EMOTIONAL RESPONSE TO MAJOR MODE MUSICAL PIECES: SCORE-DEPENDENT PERCEPTUAL AND ACOUSTIC ANALYSIS

Sergio Canazza, Giovanni De Poli, Antonio Rodà
Dep. Information Engineering
University of Padova
{sergio.canazza,giovanni.depoli,antonio.roda}@unipd.it

ABSTRACT

In the Expressive Information Processing field, some studies investigated the relation between music and emotions, proving that is possible to correlate the listeners main appraisal categories and the acoustic parameters which better characterize expressive intentions, defining score-independent models of expressiveness. Other researches take to account that part of the emotional response to music results from the cognitive processing of musical structures (key, modalities, rhythm), which are known to be expressive in the context of the Western musical system. Almost all these studies investigate emotional responses to music by using verbal labels, that is potentially problematic since it can encourage participants to simplify what they actually experiencing. Recently, some authors proposed an experimental method that makes no use of verbal labels. 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. In order to emphasize other latent dimensions, a perceptual experiment and a comprehensive acoustic analysis was carried out by using a set of musical pieces all in major mode. Results show that participants tend to organize the stimuli according to three clusters, related to musical tempo and to timbral aspects such as the spectral energy distribution.

1. INTRODUCTION

Information about music performance, structured as metadata, could further the development of new application such as automatic expressive performance or active listening, and offer a contribution to improve systems in the context of content-based retrieval, entertainment, and music education. Moreover, 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 communication of expressive content by music can be studied at three different levels, considering: (1) the expressive intentions of the performer, (2) the listeners perceptual experience, and (3) the composers message.

(1) Most studies on the performance expressiveness aim at understanding the systematic presence of deviations from the musical notation as a communication means between musician and listener (see, e.g. [1, 2]). Deviations introduced by technical constraints (such as fingering) or by imperfect performer skill, are not normally considered part of expression communication and thus are often filtered out as noise. The analysis of these systematic deviations has led to the formulation of several models (e.g., [3, 4, 5, 6]) which aim to describe where, how and why a performer modifies, sometimes unconsciously, the score notation. It should be noticed that, although deviations are only the external surface of something deeper and often not directly accessible, they are quite easily measurable, and thus widely used to develop computational models in scientific research and generative models for musical applications.

(2) These studies investigated the relation between music and emotions, showing a sort of isomorphism between musical expression and listeners affective responses. Perceptual studies proved how, generally speaking, it is possible to correlate the listeners main appraisal categories and the acoustic parameters which better characterize expressive intentions ([7, 8] for reviews).

(3) This research takes in account that part of the emotional response to music results from the cognitive processing of musical structures (key, modalities, rhythm), which are known to be expressive in the context of the Western musical system. For example, musical features such as modulation, grace notes, and harmonic progressions, are often associated with emotional responses in the verbal reports of participants [9]. Peretz, Gagnon, and Bouchard [10] demonstrated that rhythm and modality (major vs. minor) contribute to happiness or sadness. These studies are developed in [11, 12]. Generally, they analyze the elements of the musical structure and the musical phrasing that are critical for a correct interpretation of composers message.

Some studies investigated the relation between music and emotions, proving that is possible to correlate the listeners main appraisal categories and the acoustic parameters which better characterize expressive intentions, defining score-independent models of expressiveness [8].

[13] address the question of whether expressive information can be communicated (and recognized) by means of features which are not strictly related to the score. Thus, relevant musical attributes for differentiating expressions (such as articulation) can be replaced by more physical features (e.g. attack time). Professional performers of violin and flute were asked to play musical performances in order to convey different expressive intentions, described by the adjectives that lie on the affective space (happy, sad, angry and calm), and on the sensorial space (light, heavy, soft and hard). With the aid of machine learning techniques we found the audio features that are most relevant for the recognition of different expressive intentions. Using these features as coordinates, we could place the expressions on a feature space and obtain an objective measure of physical similarity. In particular, we extracted and selected a set of audio features from a set of expressive performances played by professional musicians on violin and flute. These features were tested and confirmed by the leave-one-out cross validation, and they can be grouped according to local audio features (using non overlapping frames of $46ms$ length), and event features (using sliding windows with $4s$ duration and $3.5s$ overlap).

A common characteristic of almost all these studies was to investigate emotional responses to music by using verbal labels. The use of verbal labels is potentially problematic since it can encourage participants to simplify what they actually experience [14] and the subjects responses may be conditioned by the different semantic nuances of the same word. Recently, Bigand [12] investigates the emotion conveyed by musical pieces, carrying out some perceptual experiments without making use of verbal labels. Musically trained and untrained listeners were asked to listen 27 different musical excerpts and to freely 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, 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).

Though in his paper Bigand refers to a hypothetical third axis, this aspect is not discussed in detail. Since much of the variance in the results of the Bigand's experiment is due to the mode (major or minor) of the musical pieces, we planned an experiment to investigate other secondary aspects of the relation between music and emotions. Musically trained and untrained listeners were asked to listen the 23 different musical excerpts, all in major mode, and to group those that conveyed similar subjective emotions (see Sec. 2.1). The statistical analysis of the responses (see Sec. 2.2), showed that the listeners organized the musical stimuli in three clusters. In order to investigate the nature of these associations (both the four Bigand's clusters and the three clusters of the new experiment), 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 musical and the affective domains, in order to emphasize the existence of secondary factors that characterize the perception of emotion in music. To this end, we have selected musical pieces only in a major mode, a parameter largely related to the axis of the affective valence.

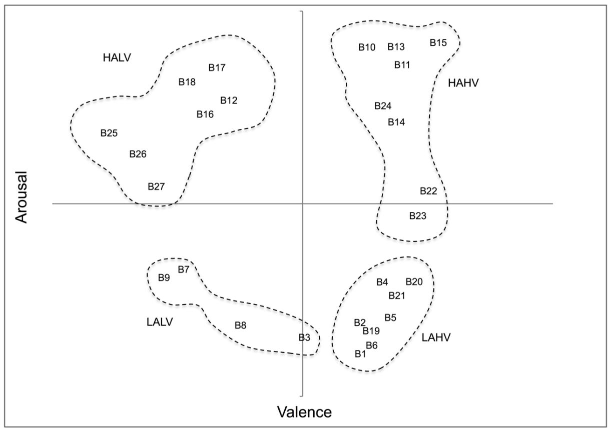

Figure 1. The 27 excerpts of the experiment in Bigand [12], mapped on a two-dimensional space. Dashed lines represent the four affective clusters: high arousal and high valence (HAHV); high arousal and low valence (HALV); low arousal and high valence (LAHV); low arousal and low valence (LALV). Figure adapted from [12].

2. PERCEPTUAL EXPERIMENT IN MAJOR MODE

An experiment has been carried out in order to emphasize the existence of secondary factors that characterize the perception of emotion in music. In the previous section, it has been noted that much of the variance in the results of the Bigand's experiment is due to the modality (major or minor) of the musical pieces. To reduce the effect of this component, we decided to follow the same experimental method, but applying it to musical pieces only in major mode.

2.1 Materials and method

For the experiment, 23 musical excerpts have been chosen as follows: 11 pieces are taken from Bigand in [12], selecting those in a major mode, that are numbered 1, 4, 5, 6, 11, 13, 14, 15, 20, 21, and 23; 12 other pieces were chosen by the Western music repertoire, from XVII to XX century. In particular, the added excerpts are all in a major mode and have been chosen to be representative of various compositional styles. They correspond either to the beginning of a musical movement, or to the beginning of a musical theme or idea. The duration of the excerpts is on average of 30s.

The procedure follows the one already used in [12]. The experiment was conducted using an especially developed software interface. Participants were presented with a visual pattern of 23 loudspeakers, representing the 23 excerpts in a random order. They were required first to listen to all of these excerpts and to focus their attention on the emotional experience of the listening. They were then asked to look for excerpts that induced a similar emotional experience and to drag the corresponding icons in order to group these excerpts. They were allowed to listen to the excerpts as many times as they wished, and to regroup as many excerpts as they wished. The experiment were performed by a total of 40 participants. Of these, 20 did not have any musical experience and are referred to as non-musicians; 20 were music students for at least five years and are referred to as musicians. The duration of the test was about 30 minutes.

2.2 Results

Participants have formed an arbitrary number N of groups. Each group G_k contains the stimuli that the a subject thinks similar (i.e., that induces a similar emotive experience). The dissimilarity matrix A is defined by counting how many times two excerpts i and j are not included in the same group:

$$A[i,j] = \begin{cases} A[i,j]+1 & \text{if} i \in G_k \wedge j \notin G_k \\ A[i,j] & \text{otherwise} \end{cases} \quad (1)$$

$\forall i, j = 1, ..., 23$ and $\forall k = 1, ..., N$.

The dissimilarity matrix was analyzed by means of a Multi-Dimentional Scaling (MDS) method. The location of the 23 excerpts along the two principal dimensions is represented in Figure 2. The excerpts that are close in this space are those evaluated to be more similar by the subjects. The musical pieces coming from the Bigand's experiment have maintained their original numeration (with a 'B' before the number); the other pieces, instead, have been labeled without any letter. Moreover, the MDS solution was compared with a cluster analysis performed on the dissimilarity matrix. The three main clusters are marked in Figure 2 by means of dotted lines.

In Bigand's experiment, the selected excerpts were grouped in two clusters (see Fig. 1): LAHV and HAHV. It can be noted that the excerpts B1, B4, B5, B6, and B21 are still grouped together in one cluster, named Low-Arousal (LA) cluster. Differently, the other Bigand's excerpts are divided into two clusters, named High-Arousal clusters (HA1 and HA2). In short, the first dimension of the MDS is related with the arousal dimension of Figure 1, whereas the second dimension does not seem connected to any of the axes identified by Bigand in his study.

3. ACOUSTIC ANALYSIS OF PREVIOUS EXPERIMENT

3.1 Feature extraction

In order to relate subjects' answers with musical features, we carried out a detailed acoustic analysis of the musical stimuli both of Bigand's and present experiment. A set of acoustic features were calculated for each excerpt. The set was chosen among those features that in previous listening experiments [15] were found to be important for discriminating different emotions and were also used to classify

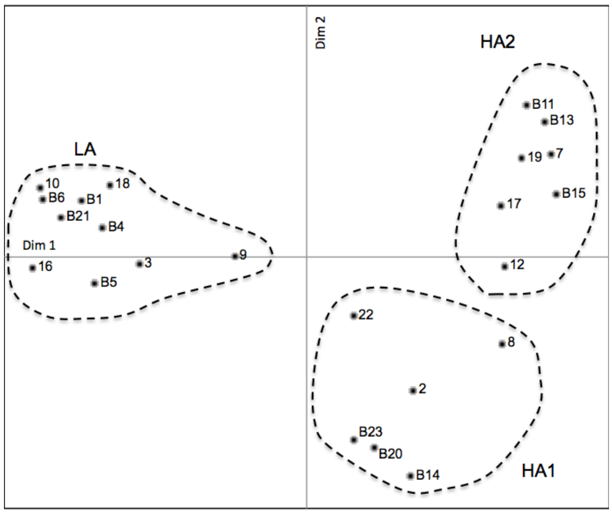

Figure 2. MDS analysis on experiment data. Dashed lines represent the outcome of the cluster analysis.

the style [16] and the expressive content in musical performances [17] and [13]. 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. See Tab. 1 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 magnitude. It is related with the impression of 'brightness" of a sound [18], 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 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 [19], that proposed an estimation of the dissonance degree between two sinusoids, depending on the ratio of 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 [20]; 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 as the average of the piece; m) *Mode* 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 subset of features related both to the four clusters of the Bigand's experiment and the three clusters of the experiment in major mode. 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) [21] is applied with reference to a linear classifier. The feature selection procedure was applied twice. The first time we selected the set of features that classify the 23 excerpts, with a minimum error rate, following the classes specified by the four clusters HAHV, LAHV, HALV, and LALV. The SFS process selected the following four features, in order of selection: *Tempo*, *Mode*, *Centroid*, and *RMS*. The minimum error rate is 18%. Then, we selected the set of features that classify, with a minimum error rate, the 23 excerpts following the classes specified by clusters LA, HA1, and HA2. The SFS process selected the following three features, in order of selection: *Tempo*, *Rolloff*, *Zerocross*. The minimum error rate is 23%.

3.2 Results

Tab. 2 shows the mean values of the four features selected for the Bigand's clusters, calculated for each excerpt of his experiment. 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 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 mode; 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, $p > 0.05$).

cluster	excerpt	Tempo [bpm]	Mode	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

Table 2. Acoustic features related to the clusters resulting from the Bigand's experiment.

Tab. 3 shows the mean values of the three features selected for the experiment in major tonality, calculated for each excerpt. The excerpts belonging to the cluster with low arousal (LA) are characterized, with a few exceptions, by a low value of *Tempo*. On the contrary, the clusters with high arousal (HA1 and HA2) are characterized by a high value of *Tempo*. In particular, the mean value among the excerpts is $59bpm$ for LA, $93bpm$ for HA1, and $97bpm$, for HA2. The ANOVA test shows that these differences are statistically significant ($F = 8.3$ on 2 and 20 df, $p < 0.01$). On the contrary, no significant difference exists between the *Tempo* of HA1 and HA2 clusters. It means that *Tempo* feature is related to the dimension 1 (LA versus HA1 and HA2), but it is not related to the dimension 2 (HA1 versus HA2).

As concern the *Rolloff* feature, significant difference exists among the mean values of the three clusters ($F = 9.8$ on 2 and 20 df, $p < 0.01$): $1923Hz$ for LA, $2225Hz$ for HA1, and $3828Hz$ for HA2. Considering the clusters two by two, the difference between LA and HA1 is not significant, while it is significant between HA1 and HA2. This result means that the dimension 2 can be related to *Rolloff* feature.

Finally, the mean values of $Zerocros$ feature are 585 for cluster LA, 732 for HA1, and 1150 for HA2 ($F = 11.5$ on 2 and 20 df, $p < 0.01$). Similar to *Rolloff*, the difference is not significant between LA and HA1, while it is significant between HA1 and HA2. Tables 4 and 5 summarize qualitatively the results of the two acoustic analyses.

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	$\frac{\sum_{f=1}^{M} \mathrm{I}\{\mathrm{rms}(x(f))<\mathrm{rms}(x)\}}{M}$

Table 1. 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.

cluster	excerpt	Tempo [bpm]	Rolloff [Hz]	Zerocross
LA	B1	61	2372	521
	3	52	1868	516
	B4	53	1747	938
	B5	53	1106	370
	B6	50	1028	443
	9	78	2289	576
	10	54	2707	468
	16	56	1069	449
	18	60	2560	713
	B21	76	2487	852
HA1	2	120	3210	784
	8	98	3234	1044
	B14	102	1799	655
	B20	103	1582	735
	22	76	1741	498
	B23	59	1786	675
HA2	7	84	2714	817
	B11	53	4972	1650
	12	104	4367	1083
	B13	103	3177	972
	B15	145	2495	827
	17	72	3229	1121
	19	116	5841	1579

Table 3. Acoustic features related to the clusters resulting from the experiment in key major.

Cluster	Mode	Tempo
LALV	-	-
LAHV	+	-
HALV	-	+
HAHV	+	+

Table 4. Relation among clusters and selected features in the Bigand's experiment.

Cluster	Tempo	Rolloff	Zerocross
LA	-	-	-
HA1	+	-	-
HA2	+	+	+

Table 5. Relation among clusters and selected features in the major mode experiment.

4. CONCLUSIONS

An experiment has been carried out in order to emphasize the existence of secondary factors that characterize the perception of emotion in music. To this end, we have selected musical pieces only in a major mode, a parameter largely related to the axis of the affective valence. The results show that participants tend to organize the stimuli according to three clusters. The meaning of these clusters 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 found clusters. The analysis of the acoustic features on one hand confirms the results of previous research [22][12], i.e. the main parameters that characterize the affective responses to music are *Tempo* and *Mode*. On the other hand, it gave rise to other aspects that affect the emotional perception of music, such as timbral elements related to the spectral energy distribution.

Interesting similarities are further recognizable with the results of score-independent studies (see [8, 13]) which explored the relation between timbral parameters and musical expression, suggesting the existence of a common level of representation for music expressiveness both in score-dependent and score-independent contexts.

5. REFERENCES

[1] N. P. McAngus Todd, "The kinematics of musical expression," *Journal of the Acoustical Society of America*, vol. 97, pp. 1940–1949, 1995.

[2] C. Palmer, "Music performance," *Annual Review of Psychology*, vol. 48, pp. 115–138, 1997.

[3] B. H. Repp, "Diversity and commonality in music performance: an analysis of timing microstructure in Schumann's "Träumerei"," *Journal of The Acoustical Society of America*, vol. 92, pp. 2546–2568, 1988.

[4] ——, "Expressive timing in Schumann's "Träumerei:" an analysis of performances by graduate student pianists," *Journal of The Acoustical Society of America*, vol. 98, pp. 2413–2427, 1995.

[5] ——, "A microcosm of musical expression: I. quantitative analysis of pianists' timing in the initial measures of chopin's etude in E major," *Journal of The Acoustical Society of America*, vol. 104, pp. 1085–1100, 1998.

[6] ——, "A microcosm of musical expression: II. quantitative analysis of pianists' dynamics in the initial measures of chopin's etude in E major," *Journal of The Acoustical Society of America*, vol. 105, pp. 1972–1988, 1999.

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

[8] S. Canazza, G. De Poli, A. Rodà, and A. Vidolin, "An abstract control space for communication of sensory expressive intentions in music performance," *Journal of New Music Research*, vol. 32, no. 3, pp. 281–294, 2003.

[9] J. A. Sloboda, "Music structure and emotional response: Some empirical findings," *Psychology of music*, vol. 19, pp. 110–120, 1991.

[10] I. Peretz, L. Gagnon, and B. Bouchard, "Music and emotion: Perceptual determinants, immediacy and isolation after brain damage," *Cognition*, vol. 68, pp. 111–141, 1998.

[11] A. Gabrielsson and E. Lindstrom, "The influence of the musical structure on emotional expression," in *Music and emotion. Theory and research*, P. N. Juslin and J. A. Sloboda, Eds. Oxford University Press, 2001, pp. 223–249.

[12] E. Bigand, S. Vieillard, F. Madurell, J. Marozeau, and A. Dacquet, "Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts," *Cognition and Emotion*, vol. 19, no. 8, pp. 1113–1139, 2005.

[13] L. Mion and G. De Poli, "Score-independent audio features for description of music expression," *IEEE Trans. Speech, Audio, and Language Process*, vol. 16, no. 2, pp. 458–466, 2008.

[14] K. Scherer, "Affect bursts," in *Emotions: Essays on emotion theory*, S. van Goozen, N. E. van de Poll, and J. A. Sergeant, Eds. Erlbaum, 1994, pp. 161–196.

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

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

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

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

[19] R. Plomp and W. J. M. Levelt, "Tonal consonance and critical bandwidth," *Journal of the Acoustical Society of America*, vol. 38, no. 4, pp. 548–560, 1965.

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

[21] A. Whitney, "A direct method of nonparametric measurement selection," *IEEE Trans. Comput.*, vol. 20, no. 9, pp. 1100–1103, 1971.

[22] L. Gagnon and I. Peretz, "Mode and tempo relative contributions to happy–sad judgements in equitone melodies," *Cognition and Emotion*, vol. 17, no. 1, pp. 25–40, 2003.