveness has been developed on the base of results of analyses. The model allows to obtain different performances, by modifying the acoustic parameters of a given neutral performance. The modification is performed by algorithms which use the hierarchical segmentation of the musical structure. Opportune envelope curves are applied, for every hierarchical level, to the principal acoustic parameters. Level's self-similarity is the main criteria for the envelope curves construction. The rendering steps can be realized both with synthesis and post-processing techniques.

1 Introduction

Different musicians, even when referring to the same score, can produce very different performances. The score carries information such as the rhythmical and melodic structure of a certain piece, but there is not yet a notation able to describe precisely the temporal and timbre characteristics of the sound. The conventional score is quite inadequate to describe the complexity of a musical performance so that a computer might be able to perform it. Whenever the information of a score (essentially note pitch and duration) is stored in a computer, the performance sounds mechanic and not very pleasant. The performer, in fact, introduces some micro-deviations in the timing of performance, in the dynamics, in the timbre, following a procedure that is correlated to his own experience and common in the instrumental practice. It is exactly for this great variety in the performance of a piece that it is difficult to determine a general system of rules for the execution. An important step in this direction was made by Sundberg and co-workers [1]. They determined a group of criteria which, once applied to the generic score, can bring to a *musically correct* performance. Further on, the performer operates on the microstructure of the musical piece not only to convey the structure of the text written by the composer, but also to communicate his own feeling or expressive intention. Quite a lot of studies have been carried on in order to understand how much the performer's intentions are perceived by the listener, that is to say how far they share a common code (see [2] for references). Gabrielsson & Juslin [2] in particular, studied the importance of emotions in the musical message. In this context, we tried to understand the way an expressive intention can be communicate to the listener and we realized a model able to explain how it can be possible to modify the performance of a musical piece in such a way that it may convey a certain expressive intention. A group of sensorial adjectives was chosen (hard, soft, light, heavy, bright, dark) which should inspire a certain expressive idea to a musician. A musician, inspired by appropriate adjectives, produces, systematically, different performances of the same piece. Perceptual analysis proved that the audience can indeed perceive the kind of intention he wanted to convey. Acoustic analysis confirmed that there are micro-deviations in the note parameters. We outlined models to connect such deviations with the intention wanted. Following the analysis-by-synthesis method, some musical synthesis were produced to verify and develop a model of musical expressiveness.

This paper, starting from the results of the acoustic and perceptual analysis, presents the design of a model able to add expressiveness to automatic musical performance. These studies on the model of musical performance are interesting not only from a scientific point of view, but also from an practical one, both in the field of automatic musical performance and in general in the multimedia systems.

This research was supported by Telecom Italia, under a research contract Cantieri Multimediali.

2 Perceptual Analysis

Perceptual analyses were carried out to observe the listeners' judgment categories and to verify if performers succeeded to convey to the listeners the expressive intentions.

Seven different interpretations (*light*, *heavy*, *soft*, *hard*, *bright*, *dark*, and *normal*) of a fragment of theme of Mozart's K622 Concert for Clarinet, were performed by a professional clarinet player. 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 minimize 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 an experiment [4] 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.

The subjects were asked to describe the performances along 17 scales of evaluation adjectives of sensorial nature: **black** (nero), **oppressive** (greve), **serious** (grave), **dismal** (tetro), **massive** (massiccio), **rigid** (rigido), **mellow** (soffice), **tender** (tenero), **sweet** (dolce), **limpid** (limpido), **airy** (aereo), **gentle** (lieve), **effervescent** (spumeggiante), **vaporous** (vaporoso), **fresh** (fresco), **abrupt** (brusco), **sharp** (netto).

This list of adjectives did not contain those used in the performances and did not include their opposites.

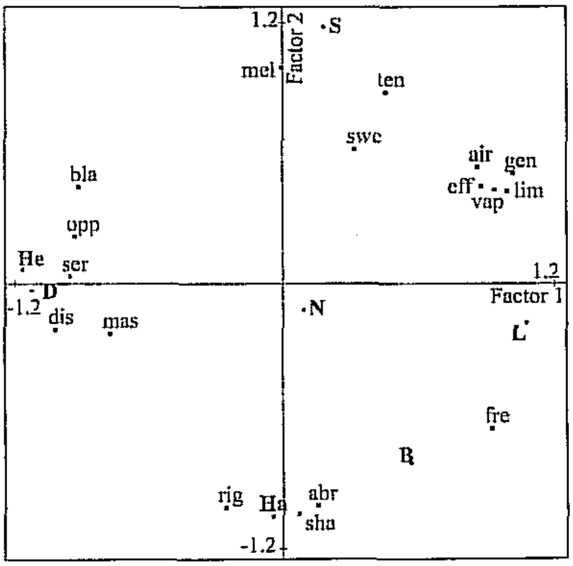

Fig. 1: factor analysis on the adjectives. The first factor explains 60% of the total variance, the second 27.2%.

They were chosen so as to offer the subjects a exhaustive sampling of a semantic space.

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, 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. 1 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.

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.

3 Acoustic Analysis

The principle aim was to identify the relationship in the physical parameters 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 acoustic analyses have been carried out on various musical pieces using different instruments and performers. Up to now, performances involving the clarinet [3], the violin [4] and the piano [5] have been analyzed.

Not all the results can be compared as some of the parameters measured were defined differently in the studies. Besides, quantitative comparisons are sometimes not very significant as the absolute values of parameters depend on the technical characteristics of the instrument used. Nevertheless, it is possible to make a qualitative comparison at least as far as the mean tempo (MM), legato (L), note attack time (DRA) and brightness (BR) of spectrum are concerned. Our main aim in this study is to determine, at least at a general level, those common strategies used by musicians to communicate their expressive intentions. Table 1 and 2 show the tendencies of parameters measured in the clarinet and violin performances and table 3 shows the qualitative results of parameters measured in piano performances. In the second case, it was only possible to compare MM and Legato parameters, considering the different technical characteristics of piano. It can be seen how, notwithstanding some differences, the pieces referring to the various expressive intentions have a similar behaviour in the different experiments. For instance, the adjective bright induced the musicians to perform their piece with a quicker metronome, a lesser legato, and a shorter attack time. In the piano, in fact, a high key velocity means a quicker attack. The main differences among the experiments have to do with a different choice in the expressive resources used, but not with a different use of these resources. In the soft version, for instance, the clarinet performer played with the values of the MM (low), DRA (high) and BR (low) parameters significantly different from the other versions. The violinist played in the same way as far as the BR (low) parameter is concerned, but unlike the clarinettist, he modified the MD-MA (difference between the amplitude at the start of decay and end of attack) and UDR (ratio between upbeat and downbeat duration) parameters. It is worth noting that a high MD-MA value means an amplitude profile slowly raising, while a high legato value in piano, together with a low key velocity, leads to an equivalent qualitative result. The only conflicting result regards the heavy piece performed by the clarinet player with a different use of the parameter L. In this case, it seems that the clarinettist used a quick note attack time and a slow metronome, causing in the listeners a sense of heavy locomotion; but the violinist and the pianist tried to communicate a sense of effort in moving things.

	N	Ha	S	He	L	B	D
MM		high		low	high	high	
L				high	low	low	
DRA		low			high	low	low
BR		high	low	high	low		low
UDR			low		high		
MD-MA			high		low		high
VR						high	low

Tab. 1: Behavior of statistically significant parameters on varying expressive intentions in violin performances, Arcangelo Corelli's Violin Sonata in A Major, V Op. [4]

	N	Ha	S	He	L	B	D
MM			low	low	high	high	
L	high			low	low	low	
DRA		low	high	low		low	
BR		high	low	high	low		

Tab. 2: Behavior of parameters on varying expressive intentions in clarinet performances Mozart K622 [3]

	N	Ha	S	He	L	B	D
MM				low	high	high	low
L	low		high	high	low	low	high
KV		high	low	high	low	high	

Tab. 3: Behavior of parameters on varying expressive intentions in piano performances, Mozart k622 [5]

4 Architecture of the model

The researches we have been making [3] and [4] prove that the performance of a piece following a certain expressive intention can be described observing which variations take place with reference to a neutral and a nominal performance of the same piece. By *nominal performance* we mean the mechanic performance of the score where the metrical durations (the score) are accurately observed and by *neutral. performance* we mean a literal human performance of the score without any expressive intention or stylistic choice.

The model developed [6] is able to obtain an expressive intention, transforming a neutral performance both with reference to the score and the acoustic signal itself. It must be underlined that our approach provides for the adoption of hierarchical structures similar to the spoken language ones (words, phrases), in the musical language. Once these structures are recognized, it is possible to modify the parameter of a group of notes (for example metronome or intensity) following a certain curve. Such curve describes the characteristic of the musical gesture on the group of notes. It is therefore convenient to describe (appropriately codified) the information about the neutral performance and the nominal performance (i.e. the score), the variations to be applied on the expressive traits of the single note (timing, intensity, timbre), the subdivisions of the piece into expressive units (words, semi-phrases, phrases) characterized by curves that modify one or more parameters of the notes that constitute them.

To this aim, we propose a new representation of the score, where the fundamental components and parameters of a musical piece are highlighted. Moreover it is provided with a number of controls on expressive parameters that allow the model to operate on the piece. Later we shall refer to this new score as metascore. The metascore is a file where the information about both a nominal performance and a neutral performance are codified. The parameters of the neutral performance are expressed as deviations from the nominal performance. The performances are read by a MIDI file and transcribed in the metascore. From the parameters of the

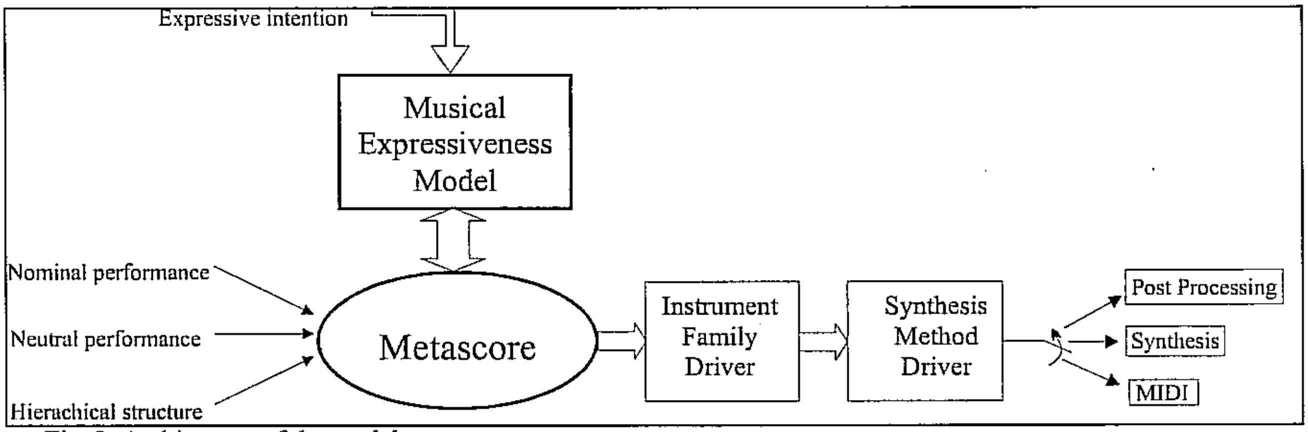

Fig. 2: Architecture of the model.

MIDI protocol, the model computes new parameters that describe the attack, the intensity, the spectral characteristics and other physical attributes of each note. Thus, the basic parameters of each note are immediately accessible. Each of these parameters is expressed in perceptual scale, in which the unit represents the difference between two perceptual levels (e.g. the difference between *f* and *ff* for loudness).

These parameters are independent from the particular musical instrument that will play the score. The importance of the notes in the piece can be specified by parameters such as the accent and the elasticity. This last parameter takes in account that expressive deviations are not made to the same extent in each note because the score sets technical and structural limits that prevent the musical phrase from being distorted [4]. The metascore also requires, as input, a description of the hierarchical structure of the piece (fig. 2). As already mentioned, we should not see the notes of a piece as units independent one from another: each piece has its own structure where the notes gathers to form the words or phrases of a musical discourse. For instance it is possible to set the metronome profile and the intensity on groups of notes following the division in words, semi-phrases and phrases. At the phrase level it is possible to specify an adjective inspiring the performance.

With reference to the adjectives and considering other factors such as elasticity, the model computes the deviations from neutral performance in order to render expressive synthesis. Output of this first computation is still independent from the instrument that will play the score. It is then necessary to adapt the modify metascore to a particular family of instruments and then to a specific synthesis method (fig. 2). The modular structure of the system defines an open architecture, where the rendering steps can be realized both with synthesis and post-processing techniques.

Different synthesis techniques, like FM or Wavetable, have been explored. Expressive synthesis of pieces belonging to different musical genres (European classical, European ethnic, Afro-American) verified the

Fig. 3 Violin original: **neutral** performance a). Violin obtained through post-processing, using time-frequency techniques, from the neutral performance: **hard** performance b); soft performance c).

generalization of the rules used in the model.

As example of a piece codified and performed through the model we present A. Corelli's Violin Sonata in A Major, V Op (score in fig. 3). We shall show now the graphics of amplitude envelops of some clarinet performances obtained (through post-processing) thanks to the controls given by the model. In figure 3a the neutral performance is shown. In figure 3b, the hard performance and in figure 3c the soft performance obtained using time-frequency techniques in order to bring about the transformations calculated by the model.

5 Conclusions

In this paper has been presented an overview of a study about expressivity in music. Studies on musical expressiveness ([3], [4], [5], and [6]) made clear which are the choices made during performance in order to give a certain expressive intention. A new coding for the score (the metascore) suitable to the automatic performance of a musical piece was studied. The model was provided with a series of controls working on the single note. Besides, special attention was given to the importance of working on groups of notes, hierarchically ordered, and significant for the performance of the piece. The metascore thus obtained is not dependent on the instrument. The model processes this metascore in order to particularize it to a particular instrument family. Besides the fact that it was developed mainly for western classical music, the model showed a general validity in its architecture, even if it needs some tuning of the parameters.

References

[1] Friberg, A. "Generative Rules for music performance: a formal description of a rule system". *CMJ*, 15(2), pp. 49-55. 1991

[2] Gabrielsson, A., Juslin, P. "Emotional expression in music performance". *Psychol. of music*, 24. pp 68-91. 1996

[3] Canazza S., De Poli G., Rinaldin S., & Vidolin A. Sonological analysis of clarinet expressivity. In: M. Leman (Ed.) "Music, Gestalt, and Computing ". Berlin: Springer-Verlag. pp 431-440. 1997

[4] Canazza S., De Poli G., Roda' A., & Vidolin A. "Analysis and synthesis of expressive intentions in musical performance". In *Proc. of the ICMC 1997*. pp. 113-120. Tessaloniki: ICMA. 1997

[5] Battel G.U., Fimbianti R. "Analysis of expressive intentions in pianistic performances". In *Proc. of the Int. Kansei Workshop 1997*. Genova: Associazione di Informatica Musicale Italiana. pp. 128-133. 1997

[6] Canazza S., De Poli G., Di Sanzo G., Vidolin A. "A model to add expressiveness to automatic music performance". In *Proc. of the ICMC 1998*. Ann Arbouri: ICMA. 1998 [in press]