ON THE EXPRESSIVE GESTURES
LOOKING FOR COMMON TRAITS BETWEEN MUSICAL AND PHYSICAL DOMAIN

Sergio Canazza * [a], Giovanni De Poli[a] and Antonio Rodà[b]

[a]*Dept. of Information Engineering, University of Padova, Italy*
[b]*AVIRES Lab., University of Udine, Italy*

ABSTRACT

This paper investigates the relations between music contents and other non-verbal ways of expression. In particular, the idea that music is, among other things, a performing art and it is composed by what we commonly call musical gestures, suggested us to study if and how some aspects of music expression can be associated with the properties of a physical gesture. An experiment was carried out in order to verify whether subjects are able to associate musical excerpts with physical properties, such as elasticity, inertia, and friction, represented by means of a set of computer generated haptic stimuli. The comparison with a previous study on the affective response to the same set of excerpts allows us to point out relations among the physical, the musical, and the affective domains.

Keywords: Expressive information processing, Musical and physical gestures, Audio analysis, Perceptual analysis.

1. INTRODUCTION

Nowadays, the study of music is not limited to the artistic field. Indeed, the power of music to arouse in the listener a rich set of sensations, such as images, feelings, or emotions, can have many applications. In the information technology field, a musical signal can contribute to the multimodal/multisensory interaction, communicating events and processes, providing the user with information through sonification, or giving auditory warnings. In this sense, sound design requires great attention and a deep understanding of the influence of musical parameters on the user's experience. The relation between music and emotions has been largely investigated by the scientific community (see [1] for a review). Recently, Bigand [2] investigates the emotion conveyed by mu-

*Corresponding author: Via Gradenigo 6/B, 35131 Padova (Italy), sergio.canazza@unipd.it

sical pieces, carrying out some experiments in which the participants were encouraged to focus on their own emotional experience. Musically trained and untrained listeners were asked to listen to 27 different musical excerpts and to group those that conveyed similar subjective emotions. By means of the multidimensional scaling method (MDS), a two-dimensional space was found to provide a good fit of the data, with arousal and emotional valence as the primary dimensions (Fig. 1). In particular, the excerpts resulted grouped in four clusters, hereafter named *affective clusters*, characterized by i) high arousal and high valence (HAHV), ii) low arousal and high valence (LAHV), iii) high arousal and low valence (HALV) and iv) low arousal and low valence (LALV). However, music experience is evidently not limited to the emotions. Music experience is a very complex issue, that can be described in manifold ways. Often, musicians and listeners talk about music using terms borrowed from different sensorial modalities: e.g., a music piece can be described by words belonging to the tactile domain, such as *hard*, or to the visual domain, such as *bright*. Canazza et al. [3] investigated the relations between musical parameters and sensorial adjectives by means of some perceptual experiments. Camurri et al. [4] defined a multi-layer model to represent common characteristics of different sensorial domains, such as sounds and physical gestures.

We move from the assumption that the numerous ways to describe the characteristics of a music piece are not mutually exclusive, but rather complementary points of view of the same complex experience. Each description is a metaphor that allows to represent particular aspects of the musical experience, without totally representing that experience. This paper aims at investigating the relations between music contents and non verbal ways of expression. In particular, the idea that music is, among other things, a performing art and it is composed by what we commonly call musical gestures, suggest us to study if and how some aspects of music expression can be associated with physical properties, such as elasticity, inertia, and friction. In a previous work [5] the present authors carried out an experiment using a set of simple monophonic musical excerpts and three computer generated haptic stimuli (see Sec. 2.1), called attractors, simulating three different physical properties. The theme from Händel's Sonata HWV 379 in E Minor Op. 1 No. 1 (Adagio) and the traditional song *Twinkle Twinkle Little Star* were played by a solo instrument (violin, flute, and guitar) several times in order to convey expressive intentions Happy, Sad, Angry and Calm (affective metaphor), Light, Heavy, Soft, and Hard (sensorial metaphor). Participants were asked to listen to each musical excerpt and to associate them with one of the three attractors. The results showed that subjects are able to consistently associate musical excerpts with the haptic stimuli. In order to verify if this result can be generalized to complex polyphonic musical pieces, we carried out an experiment (see Sec. 2) with the same 27 musical stimuli used by Bigand [2]. Participants were asked to listen to each musical excerpt and to associate it with one of three computer generated haptic stimuli (see Sec. 2.1). The statistical analysis of the responses (see Sec. 2.2), showed that the listeners organized the musical stimuli in five clusters, later on named *haptic clusters*, each one associated with one or a combination of the three attractors. In addiction, in order to investigate the nature of these associations (both the four affective clusters and the five haptic clusters), we carried out a detailed acoustic analysis of the musical stimuli (see Sec. 3). This analysis allowed us to relate the subjects' answers with the music features and to identify relations among the physical, the musical, and the affective domains.

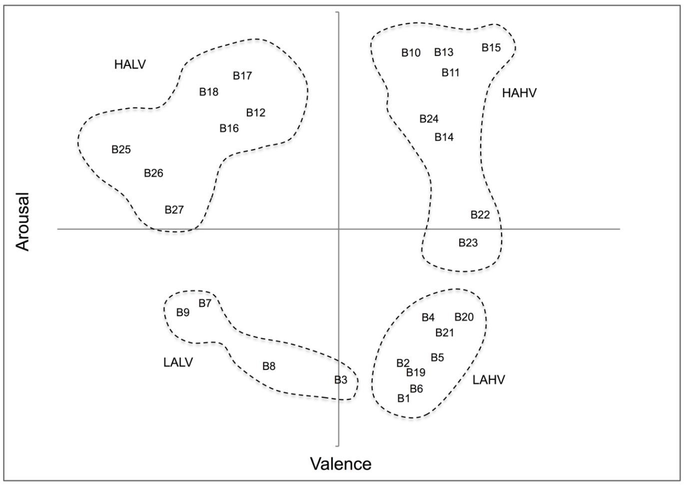

Figure 1: The 27 excerpts of the experiment in Bigand [2], mapped on a two-dimensional space. Dashed lines represent the four affective clusters: high arousal and high valence (HAHV); high arousal and low valence (HALV); low arousal and high valence (LAHV); low arousal and low valence (LALV). Figure adapted from [2].

2. PERCEPTUAL EXPERIMENT

We carried out a perceptual experiment to verify if some aspects of the musical expression can be associated with basic physical properties, such as elasticity, inertia, and friction. Participants were asked to listen a set of musical excerpts and to associate them with one of three haptic stimuli.

2.1. Materials and apparatus

A set of musical excerpts and a set of haptic stimuli, called attractors, were used in the experimental setup. The musical excerpts are the same used in Bigand [2]: 27 musical stimuli with a polyphonic structure selected from recordings of the Western music repertoire, from XVII to XX century. The set of attractors is composed by three haptic stimuli generated by means of a *Phantom Omni*[1], a haptic device with six degrees of freedom, controlled in order to simulate the basic effect of a mechanical system composed by mass, spring, and damper: (i) an elastic force, with elastic constant K_{el} (Elasticity - E); (ii) a viscous damper, with damping coefficient μ_V (Friction - F); (iii) an inertial mass m (Inertia - I).

The excerpts were represented on a computer screen by means a visual interface implemented using the real-time sound synthesis program PD (Pure Data)[2]. The interface consists on three buttons displayed on the top of the screen, associated with the three attractors, and on a set of buttons listed in column associated with the 27 musical excerpts, which are presented (in random order) to the participants. Participants were allowed to listen to the excerpts and to try the attractors as many time as they wished just by pressing the corresponding button. Each excerpt was also associated with a radio button where the participants could only select one choice (a, b, or c)

[1] http://www.sensable.com/haptic-phantom-omni.htm

[2] http://puredata.info

employing a three-alternative forced-choice (3AFC) method. A total of 21 subjects participated to the experiment (7 male, 14 female). Of these, 3 subjects had a professional musical training for 5 years at least; 7 subjects played an instrument; 11 subjects did not have any musical training. Participants were of Italian nationality and were aged from 22 to 53 years (30 years average). The duration of the test was about 20 minutes.

2.2. Results

Tab. 1 shows the contingency table of the subjects' responses with rows representing the 27 musical excerpts and columns representing the 3 haptic attractors. The Pearson's Chi-squared test denoted a strong relation between musical excerpts and attractors ($\chi^2 = 205.2$, $df = 52$, $p < 0.001$) and confirmed that subjects are able to distinguish the different haptic attractors and to use them to classify the musical excerpts. In order to verify if the haptic stimuli can be related to the affective clusters, a new contingency table with 4 rows was obtained by grouping the musical excerpts following the composition of the affective clusters (reported in the first column of the Tab. 1). A $\chi^2 = 135.0$ ($df = 6$, $p < 0.001$) confirmed that subjects recognized a relation among clusters and attractors. In particular, significative relation was found between the HAHV cluster and the Elasticity attractor ($\chi^2 = 35.9$, $df = 1$, $p < 0.001$), between the LALV cluster and the Inertia attractor ($\chi^2 = 56.2$, $df = 1$, $p < 0.001$), and between the LAHV cluster and the Inertia attractor ($\chi^2 = 29.9$, $df = 1$, $p < 0.001$). On the contrary, no statistically significative relation has been found with the Friction attractor. The contingency table was submitted to Simple Correspondence Analysis in order to graphically represent the degree of association between musical stimuli and attractors. Finally, we carried out a k-means analysis in order to identify clusters of stimuli. The results are mapped in Fig. 2 and show that the stimuli can be grouped in five clusters. Three of them include one of the attractor and in the next we will refer to these cluster with the name of the attractor: cluster Inertia (I), cluster Elasticity (E), and cluster Friction (F). The others are in the middle between two attractors, so they have been named the cluster Inertia-Friction (IF) and the cluster Elasticity-Friction (EF).

The comparison between the Figs. 1 and 2 shows that the same 27 stimuli have been organized differently in the two experiments (the affective and the haptic one). Although some relations can be found among the affective and the haptic clusters (e.g., all the excerpts of the cluster LALV are mapped inside the cluster I, all the excerpts of the cluster E are mapped inside the clusters HAHV and HALV), many differences can be identified in the two placements. To explain these differences, we hypothesized that the two experimental setups induce the listeners to classify the excerpts on the base of different musical features: in the Bigand's experiment the listeners were asked to focus on their own affective response so that would be reasonable to hypothesize that they paid more attention to the musical features related to the affective domain; differently, in our experiment the subjects were asked to pay attention on the haptic response, so they may have been focused on different features. In order to verify this hypothesis we carried out an acoustic analysis of all 27 excerpts, as reported in the following section.

Table 1: Contingency table of the subjects' responses.

Cluster	Excerpt	Friction	Elasticity	Inertia
HALV	B12	7	13	1
	B16	9	8	4
	B17	4	15	2
	B18	9	8	4
	B25	9	5	7
	B26	12	5	4
	B27	5	8	8
HAHV	B10	9	9	3
	B11	12	3	6
	B13	9	12	0
	B14	7	12	2
	B15	6	14	1
	B22	7	11	3
	B23	7	11	3
	B24	6	13	2
LALV	B3	3	3	15
	B7	3	3	15
	B8	3	2	16
	B9	3	5	13
LAHV	B1	3	3	15
	B2	8	8	5
	B4	9	1	11
	B5	8	1	12
	B6	4	1	16
	B19	10	4	7
	B20	9	3	9
	B21	7	3	11

3. ACOUSTIC ANALYSIS

3.1. Feature extraction

In order to relate the subjects' answers with the musical features, we carried out a detailed acoustic analysis of the musical stimuli. A set of acoustic features were calculated for each excerpt. The set was chosen among those features that in previous listening experiments [6] were found to be important for discriminating different emotions and were also used to classify the style [7] and the expressive content in musical performances [8] and [9]. We computed the features using non-overlapping frames (of 46-ms length), and then we considered their mean value within sliding windows (with 4-s duration and 3.5-s overlap). The window size allows to include a reasonable number of events, and it roughly corresponds to the size of the echoic memory. In total, we collected a set of 13 audio features on about 1700 different windows. See Tab. 2 for a formal description of the features. The features are: a) *Zerocross* consists in counting the number of times the audio signal changes sign. It can be considered as a simple indicator of noisiness; b) *RMS* takes into account the global energy of the signal, computed as the root average of the square of the amplitude (root-mean-square); c) *Centroid* is the first moment of the spectral amplitude. It is related with the impression of 'brightness" of a sound [10], because a high centroid value means that the sound energy is concentrated at the higher frequencies; d) *Brightness* measures the amount of energy above the frequency of 1000 Hz. The result is expressed as a number between 0 and 1; e) *Spectral ratios* (SRs) over different frequency bands of of the spectrum are other useful indications of the spectrum shape. The spectrum is divided in three regions: below 534 Hz (SRl), from 534 to 1805 Hz (SRm), and above 1805 Hz (SRh); f) *Rolloff* is the frequency such that the 85% of the total energy is contained below that frequency. It is related to the "brightness" of the sound; g) *Roughness* is calculated starting from the results of Plomp and Levelt [11], that proposed an estimation of the dissonance degree between two sinusoids, depending on the ratio of

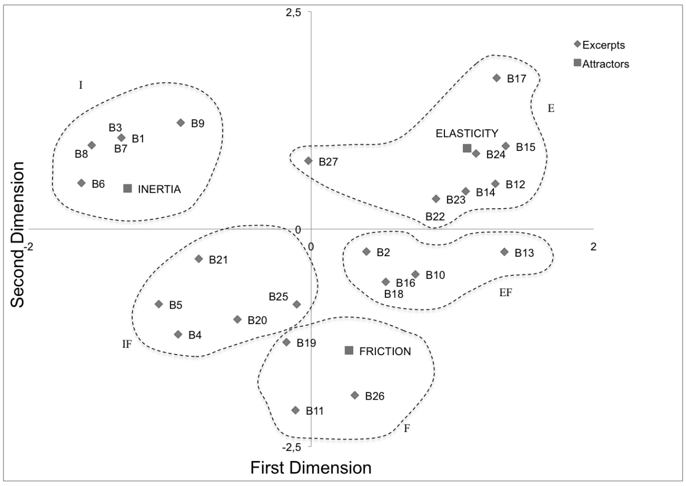

Figure 2: Correspondence analysis on experiment data. Dashed lines represent the outcome of the cluster analysis.

their frequency. The total roughness for a complex sound can be calculated by computing the peaks of the spectrum, and taking the average of all the dissonance between all possible pairs of peaks [12]; h) *Spectralflux* is the distance between the spectrum of each successive frame; i) *Lowenergy* is the percentage of frames showing less-than-average energy. It is an assessment of the temporal distribution of energy, in order to see if it remains constant throughout the signal, or if some frames are more contrastive than others; l) *Tempo* is the musical velocity of the performance. Since many of the 27 excerpts have a complex polyphonic structure, it is not easy to have a good estimation of this feature using an automatic routine. Then, the $Tempo$ of each excerpt was estimated by means of the manual annotations of an expert; m) *Modality* is a basic aspect of the musical structure. In Western tonal music there are two modes, named major and minor mode. Also in this case, we used the annotations of an expert who analysed the musical sheets.

Starting from the calculated features, we selected the set of features related both to the affective and the haptic clusters. The feature selection procedure consists in finding the audio features that give the highest classification ratings. A wrapper approach based on sequential feature selection (SFS) [13] is applied with reference to a linear classifier. The feature selection procedure was applied twice. On the first time we selected the set of features that classify the 27 excerpts, with a minimum error rate, following the classes specified by the four affective clusters. The SFS process selected the following four features, in order of selection: $Tempo$, $Modality$, $Centroid$, and RMS. The minimum error rate is 18%. Then, we selected the set of features that classify, with a minimum error rate, the 27 excerpts following the classes specified by the five haptic clusters. The SFS process selected the following three features, in order of selection: $Tempo$, $Rolloff$, $Zerocross$. The minimum error rate is 35%.

3.2. Results

Tab. 3 shows the mean values of the four features selected for the affective clusters, calculated for each excerpt. The excerpts belonging to the clusters with high arousal (i.e. HAHV and HALV) are characterized, with a few exceptions, by a high value of $Tempo$. In particular, the mean value

Table 2: List of the acoustic features. The signal x is blocked in M frames of N samples. Let be $x(f,n)$ the signal amplitude of the sample n at the frame f; $X(f,k)$ the spectrum magnitude of the bin k at the frame f and $F(f,k)$ the center frequency of that bin; k_{f_t} the bin corresponding to the frequency f_t; $\mathrm{I}\{A\}$ the indicator function equal to 1 if A is true and 0 otherwise; $\mathrm{sign}(x)$ a function equal to 1 if $x \geq 1$ and 0 otherwise; $\mathrm{rms}(x(f))$ the RMS value over the frame f and $\mathrm{rms}(x)$ the RMS value over the entire signal x.

RMS	$\sqrt{\frac{1}{n}\sum_{n=1}^{N} x(f,n)^2}, f = 1, ..., M$
Zerocross	$\sum_{n=1}^{N-1} \mathrm{I}\{\mathrm{sign}(x(f,n)) \neq \mathrm{sign}(x(f,n+1))\}, f = 1, ..., M$
Centroid	$\frac{\sum_{k=1}^{N} F(f,k)X(f,k)}{\sum_{k=1}^{N} X(f,k)}, f = 1, ..., M$
Brightness	$\frac{\sum_{k=k_{1000}+1}^{N} X(f,k)}{\sum_{k=1}^{N} X(f,k)}, f = 1, ..., M$
SRl	$\frac{\sum_{k=1}^{k_{534}} X(f,k)}{\sum_{k=1}^{N} X(f,k)}, f = 1, ..., M$
SRm	$\frac{\sum_{k=k_{534}+1}^{k_{1805}} X(f,k)}{\sum_{k=1}^{N} X(f,k)}, f = 1, ..., M$
SRh	$\frac{\sum_{k=k_{1805}+1}^{N} X(f,k)}{\sum_{k=1}^{N} X(f,k)}, f = 1, ..., M$
Rolloff	$f(k_{85})$, where $k_{85} = \min(k_0) : \frac{\sum_{k=1}^{k_0} X(f,k)}{\sum_{k=1}^{N} X(f,k)} > 0.85, f = 1, ..., M$
Spectralflux	$\sqrt{\sum_{k=1}^{N} [X(f+1,k) - X(f,k)]^2}, f = 1, ..., M-1$
Lowenergy	$\frac{\sum_{f=1}^{M} \mathrm{I}\{\mathrm{rms}(x(f))<\mathrm{rms}(x)\}}{M}$

among the excerpts of HALV is $127bpm$, HAHV is $100bpm$, LAHV is $63bpm$, and LALV is $47bpm$ ($F = 11.2$ on 3 and 23 *df*, $p < 0.001$), where *bpm* stands for beats-per-minute. The excerpts belonging to the clusters with low valence (i.e. HALV and LALV) are characterized by a minor modality; all excerpts except number 25 and 26 that are atonal pieces. On the contrary, all the excerpts but one of the HAHV cluster have a major mode and the excerpt 24, taken from a Stravinsky's composition, has an uncertain tonality based on two superposed major chords. The excerpts of the LAHV cluster are mostly characterized by a major mode. A Chi-squared analysis showed that modality is significatively related with the valence factor ($\chi^2 = 14.9$, $df = 2$, $p < 0.001$). In regard to the other two selected features, a high $Centroid$ value characterizes the clusters with high valence (the average value is $1588Hz$ for HAHV, $1573Hz$ for LAHV, $1426Hz$ for HALV, and $1348Hz$ for LALV), whereas a high RMS value distinguishes the clusters with high arousal from the others (the average value is 0.094 for HAHV, 0.080 for LAHV, 0.098 for HALV, and 0.057 for $LVLH$). However, for both these features, the differences are not statistically significative ($F < 0.9$ on 3 and 23 *df*, $p > 0.05$).

The Tab. 4 shows, for each excerpt, the mean values of the selected features. $Tempo$ is, also in this case, the first selected feature. On average, the cluster I is characterized by a value of $50bpm$, cluster E by a value of $113bpm$, cluster F by $80bpm$, cluster EF by $115bpm$, and cluster IF by $85bpm$ ($F = 4.7$ on 4 and 22 *df*, $p < 0.01$). In regard to the $Rolloff$ feature, the mean value over the cluster I is $1902Hz$, the cluster E is $2024Hz$, the cluster F is $3730Hz$, the cluster IF is $1954Hz$, and the cluster EF is $3015Hz$ ($F = 4.4$ on 4 and 22 *df*, $p < 0.01$). Finally, the mean values of $Zerocross$ are 571 for cluster I, 734 for cluster E, 1136 for cluster F, 815 for cluster IF, and 838 for cluster EF ($F = 3.3$ on 4 and 22 *df*, $p < 0.05$).

Table 3: Acoustic features related to the affective clusters.

cluster	excerpt	Tempo [bpm]	Modality	Centroid [Hz]	RMS
HAHV	10	109	major	1643	0.075
	11	53	major	2684	0.091
	13	103	major	1737	0.080
	14	102	major	1141	0.151
	15	145	major	1473	0.067
	22	103	major	1376	0.060
	23	59	major	1047	0.053
	24	123	undetermined	1603	0.174
LAHV	1	61	major	1694	0.086
	2	77	minor	2322	0.067
	4	53	major	1288	0.089
	5	53	major	1075	0.108
	6	50	major	1078	0.086
	19	65	minor	2091	0.105
	20	65	minor	1345	0.051
	21	76	major	1691	0.045
HALV	12	157	minor	1097	0.061
	16	142	minor	1074	0.220
	17	149	minor	1844	0.045
	18	144	minor	1760	0.174
	25	151	undetermined	1725	0.056
	26	88	undetermined	1487	0.054
	27	58	minor	997	0.079
LALV	3	40	minor	1106	0.018
	7	48	minor	1034	0.088
	8	50	minor	1615	0.073
	9	51	minor	1634	0.048

The Tab. 5 qualitatively summarizes the relations among the haptic clusters and the selected features. The cluster I is characterized by a low value in all the features, the cluster E by a high $Tempo$, and the cluster F by a high value of $Rolloff$ and $Zerocross$. In regard to the clusters IF and EF, they seem to have characteristics that are intermediate between two other clusters. In particular, IF is characterized by low $Tempo$ and $Rolloff$, as cluster I, and high $Zerocross$ as cluster F; EF is characterized by high $Tempo$ as cluster E, and high $Rolloff$ and $Zerocross$ as cluster F.

4. CONCLUSIONS

An experiment was carried out to verify if it is possible to identify relations between physical and musical domain. The results confirm that subjects are able to associate haptic stimuli simulating some basic physical properties with polyphonic musical stimuli in a consistent way. The nature of that association has been investigated by means of an in-depth acoustic analysis, that revealed a significative correlation between some musical/acoustic features and the subject's responses: $Tempo$, $Rolloff$ (a feature related to the brightness of the sound), and $Zerocross$ (related to the noisiness of the sound) are the parameters selected to be the most representative of the haptic clusters. The analysis of the acoustic features related to the affective clusters confirms the results of previous researches [14], i.e. the main parameters that characterize the affective responses to music are $Tempo$ and $Modality$. In regard to relations between haptic and affective metaphors, significative relations were found between the $Elasticity$ and $Inertia$ clusters and the arousal dimension. The differences between the haptic and the affective clusters support the idea that the musical experience can be represented by means of several different metaphors, each one focused on different aspects of that experience. Indeed, different experimental setups induced the listeners to classify the excerpts on the base of different musical features: in the Bigand's experiment the

Table 4: Acoustic features related to the haptic clusters.

cluster	excerpt	Tempo [bpm]	Rolloff [Hz]	Zerocross
I	1	61	2313	532
	3	40	1366	499
	6	50	1014	460
	7	48	1415	521
	8	50	2560	629
	9	51	2747	785
E	12	157	1906	602
	14	102	1794	656
	15	145	2436	832
	17	149	3281	959
	22	103	1477	747
	23	59	1309	690
	24	123	2282	883
	27	58	1707	505
F	11	53	5037	1663
	19	65	3659	937
	26	88	2494	807
IF	4	53	1740	956
	5	53	1157	389
	20	65	1474	640
	21	76	2531	879
	25	151	2870	1212
EF	2	77	4066	920
	10	109	2921	889
	13	103	3082	947
	16	142	1759	565
	18	144	3246	871

Table 5: Relation among the haptic clusters and the selected features.

Cluster	Tempo	Rolloff	Zerocross
I	–	-	–
E	++	-	-
F	-	++	++
IF	-	-	+
EF	++	+	+

listeners were asked to focus on their own affective response so they paid more attention to those musical features related to the affective domain ($Tempo$ and $Modality$); differently, in our experiment the subjects were asked to pay attention on the haptic responses, so they have been focused, beside $Tempo$, on features such as $Rolloff$ and $Zerocross$, related to timbric aspects.

4.1. Application scenario

Technology-mediated music access and fruition is more and more becoming an interactive process. Users need interactive tools for managing and browsing a huge amount of music files. Moreover, games like Singstar^{TM}, Ultrastar^{TM}, Guitar-Hero^{TM} (and its open source version Frets on fire^{TM} or its piano version Synthesia^{TM}) and Rock Band^{TM} allow the users to actively participate in the music production process by playing simplified models of musical instruments or moving their hand over an invisible guitar. Almost all new mobile devices (phones, media players) and video-game consoles implement interfaces for music content access and fruition. While a traditional monitor and mouse interface allows a linguistic based interaction, the reduced dimension of the mobile devices requires novel strategies: touch sensitive screens, position and movement sensors allow non-linguistic communication and action-based interaction modalities. Unfortunately, these systems support only limited interaction schemes: e.g., in the musical games the users is only required to press buttons or perform movements on the right time and no effective relation is defined between the gesture properties and the musical features. Although our work is at a basic research level, we believe that a deeper understanding of the relations among musical and physical stimuli may improve the design of multisensorial interfaces, towards an effective mediation technology for music content access and fruition, capable of sensing and responding appropriately to the users actions.

REFERENCES

[1] Juslin, P. N and Sloboda, J. A. *Music and emotion. Theory and research.* Oxford University Press, 2001.

[2] Bigand, E, Vieillard, S, Madurell, F, Marozeau, J, and Dacquet, A. Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. *Cognition and Emotion*, 19(8):1113–1139, 2005.

[3] Canazza, S, De Poli, G, Rodà, A, and Vidolin, A. An abstract control space for communication of sensory expressive intentions in music performance. *Journal of New Music Research*, 32(3):281–294, 2003.

[4] Camurri, A, De Poli, G, Leman, M, and Volpe, G. Communicating expressiveness and affect in multimodal interactive systems. *IEEE Multimedia*, 12(1):43–53, 2005.

[5] Mion, L, De Poli, G, and Rodà, A. Toward an action based metaphor for gestural interaction with musical contents. *Journal of New Music Research*, Accepted for pubblication, 2009.

[6] Juslin, P. N. Communicating emotion in music performance: A review and a theoretical framework. In Juslin, P. N and Sloboda, J. A, editors, *Music and Emotion: Theory and Research*, pages 305–333. New York: Oxford Univ. Press, 2001.

[7] Dannenberg, R, Thorn, B, and Watson, D. A machine learning approach to musical style recognition. In *Proc. Int. Comput. Music Conf. (ICMC97)*, pages 344–347, San Francisco, CA, 1997.

[8] Friberg, A, Schoonderwaldt, E, Juslin, P, and Bresin, R. Automatic real-time extraction of musical expression. In *Proc. Int. Comput. Music Conf. (ICMC02)*, pages 365–367, Goteborg, Sweden, 2002.

[9] Mion, L and De Poli, G. Score-independent audio features for description of music expression. *IEEE Trans. Speech, Audio, and Language Process*, 16(2):458–466, 2008.

[10] Schubert, E, Wolfe, J, and Tarnopolsky, A. Spectral centroid and timbre in complex, multiple instrumental textures. In University, N. W, editor, *Proceedings of the International Conference on Music Perception and Cognition*, Illinois, 2004.

[11] Plomp, R and Levelt, W. J. M. Tonal consonance and critical bandwidth. *Journal of the Acoustical Society of America*, 38(4):548–560, 1965.

[12] Sethares, W. A. *Tuning, Timbre, Spectrum, Scale*. Springer-Verlag, 1998.

[13] Whitney, A. A direct method of nonparametric measurement selection. *IEEE Trans. Comput.*, 20(9):1100–1103, 1971.

[14] Gagnon, L and Peretz, I. Mode and tempo relative contributions to happy–sad judgements in equitone melodies. *Cognition and Emotion*, 17(1):25–40, 2003.