Proceedings of ICAD 04-Tenth Meeting of the International Conference on Auditory Display, Sydney, Australia, July 6-9, 2004

EMOTIONFACE: PROTOTYPE FACIAL EXPRESSION DISPLAY OF EMOTION IN MUSIC

Emery Schubert

School of Music and Music Education
University of New South Wales
Sydney NSW 2052
AUSTRALIA
E.Schubert@unsw.edu.au

ABSTRACT

EmotionFace is a software interface for visually displaying the self-reported emotion expressed by music. Taken in reverse, it can be viewed as a facial expression whose auditory connection or exemplar is the time synchronized, associated music. The present instantiation of the software uses a simple schematic face with eyes and mouth moving according to a parabolic model: Smiling and frowning of mouth represents valence (happiness and sadness) and amount of opening of eyes represents arousal. Continuous emotional responses to music collected in previous research have been used to test and calibrate *EmotionFace*. The interface provides an alternative to the presentation of data on a two-dimensional emotion-space, the same space used for the collection of emotional data in response to music. These synthesized facial expressions make the observation of the emotion data expressed by music easier for the human observer to process and may be a more natural interface between the human and computer. Future research will include optimization of *EmotionFace*, using more sophisticated algorithms and facial expression databases, and the examination of the lag structure between facial expression and musical structure. Eventually, with more elaborate systems, automation and greater knowledge of emotion and associated musical structure, it may be possible to compose music meaningfully from synthesized and real facial expressions.

1. INTRODUCTION

The ability of music to express emotion is one of its most fascinating and attractive characteristics. Measuring the emotion which music can express has, consequently, occupied thinkers and researchers for a long time. One of the problems requiring consideration is how to measure emotion. There have been three broad approaches: physiological measurement (such as heart rate and skin conductance), observational measures (documenting the listeners physical postures and gestures made while listening) and cognitive self-reporting. Physiological measures tend to tap into changes that are reflective of the arousal dimension of emotion [1]. Few studies have shown that they can reliably differentiate, for example, between happy and sad emotional responses. Observational methods are rarely found because they are fairly complex and expensive to implement. One of the most important examples of such observational methodology is in the coding of facial expressions [eg. 2], though this approach is yet to be applied to the analysis of the music listener's face. In both of these methodologies the measurement is restricted to an emotion experienced by the listener. It seems unlikely that physiological and observational approaches could indicate the emotion the listener identifies as being in the music (for more information on the distinction between perceived and experienced emotion in music see [3]).

The most common way of measuring emotional responses to music has been through cognitive self-report, where the listener verbally reports the emotion perceived in the music. The self-report approach has been subdivided into three types of response formats: open-ended, checklist and rating scale. Typically, with each approach participants are asked to listen to a piece of music and make a response at the end of the piece. Since the 1980s researchers have had easier access to computer technology which allows emotional observations about unfolding music to be tracked continuously. For this process, Schubert has argued that the best approach is to use rating scales [4]. He proposed a method of collecting self-reported emotions by combining two rating scales on a visual display. The rating scales should be reasonably independent and explain a significant proportion of variation in emotional response. Several researchers have identified the dimensions which fulfill these criteria as being valence (happiness versus sadness) and arousal (activity versus sleepiness) (eg. [10]). The dimensions have been combined at right angles on a computer screen, with a mouse tracking system which is synchronised with the unfolding music [5, 6].

One of the applications of tracking emotional response to music in this way is that pedagogues, researchers, musicians and listeners in general can examine the two dimensional emotion space expressed by music according to the sampled population. In the past [6], the visual interface has been the same emotion space used for collecting data from individual participants. The present paper describes a method of displaying the emotion expressed by music using continuously synthesized facial expressions.

2. FACIAL EXPRESSION LITERATURE

Since Darwin's work on emotion [7] we have had a good understanding of how facial expressions communicate emotional information. Humans are highly sensitive to nuances in such facial expressions (e.g. 8, 9) and there is strong evidence that the emotion communicated by facial expressions can be understood universally. This corpus of available emotional expressions in the human face has been documented and decoded largely through the work of Ekman and Friesen [2]. Their taxonomy allows the meaningful reduction of emotion into 6 prototypical, basic emotions. These basic emotions can be translated onto a continuum using a dimensional model of emotion [10]. The eyes,

eyebrows and mouth are the main parts of the face which signal emotional messages, what Fasel and Luettin [11] refer to as 'intransient facial features'. Further, eye shape is more important than mouth shape in activating the high arousal emotion of fear [12], and therefore has an important connection with the arousal component of emotional expression. In simple animations the valence of emotion is easy to detect through the shape of the lips (concave up for happy expression, and concave down for sad expressions).

It should therefore possible to synthesize a simple, schematic face with easily recognizable emotional expressions using appropriately shaped curves to represent eye size and mouth shape. Transforming two-dimensional emotion data (valence and arousal) into mouth shape and eye size respectively was viewed to be a logical starting point for providing synthesized, visual display of emotion which a human can understand. The next section describes an algorithm used to draw such a face dynamically as music unfolds (using already gathered subjective arousal and valence data from a previous study using second by second median responses of 67 participants with a fairly high degree of musical training and experience [13]).

3. FACIAL EXPRESSION ALGORITHM

The aim of the prototype schematic *EmotionFace* interface was to produce a visually and algorithmically simple schematic face able to communicate a spectrum of facial expressions along the arousal and valence dimensions. While such a model is fairly simple and more sophisticated algorithms are available for manipulating facial expressions [14], the present realization extracts some of the basic principles which exist in the literature and applies them using only two parabolic functions.

One parabola represents the arousal as expressed by eye opening. First, the lower half of one eye is calculated according to the formula:

$$f_{lower\_eye}(x) = k_a(x - e/2)(x + e/2)/a \qquad (2)$$

where a is the median of perceived arousal value (gathered in [13]) with the addition of 100 (the addition of 100 is ensure that the parabola is always concave up, because a can have negative values as large as -100). Arousal appears in the denominator because large values of a need to make the parabola narrower and, in effect, increase the eye opening size. The roots of the parabola are fixed at the horizontal eye lines and eye widths, as shown in Figure 1. The width of an eye is, therefore, set to e, with the roots of the conjugate pair being half of e on either side of the centre. k_a is a calibration constant. In the present instantiation of the interface, the author estimated all calibration constants. k_a was set so that for small values of arousal, the eyes would appear to be in a neutral (partially opened) position, but for large negative

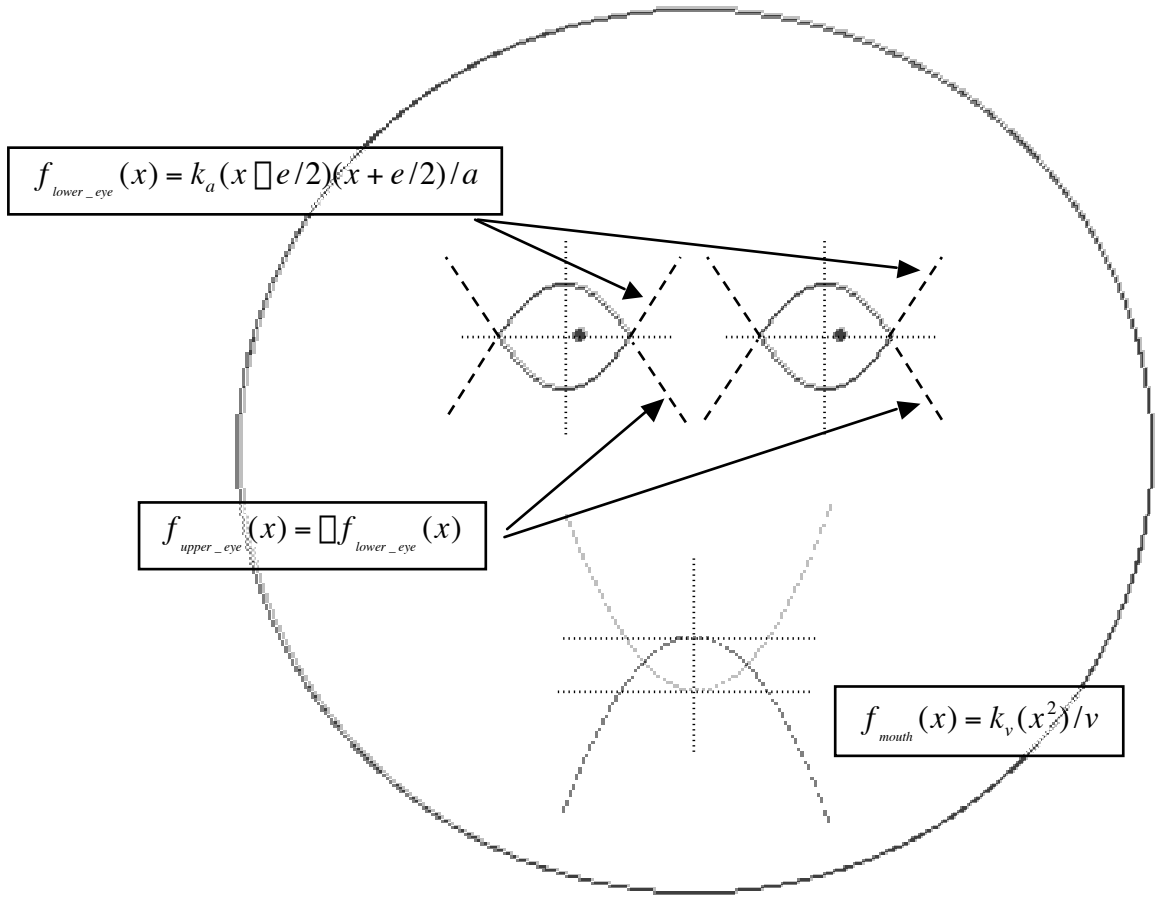

Figure 1. General anatomical/algorithmic structure of EmotionFace. The code was implemented in Hypertalk (the scripting language used in Hypercard for Macintosh). Arousal and valence data are read from a file which is synchronized with an audio CD track.

values, the eyes would appear closed (or almost closed), as if sleeping. Once the lower eye is calculated within the boundary of $—e/2 < x < e/2$, it is copied and placed in the appropriate locations based on the eye centre grids, (shown in Figure 1 as a '+' over each eye). The parabolas are then flipped, as indicated in the *upper_eye* function in Figure 1.

The mouth is represented by another parabola whose vertex is fixed at the origin according to the general form:

$$f_{mouth}(x) = k_v(x^2)/v \qquad (2)$$

As positive valence, v, increases, the mouth function deepens in a concave-up position, giving the appearance of a growing smile. When the valence becomes negative, the function flips to concave down, giving the appearance of a frown. For the discontinuity at $v = 0$, the asymptotic limit is assumed, and a straight, horizontal line is displayed (i.e. neither concave up, nor concave down). k_v is a constant used for calibrating the mouth shape. An additional calibration (not shown mathematically, but indicated visually in Figure 1) is the position of the x-axis, and therefore the vertex. As the length of the parabola increases for increasing values of $|v|$ more space is required to draw the parabola, and to look more believable (the parabolas shown for the mouth in Figure 1 demonstrate the most extreme values of negative and positive valence, values which are rarely approached in median of subjective response to musical stimuli). Therefore, as the positive value of v rises, the x-axis is adjusted by a gradual, though small amount of lowering. Similarly, as the valence becomes more negative, the x-axis is shifted upwards in small, gradual increments.

The face and eyes are drawn within a circle representing the outline of the head. The circle was placed within a square boundary of 300 by 300 pixels. From this constraint the other constants (k_a and k_v) and axis positions were calculated. Valence and arousal values were synchronized with an audio-CD playing the music corresponding to the gathered emotion data. The audio-CD track time elapsed was read by an external function written by Sudderth [151]

4. SAMPLE OUTPUTS

The algorithm was applied to data from an earlier study in which arousal and valence data were already collected [13]. The samples shown here were selected to exemplify parts of the music where extreme emotional responses occurred. The first example (Figure 2) shows one of the lowest valence points occurring in the slow movement of *Concierto de Aranjuez* by Rodrigo, which occurs around the 263rd second of the piece in the recording used. The mouth shape is a negative parabola because the valence is negative (-32 on a scale of –100 to +100), reflecting the frown, and the eyes are in a roughly neutral position, though slightly closing because of the small, negative valence (-7, also on a scale of –100 to +100).

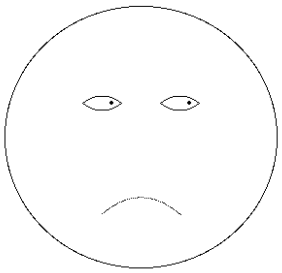

Figure 2. *EmotionFace display at the 263rd second of the Aranjuez concerto, where arousal was –7 and valence was –32 (each on a –100 to +100 scale).*

Figure 3 shows the dynamic progression of the face at the opening of Dvorak's *Slavonic Dance No. 1* Op. 46, which commences with a loud, sustained chord. *EmotionFace* always commences a piece in the neutral position (approximately 0 valence and arousal: The data upon which the facial expressions were calculated for the Dvorak can be seen in Table 1). While there is known to exist some time lag between musical activity and associated emotional response [4], the startle of the loud beginning of this piece (see score in Figure 3) promptly leads *EmotionFace* to a wide eye opening, before the valence of the music is noticeably altered. After a few seconds, when the *furiant* has commenced in the major key, the valence increases, as reflected in the growing, concave up, parabolic smile, most noticeably at about the 6th second. At the sixth second there is a noticeable visual indication of a positive valence expression.

Time (seconds)	Arousal (-100 to +100)	Valence (-100 to +100)
0	1	1
1	8.5	2
2	50	2
3	68	2
4	73	5
5	76	11
6	82	21
7	81	25
8	85	32
9	85	34
10	85	37
11	86	43

Table 1: Sample by sample median values of continuous ratings of subjectively determined arousal and valence expressed by Dvorak's Slavonic Dance, shown in Figure 2. Rated by 67 participants in from an earlier study [13].

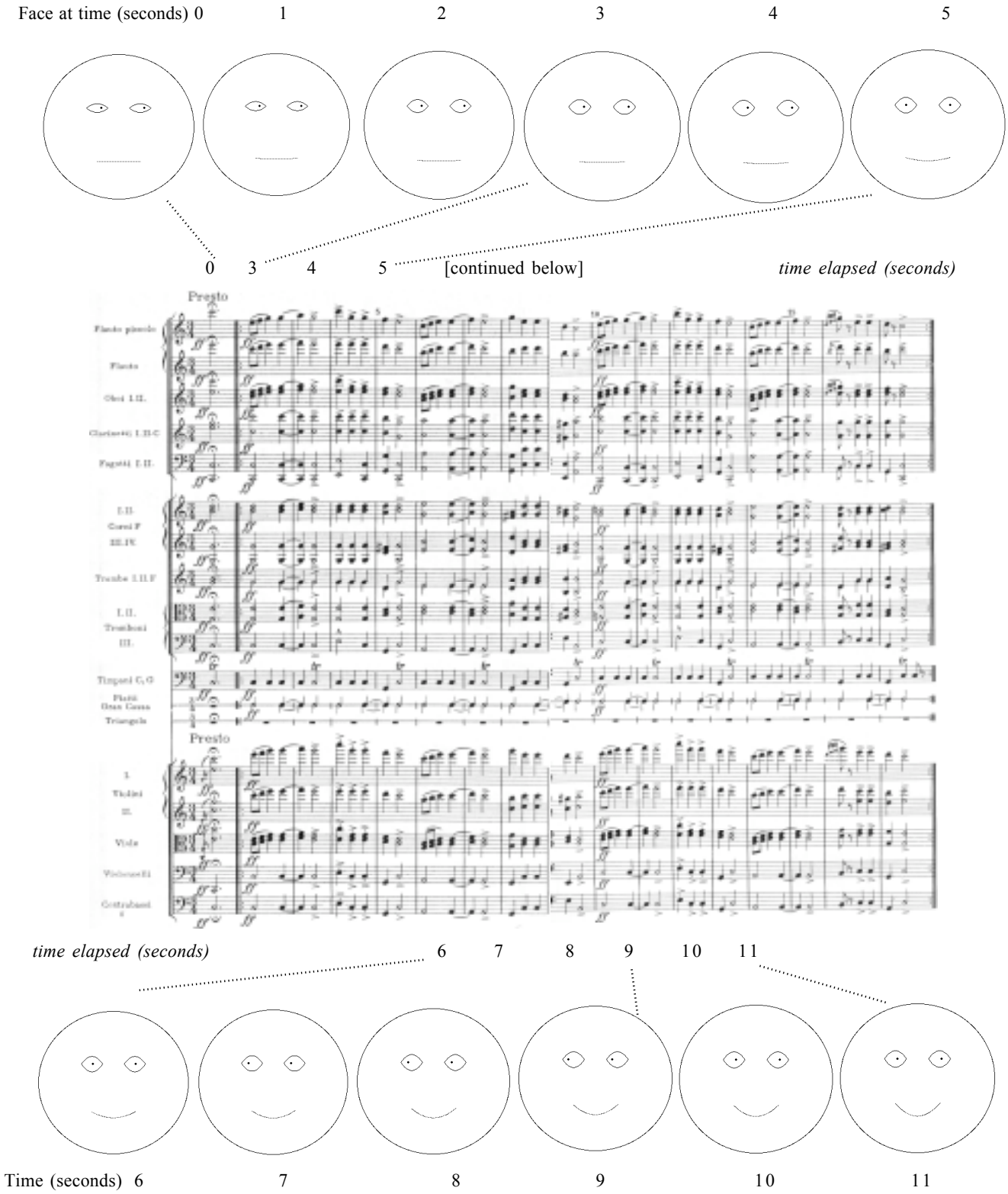

Figure 2. *EmotionFace* screen shots for the first 11 seconds of *Slavonic Dance No. 1* Op 46 by Dvorak. Each face drawn corresponding to each second of music. The second half-dozen screen shots are shown below the musical score for ease of viewing. Musical score source: Antonin Dvorak *Slavonic Dances No. 1, Op. 46, in full score*. Dover Publications, New York, (1987), pp. 1-2.

5. CONCLUSION

The *EmotionFace* interface provides an alternative, intuitive method of displaying emotion expressed by music. The approach provides another tool for examining dynamic and time dependent emotion responses to music. In some respects it provides a more meaningful display than a two dimensional plot of the arousal and valence because the human's strong affinity toward the interpretation of facial expressions. The method may have applications for pedagogues by teaching students about the kinds of emotion that music can express. On a more trivial level it could be used to accompany music on people's audio reproduction systems. If this is to occur, a database of emotional responses to many pieces of music needs to be gathered. More serious future work needs to address the lag structure between the emotion expressed by the music and when it is noticed by the listener. For example, in the Dvorak excerpt described, there is a fairly sudden increase in arousal response almost immediately (in about one or two seconds) after the piece commenced. However, Schubert & Dunsmuir [16] demonstrated that the typical delay between music and emotion is around 3 seconds. Should the facial model reflect this dynamically varying delay between causal musical features and emotional response, or should it be tied directly (instantaneously) to the musical features? Further work will also examine alternative algorithms for displaying facial expressions, or the use of a database of standardized emotional expressions.

Eventually, it may be possible to extract emotional information directly from the musical signal. This is most likely to occur when subjective measurements can be modeled with musical features alone [17], and when these musical features can be automatically extracted in real time. Alternatively, it may become possible to compose pieces of music based on facial expressions. With our current knowledge of the relationship between arousal and valence in both facial expression and in music, the results would most likely be quite primitive. However, in years to come, the prospect of facially produced music composition may become a viable proposition.

6. ACKNOWLEDGEMENT

This research was supported by an Australian Research Council Grant ARC-DP0452290. I am grateful to Daniel Woo from the School of Computer Science and Engineering at the University of New South Wales for his assistance in the preparation of this paper.

7. REFERENCES

[1] Radocy, R. E. & Boyle, J. D., *Psychological foundations of musical behaviour* (2nd ed.), Springfield, IL: Charles C. Thomas (1988).

[2] Ekman, P., & Friesen, W.V., Constants across cultures in the face and emotion, *J. Personality Social Psychol. 17 (2)* (1971), 124–129.

[3] Gabrielsson, A. Emotion perceived and emotion felt: Same or different? *Musicae Scientiae. Spec Issue, 2001-2002* (2002), 123-147.

[4] Schubert, E., "Continuous Measurement of Self-Report Emotional Response to Music", in P. Juslin and J. Sloboda (Eds.), *Music and Emotion: Theory and Research*, Oxford University Press, Oxford (2001), pp. 393-414.

[5] Madsen, C. K., "Emotional response to music as measured by the two-dimensional CRDI", *Journal of Music Therapy, 34* (1997), 187-199.

[6] Schubert, E., "Measuring Temporal Emotional Response to Music Using the Two Dimensional Emotion Space", *Proceedings of the 4th International Conference for Music Perception and Cognition*, Montreal, Canada (11-15 August) (1996), 263-268.

[7] Darwin, C., *The Expression of the Emotions in Man and Animals,* University of Chicago Press, Chicago (1965/1872).

[8] Adolphs, R. et al., Cortical systems for the recognition of emotion in facial expressions *Journal of. Neuroscience. 16* (1996). 7678–7687

[9] Davidson, R. J. & Irwin, W. The functional neuroanatomy of emotion and affective style *Trends in Cognitive Sciences, 3(1)* (1999), 11-21.

[10] Russell, J. A., Affective space is bipolar. *Journal of Social Psychology, 37* (1979), 345-356.

[11] Fasel, B. & Luettin, J., Automatic facial expression analysis: a survey, *Pattern Recognition 36* (2003), 259 – 275.

[12] Morris, J. S., de Bonis, M. & Dolan, R. J., Human Amygdala Responses to Fearful Eyes, *NeuroImage, 17 (1)* (September 2002), 214-222.

[13] Schubert, E., Measuring Emotion Continuously: Validity and Reliability of the Two Dimensional Emotion Space, *Australian Journal of Psychology, 51* (1999), 154-165.

[14] Du, Y & Lin, X., Emotional facial expression model building, *Pattern Recognition Letters, 24(16)* (2003), 2923-2934.

[15] Sudderth, J. (1995). *CoreCD (Version 1.4)* [computer software]. Core Development Group, Inc. (1995).

[16] Schubert, E. & Dunsmuir, W., Regression modelling continuous data in music psychology, in Suk Won Yi (Ed.), *Music, Mind, and Science,* Seoul National University Press (1999), pp. 298-352.

[17] Schubert, E., Modelling emotional response with continuously varying musical features. *Music Perception, 21(4)* (2004), 561-585.