Automatic Performance of *Black and White n.2*: The Influence of Emotions Over Aleatoric Music

Stefano Baldan, Adriano Baratè, and Luca A. Ludovico

LIM - Laboratorio di Informatica Musicale
Dipartimento di Informatica
Università degli Studi di Milano
Via Comelico 39/41, I-20135 Milano, Italy
{baldan,barate,ludovico}@dico.unimi.it

Abstract. *Black and White n.2* is a piece of aleatoric music by Franco Donatoni. Conceived as a set of 120 exercises for piano, it uses a non-conventional way to encode the score. Some elements of the composition are left to chance, thus they should be extemporarily determined by the performer. In this work, human choices are performed by an automatic system. The algorithms designed and described here extract emotion-related information from a video input and consequently create in real time an instance of the piece. Finally, the paper presents the case study of *Black and Byte*, an application implemented to test such algorithms.

Keywords: aleatoric music, automatic composition, emotions

1 Introduction

The relationships between music and emotions is a very rich and complex matter, and a theoretical discussion of such a subject goes beyond the goals of the present paper. In the computer field, this problem has been addressed for instance in [1], [2] and [3].

This work narrows the field by focusing on the relationships between music performance and emotions. Even in a traditional context, such as an evening at the concert hall, a music performance is influenced by the feelings of the performers and it conveys emotions to the audience. From this point of view, in many contemporary music pieces and multimedia installations new frontiers have been explored, making the listener become the protagonist of the performance. For instance, the emotions and behaviours of the audience during a performance can be captured in order to influence the performance itself in real time.

Our work is strictly related to the latter aspect. The goal is automatically rendering an aleatory composition for keyboard (see Section 2) through a computer-based system able to solve its non-determinism. In order to generate an automatic performance, all the values of aleatory variables must be computed. In general it could be sufficient to generate sequences of pseudo-random numbers, if necessary under certain conditions. But for our purposes at least two further constraints must be considered:

9th International Symposium on Computer Music Modelling and Retrieval (CMMR 2012)
19-22 June 2012, Queen Mary University of London

1. The aleatory aspects of the performance should not be fully determined by chance, but influenced in real time by emotion-related contents. In particular, our algorithms take both a score and a video as input. Some features of the video are automatically extracted and evaluated in order to add a coherent soundtrack based on the score;
2. The resulting performance for keyboard instruments, after a transcription in conventional notation, should be playable by a medium-skilled human performer. This aspect is not trivial, since generated chords have to respect the fingering rules explicitly indicated in the score, and to consider hand posture and comfort.

All these items will be discussed in detail in the following sections.

2 F. Donatoni's *Black and White n.2*

Black and White n.2 [4] is a collection of 120 pieces written by the Italian composer Franco Donatoni (9 June 1927 - 17 August 2000). They can be played on any keyboard instrument, including piano, harpsichord, celesta, electronic keyboard, etc. Versions for 2 and 3 keyboard instruments have been conceived and performed as well. The subtitle of the composition is *Esercizi per le 10 dita*, namely 10-finger exercises, and this aspect will be fundamental to understand Donatoni's notation, as explained below. This piece belongs to the genre known as *aleatoric music* [5], since some primary parameters of the composition are not predetermined, but their values depend on random processes or extemporary decisions made by the performer.

In the preface to the score, the author briefly explains the simple set of rules to read the score, whose conventions significantly differ from Common Western Notation (CWN). In fact, the two staves usually assigned to traditional keyboard notation (i.e. the grand staff) in this case do not carry pitch and rhythm, but rather finger-related information. Only lines are used, and each line corresponds to a specific finger. For the right hand, the lower line corresponds to the thumb and the upper line to the little finger, and vice versa for the left hand. Consequently, the typical symbols of a traditional score are not present: no time nor key indication, no bars nor rhytmic values, etc.

As mentioned before, a key constraint is the use of one or more keyboard instruments. Given this hypothesis, the rules to read the score are few:

- The association between symbol positions over lines and fingers is fixed;
- About the colour of the symbols, each symbol can be either filled or empty. This graphical convention forces the performer to play either a black or a white key respectively;
- The symbols that can be placed over this sort of staff have either a circle or a square shape, and this aspect is related to dynamics. At the beginning of the performance the player decides if circles should correspond to the *ppp* dynamic indication (i.e. softest possible), thus squares would correspond to *fff* (i.e. loudest possible), or vice versa;

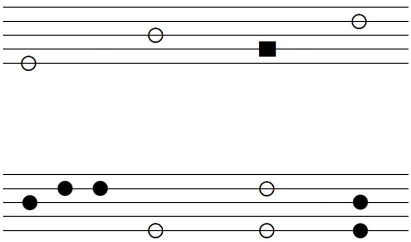

Fig. 1. A score excerpt which follows Donatoni's *Black and White n.2* set of rules.

- The concept of chord is associated to the vertical alignment of such symbols, possibly spanning over the two staves;
- Finally, arrows pointing up or down can be specified for each chord, even a degenerate 1-note chord. The meaning of either an upward or a downward arrow is using preferably either the higher or the lower octaves of the keyboard respectively.

The application of this set of rules clearly leaves many music parameters to the determination of the performer. As a consequence, for a given score a number of different performances is possible. In particular, the following parameters can be considered as degrees of freedom:

- Metronome tempo and time indication are not present. Consequently, the piece is not organized into bars, nor into other regular rhythmical grouping;
- All rhythmic aspects (note durations, note density, articulation, etc.) are not specified. As regards articulation, Donatoni allows the use of *legato*, *staccato* and *tenuto*, as well as a free use of sustain pedal;
- Specific pitches are not indicated. Score information about pitches forces only the use of either white or black keys in a given number, and provides suggestions about the pitch range to use (i.e. the lower, middle or higher octaves over the keyboard).

3 Extraction of Emotional Features from Video

As Donatoni explicitly states in the preface to *Black and White n.2*, this composition is only partially defined by a limited set of rules, while many other aspects are left to the extemporaneous interpretation and execution by the single performer. Such a work can therefore be heavily influenced by the mood of the musicians who are playing it, and on the other side it can also be heavily modified as regards melody, rhythm and harmony in order to transmit a certain kind of emotion.

The great affective versatility of *Black and White n.2* could be exploited to automatically generate consistent soundtracks for arbitrary video footage. In this paper we will propose simple yet effective methods to extract affective features from motion videos, both in online and offline scenarios. Then a possible mapping of the rules of Donatoni's composition will be provided in order to synchronize an automatic musical performance with the visual information in terms of timing and emotions.

3.1 The Affective Model: Related Work

Before extracting the emotional features out of the video footage, we must find a way to describe, analyze and measure those ephemeral entities we call "emotions". To face this non-trivial problem, an interdisciplinary research in heterogeneous fields is required, including Music Psychology and Music Information Retrieval. A review of some of the most relevant contributions makes two different approaches emerge: *categorical analysis* versus *dimensional analysis.*

In the former case, emotions are defined in a discrete way and classified as belonging to a few basic categories, such as "love", "hate", "joy", "sorrow" and so on (see [6] for further details). As stated in [7], one of the difficulties in the automatic recognition of emotions is labeling of the data. Many experiments have been conducted in this sense, starting either from symbolic contents or from audio objects. It is worth citing the data labeling proposed by Hevner in 1936, which consists of a circle of 8 classes where not all adjectives in a single group are synonims [8]. A more recent work in this context is [9], which starts from the previous one and proposes 13 classes each labeled by one, two or three adjectives. It should be clear that adding classes and dimensions means moving from a discrete to a continuous description. For our purposes, one of the main drawbacks of categorical analysis lies in its discrete nature, which does not catch the subtle nuances of human feelings in a satisfactory way. Moreover, this model describes emotions qualitatively, making it difficult to map them onto the quantitative parameters used to generate an automatic musical performance, like note pitches, beats per minute and so on.

In dimensional analysis, on the contrary, emotions are defined inside a multidimensional, continuous space. An interesting work that applies this kind of analysis to emotions detection in speech is [10], which adopts the activation dimension. In the field of affective retrieval of information, the Arousal-Valence (AV) model proposed in [11] is still considered up-to-date and it is one of the most used. For example, in [12] it is applied to problems of music classification by affective contents. The model defines a two-dimensional space where emotions are classified in terms of their level of arousal (calm versus excited) and pleasantness (positive versus negative). The dimensional approach gives the possibility to express the subtle nuances of human feelings through continuous variations in the chosen multidimensional space. Since it is based on quantitative features, it also allows an easy mapping onto the musical parameters used to generate the automatic performance. For these reasons, the present work applies a dimensional analysis technique, and in particular the AV model.

3.2 Arousal and Valence in Videos

After choosing an affective model, the next step is identifying the emotional features in motion video information, namely which parameters are relevant to the arousal and valence perception of what we see. In a previous work [13], Zhang proposes an affective categorization of musical videoclips based on five key features: sequence changes, motion, lighting, color saturation and color energy. The first two features are related to the arousal dimension, while the remaining ones are related to the valence dimension.

The work shows how a high number of sequence changes and a high amount of moving objects, camera zooming and panning effects result in a sense of excitation, whereas few scene cuts and static planes give a sense of quietness. On the other side, vivid colours and bright images are common in calm or joyful videos, while faded and dark images are often used to convey a sense of sadness, fear or anger.

In the proposed affective analysis of musical videoclips, also audio clues are taken into account. From this point of view, the key features are zero-crossing rate, tempo and beat strenght for a classification along the arousal dimension, and rhythm regularity and pitch as regards the valence dimension. While the work by Zhang focuses on the extraction of these parameters out of the existent soundtracks of musical videoclips, our goal is exactly the opposite, namely to synthesize a soundtrack whose audio clues are consistent with the ones extracted by the visual information. There are some significant differences in our approach: while the work of Zhang focuses on the offline affective analysis of a database of MPEG musical videoclips, we want to generate soundtracks for arbitrary video footage, stored in different encodings or even acquired in real time by a webcam or a RTP[1] stream.

3.3 Algorithms for Video Analysis

The algorithms used to detect sequence changes takes inspiration from the abrupt scene change detection described in [14]. It is a method based on inter-frame motion intensities, which can also be directly employed to detect motion in videos. The metric used to compute inter-frame motion intensities is the absolute difference between consecutive frames, which is described by the formula:

$$D = \sum_{x=0}^{W} \sum_{y=0}^{H} |r_1(x,y) - r_0(x,y)| + |g_1(x,y) - g_0(x,y)| + |b_1(x,y) - b_0(x,y)|$$

where W and H are respectively the width and height of both frames, $r_1(x,y)$, $g_1(x,y)$ and $b_1(x,y)$ are the red, green and blue values of the pixel at coordinates (x,y) inside the current frame, and $r_0(x,y)$, $g_0(x,y)$ and $b_0(x,y)$ are their corresponding values inside the previous frame. Scene-change detection is achieved by keeping trace of the maximum motion intensity occurring inside a local window of consecutive frames, and comparing the motion intensity of every frame

[1] RTP stands for Real Time Protocol, which is described in RFC 3550.

against this value. If the result for the current frame is n times the computed maximum, then a sequence change is detected. According to Yeo, values of n between 2 and 3 have proven to give good results. The local moving frame-window prevents the detection of false positives such as camera flashes or rapid panning and zooming of the scene. A window of 25 frames for a 25 fps video, for example, means that there cannot be two consecutive sequence changes within a second. Once sequence changes are detected, the number of shots per second can be determined dividing the number of shots in the video by its duration in seconds. In the present work this value is computed locally instead for the whole video, choosing a window of 5 sequence shots, as we want to capture in real time the local variations of this feature inside each video stream.

Valence features can be easily obtained by converting pixel values from the RGB into the HSB color space.[2] Brightness is computed in a straightforward way summing the brightness of pixels in each single frame. Similarly, saturation is computed summing the saturation of pixels in each single frame. Color energy instead is a composite parameter obtained from the combination of the two former values with the standard deviation of the pixel hue values. The underlying concept is that colorful videos often contain many different tonalities, thus yielding very high standard deviation values for the hue; on the contrary, faded videos are more likely to contain a limited amount of different colors, thus yielding low standard deviation values for this parameter.

All the key features are normalized for convenience between 0 and 1: the value of shots per second is already in this range if a window of one second (usually 25 or 30 frames, depending on the standards and formats adopted) for scene change detection is chosen. On the other side, inter-frame motion is divided by the maximum possible absolute difference between frames, which for a 24 bit RGB color video is:

$$\hat{D} = \frac{D}{W * H * 255 * 3}$$

Similarly, lighting and saturation are divided by their maximum possible values inside a single frame, which for a 24 bit HSB color video is:

$$\hat{L} = \frac{L}{W * H * 255}$$

$$\hat{S} = \frac{S}{W * H * 255}$$

Finally, the standard deviation of the hue is normalized using the formula:

$$\hat{\sigma}_h^2 = \begin{cases} 4\frac{\sigma_h^2}{h_{max}}, & \text{if } \sigma_h^2 < \frac{h_{max}}{4} \\ 2 - 4\frac{\sigma_H^2}{h_{max}}, & \text{otherwise} \end{cases}$$

[2] The RGB acronym represents an additive color model in which red, green, and blue light are added together to reproduce colors. HSB is another color model, based on the image attributes called hue, saturation, and brightness.

The final arousal value is obtained by a user-defined weighted average between shots per second and motion intensity. Similarly, the valence feature is computed by a user-defined weighted average between lighting, saturation and hue standard deviation. Both arousal and valence can be then amplified and clipped to the maximum value of 1. These manual adaptations are included in order to make the system adaptable to changes in lighting, input devices, nature of the video footage and so on.

4 Mapping Emotional Features onto a Score

The final goal of our work is the extraction of emotions from video captures in order to drive a computer-based performance of *Black and White n.2.* As discussed in Section 3, the algorithm to extract emotion values from a video employs a 2-axes classification, based on arousal and valence values respectively. Now we will explain which music features can be used, and how, in order to convey emotions and consequently to create an adequate soundtrack for motion videos. Inspiring works from this point of view are [15] and [16]. Please note that, in order to establish adequate mappings between music-conveyed emotions and musical features, psychology and neuropsychology studies must be considered, too (e.g. [17] and [18]).

Some choices specific for our implementation will be detailed in Section 5.

4.1 Rhythm

Starting from mentioned research, rhythmic aspects have been mapped onto the arousal dimension.

Note durations are not expressed as in traditional music theory (e.g. crotchets, quavers, etc.), but as absolute time intervals whose value belongs to a continuous range. No time indication or BPM[3] value is provided by the score, and no one is introduced by our algorithms.

Other aspects related to rhythm are articulation marks and pedals. The preface to Black and White n.2 explicitly cites the possibility to introduce articulations and pedals in a performance, even if the corresponding marks do not belong to allowed score symbols. As regards in-use articulation signs, they usually include *staccatissimo*, *staccato*, *martellato*, *marcato*, *tenuto*, and so on. Besides, a *legato* effect is usually indicated through slurs. Finally, the damper pedal (if present on the keyboard instrument) represents the maximum level of sustain, as all notes played will continue to sound until the pedal is released. All these aspects, which are allowed by Donatoni but not encoded inside the score, are left to improvisation and to the performer's feelings. In our opinion, the list provided before should correspond to a progressive reduction of the arousal level, ranging from *staccatissimo* to *legato* with sustain pedal.

[3] BPM stands for Beats Per Minute.

4.2 Harmony

Harmony, here intended as a sequence of chords, is one of the most important, but also one of the most difficult music dimensions to map onto the mentioned axes. Problems arise for a number of different reasons.

First, the function of a chord changes depending on the surrounding context, so that an evaluation should involve not only the chord itself, but also the harmonic path it has been inserted into. For example, a major triad is usually considered happier and brighter than a minor one, but in a tonal context a minor triad built on the second degree of a major scale (e.g. [D,F,A] in C major) does not necessarily convey a sense of sadness.

Other chords are intentionally ambiguous and their meaning is clarified only by the surrounding harmonic path. A very significant example is the *incipit* of the *Allegro non molto* from Violin Concerto in F minor RV 297 “L’inverno” by A. Vivaldi. The first chord (see Figure 2) is incrementally built by superimposing higher and higher voices: initially it sounds affirmative and stable (the F-minor tonic alone), then ambiguous (major second), misleading (minor sixth), frightful (perfect fourth), and finally clear thanks to the resolution of the dissonance on a diminished seventh chord.

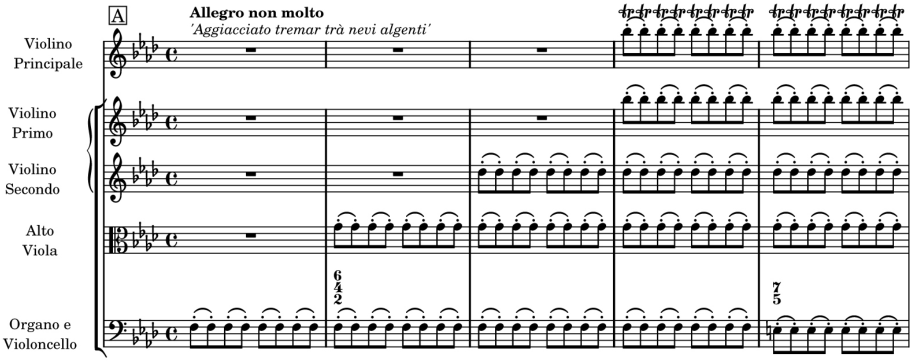

Fig. 2. *Incipit* of the *Allegro non molto* from Violin Concerto in F minor RV 297 “L’inverno” by A. Vivaldi.

Even those chords that present a commonly-accepted affective value in Western culture do not have an implicit nor a universal value, above all in a non-tonal context. An interesting survey about harmony and chord functions in the twentieth-century music can be found in [19].

Obviously, an exhaustive discussion of the matter goes beyond the goals of the present paper. Now we will briefly explain how the problem has been solved in our work from a practical perspective. Let us recall the two key concerns:

1. It is necessary to find an efficient and effective way to map in real time emotional features onto AV axes. In our work, the harmonic path is not predetermined: the performer has to chose pitches extemporarily, as the author himself recommends. A reductive but practical solution is considering chords as isolated entities, thus ignoring the harmonic context;
2. Fingering indications are one of the few contraints provided by the score, and they cannot be ignored. As a consequence, not all chords that prove to be adequate as regards their affective characteristics can be employed, but only those which support a given fingering.

Our approach can be described as a four-steps process.

First, the system is provided with a set of chord models to use, including their possible inversions. The concept of "model" means encoding halftone distances from the root note, instead of any possible combination of pitches. An easy-to-implement algorithm can produce any instance of a model starting from each available pitch.

Then, each chord model is put in correspondence with two ranges of continuous values, on the arousal and on the valence axis respectively, thus forming a rectangle. In this way, any valid point of the AV plane is covered by a variable number of overlapping rectangles, corresponding to all the chord models that can convey those given arousal and valence senses.

After defining the set of suitable chord models for a given pair of AV values, a further selection is made on the base of fingerings. In fact, not all the chords may support a particular hand position.

Finally, it is necessary to verify if the selected chord model has at least one instance that corresponds to the black/white keys configuration indicated by the score.

If one of the mentioned steps fails, as there exist no candidates having the required chracteristics, a backtracking technique is used to select a new candidate. For example, if a major triad, a dominant seventh and a diminished seventh cover the current point of the AV plane and the score requires a 4-fingers chord, let the second chord be initially selected as the candidate. After verifying that the required black/white layout cannot be instanced starting from the dominant seventh model, the algorithm selects the diminished seventh.

Please note that a formal check must be conducted on the chord-models set to verify the complete covering of the AV plane using all possible fingerings and black/white-key configurations.

As regards our implementation, details about chord models and single-chord mappings onto the AV plane will be provided in Section 5.

5 Case study: *Black and Byte*

Black and Byte is the application designed and implemented to test the algorithms described in Sections 3 and 4.

The interface allows to open a score in plain-text format. The file provides a score view on a single staff system, which is usually defined "scroll view" in music editing software. The two "staves" corresponding to right and left hand are separated by an empty line. Inside such a document, the allowed symbols are: empty circle ○, filled circle ●, empty rectangle □, filled rectangle ■, up arrow ↑, and down arrow ↓. Whitespaces and tabs are supported as well, but they have no musical meaning. Other symbols are not managed, so they are ignored by the parser.

As regards video input, it is possible either to load an available media file or to acquire motion images from a webcam or a stream in real time. The audio track (if available) is ignored. Since the duration of the performance is not known in advance (this is one of the aspects left to the performer's will), score is read in a circular way and a loop is performed to sonorize the entire video. At each iteration, music parameters are recalculated according to Donatoni's rules.

Some features of the prototype are not hard-coded, so they can be configured by the user. It is worth citing the chord-model list, which deeply influences the results of our algorithms. At present, the total amount of supported chord models is limited to those typical of traditional harmony, namely:

1. 10 bichords (minor second, major second, minor third, major third, diminished fourth, perfect fourth, augmented fourth, diminished octave, perfect octave, augmented octave) and their 10 inversions;
2. 7 triads (major, minor, diminished, augmented, suspended second, suspended fourth, and flat fifth) and their 14 inversions;
3. 7 sevenths (major, minor, dominant, diminished, half-diminished, minor/major, and augmented/major) and their 21 inversions;
4. 7 ninths (ninth, minor ninth, flat ninth, minor flat ninth, augmented ninth, nine-six, minor nine-six) and their 28 inversions.

Inverted chords are automatically computed from the model of the "parent" chord, namely the root-position chord. Since chord-model list has not been hard-coded inside the prototype, this set could be easily extended, for instance supporting elevenths and thirteenths, post-tonal chords or micro-tonal intervals.

The mapping of chords onto the AV plane is clearly subjective. In our approach:

- Consonance/dissonance among chord notes correspond to a low/high level of arousal, respectively. This parameter is computed by evaluating the intervals inside the chord, namely the relationship between the root element and the following notes;
- The belonging of the chord to the minor/major tonal area corresponds to a low/high level of valence.

For instance, a major triad is very consonant and clearly belongs to the major area, so it has a low value for arousal and a high value for valence: it can convey a sense of brightness, grace, quietness, solemnity, etc. On the other side, ambiguous chords - e.g. empty fifths - present a neutral value on both dimensions: they can represent bore, doubt, indefiniteness, etc.

The interface presents two key elements: a media player where the motion video is loaded, and a panel containing the original score. As mentioned before, the creation of the performance is extemporary, and it is computed in real time depending on video analysis. While the performance is advancing, the interface shows the corresponding transcription in CWN together with fingering indications. In this way, on one side the instances of Donatoni's rules can be verified, and on the other side the resulting score can be played also by a human performer.

A component not directly related to the real-time performance, but useful to show video-analysis results and to understand the consequent sonorization, is the panel containing the AV plane representation. A small circle defines the position of the current point along the arousal and valence axes.

6 Conclusions

In this paper we have proposed a process to sonorize motion videos in real time, starting from 1) an on-the-fly analysis of video contents and 2) a given score encoded according to Donatoni's rules.

One of the goals was exquisitely theoretical: testing dimensional-analysis results and determining efficient and effective algorithms to apply those results to aleatoric music. Besides, a number of practical applications can exist, e.g. in multimedia installations, emotion-based sonorization of videos, and real-time control of aleatory performances through face recognition or other gestures.

Acknowledgments. This work has been partially funded by the Enhanced Music Interactive Platform for Internet User (EMIPIU) project.

References

1. Yang, D., Lee, W.S.: Disambiguating music emotion using software agents. In: Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR04), pp. 52–58 (2004)
2. Yang, Y.H., Lin, Y.C., Su, Y.F., Chen, H.H.: A regression approach to music emotion recognition. IEEE Transactions on Audio, Speech, and Language Processing, vol. 16(2), pp. 448–457. IEEE Press, New York (2008)
3. Livingstone, S.R., Brown, A.R.: Dynamic response: real-time adaptation for music emotion. In: second Australasian conference on Interactive entertainment, Proceedings of the, pp. 105–111. Creativity & Cognition Studios Press (2005)
4. Donatoni, F.: Black and white II - Esercizi per le dieci dita per strumenti a tastiera. Suvini Zerboni, Milano (1968)
5. Meyer-Eppler, W.: Statistic and Psychologic Problems of Sound. Die Reihe No. 1: Electronic Music, pp. 55–61 (1958)
6. Plutchik, R.: The Nature of Emotions. American Scientist, vol. 89(4), pp. 344–350 (2001)
7. Wieczorkowska, A., Synak, P., Lewis, R., W. Raś, Z.: Extracting emotions from music data. Foundations of Intelligent Systems, pp. 456–465. Springer (2005)

8. Hevner, K.: Experimental studies of the elements of expression in music. American Journal of Psychology, n. 48, pp. 246-268 (1936)
9. Li, T., Ogihara, M.: Detecting emotion in music. In: 4th International Conference on Music Information Retrieval ISMIR, Proceedings of the, pp. 239–240 (2003)
10. Tato, R., Santos, R., Kompe, R., Pardo, J.M., Emotional space improves emotion recognition. Seventh International Conference on Spoken Language Processing (2002)
11. Schlosberg, H.: Three dimensions of emotion. Psychological review, vol. 61(2), pp. 81–88. American Psychological Association (1954)
12. Oliveira, A., Cardoso, A.: Towards bidimensional classification of symbolic music by affective content. In: International Computer Music Conference, Proceedings of the (2008)
13. Zhang, S., Huang, Q., Jiang, S., Gao, W., Tian, Q.: Affective visualization and retrieval for music video. IEEE Transactions on Multimedia, vol. 12(6), pp 510–522. IEEE Press, New York (2010)
14. Yeo, B.L., Liu, B.: A unified approach to temporal segmentation of motion JPEG and MPEG compressed video. In: Multimedia Computing and Systems, Proceedings of the International Conference on, pp. 81–88. IEEE Press, New York (1995)
15. Gabrielsson, A., Lindström, E.: The influence of musical structure on emotional expression. Oxford University Press (2001)
16. Wu, T., Jeng, S.: Extraction of segments of significant emotional expressions in music. In: 2006 International Workshop on Computer Music and Audio Technology, Proceedings of the, pp. 76–80 (2006)
17. Marin, O.S.M., Perry, D.W.: Neurological aspects of music perception and performance. Academic Press (1999)
18. Simpson, J.A.: Music and the Brain - Studies in the Neurology of Music. Journal of Neurology, Neurosurgery & Psychiatry, vol. 40(7). BMJ Publishing Group Ltd. (1977)
19. Persichetti, V.: Twentieth-century harmony: creative aspects and practice. WW Norton (1961)