Modeling Affective Responses to Music using Audio Signal Analysis and Physiology

CONFERENCE PAPER · JUNE 2015

DOI: 10.13140/RG.2.1.4902.8969

2 AUTHORS:

Konstantinos Trochidis

Singapore University of Technology and De…

Simon Lui

Singapore University of Technology and De…

Modeling Affective Responses to Music using Audio Signal Analysis and Physiology

Konstantinos Trochidis[1] and Simon Lui[1],

[1] Department of Information Systems Technology and Design
Singapore University of Technology and Design
Konstantinos@sutd.edu.sg, Simon_Lui@sutd.edu.sg

Abstract. A key issue in designing personalized music affective applications is to find effective ways to direct emotion by music selection with appropriate combination of acoustic features. The aim of this study is to understand the dynamic relationships between acoustic features, physiology and affective states. To model these relationships we used a multivariate approach including continuous measures of emotions from behavioral, subjective and physiological responses. Classical music excerpts taken from opera overtures were used as stimuli to induce emotional variations across time between neutral and intense emotional states. Continuous ratings of arousal and valence along with cardiovascular, respiratory, skin conductance and facial expressive activity were recorded simultaneously. Results show that parts of the music with higher loudness and pulse clarity induced higher ratings of arousal, sympathetic activation and increased cardiorespiratory synchronization. In contrast, pleasant and calming parts with major mode and prominent key strength induced higher ratings of valence, parasympathetic activation and increased facial activity.

Keywords: musical emotion, emotion recognition, acoustic features, physiological responses, affective computing

1 Introduction

Music by its nature has the ability to communicate strong emotions in everyday experiences. Emotions expressed or induced by music is one of the central aspects in music listening and is the main reason why music appeals to people. Given the important role of emotion in music listening, there has been intensive research on the field during the last two decades, which contributed to important developments. A large number of research approaches have been used to gain a better understanding of features and processes related to musical emotions.

Musical characteristics, such as tempo, mode, loudness, pitch, timbre, and so on, are inherent properties of the musical structure, and their influence on emotional responses to music has been shown [1]. Many studies explored the relationship between acoustic features and musical emotions [2], [3], [4]. Most of them try to extract low- and high-level acoustic features representing various music descriptors (timbre, dynamics, pitch, melody, harmony) and correlate them with emotional ratings from participants. On the other hand, there is a large amount of studies establishing the relationship between physiological changes and musical emotions during music listening [5]. Research on physiological effects of music includes mainly changes in heart rate, respiration, skin conductance and muscle tension [6], [7], [8], [9]. Most of the existing studies consider acoustic features and physiological responses separately. Few studies combined both modalities to improve emotion recognition, which indicated that merging acoustic and physiological modalities substantially improves prediction of participants' ratings of felt emotion [10], [11]. The relationship, however, between music features, physiological responses and affective states remains unexplored.

Aims of the study

The primary aim of this study is the development of an accurate and robust model for mapping the relationships between physiology and affective states. A multivariate approach is used including continuous measures of emotions from behavioral, subjective and physiological responses from listeners. To this end, instead of using only monovariate physiological measures, including skin conductivity, heart or respiration rate, bivariate measures related to cardio-respiratory synchronization were also employed. These measures reflect the coherence within the peripheral physiological system during emotion elicitation and therefore, can be used for more accurate emotion detection.
A second aim of this paper is to build a model that links musical features, physiology and affective states in order to systematically investigate their underlying relationships. In this context, we focused on 5 musical characteristics: loudness, brightness, pulse clarity, mode and key clarity, which are known to explain a large amount of musical variability [12]. Furthermore, we incorporated physiological and behavioral responses to effectively validate the responses found by subjective feeling reports from listeners.
Combining audio and physiology modalities improves music selection based on music retrieval and evaluation of the system by using affective detection and physiology. This approach can be incorporated in the development of robust affective music applications to enhance personalization and direct users current affective state.

2 Methods

Participants

Twenty-eight non-musicians students from SUTD were recruited as participants (14 males and 14 females). The mean age was 25 years old. All participants reported no hearing problems and that they liked listening to Classical music. In accordance with the requirements of Singapore University of Technology and Design Research Ethics Board, which certified this study, written informed consent was obtained from each participant prior to the experiment. No symptoms of cardiovascular, mental, or neurological disorders were reported by any of the 28 participants.

Stimuli and Apparatus

Two musical excerpts with duration of 10 minutes each taken from Classical music overtures was selected as stimuli (William Tell by Rossini and Prince Igor by

Borodin). The excerpts were selected to elicit a variety of emotional reactions along both the two-dimensional emotion space formed by the dimensions of arousal and valence. The stimuli were presented to the participants in a randomized order. The experimental session was programmed and run using an emotion tracking application programmed on a Macintosh workstation.

By moving the index finger of their dominant hand using a track-pad to move to a two dimensional emotion space from left to right, participants were instructed to indicate how positive the effect of the music was (left = negative and unpleasant; right = positive and pleasant).

By moving their finger from top to bottom, participants indicated the degree of their emotional arousal while listening to the music (top = excited; bottom = calm). Participants were instructed to rate their current emotional state on both dimensions simultaneously, with the finger position at each moment reflecting their emotional response to the piece as they were listening. They were also asked not to rate emotions recognized, but only their own emotional response.

The stimuli were presented over Sony BM6A headphones. Audio signals were sampled at 44.1 kHz with 16-bit resolution. We used two ProComp SA9309M sensors (Thought Technology, Inc., Montreal, QC) to measure skin conductance (galvanic skin response, or GSR), which were attached to the index and ring fingertips of the non-dominant hand. ECG signals were recorded with three electrodes attached to the participant's chest in a triangular configuration (three leads situated on the upper left and upper right chest, and on the left lower ribcage). Recording of respiratory signals was achieved using a ProComp SA9311M respiration stretch sensor (Thought Technology, Inc., Montreal, QC) strapped around the chest just below the pectoral muscles with Velcro. Expressive muscle activations were measured using one electromyography (EMG) electrode (MyoScan-Pro surface EMG sensors) placed on the zygomaticus major (associated with smiling) muscles. EMG electrode was placed on the side of the face contralateral to the dominant hand (with positive and negative electrodes aligned with the respective muscles and the reference electrodes placed on the cheek bone). The signals were collected using the ProComp Infiniti Unit. Signal amplitudes were registered in 16-bit integer format on a hard drive at a rate of 256 Hz with time stamps.

Procedure

To begin, the experimenter first attached the sensors to the participant's skin. Once comfortably installed, they were instructed to sit back and relax and to stay still and awake during the experiment. Once the experimenter left the room, each participant was introduced to the experiment by reading instructions of the experimental procedure in the screen from the emotion tracking application run on the computer. Subsequently, the music was presented to participants randomized in the following manner. Before and between the music pieces, a two-minute physiological baseline activity without any stimulation was recorded. Before beginning the experiment, a practice trial of one minute using a musical excerpt by Bela Bartok was presented to familiarize the participants with the experimental task and the continuous rating procedure.

3 Audio feature extraction

A theoretical selection of musical features was made based on musical characteristics such as dynamics, timbre, harmony, rhythm using the MIR Toolbox for MATLAB [13]. A total of 5 descriptors related to these features were thus extracted from the musical excerpts.

Dynamics

We computed the RMS amplitude to examine whether the energy is evenly distributed throughout the signals, or to determine whether certain time points are more contrasted than others.

Timbre

Brightness was computed as the relative amount of spectral energy above a certain cut-off frequency (1500 Hz).

Tonality

The signals were also analyzed according to their harmonic characteristics. A chromagram representing the distribution of pitch-classes is created. The key strength computed the cross-correlation of the chromagram with each possible major or minor key. The key clarity is the key strength of the key with the highest key strength out of all 24 keys. To model the mode of each piece, a computational model that distinguishes major and minor excerpts was employed. It calculates an overall output that continuously ranges from zero (minor mode) to one (major mode) [13].

Rhythm

The Pulse Clarity, a measure of the rhythmical and repetitive nature of a piece, was estimated by the autocorrelation of the onset detection curve of the signal.

4 Data analysis of physiological and behavioral measures

Preprocessing of all physiological and behavioral continuous signals recorded with a sample rate of 256 Hz was done in MATLAB (The Mathworks, Version 7.14.0.739). For each piece we extracted the physiological data using a frame size of 30 secs with an overlap of 10 secs. All signals were filtered in order to remove extraneous information, using a linear-phase filter based on the convolution of a 4th-order Butterworth filter impulse response. Subsequently, ECG signals were band-pass filtered (1-30 Hz), respiration activity was low-pass filtered (1 Hz), and skin conductance level was low-pass filtered (0.3 Hz). We subtracted all the physiology values in the music listening condition from the baseline physiology values, because we were interested in comparing the change in reactivity from the baseline condition during music listening.

Heart-rate variability analysis

To quantify HRV, a series of R peaks were derived from the ECG signal. The ECGs were further filtered (passband was set to 1-30 Hz) to remove power line noise, baseline wander and muscle noise, but to preserve most of the spectral components of the QRS complexes. To avoid phase shifting, the signal peak filtering was performed with a linear-phase filter constructed from the convolution of a 4th-order Butterworth filter impulse response convolved with itself in reverse time. An adaptive derivative-based algorithm to detect the QRS complex was applied to construct the IBI time series consisting of inter-beat intervals that were subsequently obtained as differences between successive R-peak occurrence times. The HRV analysis measured six parameters: two time-domain measures and four frequency-domain measures. The two time-domain measures were mean heart rate (HR) and the square-root difference in the R-R interval (RMSSD). Frequency-domain measures included mean high-frequency power (absHF), mean low-frequency power (absLF), and the ratio of low-frequency to high-frequency power (LF/HF ratio). We obtained frequency-domain measures from the IBI time series by applying Fast Fourier Transform (FFT) over the time series. The calculated frequency spectrum was divided into two spectral bands: low frequency (LF, 0.04-0.15 Hz), and high frequency (HF, 0.15-0.4Hz).

EMG analysis

The MyoScanPro EMG sensor automatically converted the signal to a root mean square signal (after an internal analog rectification), which was therefore not preprocessed any further (capturing EMG activity at frequencies up to 500 Hz).

Continuous ratings analysis

We removed any individual differences in scale use by individual range normalization dividing each participant's rating by his or her individual range of ratings over the entire experiment and then subtracting each participant's resulting minimum rating value over the entire experiment, creating a range for each participant from 0 to 1.

Cardio-respiratory coupling analysis

In previous studies heart and respiration rates are considered separately as monovariate measures of physiological activity during music listening. In our study, the cardiorespiratory synchronization is introduced as a bivariate measure of their interaction that better reflects the emotional state. Considering the cardiac and respiratory system as two coupled oscillators, we denote the phase of the heart as ϕ_h and the phase of respiration rate as ϕ_r respectively.

Assuming m heartbeats occur within n heart or respiratory cycles, synchronization is understood as phase locking of the corresponding phases given by:

$$|m\phi_h - n\phi_r| < const \tag{1}$$

where m and n are integers denoting the cardiac and the respiratory cycles, respectively. In other words, if the instantaneous phase difference between the oscillators remains constant within a given threshold, the oscillators are considered to be synchronized in their phase modulations.

To assess and quantify the cardio-respiratory synchronization, the synchronization γ index was employed which measures the degree of phase locking (synchronization) and calculated as:

$$\gamma = \left\| \frac{1}{M} \sum_{j=1}^{M} e^{i\theta(tj)} \right\|^2 \qquad (2)$$

where $\theta(t_j) = m\phi_h(t_j) - n\phi_r(t_j)$ is the phase difference, t_j is the time point for $j = 1,...,M$, and M is the number of sampling data points in the given time interval. This method is based on a mathematical model proposed in earlier studies [14].

If both oscillators are synchronized, γ = 1, whereas the completely desynchronized case yields γ = 0.

To detect synchronization for different $m:n$ ratios, this method has to be applied for each desired $m:n$ ratio. In practice, synchronization is present if γ exceeds a pre-defined threshold $\gamma \geq thres_\gamma$.

The study was carried out for the following $m:n$ ratios: n = 1; m = 2, … 8 and n = 2; m = 5, 7, 9, 11, 13.

5 Results

For testing of significance on listener's responses, we employed a hierarchical linear modeling approach using the MIXED procedure in SPSS Statistics (IBM, Version 21). Estimation of parameters was based on restricted maximum likelihood. Beside fixed effects coefficients, the models included an intercept and a first-order autoregressive residual covariance structure (AR1).

William Tell – Gioachino Rossini

Subjective Arousal Scores: There was a positive effect of loudness (b = 4.96, t = 27.637, $p < 0.01$), brightness (b = .93, t = 13.718, $p < 0.01$) and pulse clarity (b = .7, t = 26.304, $p < 0.01$) features on arousal. Participants responded with increased arousal ratings in louder, brighter and prominent beat sections of the pieces. On the other hand, there was a negative effect of mode (b = -.85, t = -12.690, $p < 0.01$) indicating that as the piece became less major the ratings reported were more arousing.

Subjective Valence Scores: There was a positive effect of key clarity (b = .19, t = 2.546, $p < 0.01$) and mode (b = .42, t = 6.991, $p < 0.001$) features on valence. On the other hand, there was a negative effect of pulse clarity (b = -.15, t = -5.676, $p < 0.01$)

and loudness (b = -.83, t = -4.496, $p < 0.01$) indicating that as the piece became less beat prominent and less loud the ratings reported were more positive.
EMG Zygomaticus Scores: There was a positive main effect of pulse clarity (b = .75, t = 27.637, $p < 0.01$) and mode (b = .35, t = 4.427, $p < 0.01$) features associated with increased facial expressive activity.
SCL Scores: There was a positive effect of pulse clarity (b = .25, t = 2.469, $p < 0.01$). No other effects were found.
Cardiorespiratory coupling: For the cardiorespiratory data there was a positive effect of pulse clarity (b = 1.73, t = 2.021, $p < 0.05$) and loudness (b = 19.32, t = 3.337, $p < 0.01$). Participants responded to increases in loudness and pulse clarity with higher prominence of 3:1 cardiorespiratory synchronization ratio.
HRV scores: There was a negative effect of pulse clarity (b = -.11, t = -5.237, $p < 0.01$) with increased parasympathetic activity in less beat prominent parts of the music.

Prince Igor – Alexander Borodin

Subjective Arousal Scores: There was a positive effect of pulse clarity (b = .27, t = 6.224, $p < 0.01$) and loudness (b = 6.03, t = 32.887, $p < 0.01$) features. Participants responded with increased arousal ratings in loud and more prominent beat sections of the pieces. On the other hand, there was a negative effect of mode (b = -.21, t = -2.369, $p < 0.05$) indicating that as the piece became less major the ratings reported were more arousing.
Subjective Valence Scores: There was a positive effect of key clarity (b = .73, t = 10.148, $p < 0.01$) and mode (b = .17, t = 2.386, $p < 0.05$) features. Participants responded with increased positive ratings in more major and key prominent parts of the pieces. On the other hand, there was a negative effect of loudness (b = -1.64, t = -7.746, $p < 0.01$) indicating that as the piece became less loud the ratings reported were more positive.
SCL Scores: There was a positive effect of pulse clarity (b = .47, t = 4.559, $p < 0.01$). No other effects were found.
Cardiorespiratory coupling: For the cardiorespiratory data there was a positive effect of loudness (b = 13.57, t = 2.673, $p < 0.01$) and pulse clarity (b = 2.46, t = 2.208, $p < 0.05$). Participants responded to increases in loudness and pulse clarity with higher prominence of 3:1 cardiorespiratory synchronization ratio.
HRV scores: There was a positive effect of mode (b = .02, t = 2.302, $p < 0.05$) with increased parasympathetic activity in major parts of the music.

6 Discussion

The results show that all music features (loudness, pulse clarity, brightness, mode and key clarity) influence both subjective and physiological measures of emotion. It is shown that arousal is positively related to loudness, pulse clarity and brightness, whereas valence is positively related to key clarity and mode. The relationships found between music features and emotion dimensions are in agreement with existing studies [1]. Minor mode is known to be associated to negative valence. In our study we found that minor mode (less major) is related to higher arousal. This finding has been also observed in previous studies [15]. A possible explanation for this result is the long duration music pieces used in our study where the music features are present and evolve in time.

In relation to physiological measures, it was found that SCL is positively associated to pulse clarity, i.e. beat prominent segments of music. Additionally, we found that EMG activity relates positively with both pulse clarity and mode, whereas HRV increases with increasing mode. Overall, the results are consistent with existing studies [8]. Inour study, instead of using monovariate measures of heart and respiration rates we employed the cardiorespiratory synchronization as a bivariate physiological measure of affect. We found that cardiorespiratory synchronization increases with increasing pulse clarity and loudness. During strong beat and loud music segments, the heart and respiration oscillators become stronger phase locked. In addition to its importance as a bivariate physiological measure of affect, cardiorespiratory synchronization may be related to rhythmic entrainment mechanism of emotion elicitation [12]. Rhythmic entrainment is the process where an emotion is induced by a piece of music because the strong external rhythm of music interacts with and synchronizes to internal heart or respiration rhythms producing increased arousal.

7 Conclusions

The aim of this paper was to investigate the dynamic relationships between acoustic features, physiology and affective states. The results are consistent with existing studies and show that parts of the music with higher loudness and pulse clarity induced higher ratings of arousal, sympathetic activation and increased cardiorespiratory synchronization. In contrast, parts within the pieces with major mode and prominent key strength induced higher ratings of valence, parasympathetic activation and higher levels of facial activity. A novelty of our study is the employment of cardiorespiratory synchronization as a bivariate physiological measure of affect, which may be related to rhythmic entrainment mechanism of emotion elicitation. Future work will examine methods to capture response coherence among different emotional components including behavioral, physiological and neural responses. Practical implications of this work may include the development of affective music retrieval systems, design of music-computer interfaces and music therapy applications.

Acknowledgments. This work is supported by the SUTD-MIT International design center (IDC) Research Grant (IDG31200107 / IDD11200105 / IDD61200103). The authors thank Hải Lê Vũ for valuable technical assistance during the experiments.

References

1. Gabrielsson, A., Lindstroem, E.: The role of structure in the musical expression of emotions. In: Juslin PN, Sloboda JA (eds) Handbook of music and emotion: theory, research, applications, pp 367–400. Oxford University Press, New York (2010)

2. Eerola, T., Lartillot, O., and Toiviainen, P.: Prediction of Multidimensional Emotional Ratings in Music from Audio Using Multivariate Regression Models. In: Proceedings of the International Conference for Music Information Retrieval (ISMIR), (2009)
3. Fornari, J., & Eerola, T.: The pursuit of happiness in music: retrieving valence with contextual music descriptors. In Computer music modeling and retrieval. Genesis of meaning in sound and music (pp. 119-133). Springer Berlin Heidelberg (2009)
4. Trochidis, K., Delbé, C., & Bigand, E.: Investigation of the relationships between audio features and induced emotions in Contemporary Western music. In Proceedings of the 8th Sound and Music Computing Conference (2011)
5. Hodges D.A: Psychophysiological responses to music. In: Juslin PN, Sloboda JA (eds) Handbook of music and emotion: theory, research, applications, pp 279–311, Oxford University Press, New York (2010)
6. Krumhansl, C.: An explanatory study of musical emotion and psychophysiology. Canadian Journal of Experimental Psychology 51(4), 336–352 (1997)
7. Gomez, P., Danuser, B.: Relationships between musical structure and physiological measures of emotion. Emotion 7(2), 377–387 (2004)
8. Nyklicek, I., Thayer, J., Van Doornen, L.: Cardiorespiratory differentiation of musically-induced emotion. Journal of Psychophysiology 11, 304–321 (1997)
9. Khalfa, S., Roy, M., Rainville, P., Dalla Bella, S., & Peretz, I: Role of tempo entrainment in psychophysiological differentiation of happy and sad music? International Journal of Psychophysiology, 68 (1), 17-26 (2008)
10. Coutinho, E., & Cangelosi, A.: Musical emotions: predicting second-by-second subjective feelings of emotion from low-level psychoacoustic features and physiological measurements. Emotion, 11(4), 921 (2011)
11. Trochidis, K., Sears, D., Trân, D. L., & McAdams, S.: Psychophysiological measures of emotional response to Romantic orchestral music and their musical and acoustic correlates. In: From Sounds to Music and Emotions (pp. 44-57). Springer Berlin Heidelberg (2013)
12. Juslin, P.N, Sloboda, J.A (eds): Handbook of music and emotion: theory, research, applications. Oxford University Press, Oxford (2010)
13. Lartillot, O., Toiviainen, P.: MIR in Matlab (II): A Toolbox for Musical Feature Extraction From Audio. In: International Conference on Music Information Retrieval, Vienna (2007)
14. Rosenblum, M., Pikovsky, A., Kurths, J., Schäfer, C., & Tass, P. A.: Phase synchronization: from theory to data analysis. Handbook of biological physics, 4, 279-321 (2001)
15. Van der Zwaag, M., Westernik, H.D.M. and Van der Broek, E: Deploying music characteristics for an affective music player. 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, Amsterdam,Netherlands (2009)