Kinematics-energy space for expressive interaction in music performance

Sergio Canazza, Giovanni De Poli, Antonio Rodà, Alvise Vidolin, Patrick Zanon
CSC – DEI, University of Padova
Via Gradenigo 6/a, 35131 Padova, Italy
{canazza, depoli, vidolin, patrick}@dei.unipd.it, ar@csc.unipd.it
http://www.dei.unipd.it/~musica/

Abstract

A musical interpretation is often the result of a wide range of requirements on expressiveness rendering and technical skills. Aspects that are indicated with the term expressive intention and which refers to the communication of moods and feelings are being considered more and more important in performer-computer interaction during music performance.
Recent studies demonstrated that by modifying opportunely systematic deviations introduced by the musician it is possible to convey different sensitive content like expressive intentions and emotions. We present an abstract space, that can be used for the user interface, which represents, at an abstract level, the expressive content and the interaction between the performer and the expressive engine. This space was derived by multidimensional analysis of perceptual tests on various professionally performed pieces ranging from western classical to popular music. This space reflects how the musical performances are organized in the listener's mind.

1 Introduction

It is well known that music can be used to communicate feelings and emotions; but this aspect is not yet well studied from the scientific point of view. In music the communication chain between the transmitter and the receiver has a different number of elements, depending on the musical repertory. In tonal western music it includes: the composer, the score, the performer, the acoustic signal and the listener; in electro-acoustic music the composer can work directly with the sound, using an opportune support (typically magnetic tape), avoiding the role of the performer. On the other hand in improvised music, and this is particularly true for the jazz repertory, the role of the composer is merged with the performer one, so there is no need of a written score. Few studies empirically investigated all the elements of musical communication; in this work the attention is focused on the communication between the performer intentions and the listener experience, with special regard to the communication of emotions, and, more generally, of expressive content. Music can express emotions in many different ways: emotions can be linked to a particular situation, can be generated by deviations from listener's expectations or reflect the emotional structure of the performer and of the listener. Obviously these categories are not mutually exclusive: a generic musical event can involve more than one of these possibilities. This work is focused on the communication of expressive content between performer and listener.

Many studies were developed about the perception of emotions in music, most of all on tonal western music [10], [13], [1]. These studied demonstrated that even only a short musical phrase of a composition can reflect a large degree of different emotions: tests were conducted asking to a number of invited listeners to judge different composition with the help of a list of adjectives. Another approach regards the study of the emotions related to the systematic variations of rhythmic-harmonic-melodic structures of different short musical pieces or simple tones sequences purposely written [14]. But no one of these studies investigated different properties of expressiveness which could emerge from the same composition, when performed according to different expressive intentions. In fact they studied only the different expressive properties which can be read in the written score, like pitch and note duration, but, on the other hand, the (micro)variation which a musician introduced in his/her performance are not considered. There can be deviations in timing, in dynamics, in timbre, in tempo from the indications written on the score; they are, generally, different according to the musical genre, to the used instrument and also to the performer. There are implicit rules linked to different musical genres which usually are transmitted orally and used in the instrumental practice; moreover the notation on the score can sensibly vary in different musical genres and historical periods. However, even if the same score is used, different players can considerably change the performance.

To complete the communication channel from performer to listener it is also necessary to study the listener experience, with the aim of observing how performer intentions are captured by the listener, that is if they share a common coding of musical emotions. This approach was already suggested by Seashore [15], who asserted that the psychophysics relations between the performer and the listener are fundamental for the comprehension of the microstructures of the musical performance. Some interesting studies approach this aim as, for instance, the papers on music perception by Sloboda [16]; besides, the importance of emotional aspects in the musical contest was deeply studied by Gabrielsson [8] and [9].
This paper presents an abstract space for expressive interaction in music performance, derived by perceptual analyses of interpretations of a musical phrase played 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.
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, human-machine interface and, more generally, in multimedia systems.

2 Method

As material, some musical performances, linked to certain adjectives that had been chosen to suggest and stimulate different interpretations of a musical phrase, were recorded. Sensorial type adjectives were chosen since this type seemed more suitable for sonological interpretation. Emotional type adjectives were, for the present, avoided in that it seemed more opportune to limit the overall semantics under examination. Thus, the choice of adjectives fell on the following: light, heavy, soft, hard, bright and dark and as such, each had its opposite (soft vs. hard) in order to deliberately induce contrasting performances on the part of the musician. A *natural* performance (a interpretation driven by musical praxis) was also recorded. Seven different interpretations of a fragment of Mozart's K622 Concert for Clarinet, were performed by a professional clarinet player and recorded in monophonic digital form at 16 bits and 44100 Hz at the CSC, Padua University.
Initially the focus was on the judgement dimensions used by two groups of subjects called in to listen to the various interpretations of the same musical piece.
The subjects were asked to describe along two adjective scales (different in the two experiments). The adjectives used in the scales will be described within the single experiment. The subjects were given sheets on which each adjective was represented by a line. On this line, the subjects marked on a scale, ranging from nil to extremely, how much of the quality represented by the adjective that they felt the stimulus possessed. The recorded performances were played over loudspeakers to all subjects in one session (experiment 1) and in two sessions (experiment 2). The playback was stopped after the first representation of each stimulus and after a while, the experimenter played the repetition of the same sequence. The experimenter then checked that all subjects had finished their responses before he played the following stimulus. Each experiment took about 20 min. The marks on the lines were measured and submitted to factor analysis.
A double factor analysis was applied to this measure of perceptive order and using, in the first instance, the performance then the adjectives, as variables in the analysis. The performance was used in order to see how the listeners organized the pieces in their own minds, and thus learn how many dimensions could, in reality, be determined. Then the matrix of the subjects' replies is been transposed: from the factor score, it was possible to determine into which sector of the semantic space, as defined by the adjectives, the performance had been placed. A MultiDimensional Scaling (MDS) type analysis was used to verify by how much the various musical interpretations could, effectively, be distinguished one from the other. A Cluster Analysis, as well as the previous ones, was also used in order to see if there was any similarity between the two listening groups and based on the cumulative data emerging from the tests carried out on the two groups of subjects. The results deduced from the statistical analysis were, finally, used to try to construct a model that represents a musically expressive space. The validity of this space was verified by constructing some musical synthesis. The results showed that the performer's intentions and the listeners' impressions, in general, agreed.

3 Experiment 1

Two tests were carried out, respectively, on a group of twelve musicians who had studied at the Pollini Conservatory in Padua and a group of twelve subjects who had no specific musical preparation. Both groups were asked to describe the piece of music heard along a graduated scale of the six adjectives mentioned above. An agreement index was calculated (Robinson's A) to verify the consistency of the replies. This coefficient indicate the degree of agreement expressed by the 12 trained musicians regarding the distribution of the variable *performances*:

$$\overline{A} = 1 - \frac{\sum_i \sum_j (x_{ij} - \mathbf{x}_i)^2}{\sum_i \sum_j (x_{ij} - \mathbf{x})^2} \qquad (1)$$

where x_{ij} is the value attributed to the adjective i by the subject j; $\mathbf{x}_i$ is the average of the values given to adjective i, while $\mathbf{x}$ is the overall average. There was a surprising degree of agreement between the subjects, despite the highly subjective nature of the question, reaching a value of 0.592.

From the MDS it was seen that the listeners had, effectively, arranged the pieces separately in their own minds and had tended to place them at an equal distance from the naturally played piece. The Euclidean distances deduced from the replies are shown in the triangular matrix in Tab. 1.

	Bright	Dark	Hard	Light	Soft	Natural	Heavy
Bright	0						
Dark	2.65	0					
Hard	3.17	2.06	0				
Light	0.85	1.57	2.65	0			
Soft	2.65	2.16	4.08	1.68	0		
Natural	2.06	0.85	2.16	0.85	1.92	0	
Heavy	2,65	0.85	1.68	1.68	2.65	0.85	0

Tab. 1: Matrix of the Euclidean distances

From the results of the factor analysis applied to the data, it was clear that the subjects had placed the performances along only three axes. The first three factors, in fact, explained 85.2% of the total variance and were the only ones having eigenvalues greater than 1. Tab. 2 shows the percentage variance explained by the factors and their respective eigenvalues. The factor loadings, after Varimax rotation, are listed in Tab. 3.

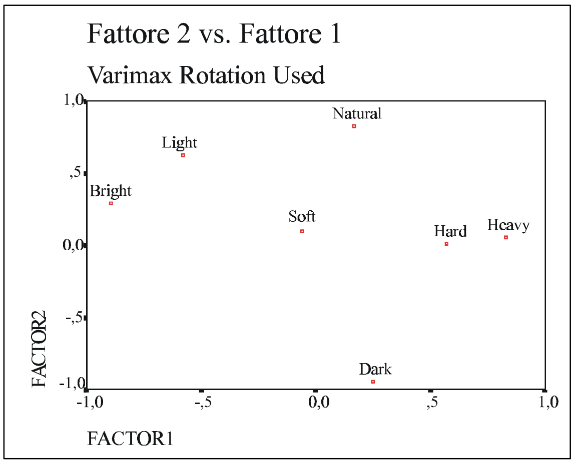
Factor	Eigenvalue	Pct of Var	Cum Pct
1	3.33	47.5	47.5
2	1.48	21.1	68.7
3	1.16	16.5	85.2
4	0.73	10.4	95.6
5	0.31	4.4	100
6	0.00	0.0	100
7	0.00	0.0	100

Tab. 2: Eigenvalue and variance explained by the factors.

	Factor 1	Factor 2	Factor 3
Bright	-0.89	0.29	-0.02
Dark	0.25	-0.94	0.17
Hard	0.57	0.01	-0.61
Light	-0.58	0.63	0.51
Soft	-0.06	0.10	0.88
Natural	0.17	0.83	0.45
Heavy	0.83	0.06	-0.16

Tab. 3: Factor loadings after the Varimax rotation.

The three dimensional space (fig. 1a and 1b) so obtained was, therefore, a model of expressiveness, by which the subjects arranged the pieces in their own minds. Fig. 1b shows the arrangement of the pieces along factor 1 and 3. Acoustical analysis of the performances [2] shown that factor 1 is closely correlated with Tempo and can be interpreted as 'Kinematics' factor; while the factor 3 was connected to attack time, Legato/Staccato, Intensity and can be interpreted as 'Energy' factor. The natural vs. dark were placed at the extreme ends of the factor 2 and appeared to depend on the expressiveness of the score structure (in fact, for example, 'dark' performance appear very unnatural for Mozart melody).



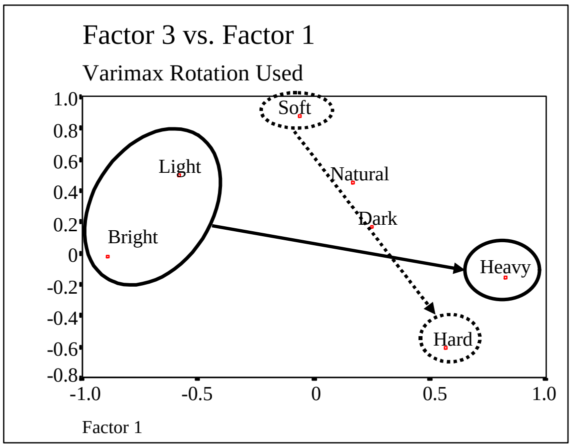


Fig. 1: Graph of factor loadings deduced from factor analysis.

From the factor analysis limited to the data obtained from the test with non musicians, on the other hand, the pieces were subdivided along one single judgement dimension. A dominant factor (60.2% of the total variance), in fact, is been seen, and which tended to divide the pieces into two groups only, separating the soft, light and dark from the rest.
Even the dendrogram deduced from the cluster analysis (see Fig. 2) showed the clear distinction between the group of trained musicians (subj.1 - subj.12) which was highly cohesive, or rather, a small distance index and the group of non musicians (subj.13 - subj.24) which showed a greater variance in the judgements given.

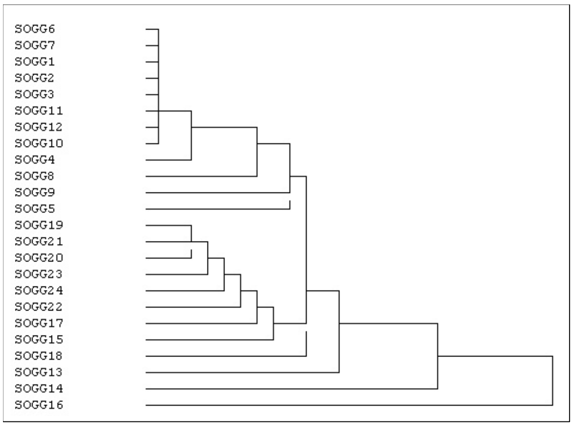


Fig. 2: Cluster analysis dendrogram using the centroid method. The ascissa represents the distance index between the subjects.

4 Experiment 2

A new test carried out on the same musical performances but giving the two groups of subjects a wider semantic choice, allowed for a more detailed study of the listeners' judgement categories. The seventeen new adjectives, again of a sensorial nature, were chosen so as to offer the subjects an exhaustive sampling of a semantic space and here, the list of chosen adjectives did not include their opposites. This new list of adjectives (which did not contain those in the previous list) included the following: nero (black), greve (oppressive), grave (serious), tetro (dismal), massiccio (massive), rigido (rigid), soffice (mellow), tenero (tender), dolce (sweet), aereo (airy), lieve (gentle), spumeggiante (effervescent), vaporoso (vaporous), fresco (fresh), brusco (abrupt), netto (sharp).
Once again, the experiment consisted of two tests, the first using 12 subjects trained at the Conservatory (eight of whom had not taken part in the first experiment), and twelve non musicians (none of whom had taken part in the first experiment).

The Robinson's agreement index is 0.5121. The Euclidean distances deduced from the replies are shown in the triangular matrix in Tab. 4. From their plot (see fig. 3), it can be notice how the listeners had tended to place the stimuli at an equal distance from the naturally played piece. In fact, natural performance is placed on the centre of the space, while other performances are disposed circularly following the order bright, light, soft, heavy, dark and hard.

	Bright	Dark	Hard	Light	Soft	Natural	Heavy
	1	2	3	4	5	6	7
Bright	,00						
Dark	2,74	,00					
Hard	1,72	1,38	,00				
Light	1,35	3,04	2,55	,00			
Soft	2,49	2,81	3,01	1,38	,00		
Natural	1,33	1,64	1,36	1,35	1,65	,00	
Heavy	2,94	0,28	1,72	3,14	2,81	1,72	,00

Tab. 4: Matrix of the Euclidean distances

The factor analysis of the performance confirmed the same two factors which had emerged from the previous experiment. From the factor analysis carried out using the adjectives as the variables, it was seen that the subjects clearly recognised the performer's intentions.

	Factor 1	Factor 2
Bright	0.89	0.04
Dark	-0.76	0.55
Hard	-0.03	0.82
Light	0.64	-0.61
Soft	-0. 02	-0.87
Natural	0.65	0.18
Heavy	-0.82	0.48

Tab. 5: Factor loadings after the Varimax rotation.

Fig. 3 shows the graph of the factor loadings and, by means of the factor scores, it was possible to insert the performances into the factor space deduced from the adjectives (see [11]). Similar interpretation of fig. 1b can be applied to this space. X-axis heavy-light is related to Kinematics and Y-axis is related to Energy. Sonological analysis on performances of different musical pieces played with the same set of expressive intentions on different instruments were conducted and seems to confirm the interpretation of the two main axes ([6], [7]).
Thus we will use this interpretation of kinematics-energy space as indication of how listeners organized the performances in their own minds. Using these two

factors as axis an abstract control space can be defined, and - for its construction - it is quite similar to the perceptive one.

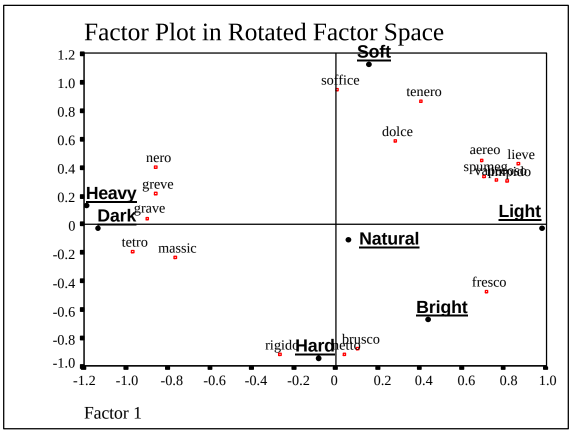


Fig. 3: Factor analysis on the adjectives. The first factor explains 49.6% of the total variance, the second 28.5%.

5 Kinematics-energy space for expressive interaction

With the previous experiments we were able to devise an interpretation of an abstract space for expressive interaction in music performance. In order to verify the validity of the results obtained and to find more generalized relationship between the two axes of the space and sonological parameter, the analysis-by-synthesis method was, finally, used. Therefore, a system to generate automatic expressive music performance was developed and the abstract control space (obtained by the perceptual analyses) is employed as user interface to control, at an abstract level, generated performances.

Analyses shown that to render a particular expressive intention, the performer uses different strategies depending on personal choices, on the score structure, and on expressive controls offered by the instrument played. The system have been developed using the results of perceptual and acoustic analyses made on professional performances. By statistical analysis, data reduction and a fine tuning based on synthesis and evaluation we found a model for mapping points in the Energy-Kinetics space to deviations of the main expressive parameters as Tempo, Legato, Intensity ([4], [5]). Thus, to render the expressive content, the model use a reduced set of controls, which has been demonstrated to be more representative and independent of the instrument and the score. The system calculates these parameters and determines the deviations that must be applied to the score for reproducing a particular expressive intention.

Fig. 4 shows a mapping surface between points of the abstract control space and the parameter *Ktempo*, (a value greater than one stands for a *rallentando*, while a value lower than one stands for an *accelerando*). On the basis of movements of the pointer on the x-y plane, the variations of the parameter *Ktempo* to be applied to the performance are thus computed.

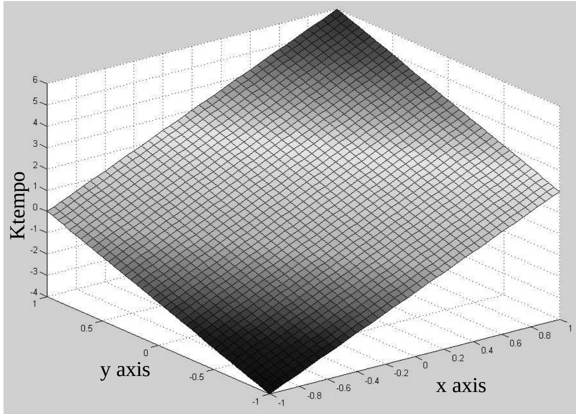


Fig. 4: mapping surface between points of the Energy-Kinetics space (x-y plane) and the parameter *Ktempo* (z axis).

It can be notice that by a suitable choice of the internal parameters of the model, the performance resulting from continuous movements in the abstract space renders in a coherent and smoothly way the intended intermediate expressive intention. Moreover, the continuous character of the space allows a conceptual interpretation of every point and trajectory in the abstract control space.

6 Discussion and conclusion

Various studies on performance have led to suggesting models that could render synthesized music less monotonous and mechanical. An attempt has been made, here, to go beyond this, in that we tried to identify an expressive model that could render different expressive intentions of a human performer. In order to do so, measurements of a perceptive nature were used. By analysing these results, two quite distinct expressive directions were observed, one related to the 'kinematics' parameters and the other related to the 'energy' of the pieces. The tests, in fact, showed two main factors that can be related to sonological variables. From a musical point of view, in fact, the first axis sets rapid Tempo (bright and light pieces) against slow (heavy) or moderate (soft) Tempo. The second factor is, on the other hand, mainly correlated to Energy related parameters as Intensity. However, it should be remembered that a correspondence of the *one on one* type between the deduced factors and the sonological variables was not demonstrated. Studies on expressiveness have, in fact, shown (see [12]) that some physical parameters depend on others, such as, for example, the expressive variations of the temporal microstructure with respect to the overall Tempo of the piece. Interpretation of the factors can not, therefore, exclude this. All the tests carried out up to the present lead, in any case, to the pieces being arranged along these two axes which have, consequently, been interpreted as two expressive types.

It is interesting to note how it was also possible to insert the automatic generated performances along these two directions, judging from the subjects' replies [3]. This would demonstrate that the performer's expressive intentions had been captured and that they could be applied to other scores.

Therefore, a system to generate automatic expressive music performance based on the abstract space and analysis results was developed. It confirmed the validity of the kinematics-energy space both of perceptual and rendering point of view.

This model allows to realize an expressive multimodal environment in which a generic media can be used as input to influence a musical performance. The application and the mapping between video and audio was realized in the framework of the IST research project MEGA.

References

[1] Brown R. W. 1981. "Music and language". In Documentary report of the Ann Arbor symposium. Pp. 233-265. Reston, VA.

[2] Canazza S., De Poli G., Rinaldin S., Vidolin A. 1997b. "Sonological analysis of clarinet expressivity". In: M. Leman (ed.) Music, Gestalt, and Computing - Studies in Cognitive and Systematic Musicology. Berlin, Heidelberg: Springer-Verlag.

[3] Canazza S., De Poli G., Vidolin A. 1997a. "Perceptual analysis of the musical expressive intention in a clarinet performance". In: M. Leman (ed.) Music, Gestalt, and Computing - Studies in Cognitive and Systematic Musicology. Berlin, Heidelberg: Springer-Verlag.

[4] Canazza S., De Poli G., Drioli C., Rodà A., Zamperini F. (2000). "Real-time morphing among different expressive intentions in audio playback". Proc. of ICMC, Berlin, 27 august-1 september pp. 356-359.

[5] Canazza S. De Poli G., Drioli C., Rodà A., Vidolin A. (2000). "Audio morphing different expressive intentions for Multimedia Systems". IEEE Multimedia, July-September, Vol. 7, N° 3, pp. 79-83.

[6] De Poli, G., Rodà, A., Vidolin, A., 1998. "Note-by-note analysis of the influence of expressive intentions and musical structure in violin performance". Journal of New Music Research, Vol. 27, N. 3, pp. 293-321.

[7] Battel G. U., Fimbianti R., 1997. "Analysis of expressive intentions in pianistic performances" in Proceedings of KANSEI - The Technology of Emotion AIMI International Workshop, pp. 128-133, Genova.

[8] Gabrielsson A. 1997. "Music performance". In: D. Deutsch (Ed.), The psychology of music. New York: Academic Press.

[9] Gabrielsson A. Juslin P. 1996. "Emotional expression in music performance". Psychology of music. 24: 68-91.

[10] Hevner K. 1935. "Expression in music: a discussion of experimental studies and theories". Psychological review. 42: 186-204.

[11] Repp B.H. 1990. "Patterns of expressive timing in performances of a Beethoven minuet by nineteen pianists." Journal of Acoustical Society of America. 2: 622-641.

[12] Repp B.H. 1995. "Quantitative effects of Global Tempo on Expressive timing in music performance: some perceptual evidence." Music perception. 13(1): 39-57.

[13] Rigg M. G. 1964. "The mood effects on music: a comparison of data from earlier investigations". Journal of Psychology. 58: 427-438.

[14] Scherer K. R. 1991. "Emotion expression in speech and music". In J. Sundberg (Ed.) Music language speech and brain. Pp. 146-156. London: MacMillan Press.

[15] Seashore H. G. 1937. "An objective analysis of artistic singer". In C. E. Seashore (Ed.) Objective analysis of musical performance, Vol IV. Pp. 12-157. Un. of Iowa.

[16] Sloboda J. A. 1985. "Expressive skill in two pianist". Canadian Journal of psychology. 39(2): 273-293.