

Psychophysiological measures of emotional response to Romantic orchestral music and their musical and acoustic correlates

Konstantinos Trochidis, David Sears, Dieu-Ly Tran, Stephen McAdams

CIRMMT, Department of Music Research, McGill University
Konstantinos.Trochidis@mail.mcgill.ca, David.Sears@mail.mcgill.ca, Dieu-Ly.Tran@mail.mcgill.ca, smc@music.mcgill.ca

Abstract. This paper focuses on emotion recognition and perception in Romantic orchestral music. The study seeks to explore the relationship between perceived emotion and acoustic and physiological features. Seventy-five musical excerpts are used as stimuli to gather psychophysiological and behavioral responses of excitement and pleasantness from participants. A set of acoustic features ranging from low-level to high-level information was derived related to dynamics, harmony, timbre and rhythmic properties of the music. A set of physiological features based on blood volume pulse, skin conductance, facial EMGs and respiration rate measurements were also extracted. The feature extraction process is discussed with particular emphasis on the interaction between acoustical and physiological parameters. Statistical relations between audio, physiological features and emotional ratings from psychological experiments were systematically investigated. Finally, a step-wise multiple linear regression model is employed using the best features, and its prediction efficiency is evaluated and discussed. The results indicate that merging the acoustic and psychophysiological modalities substantially improves the emotion recognition accuracy.

Keywords: musical emotion, music perception, feature extraction, music information retrieval, psychophysiological response

1 Introduction

The nature of emotions induced by music has been a matter of much debate. Preliminary empirical investigations have demonstrated that basic emotions, such as happiness, anger, fear, and sadness, can be recognized in and induced by musical stimuli in adults and in young children [1]. The basic emotion model, which claims that music induces four or more basic emotions, is appealing to scientists for its empirical efficiency. However, it remains far from compelling for music theorists, composers, and music lovers because it is likely to underestimate the richness of emotional reactions to music that may be experienced in real life [2]. The question of whether emotional responses go beyond four main categories is a central issue for theories of human emotion [3]. An alternative approach to discrete emotions is to stipulate that musical emotions evolve continuously along two or three major psychological dimensions [4]. There are an increasing number of studies investigating

theoretical models in relation to music, the underlying factors and the mechanisms of emotional responses to music at behavioral [5, 6] and neurophysiological levels [7]. Many studies try to investigate the relationships between physiological features, such as electrocardiogram (ECG), electromyogram (EMG), skin conductance response (SCR) and respiration rate (RR), and emotional responses to music [9, 10, 11]. On the other hand, numerous studies explore the relationships between acoustic features and musical emotion [12, 13, 14]. Most of them try to extract a set of low- and high-level acoustical features representing various music descriptors (rhythm, harmony, tonality, timbre, dynamics) and correlate them with emotional ratings from participants.

The main aim of this paper is to implement an approach for music emotion recognition and retrieval based on both acoustic and physiological features. Our model is based on a previous study [15], which investigated the role of physiological response and peripheral feedback in determining the intensity and hedonic value of the emotion experienced while listening to music. Results from this study provide strong evidence that physiological arousal influences the intensity of emotion experienced with music and affects subjective feelings. Using this fusion model, we systematically combine structural features from the acoustic domain with psychophysiological features in order to further understand their relationship and the degree to which they affect subjective emotional qualities and feelings in humans.

2 Methods

2.1 Participants

Twenty non-musicians ($M = 26$ years of age) were recruited as participants (10 females). They reported less than 1 year of training on an instrument over the past five years, and less than two years of training in early childhood. In addition, all participants reported no hearing problems and that they liked listening to Classical and Romantic music.

2.2 Stimuli

Seventy-five musical excerpts from the late Romantic period were selected for the stimulus set. The selection criteria were as follows. The excerpts had to be anywhere from 35 to 45 seconds in duration, because we wanted 30 seconds of complete music after the fade-ins and fade-outs. The music was selected by the authors from the Romantic, late Romantic, or Neo-classical period (from 1815 to 1900). However, most excerpts were selected from the Romantic and late Romantic period. These genres were selected under the assumption that music from this period would elicit a variety of emotional reactions along both dimensions of the emotion model. Each excerpt had to clearly represent one of the four quadrants of the two-dimensional emotion space formed by the dimensions of arousal and valence. Ten excerpts were chosen from a previous study [16], 21 Romantic piano excerpts from [17] and 44 from our own personal selection. Aside from the high-arousal/negative-valence quadrant, which had 18 excerpts, the other three quadrants contained 19 excerpts. Moreover, the excerpts varied in orchestration, in order to explore the effect of timbre

variation on emotion judgments. Accordingly, there were 3 conditions: orchestral (24), chamber (26), and solo piano (25).

2.3 Procedure

We measured five different physiological signals for each of the participants: facial EMGs, skin conductance, respiration rate and blood volume pulse. The electrodes were placed on the following locations: the middle finger (BVP), the index and ring fingers (SC), above the zygomaticus muscle, located roughly in the center of the cheek (EMG), and above the corrugator super cillii muscle, located above the eyebrow (EMG). The respiration belt was placed around the torso in the middle of the rib cage just below the pectoral muscles.

Before beginning the experiment, a practice trial was presented to familiarize the participants with the experimental task. After listening to each musical excerpt, participants were asked to rate their level of experienced excitement and pleasantness on Likert scales.

3 Audio Feature Extraction

3.1 Low-Level acoustical features

A theoretical selection of musical features was made based on musical characteristics such as dynamics, timbre, harmony, register, and rhythm. A total of 100 features related to these characteristics were extracted from the musical excerpts. For all features, a series of statistical descriptors was computed such as the mean, the standard deviation and the linear slope of the trend across frames, i.e., the derivative. The MIR 1.3.4 Toolbox was used to compute the various low- and high-level descriptors [18].

3.1.1 Loudness features

We computed information related to the dynamics of the musical signals such as the RMS amplitude and the percentage of low-energy frames to see if the energy is evenly distributed throughout the signals or certain frames are more contrasted than others.

3.1.2 Timbre features

Mel Frequency Cepstral Coefficients (MFCCs) used for speech recognition and music modeling were employed. We derived the first 13 MFCCs. Another set of 4 features related to timbre were extracted from the Short-term Fourier Transform: spectral centroid, rolloff, flux, flatness entropy and spectral novelty which indicate whether the spectrum distribution is smooth or spiky. The size of the frames used to compute the timbre descriptors was 0.5 sec with an overlap of 50% between successive windows.

3.1.3 Tonality features

The signals were also analyzed according to their harmonic context. Descriptors such as the Chromagram (energy distribution of the signals wrapped in the 12 pitches), the key strength (i.e., the probability associated with each possible key candidate, through a cross-correlation with the Chromagram and all possible key candidates), the tonal Centroid (a vector derived from the Chromagram corresponding to the projection of the chords along circles of fifths or minor thirds) and the harmonic change detection function (flux of the tonal Centroid) were extracted.

3.1.4 Rhythmic features

A rhythmic analysis of the musical signals was performed. Descriptors such as the fluctuation (the rhythmic periodicity along auditory frequency channels) and the estimation of notes and number of onset and attack times per second were computed. Finally, the tempo of each excerpt in beats per minute (bpm) was estimated.

3.2 High-level acoustical features

In conjunction with the low-level acoustic descriptors, we used a set of high-level features computed with a slightly longer analysis window (3s). The high-level features are characteristics of music found frequently in music theory and music perception research.

3.2.1 Pulse Clarity

This descriptor measures the sensation of pulse in music. Pulse can be described as a fluctuation of musical periodicity that is perceptible as “beatings” in a sub-tonal frequency band below 20 Hz. The musical periodicity can be melodic, harmonic or rhythmic as long as it is perceived by the listener as a fluctuation in time [19].

3.2.2 Articulation

This feature attempts to estimate the articulation from musical audio signals by attributing to it an overall grade that ranges continuously from zero (staccato) to one (legato) by analyzing a set of attack times.

3.2.3 Mode

This feature refers to a computational model that rates excerpts on a bimodal major-minor scale. It calculates an overall output that varies along a continuum from zero (minor mode) to one (major mode) [14].

3.2.4 Event density

This descriptor measures the overall amount of simultaneous events in a musical excerpt. These events can be melodic, harmonic and rhythmic, as long as they can be

perceived as independent entities by listeners.

3.2.5 Brightness

This descriptor measures the sensation of how bright a musical excerpt is felt to be. Attack, articulation, or the unbalance or lacking of partials in other regions of the frequency spectrum can influence its perception.

3.2.6 Key Clarity

This descriptor measures the sensation of tonality, or tonal center in music. This is related to the sensation of how tonal an excerpt of music is perceived to be by listeners, disregarding its specific tonality, but focusing on how clear its perception is. This scale is also continuous, ranging from zero (atonal) to one (tonal).

4 Feature extraction of physiological signals

From the five psychophysiological signals we calculated a total of 60 features including conventional statistics in time series, frequency domain and sub-band spectra as suggested in [20].

4.1 Blood volume pulse

To obtain the HRV (heart rate variability) from the continuous BVP signal, each QRS complex was detected and the RR intervals (all intervals between adjacent R waves) or the normal-to-normal (NN) intervals (all intervals between adjacent QRS complexes resulting from sinus node depolarization) were determined. We used the QRS detection algorithm in [21] in order to obtain the HRV time series. In the time-domain of the HRV, we calculated statistical features including mean value, standard deviation of all NN intervals (SDNN), standard deviation of the first difference of the HRV, the number of pairs of successive NN intervals differing by greater than 50 ms (NN50), and the proportion derived by dividing NN50 by the total number of NN intervals. In the frequency-domain of the HRV time series, three frequency bands are of interest in general; very-low frequency (VLF) band (0.003-0.04 Hz), low frequency (LF) band (0.04-0.15 Hz), and high frequency (HF) band (0.15-0.4 Hz). From these sub-band spectra, we computed the dominant frequency and power of each band by integrating the power spectral densities (PSD) obtained by using Welch's algorithm, and the ratio of powers between the low-frequency and high-frequency bands (LF/HF).

4.2 Respiration

After detrending and low-pass filtering, we calculated the Breath Rate Variability (BRV) by detecting the peaks in the signal within each zero-crossing. From the BRV time series, we computed the mean value, SD, and SD of the first difference. In the

spectrum of the BRV, peak frequency, power of two sub-bands, low-frequency band (0-0.03Hz) and high-frequency band (0.03-0.15 Hz), and the ratio of power between the two bands (LF/HF) were calculated.

4.3 Skin conductance

The mean value, standard deviation, and mean of the first and second derivatives were extracted as features from the normalized SC signal and the low-passed SC signal using a 0.2 Hz cutoff frequency. To obtain a detrended SCR (skin conductance response) waveform without DC-level components, we removed continuous, piecewise linear trends in the two low-passed signals, i.e., very low-passed (VLP) with 0.08 Hz and low-passed (LP) signal with 0.2 Hz cutoff frequency.

4.4 Electromyography (EMGs)

For the EMG signals, we calculated similar types of features as in the case of the SC signal. From normalized and low-passed signals, the mean value of the entire signal, the mean of first and second derivatives, and the standard deviation were extracted as features. The number of occurrences of myo-responses and the ratio of these responses within VLP and LP signals were also added to the feature set in a similar manner used for detecting the SCR occurrence, but with 0.08 Hz (VLP) and 0.3 Hz (LP) cutoff frequencies.

5 Results

For the 75 excerpts a step-wise multiple linear regression to predict the participant ratings based on the acoustical and physiological descriptors between the acoustical and physiological descriptors and participant ratings were computed to gain insight into the importance of features for the arousal and valence dimensions of the emotion space. Table 1 provides the outcome of the MLR analysis of the acoustic features onto excitement and pleasantness coordinates of the excerpts and Table 2 the outcome of the analysis of the acoustic and physiological features onto the same coordinates. The resulting model provides a good account of excitement with an $R^2 = 0.81$ (see Table 1) using only the acoustic features spectral fluctuation ($\beta = 0.551$), entropy ($\beta = 0.302$) and spectral novelty ($\beta = -0.245$). For pleasantness, the model provides an $R^2 = 0.44$ using only the acoustic features Mode ($\beta = 0.5$), Key Clarity ($\beta = 0.27$) and entropy of Chroma ($\beta = 0.381$).

The model using both acoustic and physiological features provides an $R^2 = 0.85$ (see Table 2) with spectral fluctuation ($\beta = 0.483$), entropy ($\beta = 0.293$), spectral novelty ($\beta = -0.239$), the std of the first derivative of the zygomaticus EMG ($\beta = -0.116$), skin conductance ratio ($\beta = 0.156$), and the maximum value of the amplitude in blood volume pulse ($\beta = -0.107$). The model provides for pleasantness an $R^2 = 0.54$ using the acoustic and physiological features Mode ($\beta = 0.551$), Key Clarity ($\beta = 0.211$), entropy of Chroma ($\beta = 0.334$), the minimum of the std of the first derivative of the zygomaticus EMG ($\beta = 0.25$), and the minimum of the blood volume pulse ($\beta = -0.231$).

Table 1. Outcome of the multiple linear regression analysis of the acoustic features onto the coordinates of the emotion space.

Excitement	β	Pleasantness	β
Fluctuation	0.551	Mode	0.5
Entropy	0.302	Key Clarity	0.27
Novelty	-0.245	Chroma Entropy	0.381

Table 2. Outcome of the multiple linear regression analysis using acoustic features and physiological features onto the coordinates of the emotion space.

Excitement	β	Pleasantness	β
Fluctuation	0.481	Mode	0.551
Entropy	0.293	Key Clarity	0.221
Novelty	-0.23	Chroma Entropy	0.334
1 diff EMGZ std	-0.11	1 diff EMGZ min	0.25
SC Ratio	-0.15	BVP min	-0.231

6 Conclusions

In the present paper, the relationships between acoustic and physiological features in emotion perception of Romantic music were investigated. A model based on a set of acoustic parameters and physiological features was systematically explored. The regression analysis shows that low- and high-level acoustic features such as Fluctuation, Entropy and Novelty combined with physiological features such as the first derivative of EMG Zygomaticus and Skin Conductance are efficient in modeling the emotional component of excitement. Further, acoustic features such as Mode, Key Clarity and the Chromagram combined with the minimum of the first derivative of EMG zygomaticus and blood volume pulse effectively model the emotional component of pleasantness. Using the existing approach merging acoustic and physiological features boosts the correlation with behavioral estimates of subjective feeling in listeners in terms of excitement and pleasantness. Results show an increase in the prediction rate of the model of 4% for excitement and 10% for pleasantness when psychophysiological measures are added to acoustic features.

Future work will explore and investigate by means of a similar model which low- and high-level acoustical and physiological features influence human judgments on semantic descriptions and perceptual qualities such as speed, articulation, harmony, timbre and pitch.

Acknowledgments. Konstantinos Trochidis was supported by a post-doctoral fellowship by the ACN Erasmus Mundus network. and a grant to Stephen McAdams from the Social Sciences and Humanities Research Council of Canada. The authors thank Bennett Smith for valuable technical assistance during the experiments.

References

1. Dolgin, K. G., & Adelson, E. H.: Age changes in the ability to interpret affect in sung and instrumentally-presented melodies. *Psychology of Music*, 18, 8--98 (1990)
2. Zentner, M., Grandjean, D., & Scherer, K.: Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion*, 8, 494--521 (2008)
3. Ekman, P.: *The nature of emotion: Fundamental questions*. New York: Oxford University Press (1994)
5. Russell, J. A.: A circumplex model of affect. *Journal of Personality and Social Psychology*, 39, 1161--1178 (1980)
6. Juslin, P. N., & Västfjäll, D.: Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, 31, 559--575 (2008)
7. Juslin, P. N., & Sloboda, J. A.: Psychological perspectives on music and emotion. In: P. N. Juslin & J. A. Sloboda (eds.), *Music and emotion: Theory and research* (pp. 361--392). New York: Oxford University Press (2001)
8. Schmidt L. A., Trainor, L. J.: Frontal brain activity (EEG) distinguishes valence and intensity of musical emotions, *Cognition and Emotion*, 15, 487--500 (2001)
9. Gomez, P., & Danuser, B.: Relationships between musical structure and psychophysiological measures of emotion. *Emotion*, 7(2), 377--387 (2007)
10. Khalfa, S., Peretz, I., Blondin, J.P., & Manon, R.: Event-related skin conductance responses to musical emotions in humans. *Neuroscience Letters*, 328, 145--149 (2002)
11. Sears, D., Ogg, M., Benovoy, M., Tran, D. L., S. McAdams, S.: Predicting the Psychophysiological Responses of Listeners with Musical Features. Poster presented at the 51st Annual Meeting of the Society for Psychophysiological Research, Boston, MA, September 14-18 (2011)
12. Eurola, T., Lartillot, O., Toivainen, P.: Prediction of Multidimensional Emotional ratings in Music from Audio Using Multivariate Regression Models, in *Proc. ISMIR* (2009)
13. Fornari, J. & Eerola, T.: Computer Music Modeling and Retrieval. Genesis of Meaning in Sound and Music, in *Lecture Notes in Computer Science*, chapter *The Pursuit of Happiness in Music: Retrieving Valence with Contextual Music Descriptors*, 5493, 119-133. Springer (2009)
14. Saari, P., Eerola, T., & Lartillot, O.: Generalizability and simplicity as criteria in feature selection: Application to mood classification in music. *IEEE Transactions in Audio, Language, and Speech Processing*, 19 (6), 1802--1812 (2011)
15. Dibben, N.: The role of peripheral feedback in emotional experience with music. *Music Perception*, 22(1), 79--116 (2004)
16. Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., & 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)
17. Ogg, M.: *Physiological responses to music: measuring emotions*. Undergraduate thesis. McGill University (2009)
18. Lartillot, O., & Toivainen, P.: MIR in Matlab (II): A Toolbox for Musical Feature Extraction From Audio, *Proceedings of the International Conference on Music Information Retrieval*, Wien, Austria (2007)
19. Lartillot, O. Eerola, T., Toivainen, P. Fornari, J.: Multi-feature modeling of pulse clarity: Design, validation, and optimization. In *Proceedings of the International Symposium on Music Information Retrieval* (2008)
20. Kim, J. and André, E.: Emotion Recognition Based on Physiological Changes in Listening Music, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(12), 2067--2083 (2008)
21. Pan, J. and Tompkins, W.: A Real-Time QRS Detection Algorithm, *IEEE Trans. Biomedical Eng.*, 32(3), 230--233 (1985)