Running head: Musical emotions: psycho-physiological investigations.

Musical emotions: predicting second-by-second subjective feelings of emotion from low-level psychoacoustic features and physiological measurements

Eduardo Coutinho[1] and Angelo Cangelosi[2]

[1]University of Liverpool, Liverpool, United Kingdom

[2]University of Plymouth, Plymouth, United Kingdom

Words: 10001

Tables: 7

Figures: 5

Correspondence address:

Eduardo Coutinho

University of Liverpool, School of Music

80-82 Bedford Street South, Liverpool L69 7WW, United Kingdom

Phone: +44 151 794 3096; Fax: +44 151 794 3141

E-mail: e.coutinho@liverpool.ac.uk

Date: February 12, 2010

Abstract

We sustain that the structure of affect elicited by music is largely dependent on dynamic temporal patterns in low-level music structural parameters. In support of this claim, we have previously provided evidence that spatiotemporal dynamics in psychoacoustic features resonate with two psychological dimensions of affect underlying judgments of subjective feelings: arousal and valence. In this article we extend our previous investigations in two aspects. Firstly, we focus on the emotions experienced rather than perceived while listening to music. Secondly, we evaluate the extent to which peripheral feedback in music can account for the predicted emotional responses, i.e., the role of physiological arousal in determining the intensity and valence of musical emotions. Akin to our previous findings, we will show that a significant part of the listeners' reported emotions can be predicted from a set of six psychoacoustic features - loudness, pitch level, pitch contour, tempo, texture and sharpness. Furthermore, the accuracy of those predictions is improved with the inclusion of physiological cues - skin conductance and heart rate. The interdisciplinary work presented here provides a new methodology to the field of music and emotion research based on the combination of computational and experimental work, which aid the analysis of the emotional responses to music, while offering a platform for the abstract representation of those complex relationships. Future developments may aid specific areas, such as, psychology and music therapy, by providing coherent descriptions of the emotional effects of specific music stimuli.

Key words: Emotion, Arousal and Valence, Physiology, Psychoacoustics, Neural Networks

Musical emotions: predicting second-by-second subjective feelings of emotion from low-level psychoacoustic features and physiological measurements

The ability of music to stir human emotions is a widely known fact. However, the manner in which music contributes to those experiences remains largely obscured. Despite the fact that during the last few decades important advances allowed for an enhanced understanding of features related to emotional experiences with music (see Juslin & Sloboda (2010) for an up-to-date compilation of perspectives and approaches to studies on music and emotion), explaining their nature, and unveiling the underlying neurobiological and psychological mechanisms supporting them, continues to be one of the great challenges to scientific research (see Juslin & Västfjäll (2008) for a comprehensive discussion).

(Musical) emotions

Emotions are complex collections of organized chemical and neural responses to specific external or internal stimuli (real or imaginary), signaling events of potential relevance to the organism, and leading to the creation of circumstances advantageous to the organism (Damasio, 2000). These responses comprise a set of interrelated subepisodes, which include at least three broad phenomena: subjective feelings (the conscious "feeling" of an emotion), behavioral changes (expressive displays, e.g., postures, gestures, facial and vocal expressions; and action tendencies, e.g., running away in the case of danger), and physiological arousal (e.g., changes in the somatic and autonomic nervous systems) (Oatley & Jenkins, 1996; Scherer & Zentner, 2001)[i].

One of the main causes of the difficulties in conceptualizing musical emotions (MEs) is their problematic integration within the framework of "real" emotions. Indeed, the goal-oriented nature usually attributed to emotions clashes with music's ability to elicit all kinds of strong and mild emotional reactions in its listeners, an aspect that has been at the very centre of a long-standing controversial discussion: Can music induce emotions? Are MEs just like other ("real") emotions?

Broadly speaking, there are three main theoretical explanations for the nature of MEs, each of which answers these questions in different ways. Some researchers claim that music cannot induce any emotions at all since it does not appear to have any goal implications (e.g., Konečni, 2003). Music may be about emotion, it may emulate emotional dynamics, and it may communicate emotional meaning, but it certainly does not induce emotions directly in listeners. Advocates of a second perspective, suggest that music can induce emotion, but only a limited set of states specific to music, and certainly not the so-called basic emotions, i.e., those emotions related to survival functions (Kivy, 1990; Scherer, 2003). Besides, they involve a set of mechanisms distinct from those associated with "real" emotions (e.g. Scherer & Zentner, 2001). From a third standpoint, some consider that music can induce emotions in its listeners, and that MEs are indeed "real" emotions since they elicit reactions of the same emotion mechanisms (subjective feelings, behavioral changes, and physiological arousal; e.g.; Juslin & Västfjäll, 2008; Panksepp & Bernatzky, 2002; Krumhansl, 1997).

In point of fact, the idea that music can elicit responses in the various subcomponents of emotional reactions (see Juslin & Västfjäll, 2008 for a thorough review) is vital if MEs are to be envisaged as "real" emotions. Admittedly, cumulative evidence suggests that musical stimuli engage our brain-mind (e.g., Koelsch, Fritz, Cramon, Müller, Friederici, 2006; Blood & Zatorre, 2001) and body (e.g., Krumhansl, 1997) in ways similar to other stimuli. Furthermore, while acknowledging that the emotions evoked by a piece of music are a function of several individual and cultural parameters (Scherer & Zentner, 2001), there is pervasive evidence of specific acoustic cues and patterns communicating similar emotions to all listeners (Gabrielsson & Lindström, 2010), and where acoustic profiles transcend cultural boundaries (Fritz, Jentschk, Gosselin, Sammler, Peretz, Turner, Friederici & Koelsch, 2009; Balkwill, Thompson & Matsunaga, 2004; Balkwill & Thompson, 1999). To some extent, listeners seem to construe emotional meaning by attending to structural aspects of the acoustic signal.

Emotional expression: the role of composed structure

The basic perceptual attributes involved in music perception (also referred to as psychoacoustic features) are loudness, pitch, contour, rhythm, tempo, timbre, spatial location and reverberation (Levitin, 2006). While listening to music, our brains continuously organize these dimensions according to diverse gestalt and psychological schemas. Some of these schemas involve further neural computations on extracted features which give rise to higher order musical dimensions (e.g., meter, key, melody, harmony), reflecting (contextual) hierarchies, intervals and regularities between the different music elements. Others involve continuous predictions about what will come next in the music as a means of tracking structure and conveying meaning (e.g., Meyer, 1956). In this sense, the aesthetic object is also a function of its objective design properties (i.e., the way musical features are combined by the composer), and so the subjective experience should be, at least partially, dependent on those features.

The investigation of the factors in musical structure contributing to the perceived emotional expression has a long history (at least back to our ancient Greek ancestors Socrates, Plato, and Aristotle), but it gained particular attention after Kate Hevner's studies during the 1930's (Hevner, 1936, 1937). Hevner was one of the first to systematically analyze which musical parameters (e.g. major versus minor modes, firm versus flowing rhythm, direction of melodic contour) are related to the reported emotion (e.g., happy, sad, dreamy, exciting). Since then, a core interest amongst music psychologists has been the isolation of the perceptible factors in music that may be causally linked to the many observed effects.

There is now robust evidence that certain music dimensions and qualities communicate similar affective experiences to listeners. Gabrielsson and Lindström's (2010) recent review of more than one hundred studies clearly shows that several music variables have consistent relationships with emotions perceived. Interestingly, the most distinct results pertain to basic variables in human audition (low-level features), namely loudness, tempo, timbre, and pitch (although pitch seems to be more ambiguous and context dependant). These features, also known as low-level features, contrast with high-level features in that the latter are culturally shaped and recognizable as traditional forms (e.g., mode, or several aspects of melody and harmony).

In our previous work (Coutinho & Cangelosi, 2009; Coutinho, 2010), we proposed that fundamental information about the emotions perceived by listeners in music is conveyed by nonlinear spatio-temporal patterns among the low-level psychoacoustic features. In order to test this hypothesis, we created a computational model sensitive to the temporal structure of psychoacoustic features that could predict the emotions perceived in music by a group of listeners. Analyses of the performance of the computational model show that a significant part of listeners' affective response can be predicted from a set of six low level features of music: loudness, pitch level, pitch variation (contour), tempo, texture and sharpness.

Emotional expression and induction: perception vs. feeling

As summarized above, a large corpus of literature has consistently reported that listeners often agree rather strongly about what type of emotion is expressed in a particular piece or even in particular moments or sections (a review of accumulated empirical evidence from psychological studies can also be found in Juslin & Sloboda, 2010), thus forcefully suggesting that the same music stimulus can evoke similar affective experiences in different listeners, occurring, to some extent, independently and consistently across individual, situational and cultural contexts.

However, the perception of emotion in music is fundamentally a sensory and cognitive process that does not necessarily mirror what a listener is actually feeling. For instance, one does not necessarily become sad while listening to a sad-sounding music piece. Indeed, emotions perceived (or expressed by the music) and felt by listeners' may differ, and there is evidence that the relationships between perceived and felt emotions are manifold (Gabrielsson, 2002).

Whereas the emotions expressed in a piece of music seems to rely mostly on the arrangement of musical features over time (Juslin & Laukka, 2004), the emotions experienced by listeners can be triggered by, and are a collective function of, many parameters, including the mood and psychological state of the listener, memories and other previous listening experiences, environmental and other situational aspects, individual preferences and attitudes, cultural conventions, among others (a systematic review of these factors and their possible influence in the emotional experience can be found in Scherer and Zentner (2001)).

Although the interaction between all these factors is still not well understood, there is nevertheless confirmation that the most common relationship between perceived and felt emotions is positive (i.e., when listeners perceive an emotion expressed musically, they often feel the same emotion; Gabrielsson, 2002), and estimated to occur in approximately 62% to 87% of all instances (Evans & Schubert, 2008). A possible explanation for the propensity to feel the emotions perceived in the music is emotional contagion, a process by means of which a listener may have the tendency to feel the emotions perceived in the music by means of either peripheral feedback or a more direct activation of the relevant emotional representations in the brain (Juslin & Västfjäll, 2008).

Emotion induction: evidence from physiological arousal

As seen, the expression and induction of emotions are separable phenomena. Moreover, induction does not necessarily follow the perception of an emotion[ii]. Due to this segregation it is important to be clear about which aspect of emotion is being studied in empirical experiments, since they may involve different psychological mechanisms. In terms of subjective feelings it is necessary to explain this distinction to participants, and ask them to report the emotion they feel rather the one they think that the music expresses. Notwithstanding, in addition to this, emotion induction can also be observed through changes in other emotion components.

One of the most common measures of experienced emotion used in music and emotion studies is physiological arousal (in this context we are only referring to measures of peripheral nervous system activity)[iii]. Physiological signals are controlled by subcortical emotion-processing parts of the brain through the release of chemicals into the blood, and the spread of neural activation in various brain areas and muscles. Through these mechanisms an emotional response has an effect on physiological activity, affecting for instance breathing, perspiration, blood pressure, heart rate, muscle contractions, among other autonomic measures (e.g., Stemmler, 2003).

If music can induce emotions, then listening to music should also have an effect on the autonomic nervous system (ANS), by eliciting responses associated with emotional processing (e.g. Juslin & Västfjäll, 2008; Trainor & Schmidt, 2003). Indeed, this seems to be the case, and several studies have provided evidence suggesting relations between emotional states induced by music and physiological activity, with heart rate, perspiration and respiratory functions among the most common measures (e.g., Gomez & Danuser, 2007; Grewe, Nagel, Kopiez, Altenmüller, 2007a, 2007b; Guhn, Hamm, & Zentner, 2007; Iwanaga, Kobayashi, & Kawasaki, 2005; Gupta & Gupta, 2005; Rickard, 2004; Khalfa, Peretz, Blondin, & Manon, 2002; Iwanaga & Moroki, 1999; Krumhansl, 1997; Harrer & Harrer, 1977). It should be noted that whether distinct autonomic signatures characterize different emotions, is a very controversial matter. Although there are similar results across studies, there are also several examples of contradictory results. Whereas some of them might be attributable to methodological differences, others cannot, and one of the main problems seems to be due to individual differences.

Peripheral feedback

An important aspect of physiological activation is its interaction with other emotion components. Currently, neurobiological models of emotion recognize not only the importance of higher neural systems on visceral activity (top-down influences) but also influences in the opposite direction (bottom-up) (see Berntson, Shafi, Knox, & Sarter, 2003 for an overview). While the top-down influences allow cognitive and emotional states to match the appropriate somato-visceral substrate, the bottom-up ones serve to bias emotion and cognition towards a desired state (e.g., guiding behavioral choice, Bechara, Damasio, & Damasio, 2003). Hence, physiological activation can also affect the subjective feeling of an emotion, a process known as peripheral feedback.

These ideas were applied to music by Dibben (2004), whom investigated the role of physiological arousal in determining the intensity and hedonic value of the emotion experienced while listening to music. Participants were asked to report the emotion “felt”, as well as the emotion “thought to be expressed”. The experiment design included two groups with different levels of induced physiological arousal: participants who exercised (aroused condition), and participants who relaxed (control). Results show that the group of participants with induced physiological arousal reported more intense emotions "felt" than those without (the increased physiological arousal intensified the dominant valence of the emotion "felt"). No effect was found for the reports of emotion "thought to be expressed". These results provide strong evidence that physiological arousal influences the intensity of emotion experienced with music, by suggesting that people may use physiological cues as a source of information about the emotion "felt" while listening to music.

Overview of the Present Study

As others (e.g., Dissanayake, 2008; Panksepp & Bernatzky, 2002; Krumhansl, 1997; Clynes, 1977), we believe that music structure elicits a response in neurological mechanisms related to motivation / emotion / cognition in the receptive listener, giving rise to profound changes in the body and brain dynamics by the same neuro-biological and psychological mechanisms involved in processing external stimuli with emotional significance (e.g., Panksepp & Bernatzky, 2002; Juslin & Västfjäll, 2008). We base this view on a vast number of empirical studies implying the existence of causal relationships between the organization of musical elements and emotional responses (intra- and cross-culturally), as well as on neuroscientific studies suggesting that cognitive attributions (e.g., Peretz, 2001; Peretz, Gagnon, & Bouchard, 1998) and mediation (e.g., Blood & Zatorre, 2001; Blood, Zatorre, Bermudez, & Evans, 1999), are not necessary for music to elicit emotions in the listener. Furthermore, the perception of emotion in music appears to be only marginally affected by factors such as age, gender or musical training (e.g. Robazza, Macaluso, & D'Urso, 1994; Bigand, Vieillard, Madurell, Marozeau & Dacquet, 2005).

In particular, we argue that music evokes emotion by creating dynamic temporal patterns to which our evolved socio-emotional brain is particularly sensitive. In our previous work (Coutinho & Cangelosi, 2009; Coutinho, 2010), we proposed that low-level acoustic patterns convey fundamental information about the emotions perceived in music. In order to test that hypothesis, we devised a computational methodology sensitive to the temporal structure of sound that could predict a significant part of listeners' perceived emotions from a set of six low level features (basic variables of human audition that are perceived similarly across individuals and shared across acoustic modalities): loudness, pitch level, pitch variation (contour), tempo, texture and sharpness.

In this article we extend that work further in two respects. First, we focus on emotions felt by listeners while listening to music (rather than perceived). Second, we include physiological features as a measure of induced emotions. Our aim is twofold: 1) To establish the importance of music structure in determining the global emotional experience with music by assuming that musical elements are organized according to an orderly complex system conveying emotional meaning (which may or may not induce emotions in its listeners) in a fairly consistent way across individuals; 2) To explore the extent to which physiological activity may be linked with self-reports of subjective feelings of emotion.

In order to achieve our aims we describe in this manuscript an interdisciplinary methodology to analyze emotional responses to music, comprising an empirical study and a computational modeling experiment. The goal of the empirical experiment is to collect continuous ratings of felt emotions while listening to music, and to simultaneously monitor subjects’ physiological activity by means of heart rate and skin conductance (together with respiration measures these variables are some of the most common autonomic indexes used in music studies). Continuous data is essential to relate the emotional dynamics (subjective feelings and physiological activation) to the temporal structure of acoustic cues. The goal of the computational study is to create an analytical tool to retrieve the relevant information from the acoustic stimuli and physiological activity, and to predict participants' subjective feelings of emotion (the data collected in the empirical experiment). We want to determine the extent to which the dynamics of emotional responses to music, and more specifically the subjective feeling component, are the result of the perception of psychoacoustic patterns with emotional meaning, as well as the self-perception of physiological activation (peripheral feedback).

The remaining of this article is organized in three main sections: In the next section (Experimental Study), the empirical study is described, and a preliminary analysis of the results is performed with the aim of establishing the relevant acoustic and physiological features to be used in the computational study. Then, in the following section we describe the modeling experiment in the "Computational Study" section, which details the procedure of creating a computational model for mimicking and predicting the subjective feelings of emotion induced in the listeners. Our hypothesis is that the spatiotemporal structure of particular psychoacoustic features (i.e., relationships among different perceptual dimensions as well as their temporal organization) encodes information that allows humans to perceive and feel emotion while listening to music, and that physiological activation is involved in this process. In the last section ("Discussion and Conclusions"), we evaluate our findings and discuss the implications of this work to emotion research.

Experimental Study

The aim of the experimental study is to collect continuous ratings of emotions felt while listening to music. We adopt a continuous response methodology to obtain listeners' responses in order to detect fine-grained temporal variations. This methodology involves collecting real-time responses to music stimuli, by allowing participants to report changes in their emotional state at any moment in the music, instead of doing so only at the end of the piece. The result is a time series depicting the dynamics of participants' ratings of emotion at every moment in the music[iv] - this is a very relevant aspect of the study, because our main hypothesis is that the temporal structure of acoustic features is a fundamental medium for communicating emotional information. Continuous measurements frameworks have been previously used in other studies investigating emotional responses to music (e.g., Grewe et al., 2007a, 2007b; Korhonen, 2004; Schubert, 2004).

Emotion is quantified here using a dimensional model. According to dimensional theorists, the subjective experience of emotion can be depicted by the combination of two or more underlying psychological "dimensions". Here we follow Russell's (1980) model, which construes emotions as linear combinations of two independent neurophysiological dimensions - arousal and valence. Arousal stands for level of activation or neurophysiological alertness in response to sensory stimuli. Valence corresponds to the continuum from negative to positive affect, and characterizes the hedonic tone of subjectively experienced emotions. It has been shown that these two dimensions account for the majority of observed variance in the emotional labeling of several types of experimental stimuli, including linguistic (Russell, 1980), pictorial (Bradley & Lang, 1994); and musical (e.g., Thayer, 1986). In addition to this, it has been shown recently, that arousal and valence are subserved by distinct neural systems during the experience of induced emotions (Colibazzi, Posner, Wang, Gorman, Gerber, Yu, Zhu, Kangarlu, Duan, Russell, 2010). It is important to note that the influential model of discrete emotions and the dimensional model are not mutually exclusively, but rather different conceptual frameworks. The dimensional model was used in this work for two main reasons: it offers a platform for the representation of a wide range of emotions not necessarily depicted by a particular emotion word (which is particularly important for musical stimuli); it is a reliable and economical model, easily adaptable for experimental studies, and suitable for continuous measurements when implemented using a computer framework.

The physiological indexes measured were heart rate (HR) and skin conductance (SC), two common measures of ANS activity, and amongst the most common measures used in music and emotions studies (e.g., Gomez & Danuser, 2007; Grewe, et al., 2007a, 2007b, 2005; Guhn, et al., 2007; Iwanaga, et al., 2005; Gupta & Gupta, 2005; Rickard, 2004; Khalfa, et al., 2002; Iwanaga & Moroki, 1999; Krumhansl, 1997; Harrer & Harrer, 1977). Heart rate (HR) is a measure of the cardiovascular system activity (measured in beats-per-minute - *bpm*). The heart muscle is under the control of the ANS (through the spinal chord and the brain stem) in such a way that the activation of the sympathetic nerves lead to an increase in HR (faster muscle contraction), whereas the activation of the parasympathetic nerves causes a decrease (e.g., Jänig, 2003). By virtue of the galvanic skin response (GSR), ANS activity causes a change in the skin's conductivity. GSR is under the strict control of the sympathetic nerves, by means of which higher activity (e.g., stress or excitement) leads to the secretion of sweat glands causing a fall in skin resistance (skin resistance, measured in Ω, is the reciprocal of skin conductance, measured in *S)*, whereas reduced activity (e.g., relaxation or meditation) leads to a rise in skin resistance (e.g., Jänig, 2003). There is evidence that these two measures are affected by emotionally evocative stimuli (e.g., Lang, Greenwald, Bradley, & Hamm, 1993), including music (e.g., Krumhansl, 1997). Furthermore, some studies suggest that distinct HR and SC signatures are related with distinct emotional qualities (e.g., Ekman, Levenson, & Friesen, 1983).

Method

Participants

Forty-five volunteers participated in the experiment. Due to failures in the recording of the self-report framework and physiological measurements, six listeners were removed from the analysis. The final list of valid data includes 39 participants (mean age: 34, std: 8, range: 20-53 years, 19 females and 20 males, 33 right handed and 6 left handed). The participant set includes listeners with heterogeneous backgrounds and musical education/practice (15 participants with less than one year or none; 14 participants with five years or more). The population includes listeners from 15 different countries and with 12 different mother tongues (all speak English). All participants in this experiment, with the exception of one, reported to be at least "occasionally" exposed to western art music. Participants also reported a high level of enjoyment of this music style (the mean rating was 4.2 out of 5).

Music Materials

The stimulus materials consisted of nine pieces of music, chosen by two professional musicians (one composer and one performer, other than the authors), attempting to illustrate the a wide range of emotional responses distributed throughout the two-dimensional emotional space (2DES) formed by arousal (vertical axis) and valence (horizontal axis), in order to induce differentiated subjective feelings of emotion in the listeners. Additionally, the pieces were chosen so as to be from the same musical genre, western art music (a style familiar to participants), and to be diverse within the style chosen in terms of instrumentation and texture.

The music used is shown in Table 1. The expected emotion produced by each piece is indicated by the labels Q_1 to Q_4, which represent the four main areas resulting from a division of the 2DES arousal/valence diagram into quadrants: Quadrant 1 (Q_1) – positive arousal and positive valence, Quadrant 2 (Q_2) – positive arousal and negative valence, Quadrant 3 (Q_3) – negative arousal and negative valence, and Quadrant 4 (Q_4) – negative arousal and positive valence.

-- Insert Table 1 here --

Procedure

Each participant sat comfortably in a chair inside a quiet room. The goal of the experiment was explained through written instructions which described the quantification of emotion and the self-report framework to be used during the listening task. Leads were attached to the participant's chest and left hand (for right-handed participants; right hand otherwise) index and middle fingers respectively for measuring HR and SC. The physiological measures were collected using the WaveRider biofeedback system (MindPeak, USA), and the signals were obtained with a sample rate of 128 Hz. The data used in the analysis was exported using WaveWare (the software interface for the WaveRider), which automatically downsamples the time series data to 1 Hz (using an average filter over a set of 128 consecutive data points). Those are the values used in our analysis.

Participants reported their emotional state by using the EMuJoy software (Nagel, Kopiez, Grewe, & Altenmüller, 2007), which consists of a computer representation of a two-dimensional emotional space (2DES). The self-report data was later synchronized with physiological data.

In the initial part of the experiment, each participant was given the opportunity to practice with the self-report framework (EMuJoy). A set of 10 pictures taken from the International Affective Picture System manual (Lang, Bradley & Cuthbert, 2005) was selected, in order to represent emotions covering all four quadrants of the 2DES (two per quadrant), as well as the neutral affective state (centre of the axis). The pictures were shown in a nonrandomized order, in order to avoid starting or finishing the picture slideshow with a scene of violence. Each picture was shown for 30 seconds, with a ten seconds delay in-between presentations. The only aim of this exercise was to familiarize participants with the use of the self-report framework.

After the practice period, participants were asked about their understanding of the experiment, and whether they felt comfortable in reporting the intended affective states with the software provided. Participants were then reminded to rate the emotions “felt” and not the ones thought to be expressed by the music. When the participant was ready, the main experiment started and the first piece was played. The pieces were presented in a randomized order, with a break of 75 seconds between each piece (unless the participant needed more time). Each experimental session lasted for about 60 minutes, including debrief, preparation and training periods. Before any physiological data was recorded, participants had 15 to 20 minutes (debrief, preparation and training period) to acclimatize and settle into the location. A baseline recording of 30 seconds was obtained for each participant immediately before the experiment started.

Data Processing

Description of the psychoacoustic measures

Our hypothesis in this experiment is that low level music structural features have causal relationships with listeners’ reports of emotion. To extract such information from the pieces of music we analyzed the perceptual experience using the same psychoacoustic variables used in our previous work (Coutinho & Cangelosi, 2009)[v], which consist of a set of six features: loudness, tempo, pitch level (power spectrum centroid), melodic contour, timbre (sharpness) and texture (multiplicity). A summary of these features, a brief description, and the aliases for use in this article is shown in Table 2.

-- Insert Table 2 here --

Self-report variables. The arousal and valence reported by each participant was recorded from the mouse movements. These values were normalized to a continuous scale ranging from -1 to 1, with 0 as neutral. The central tendency of the individual values of arousal and valence was estimated by calculating the arithmetic mean across all participants, on a second by second basis, for each piece of music.

Physiological responses. The physiological variables had to be processed to rule out the effects of individual differences on physiological levels. The same method was applied to both variables and consisted of dividing the individual HR and SC readings by the average of the 30 seconds individual baseline readings (obtained in a non-stimulus condition before the experiment started). The output of this calculation consists of the relative deviations from participants' individual baselines (represented as 1.0), allowing comparison between subjects without further calculations. These are the values for HR and SC that we report in this article.

Results

Figure 1 shows the second-by-second values of the self-reported emotional arousal and valence averaged across all participants for each piece. Each pair of values is represented by their corresponding location in the 2DES (represented as small dots). The gray squares indicate the quadrants expected to contain participants' responses ratings of emotion for each piece.

-- Insert Figure 1 here --

Overall, the classes of affective states expected to be induced by the chosen pieces correspond to the subjective feelings of emotion reported by participants. Indeed, most of the pieces (with the exception of piece 8) elicited responses in the predicted quadrants (see the description on page 15). It is noteworthy that within each piece there is a wide variability of responses, with most of the pieces containing sections that cover very different locations on the 2DES (for instance pieces 7, 8 and 9, overlap different quadrants). It is also evident that the pieces used did not elicit responses in the whole range of the 2DES, particularly in areas of negative valence (quadrants 2 and 3). Moreover, there seems to be a strong tendency for pieces to be rated within quadrant 1.

Music segments analysis

The fact that five of the nine pieces in the experiment elicited responses in more than one quadrant of the 2DES indicates the variety of affective responses that can occur within a single piece. In order to analyze intra-piece variability in more detail, a professional composer and two professional musicians were asked to divide each piece into segments by focusing on criteria related to its acoustic structure, form, and perceived affective value. The goal was to locate the moments in the pieces where salient changes in instrumentation, harmony, or rhythm, associated or not with the division of the piece into musical sections, were also linked with noteworthy changes in the perceived emotions. The segmentation points were chosen based on the common selections provided by all three professionals. The total number of segments was 27 (see Table 3).

-- Insert Table 3 here --

Two separate tests were conducted on the mean data of the segments in order to observe how the reported emotion dimensions relate to the acoustic composition of each segment and also to the physiological arousal levels (peripheral feedback). The results of both are shown in Table 4.

-- Insert Table 4 here --

Regarding the first test, we found significant linear correlations between arousal and the following sound features: loudness ($r = .60$, $p < .001$), tempo ($r = .67$, $p < .001$), pitch level ($r = .52$, $p < .001$) and sharpness ($r = .63$, $p < .001$). All have positive relationships with the level of arousal in the segments, i.e., arousal is higher in the segments with higher loudness, faster tempi, higher pitch and sharper sounds. Valence correlated with tempo ($r = .54$, $p <. 001$) and pitch level ($r = .41$, $p <. 05$).

The test of the relationships between physiological features and emotion yielded only one significant correlation, implying that increased HR relates with reports of higher arousal ($r = .46$, $p < .05$). We also verified that the changes in heart rate have clear associations with sound features, and we found that the heart rate had a propensity to be higher during louder segments ($r = .51$, $p < .01$). This liaison is consistent with the fact that higher loudness relates with higher subjective arousal. No significant correlations were found relating SC with either of the emotion dimensions.

Music segments: classification analysis

We conducted another examination of the experimental data with the purpose of searching for the combinations of sound features and physiological variables that best categorize each segment into the 2DES quadrants. The intention was to detect the contribution of physiological features to the discrimination of the affective value of each segment in order to evaluate its contribution to the subjective feeling response. To do so, we used Linear

Discriminant Analysis (LDA) (Mclachlan, 1992), a classical method of classification using categorical target variables (features that relate to or describe objects). The categories (grouping variables) chosen for our analysis were the locations of the mean arousal and valence values of each segment in the 2DES, i.e., the emotional quadrants (Q_1 to Q_4). We then tested two conditions: 1) To discriminate the affective values of each segment using only the mean levels of all sound features; 2) To discriminate the affective values of each segment using both sound and physiological feature sets. By choosing these test cases we are assessing the discriminatory power of the mean levels of sound features alone (which we expect to be elevated due to the high correlation found with arousal and valence), and also the additional contribution of the physiological cues (expected to have at least some contribution due to the association between heart rate and arousal) to that differentiation. Our intention is to estimate the relevance of physiological cues to the determination of the core affect categories for each segment.

The average values for each segment resulted in sixteen segments belonging to Q_1, four to Q_3, and 7 to Q_4 (there were no segments in Q_2). The analysis yielded two canonical discriminant functions with the first functions of each test condition explaining, respectively, 87% ($\rho = .808$, $\Lambda = .270$) and 94% ($\rho = 0.912$, $\Lambda = .118$) of the variance (the standardized canonical discriminant function coefficients are shown in Table 5). The LDA with condition 1 resulted in a correct classification of the mean levels of the sound features in 85% of the cases (Q_1: 14/16, Q_3: 3/4, Q_4: 6/7), with a cross validation of rate of 70%. The second test produced a success rate of 89% (Q_1: 16/16, Q_3: 3/4, Q_4: 5/7), with a cross validation of 78%. The separability of the groups, measured as the mean distance between group centroids, also increased from 2.4 (“Condition 1”) to 3.5 (“Condition 2”). Overall, these results indicate that the inclusion of the physiological variables leads to an improvement in the emotional classification of the music segments. This effect is nevertheless small (8% increase in the cross-validation results), and sound features hold clearly the strongest discriminatory power.

-- Insert Table 5 here --

Discussion

The empirical study aimed at collection self-report of subjective feelings, as well as ongoing physiological activity, while listening to music expressing a wide range of emotions. Generally, the emotions expected to be induced in the listeners by the chosen pieces correspond to the subjective feelings of emotion reported by participants. Even so, we noticed a tendency for pieces to be rated within quadrant 1 (increased arousal and positive valence), and the infrequent report of feelings with negative valence (quadrants 2 and 3). A possible explanation of this propensity might be that the chosen pieces elicit responses with positive valence and arousal, and that they lack stimuli with negative valence. We believe that this is an unlikely possibility considering that piece 1 has been used in other experiments and received high ratings of sadness (see for instance Krumhansl, 1997), i.e., low arousal and negative valence. Moreover, the fact that the data appears to be compressed is due to the fact that it was averaged across participants. The observations of individual time series clearly shows that individually the pieces elicited more extreme values, including responses with negative valence[vi].

Another possible cause of this apparent compression of responses to areas of positive valence may be a positive effect of music on mood, i.e., the induction of a pleasant affective state, lasting longer than an emotion, and less sensitive to be changed by a particular stimulus; the positive valence of the responses may reflect the pleasantness of listening to the pieces, which would be consistent with the fact that participants reported emotions felt. As a matter of fact, some of the participants (mostly expert musicians) reported difficulties in using the left side of the valence scale justifying it by claiming to have experienced positive sensations with all pieces. Future experiments should look in more detail at the possible factors influencing the self-report of emotion. In particular, it would be insightful to use a measure for participants' enjoyment of each piece or even of particular sections. This would allow to test the hypothesis of weather pleasure has an influence on ratings of valence, and gain some understanding on the relationship between perception and induction of emotions while listening to music[vii].

Observed the fact that five out of the nine pieces used elicited responses in more than one quadrant of the 2DES, we divided the pieces into segments in order analyze in more detail how the reported emotion dimensions relate to the acoustic composition of each segment and also to the physiological arousal levels. Overall, the results are consistent with Gabrielsson and Lindström (2010) meta-analysis of the associations between music composed features and emotion found in most studies up to this date. The most persistent associations reported, that suggests a positive relationship between loudness, tempo, timbre, and arousal, are corroborated here. Additionally, we also found that the pitch level relates positively with both arousal and valence, suggesting more complex associations with emotion. For instance, pitch may affect arousal and valence together or perhaps only one (or none) at a time, depending on the musical context. That could explain why some studies report low pitch to be related with both pleasantness and boredom (see Gabrielsson & Lindström, 2010 for further details). Tempo is another feature which relates to the valence ratings as well as to those of arousal. Such a result is also congruent with previous studies (and also suggests complex relationships with emotion), which report an association of tempo with ratings of sadness/happiness (and other emotional states which vary in both arousal and valence). Lastly, all the correlations highlighted are also concordant with our previous work (Coutinho & Cangelosi, 2009).

In relation to the physiological indexes, we found that the heart rate level was positively correlated with arousal, i.e., section that induced feelings characterized by higher aroused levels were accompanied by increased heart rate levels. This result is consistent with Krumhansl's (1997), who has shown that increased heart rate levels related to fear and happy excerpts rather than to sad ones (i.e., segments with higher arousal), Witvliet and Vrana's (1995), and Iwanaga and Moroky (1999), whose results showed increased heart rate levels for excitative music (correlated with subjective arousal). We have found that physiological cues help to differentiate more accurately the emotions induced in the listeners, however, their contribution, is much smaller than acoustic features alone.

Computational Study

In this computational experiment we follow up and extend our previous model (Coutinho & Cangelosi, 2009) based on the use of nonlinear models, such as spatiotemporal artificial (connectionist) neural networks, capable of dealing with the spatiotemporal patterns of psychoacoustic features derived from music (thus capturing static and dynamic aspects). Apart from applying the model to a new data set, we report on a novel experiment which extends the feature space used for the prediction of subjective feelings of emotion from human participants to physiological cues - heart rate and skin conductance. Therefore, in this article, we test the reliability of our model and also the possible accommodation of physiological features and their impact on its performance. We are motivated by the idea that musical emotions may exhibit time-locking variations with psychological and physiological processes. For a thorough description of the model and a discussion of the use of computational models in music and emotions studies please refer to Coutinho & Cangelosi (2010).

Framework: Artificial Neural Networks

Artificial Neural Networks (ANNs) were at first developed as mathematical models of the information processing capabilities of biological brains (McCulloch & Pitts, 1943; Rosenblatt, 1963; Rumelhart, Hinton & William, 1986). Despite the fact that they bear little resemblance to real neural networks, ANNs have acquired great popularity, especially as pattern classifiers.

This type of model paradigm is very flexible in terms of application because it offers a highly personalized definition of the model characteristics. The typical structure of an ANN consists of a set of basic informational processing units (representing biological neurons), interconnected through weighted connections (representing the weight of the synapses between neurons). The network receives information through a set of inputs (which can be one or more of the networks' processing units), activity that is then spread throughout the network according to the structure defined by the weighted connections. While in biological networks neuron activations consist of a series of pulses of very short duration, ANN were created to model the average firing rate of these spikes.

ANN topologies define the pattern of connections between the processing units, i.e. the arrangement of the different processing units (also called artificial neurons) and their interconnectivity that defines the flow of information within the model. Many topologies have been proposed over the years, aiming at tackling different problems, but there are two meta-classes that deserve to be distinguished: those purely acyclic and those comprising cyclical connections. The former are also called Feed-forward Neural Networks (FNNs), while the later are referred to as Recurrent Neural Networks (RNNs). For the purpose of this article we focus on the latter.

Recurrent neural networks involve some form of recurrence (feedback connections). Although in some cases the topological differences between FNNs and RNNs may be trivial, the implications for information processing are significantly different: while the FNN topologies only map inputs to outputs, RNN's can (ideally) map from the entire history of past inputs to the output. In point of fact, it has been shown that a RNN can approximate any measurable sequence-to-sequence mapping to arbitrary accuracy (Hammer, 2000). This is a striking property of RNNs: a kind of implicit memory of the past inputs is allowed to persist in future computational cycles, influencing the network output. As a consequence, RNNs have been extensively used in tasks where the network is presented with a time series of inputs, and are required to produce an output based on this series.

Due to their adaptability to deal with patterns distributed across space (relationships among simultaneous features) and time (memory of the past states of the features), RNNs were used in our previous work on music and emotion (Coutinho & Cangelosi 2009). These RNN models are also known as spatio-temporal connectionist models. Specifically in this article we will use a type of RNN called Elman Neural Network (ENN) (Elman, 1990). This model consists of the traditional feed-forward multilayered perceptron (MLP) (Rumelhart et al., 1986) with added recurrent connections on the hidden layer that endow the network with a dynamic memory. While the basic feedforward network can be thought of as a function that maps from input to output vectors, parameterized by the connection weights, and capable of instantiating many different functions, the ENN can map from the history of previous inputs to predict future states in the output later. The key point is that the recurrent connections allow the sequence of internal states of an ENN to hold not only information about the prior event but also relevant aspects of the representation that was constructed in predicting the prior event from its predecessor. If the process being learned requires that the current output depends somehow on prior inputs, then the network will need to "learn" to develop internal representations which are sensitive to the temporal structure of the inputs.

Procedure

The model and the optimization of its parameters are overall similar to that in Coutinho & Cangelosi (2009). Here we used the same modeling paradigm, which consists of the basic Elman network, and we maintain its architecture: five hidden (and memory) units, and two outputs (one for arousal and another for valence). The number of inputs varies according to the two simulation experiments presented below: a first one that includes six sound features (and thus six inputs), and a second one which aims at testing the effect of the additional physiological inputs (with a total of 8 inputs). The model is illustrated in Figure 2.

-- Insert Figure 2 here --

In independent simulations, two different input sets were tested: (i) the six sound features alone and (ii) the six sound features plus two units to represent both physiological features. Each simulation consisted of a set of 15 trials in which the models are trained using different initial conditions (randomized weights distributed between -0.05 and 0.05, except for the connections from the hidden to the memory layer which are set constant to 1.0)[viii]. Each trial consists of 80000 iterations of the learning algorithm, implemented using a standard back-propagation technique (Rumelhart et al., 1986). During training the same learning rate (0.075) and momentum (0.0) were used for each of the three connection matrices.

The “training set” (collection of stimuli used to train the model) includes five of the pieces used in the experiment (pieces 1, 4, 5, 6 and 8; see Table 1). The “test set”' (novel stimuli, unknown to the system during training, that test its generalization capabilities) includes the remaining four pieces (pieces 2, 3, 7 and 9). The pieces were distributed between both sets in order to cover the widest range of values of the emotional space. The rationale for this decision is that, for the model to be able to predict the emotional responses to novel pieces in an ideal scenario, it is necessary to have been exposed to the widest range of values attainable. Sets were defined so as to contain stimuli covering comparable areas of the 2DES, and to have extreme values in each variable (refer to Figure 1).

The “teaching input”' (or target values) are the average A/V pairs obtained experimentally for the training pieces. The task at each training iteration (t) is to predict the next (t+1) values of arousal and valence, from the inputs to the model. The range of values for each variable (sound features, self report and physiological variables) was normalized to a range between 0 and 1 in order to be scaled to the model.

Simulation Results

The root mean square error (*rmse*) was used to quantify the deviation of the model outputs from the values observed experimentally. For each trial the training stop-point was estimated *a posteriori* by calculating the number of training iterations so as to minimize the model output error (i.e., the *rmse*) for both training and test sets, thus avoiding the over fitting of the training set. The motivation for this approach is that if the model is able to respond with low error to novel stimuli, then the training algorithm was able to extract from the training set more general rules that relate music features to emotional ratings.

Table 6 shows the mean *rmse* across all 15 trials for each of the two simulation conditions. In addition, we also indicate the correlation between model outputs and experimental data, using the Pearson product-moment correlation coefficient (*r*).

-- Insert Table 6 here --

Both statistics are very similar for all simulations, indicating that the additional physiological inputs have a small impact in the model performance. Nevertheless, the best performance was achieved using the model with the extra physiological inputs, suggesting that HR and SC contain relevant information related to the self-report of emotion. This effect seems to be more evident for the arousal predictions, than for the valence ones. Overall, the model in simulation 2, which uses sound and physiological features as inputs, explains 78% of the total variance in arousal and 51% of the total variance in valence. For the remainder of this analysis we focus on this model.

Figures 3 and 4 portray the model predictions together with the experimental data for three sample pieces from each set, training and test, respectively.

-- Insert Figure 3 here --

-- Insert Figure 4 here --

Observing the figures it is possible to see that the model was able to capture the overall level and general fluctuations of the experimental data. The good performance for the pieces used to test the model predictions to the new set of unknown stimuli (“test set”) is particularly remarkable. A good example of this is piece 2 (see top of Figure 4): the model predicts the emotional responses to this unknown piece (equivalent to a piece heard for the first time by a subject) by explaining 94% ($r = .97$, $p< .001$) of the arousal and 86% ($r = .93$, $p< .001$) of the valence variance in the experimental data.

So as to observe the similarity between the segments mean levels predicted by the model and the ones obtained experimentally we compared both data sets in terms of the strength of the correlations between the mean levels of arousal and valence. Figure 5 shows the mean level of arousal (left) and valence (right) for each segment (defined earlier; see Table 3), for both experimental data and model predictions. The mean values of arousal and valence predicted by the model correlate significantly with the experimental data ($r_{A,A'} = .91$, $p < .001$; $r_{V,V'} = .82$, $p < .001$), meaning that the affective character of the segments was correctly predicted most of the time.

-- Insert Figure 5 here --

In order to verify if the model predictions and experimental data share similar relationships to the sound and physiological features, the mean levels of the psychoacoustic and physiological features for each segment were also compared. The results of the correlation analysis are shown in Table 7.

-- Insert Table 7 here --

As can be seen in this table, the participants' responses and model predictions exhibit similar relationships with sound features and physiological variables. Comparing the correlations between sound features and self-report dimensions in Tables 4 and 6, it can be seen that both the experimental data values (A/V) and the model outputs (A'/V') have similar structural relationships. The correlation analysis yields positive relationships between the mean level of reported arousal in the segments and loudness, tempo, mean pitch and sharpness. The other emotion dimension, valence, correlates significantly with tempo and pitch level. Regarding the correlations between physiological variables and emotional dimensions we also found similar relationships between the model and experimental data: only HR had a significant correlation with arousal ($r_{HR,A} = .46$ and $r_{HR,A'} = .38$; $p < .05$).

Discussion and Conclusions

In this study we presented an interdisciplinary methodology comprising an empirical experiment and a computational study applied to the investigation of the relationships between music structural factors and the subjective feelings of emotion experienced by listeners while listening to music. In addition to this, we tried to establish whether or not ongoing physiological activity is concomitant with the subjective feelings reported. Our goals were to determine the importance of low-level sound features in determining the global emotional experience with music, and to explore the extent to which physiological activity may affect reports of emotions felt through peripheral feedback.

We started by suggesting that music structure evokes emotion by creating dynamic temporal acoustic patterns conveying emotional meaning to its listeners. In that context, we focused on the temporal dynamics of music sequences and emotional responses, and we adopted a continuous response methodology to obtain listeners' responses as a means to detect fine-grained temporal variations of emotion ratings. Our intention was to collect empirical data at an appropriate temporal scale, so as to detect variations congruent with music structural changes.

In the analysis of segments ("Music segments analysis"), we have shown that the majority of the pieces used elicited responses in more than one quadrant of the emotion space. This intra-piece variability is particularly important since the averaging methods commonly used to compare music structural features with emotional responses can mask variations in emotional character and promotes the analysis of acoustic factors in terms of extremes (e.g., high/low, slow/fast). This method often leads to the misleading generalization of the relationships between music structure and emotion features to intermediate values, by assuming the existence of linear associations and no interaction between music variables.

In order to overcome these limitations we used a connectionist model to analyze the temporal (the dynamics of musical sequences) and spatial (the parallel contribution of various psychoacoustic factors and their interactions) interactions between sound features and the dynamics of emotional ratings. Consistent with our previous modeling work (Coutinho & Cangelosi, 2009; Coutinho, 2010b), we have shown that a group of six low-level features – loudness, tempo, pitch level, pitch contour, texture and sharpness – contain fundamental information that allows for an accurate prediction of emotional responses to music. Indeed, we were able to train a computational model to reproduce participants' responses to approximately half of the music pieces used in the experiment. Then, using the same model, we were able to predict the emotional responses for the remaining pieces.

By including physiological cues as inputs to the model, we tested the influence of peripheral feedback on self-report of emotions while listening to music. We selected heart rate and skin conductance as measures of autonomic activity, and tried to establish how they affect reports of felt emotions. Although the improvement in predictive power was rather small compared with the supremacy of sound features, the model performed better when adding heart rate and skin conductance as extra inputs. The model used information from ongoing (current and past states) physiological activity to predict subjective feelings, and consequently the results support previous work on the role of peripheral feedback in music listening (Dibben, 2004), and indicate that physiological cues may be an important path to explore in future studies.

We indentify a set of possible causes to justify the small improvement in the model performance when physiological cues were included (2.5% improvement in explained arousal and 1.3% in explained valence). Firstly, if autonomic activity is itself affected by the music structure, then much of its information is redundant for the model. Indeed, as Gomez and Danuser (2007) have shown, structural aspects of music (especially rhythmic) decisively affect physiological responses, and this could justify the fact that such information is computationally neglected. Secondly, physiological systems show characteristic temporal patterns (lags, latencies and other limitations in the way they change), which cannot be captured by our model. Thirdly, considering the fact that specialized training that promotes greater body awareness may affect the coherence between subjective reports and physiological activity (at least the coherence between subjective and cardiac aspects of emotion; Sze, Gyurak, Yuan, & Levenson, 2010), then an appropriate control of participants body awareness is necessary. Lastly, it is also important to mention the fact that the detection of physiological changes comparing the current physiological state with a baseline period is dependant on the quality of the baseline against which a change is detected and interpreted. In this sense, the baseline period used in our experiment was rather short, and could have influenced the normalization of the participants' responses.

That said, we think that it is important to consider physiological cues and the role of peripheral feedback in future studies, since understanding their dynamics may provide important information regarding the impact that music has on listeners at motivational, emotional, and cognitive levels, especially considering peripheral routes that may exert important interactions with high level cognitive processing. Indeed, it remains to establish their importance in coherently matching other emotional perception and experience, and how they may be fundamental in the link between perceived and induction of emotion. For instance, it is notable that the model performance in predicting felt emotions was similar to that in previous experiments which focused on emotions perceived (see Coutinho & Cangelosi, 2009; Coutinho, 2010b). This suggests that both aspects of emotional responses to music share a common underlying process which extracts relevant affective features from musical structure. A possible explanation may be that both phenomena are linked by means of an emotional contagion mechanism, in such a way that the perception of an emotion is then mimicked internally and leads to emotion induction (Juslin & Västfjäll, 2008). Such process may be mediated, at least partially, through peripheral feedback.

In terms of modeling technique, our proposal constitutes an advance in several respects. First, it incorporates all music variables together in a single model, which permits consideration of interactions among sound features (overcoming an important drawback of previous models, such as the ones by Schubert (2004) and Korhonen (2004)). Second, artificial neural networks, as non-linear models, enlarge the complexity of the relationships between music structure and emotional response which can be observed, since they can operate in higher dimensional spaces and are therefore not restricted to linear associations. Third, the excellent generalization performance (prediction of emotional responses for novel music stimuli) validates the model and supports the hypothesis that sound features are good predictors of the subjectively felt experience of emotion in music (at least for the affective dimensions considered). This is a particularly important aspect of the work presented here: the relationships learned from the sample set not merely a good fit of the experimental data, but rather they encode relevant information that allows to predict shared features of human emotional experiences with music. In addition to this, the set of features found, and the relationships between their structure and emotion dimensions are in agreement with other empirical studies (see Gabrielsson & Lindström, 2010).

It is also worth noting that the sound features in use by the model constitute a basic set of psychoacoustic dimensions not exclusive to music, but general to the auditory domain. This raises the possibility that the associations between acoustic features and emotions are shared with other modalities which convey emotional information through sound, such as speech prosody. Evidence supporting this hypothesis has been presented recently, suggesting the existence of similar acoustic profiles conveying emotion in speech prosody and music (Coutinho & Dibben, 2010).

There are several aspects in the work presented here that need to be addressed in future research. Perhaps the most obvious aspect is the fact that we focused on the average responses of a group of listeners, and discarded the idiosyncrasies of individual responses. The intention of doing so was to obtain the common features of emotion ratings from the population studied and to eliminate, as far as possible, the random and personal factors that influence individual responses. By modeling those responses we intended to represent the relationships between psychoacoustic features and emotional dimensions, which we believe to underlie emotional experiences with music and to be modulated (in some cases even suppressed) by individual and cultural aspects. Future work could extend this by considering individual traits and determining the relative importance of musical structure to the whole emotional experience with music. This could be achieved, for instance, using a series of case studies analyzing the responses of single listener. Another aspect to consider in future work is the way sound features convey emotional meaning. Up to now we have concentrated on creating a model relating sound features and emotion, but we did not explore the model dynamics, which can be an excellent source of information about the rules underlying input-output transformations. Finally, the physiological cues used in this study corresponded to a limited set of features. In future studies new variables could also be included, such as respiration-related features, and other properties of the variables used could be explored further (e.g. skin conductance response, heart rate variability). This may be particularly important to explore relationships between psychophysiology and emotional valence, since in virtually all music studies (as well as in this study) heart rate and skin conductance are associated with emotional arousal.

We believe that it is necessary to look in more detail to the complexity of music features and emotion in order to better understand musical emotions. This is certainly not to say that those relationships utterly define the complexity of our emotional responses to music, but we think that an important component of the musical experience, and especially its affective concomitants, emerge from the dynamic qualities of the music, the attributes of which may reflect adaptive responses to environmental cues. In this context, we believe that low-level psychoacoustic cues encode primary information about the emotions perceived and felt, and we have contributed a new methodology that brings the prediction of emotional responses to an appropriately high level. It remains nevertheless to establish weather this particular model is generalizable to other populations and music styles. Although in separate experiments we have shown that our modeling technique is applicable to other music genres (including the prediction of emotional responses to classical and film music using a model, trained on pop, dance and rock music; see Coutinho, 2010) and different populations, it remains to be established if the same underlying rules linking music structure and emotions reported are common to all of them. This is one of our main goals in the near future, together with the extension of our studies to the cross-cultural level. We believe these and other future developments will lead to fundamental advances in different areas of research since they may provide consistent descriptions of the emotional effects of particular music stimuli, which can aid specific areas, such as psychology or music therapy.

References

Balkwill, L.-L. & Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: psychophysical and cultural cues. *Music Perception*, 17, 43-64.

Balkwill, L.-L., Thompson, W. F., & Matsunaga, R. (2004). Recognition of emotion in Japanese, Western, and Hindustani music by Japanese listeners. *Japanese Psychological Research*, 46 (4), 337–349.

Bechara, A., Damasio, H., & Damasio, A. (2003). Role of the amygdala in decision-making. *Annals of the New York Academy of Sciences*, 985 (1), 356–369.

Berntson, G., Shafi, R., Knox, D., & Sarter, M. (2003). Blockade of epinephrine priming of the cerebral auditory evoked response by cortical cholinergic deafferentation. *Neuroscience*, 116 (1), 179-186.

Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., & Dacquet, A. (2005). Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. *Cognition & Emotion*, 19 (8), 1113-1139.

Blood, A., Zatorre, R.J., Bermudez, P., & Evans, A. (1999). Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. *Nature Neuroscience*, 2 (4), 382-387.

Blood, A. & Zatorre, R.J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Sciences of the United States of America*, 98 (20), 11818-11823.

Bradley, M., & Lang, P. (1994). Measuring emotion: the Self-Assessment Manikin and the Semantic Differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25 (1), 49–59.

Cabrera, D., Ferguson, S., & Schubert, E. (2007). PsySound3: Software for acoustical and psychoacoustical analysis of sound recordings. In W.L. Martens, G. Scavone, & R. Quesnel (Eds.). *Proceedings of the 13th International Conference on Auditory Display* (pp. 356-363). Montreal, Canada: McGill University.

Clynes, M. (1977). *Sentics: The Touch of Emotions*. New York: Anchor Press.

Colibazzi, T., Posner, J., Wang, Z., Gorman, D., Gerber, A., Yu, S., Zhu, H., Kangarlu, A., Duan, Y., Russell, J., & Peterson, B. (2010). Neural systems subserving valence and arousal during the experience of induced emotions. *Emotion*, 10 (3), 377-389.

Coutinho, E. (2010). Modeling psycho-physiological measurements of emotional responses to multiple music genres. In S.M. Demorest, S.J. Morrison, & P.S. Campbell (Eds.), *Proceedings of the 11th International Conference of Music Perception and Cognition (ICMPC11)* (p. 53). Seattle, WA, USA.

Coutinho, E., & Cangelosi, A. (2009). The use of spatio-temporal connectionist models in psychological studies of musical emotions. *Music Perception*, 27 (1), 1-15.

Coutinho, E. & Cangelosi, A. (2010). A Neural Network Model for the Prediction of Musical Emotions. In S. Nefti-Meziani & J.G. Grey (Eds.). *Advances in Cognitive Systems* (pp. 331-368). London: IET Publisher. ISBN: 978-1849190756.

Coutinho, E., & Dibben, N. (2010). Music, Speech and Emotion: psycho-physiological and computational investigations. In N. Dibben & R. Timmers, *Proceedings of the Conference on Interdisciplinary Musicology: Nature versus Culture (CIM10)* (pp. 47-48). Sheffield, UK.

Damasio, A. (2000). *The feeling of what happens: Body, emotion and the making of consciousness*. London: Vintage.

Dibben, N. (2004). The role of peripheral feedback in emotional experience with music. *Music Perception*, 22 (1), 79-115.

Dissanayake, E. (2008). If music is the food of love, what about survival and reproductive success? *Musicae Scientiae, S*pecial Issue, 169-195.

Dittmar, C., Dressler, K., & Rosenbauer, K. (2007). A toolbox for automatic transcription of polyphonic music. In *Proceedings of audio mostly: 2nd conference on interaction with sound* (pp. 58-65). Ilmenau, Germany.

Dixon, S. (2006). MIREX 2006 audio beat tracking evaluation: BeatRoot. http://www.music-ir.org/evaluation/MIREX/2006 abstracts/BT dixon.pdf.

Ekman P., Levenson, R.W., & Friesen, W. (1983). Autonomic nervous system activity distinguishes among emotions. *Science*, 221 (4616), 1208-1210.

Elman, J. (1990). Finding structure in time. *Cognitive Science*, 14, 179-211.

Evans, P. & Schubert, E. (2008). Relationships between expressed and felt emotions in music. *Musicae Scientiae*, 12(1), 75-99.

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., Friederici, A.D., & Koelsch, S. (2009). Universal recognition of three basic emotions in music. *Current biology*, 19 (7), 573-576.

Gabrielsson, A. & Lindström, E. (2010). The role of structure in the musical expression. In Juslin, P. & Sloboda, J. (Eds.), *Handbook of Music and Emotion* (pp. 367-400). Oxford, UK: Oxford University Press.

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

Glasberg, B.R. & Moore, B.C. (2002). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47 (1-2), 103-138.

Goldstein, A. (1980). Thrills in response to music and other stimuli. *Physiological Psychology*, 8(1), 126-129.

Gomez, P. & Danuser, B. (2007). Relationships between musical structure and psychophysiological measures of emotion. *Emotion*, 7 (2), 377-387.

Grewe, O., Nagel, F., Kopiez, R., & Altenmüller, E. (2005). How does music arouse "chills"? investigating strong emotions, combining psychological, physiological, and psychoacoustical methods. *Annals of the New York Academy of Sciences*, 1060 (1), 446-449.

Grewe, O., Nagel, F., Kopiez, R., & Altenmüller, E. (2007a). Listening to music as a re-creative process: physiological, psychological, and psychoacoustical correlates of chills and strong emotions. *Music Perception*, 24 (3), 297-314.

Grewe, O., Nagel, F., Kopiez, R., & Altenmüller, E. (2007b). Emotions over time: Synchronicity and development of subjective, physiological, and facial affective reactions to music. *Emotion*, 7 (4), 774-788.

Guhn, M., Hamm, A.O., & Zentner, M. (2007). Physiological and Musico-Acoustic Correlates of the Chill Response. *Music Perception*, 24 (5), 473-483.

Gupta, U. & Gupta, B. (2005). Psychophysiological responsivity to Indian instrumental music. *Psychology of Music*, 33 (4), 363-372.

Hammer, B. (2000). On the approximation capability of recurrent neural networks. *Neurocomputing*, 31(1-4), 107–123.

Harrer, G. & Harrer, H. (1977). Music, emotion and autonomic function. In M. Critchley & R.A. Henson (Eds.). *Music and the Brain: Studies in the neurology of Music* (pp. 202 -216). London, UK: William Heinemann Medical Books Limited.

Hevner, K. (1936). Experimental studies of the elements of expression in music. *The American Journal of Psychology*, 48 (2), 246-268.

Hevner, K. (1937). The affective value of pitch and tempo in music. *American Journal of Psychology*, 49 (4), 621–630.

Iwanaga, M., Kobayashi, A., & Kawasaki, C. (2005). Heart rate variability with repetitive exposure to music. *Biological Psychology*, 70 (1), 61-66.

Iwanaga, M. & Moroki, Y. (1999). Subjective and physiological responses to music stimuli controlled over activity and preference. *Journal of Music Therapy*, 36 (1), 26-38.

Jänig, W. (2003). The autonomic nervous system and its coordination by the brain. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.) *Handbook of affective sciences* (pp. 135-186). New York, NY, USA: Oxford University Press.

Juslin, P.N. & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code? *Psychological Bulletin*, 129 (5), 770–814.

Juslin, P. N., & Laukka, P. (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research*, 33, 217-238.

Juslin, P.N. & Sloboda, J. (Eds.). (2010). *Handbook of Music and Emotion*. Oxford, UK: Oxford University Press.

Juslin, P.N. & Västfjäll, D (2008). Emotional responses to music: the need to consider underlying mechanisms. *The Behavioral and Brain Sciences*, 31 (5), 559-575 (discussion 575-621).

Kallinen, K. & Ravaja, N. (2006). The Role of Personality in Emotional Responses to Music: Verbal, Electrocortical and Cardiovascular Measures. *Journal of New Music Research*, 33 (4), 399-409.

Khalfa, S., Isabelle, P., Blondin, J.-P., & Manon, R. (2002). Event-related skin conductance responses to musical emotions in humans. *Neuroscience letters*, 328, 145-149.

Kivy, P. (1990). *Music alone: Philosophical reflections on the purely musical experience*. Ithaca, NY, USA: Cornell University Press.

Koelsch, S., Fritz, T., Cramon, D., Müller, K., Friederici, A.D. (2006). Investigating emotion with music: an fMRI study. *Human Brain Mapping*, 27, 239-250.

Konečni, V. J. (2003). Review of P. N. Juslin and J. A. Sloboda (Eds.), "Music and Emotion: Theory and Research." *Music Perception*, 20 (3), 332-341.

Korhonen, M. (2004). *Modeling Continuous Emotional Appraisals of Music Using System Identification* (Unpublished master's thesis). University of Waterloo, Canada.

Krumhansl, C. L. (1997). An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology*, 51, 336-353.

Lang, P.J., Bradley, M.M., & Cuthbert, B.N. (2005). International affective picture system (IAPS): Technical manual and affective ratings. Gainesville, FL, USA: University of Florida, The Center for Research in Psychophysiology.

Lang, P.J., Greenwald, M.K., Bradley, M.M., & Hamm, A.O. (1993). Looking at pictures: affective, facial, visceral, and behavioral reactions. *Psychophysiology*, 30 (3), 261-273.

Levitin, D.J. (2006). *This is Your Brain on Music: The Science of a Human Obsession*. New York, NY, USA: Plume.

McCulloch, W. S. and Pitts, W. H. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115-133.

McLachlan, G. (1992). *Discriminant analysis and statistical pattern recognition*. New York, NY, USA: Wiley InterScience.

Meyer, L. (1956). *Emotion and Meaning in Music*. Chicago, IL, USA: Chicago University Press.

Nagel, F., Kopiez, R., Grewe, O., & Altenmüller, E. (2007). EMuJoy: Software for continuous measurement of perceived emotions in music. *Behavior Research Methods*, 39 (2), 283-290.

Oatley, K., & Jenkins, J. (1996). Understanding Emotions. Blackwell Publishers.

Panksepp, J., & Bernatzky, G. (2002). Emotional sounds and the brain: the neuro-affective foundations of musical appreciation. *Behavioral Processes*, 60 (2), 133–155.

Parncutt, R. (1989). *Harmony: a Psychoacoustical Approach*. Berlin, Germany: Springer-Verlag.

Peretz, I., Gagnon, L., & Bouchard, B. (1998). Music and emotion: perceptual determinants, immediacy, and isolation after brain damage. *Cognition*, 68 (2), 111-141.

Peretz, I. (2001). Brain specialization for music: new evidence from congenital amusia. *Annals of the New York Academy of Sciences*, 930 (1), 153.

Rickard, N. (2004). Intense emotional responses to music: a test of the physiological arousal hypothesis. *Psychology of Music*, 32 (4), 371-388.

Robazza, C., Macaluso, C., & D'Urso, V. (1994). Emotional reactions to music by gender, age, and expertise. *Perceptual and Motor Skills*, 79 (2), 939-944.

Rosenblatt, F. (1963). *Principles of Neurodynamics*. New York, NY, USA: Spartan.

Rumelhart, D., Hintont, G., & Williams, R. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533-536.

Russell, J. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39 (6), 1161-1178.

Scherer, K.R. & Zentner, M.R. (2001). Emotion effects of music: Production rules. In P.N. Juslin & J. Sloboda (Eds.), *Music and Emotion: Theory and Research* (pp. 361-392). Oxford, UK: Oxford University Press.

Schubert, E. (2004). Modeling perceived emotion with continuous musical features. *Music Perception*, 21, 561-585.

Stemmler, G. (2003). Introduction: Autonomic psychophysiology. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.) *Handbook of affective sciences* (pp. 131-134). New York, NY, USA: Oxford University Press.

Sze, J.A., Gyurak, A., Yuan, J.W., & Levenson, R.W. (2010). Coherence between emotional experience and physiology: Does body awareness training have an impact? *Emotion*, 10(6), 803-814.

Thayer, J. (1986). *Multiple Indicators of Affective Response to Music* (Unpublished doctoral dissertation). New York University.

Trainor, L.J. & Schmidt, L.A. (2003). Processing emotions induced by music. In I. Peretz & R. Zatorre (Eds.), *The Cognitive Neuroscience of Music* (pp. 310–324). Oxford, UK: Oxford University Press.

Tzanetakis, G., & Cook, P. (2000). Marsyas: a framework for audio analysis. *Organised Sound*, 4 (3), 169-175.

Witvliet, C. & Vrana, S. (1995). Psychophysiological responses as indices of affective dimensions. *Psychophysiology*, 32 (5), 436-443.

Zwicker, E., & Fastl, H. (1990). *Psychoacoustics: facts and models*. New York, NY, USA: Springer-Verlag.

Table 1

Pieces of music used in the experimental study. The pieces are numbered consecutively, so as to serve as aliases for reference in this article. For each piece we indicate the composer and title, its duration, a short description, and also the 2DES quadrant corresponding to the emotional response we expect it will elicit in listeners.

ID	Piece	Duration	Features
1	T. Albinoni – Adagio (G minor)	200s	Piece for strings and organ, solemn in mood, with occasional outbursts of melancholy and tragedy (Q_3).
2	E. Grieg - Peer Gynt Suite No. 1 (Op. 46): IV. “In the Hall of the Mountain King”	135s	Begins slowly evolving through low registers, with careful and quiet movements. Then, tempo gradually speeds up and the music becomes increasingly louder (Q_1).
3	J. S. Bach - Prelude and Fugue No. 15 (BWV 860): I. “Prelude” (G major)	43s	Piece evolving at a fast tempo, slowing down towards the end (Q_1).
4	L. V. Beethoven - Romance No. 2 (Op. 50, F major)	123s	Piece notated as an “adagio cantabile” and called romance for its light, sweet tone (Q_1, Q_4).
5	F. Chopin - Nocturne No. 2 (Op. 9, E flat major)	157s	Piece with a romantic character, with an expressive and dream-like melody (Q_4).
6	W. A. Mozart – Divertimento (K. 137): “Allegro di molto” (B flat major)	155s	Piece with some dance-like rhythms, and simple harmonies (Q_1).

7	C. Debussy - La Mer: II. “Jeux de vagues”	184s	Piece of variety and “color” that suggests a lively motion, conveying sensations of both bizarre and a dreamy atmospheres (Q_1, Q_2, Q_4).
8	F. Liszt - Liebesträume No.3 (S. 541, A flat)	183s	Piece composed to describe mature love (Q_3, Q_4).
9	J. S. Bach - Partita No. 2 (BWV 1004): V. “Chaconne” (D minor)	240s	Excerpt from a transcription for piano performed by Mikhail Pletnev, expressing a wide range of emotions (Q_1, Q_2, Q_3, Q_4)

Table 2

Psychoacoustic variables considered for this study and their description. All features, except tempo, which was estimated using BeatRoot (Dixon, 2006), and contour, estimated using Dittmar, Dressler & Rosenbauer's (2007) tool, were obtained using PsySound 3 (Cabrera, Ferguson & Schubert, 2007). The time series obtained were down-sampled from the original sample rates (which vary from feature to feature) to 1Hz in order to obtain second by second values. For convenience the input variables are referred to with the aliases indicated in the table throughout this paper.

Psychoacoustic Group	Feature and description	Alias
Loudness	Dynamic Loudness (Glasberg & Moore, 2002): subjective impression of the intensity of a sound (measured in sones).	L
Pitch Level	Power Spectrum Centroid: first moment of the power spectral density.	P
Pitch Contour	Melody contour: calculated using a melodic pitch extractor adequate to be used with polyphonic sounds.	C
Timbre	Sharpness (Zwicker and Fastl, 1990; usually considered a dimension of timbre): a measure of the weighted centroids of the specific loudness, which approximates the subjective experience of a sound on a scale from dull to sharp (measured in acum).	S
Tempo	Number of beats per minute (bpm)	T
Texture	Multiplicity (Parncutt, 1989): estimates of the number of tones simultaneously noticed in a sound.	Tx

Table 3

Piece segmentation details: each segment is identified by its piece number followed by a letter (only for pieces with more than one segment) indicating, in alphabetical order, the segment that it refers to (e.g. the alias for segment b of piece 1 is 1b). The location of each segment within the corresponding piece is indicated by the initial and final timestamps (the values indicated are in seconds).

Piece	Nr.	Segments				
	Segments	a	b	c	d	e
1	3	1-26	27-78	79-end	-	-
2	2	1-79	80-end	-	-	-
3	1	1-end	-	-	-	-
4	3	1-33	34-99	100-end	-	-
5	2	1-62	62-end	-	-	-
6	4	1-42	43-85	86-110	111-end	-
7	3	1-52	53-126	127-end	-	-
8	4	1-34	35-84	85-114	115-end	-
9	5	1-56	57-111	112-140	141-213	214-end

Table 4

*Correlation analysis of experimental data: psychoacoustic and physiological features are compared with the arousal and valence dimensions. The values compared are the mean levels for each segment of all pieces (*p < .01; ** p < .05).*

	A	V
Loudness	.60*	-
Tempo	.67*	.54*
Pitch level	.52*	.41**
Pitch contour	-	-
Texture	-	-
Sharpness	.63*	-
SC	-	-
HR	.46**	-

Table 5

Composition of the first discriminant function of each analysis: using only the mean levels of all sound features (condition 1); and using both sound and physiological features (condition 2). The values indicated correspond to the standardized coefficients (weights) of the canonical discriminant function.

	Condition 1	Condition 2
Loudness	-0.54	-1.00
Tempo	0.48	1.00
Pitch level	0.41	1.03
Pitch contour	0.05	0.55
Texture	-0.91	-1.03
Sharpness	0.93	0.77
SC		-1.53
HR		0.85

Table 6

Comparison between simulations 1 – using only sound features as inputs – and 2 – which uses the additional physiological inputs. The statistics shown quantify the deviation (rmse) and similarity (r) between the average outputs of the 15 trials run for each simulation and the responses of the human participants. Each emotion dimension is shown separately in order to evaluate the arousal and valence predictions individually.

Sim. ID	Inputs	*rmse*		r	
		A	V	A	V
1	L, T, P, C, Tx, S	.074	.063	.871	.705
2	L, T, P, C, Tx, S + HR, SC	.069	.063	.885	.714

Table 7

*Correlation analysis of the model predictions: psychoacoustic and physiological features are compared with the arousal and valence dimensions. The values compared are the mean levels for each segment of all pieces (*p < .01; ** p < .05).*

	A'	V'
Loudness	.53*	-
Tempo	.86*	.54*
Pitch level	.53*	.50*
Pitch contour	-	-
Texture	-	-
Sharpness	.62*	-
SC	-	-
HR	.38**	-

Figures captions

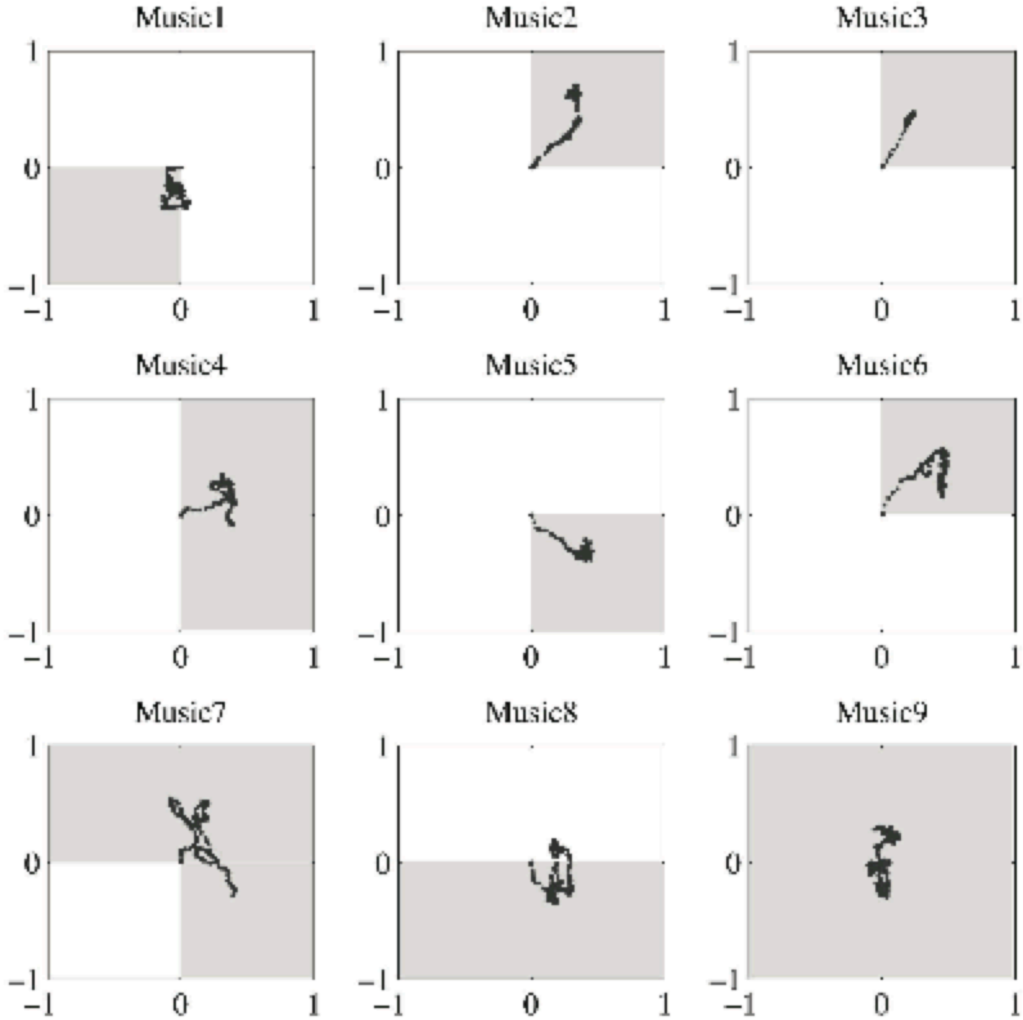
Figure 1. The figure shows the second by second values of Arousal and Valence, averaged across participants, for each piece used in the experiment. The grey rectangles indicate the areas of the 2DES, which correspond to the core affective states expected to be elicited in the listeners (see Table 1).

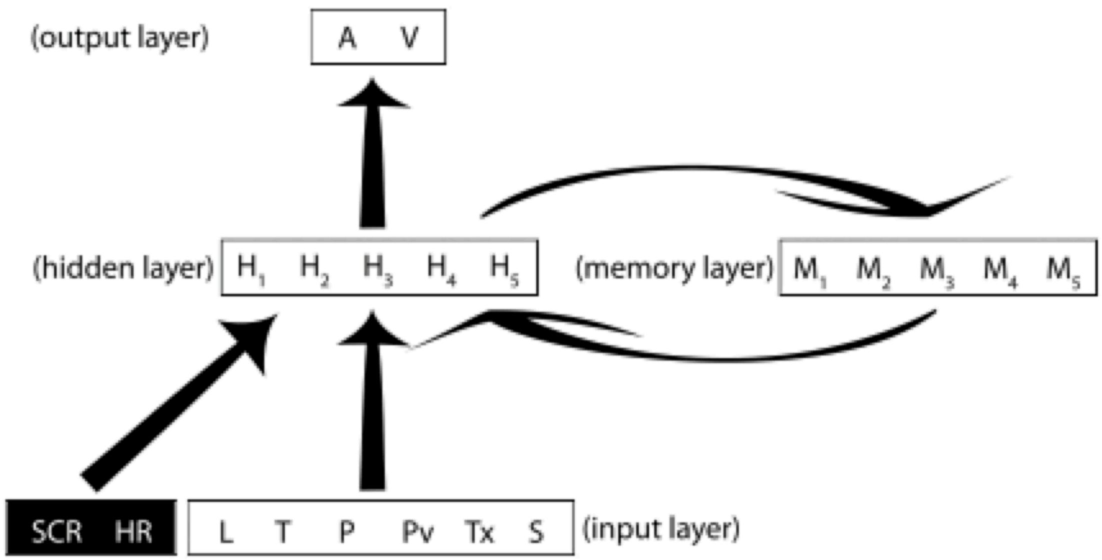
Figure 2. Model architecture of the neural network used in the simulation experiments. Input units: sound features (T, Tx, L, P, S and C) and physiological variables (SC and HR); Hidden units - H_1 to H_5; Memory (context) units - M_1 to M_5; Output units: arousal (A) and valence (V).

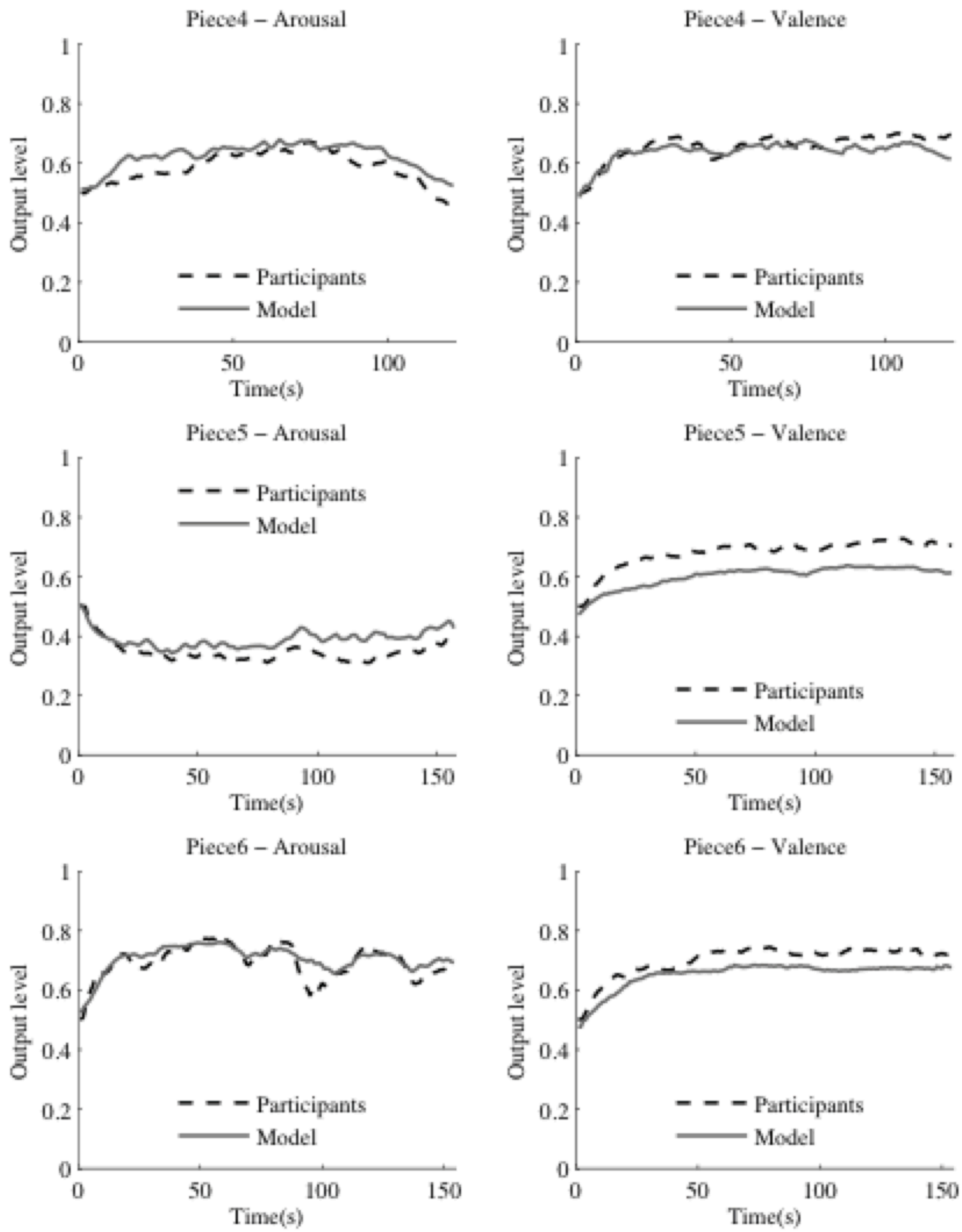
Figure 3. Comparison between the model arousal and valence outputs and experimental data for three samples pieces from the training data set (from top to bottom): Piece 4 (Beethoven - Romance No. 2), Piece 5 (Chopin - Nocturne No. 2) and Piece 6 (Debussy - La Mer, “Jeux de vagues”). The arousal and valence values shown correspond to the values used with the model, and so they are normalized between 0 and 1 (corresponding to [-1, 1] scale in the original data), with 0.5 as the neutral state value.

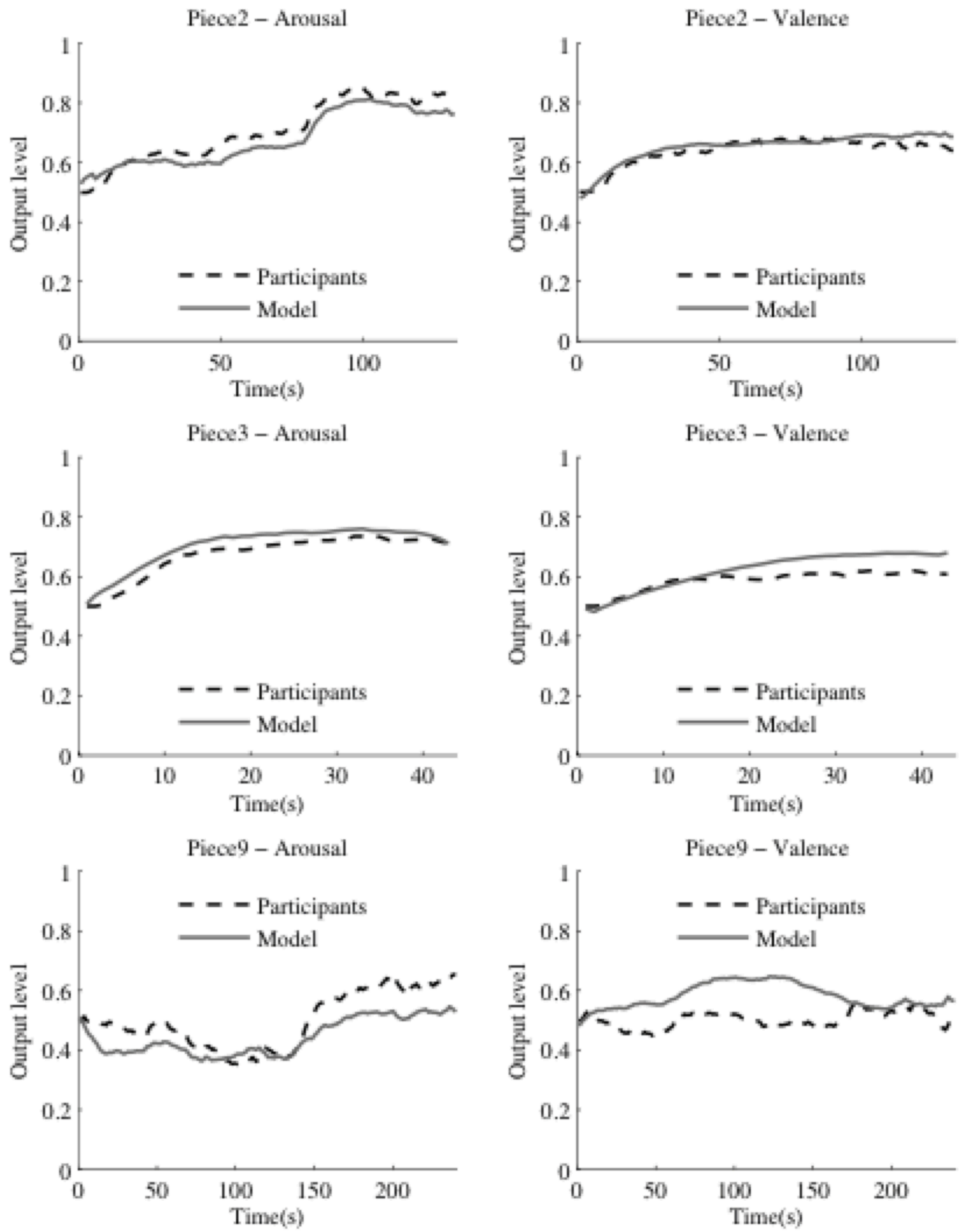
Figure 4. Comparison between the model arousal and valence outputs and experimental data for three samples pieces from the test data set: Piece 2 (Grieg - Peer Gynt Suite), Piece 3 (Bach - Prelude No. 15), and Piece 9 (Bach - Partita No. 2, “Chaconne”). The arousal and valence values shown correspond to the values used with the model, and so they are normalized between 0 and 1 (corresponding to [-1, 1] scale in the original data), with 0.5 as the neutral state value.

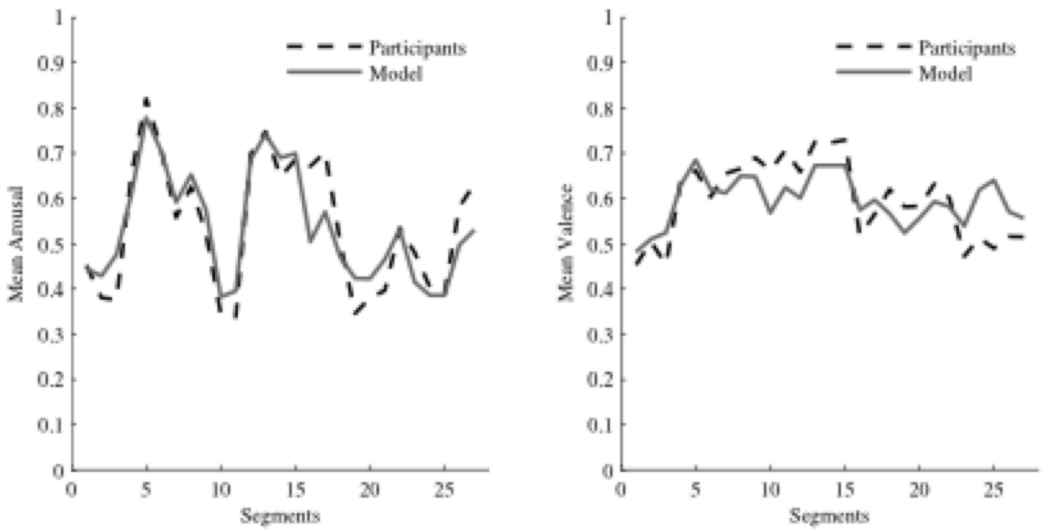
Figure 5. Comparison between experimental data and model predictions: average arousal (left) and valence (right) for each music segment as indicated in Table 3.


Music1
Music2
Music3
Music4
Music5
Music6
Music7
Music8
Music9
1
0
−1
−1
0
1


(output layer)
A V
(hidden layer)
H1 H2 H3 H4 H5
(memory layer)
M1 M2 M3 M4 M5
SCR HR
L T P Pv Tx S
(input layer)


Piece4 – Arousal
Piece4 – Valence
Piece5 – Arousal
Piece5 – Valence
Piece6 – Arousal
Piece6 – Valence
Output level
Time(s)
Participants
Model
0
0.2
0.4
0.6
0.8
1
50
100
150


Piece2 – Arousal
Piece2 – Valence
Piece3 – Arousal
Piece3 – Valence
Piece9 – Arousal
Piece9 – Valence
Output level
Time(s)
Participants
Model
0
0.2
0.4
0.6
0.8
1
10
20
30
40
50
100
150
200


Participants
Model
Mean Arousal
Segments
Participants
Model
Mean Valence
Segments
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
5
10
15
20
25

Author Notes

This research was supported by a doctoral fellowship from the Portuguese Foundation for Science and Technology (FCT) to the first author. We would also like to express our gratitude to Dr. Nicola Dibben for her pertinent insights, suggestions, and help while preparing this manuscript.

Notes

[i] Another subcomponent often considered as part of an emotion is cognitive appraisal (e.g., Scherer, 1999). Appraisal accommodates the process of evaluation of a stimulus through relatively low level and involuntary cognitive processes, but also through reasoning. The result of this evaluation determines the significance of the stimulus for the individual, and triggers appropriate responses in the form of emotion, based on a principle of attraction-aversion tendency.

[ii] This is true not only for music, but also for other domains of emotional expression. For instance, when perceiving sadness in a fellow-being, one does not necessarily become sad.

[iii] The study of physical and motoric responses to music has a long history; see for instance Dainow (1997) for a review of studies dating back to the 18th century.

[iv] This method does not imply that participants will constantly use the interface to express their feelings (in fact they may not even move the mouse if they feel that there was no change in their emotional state), but it allows them to do so anytime they intend to report a change in their emotional state.

[v] From the original set used in our previous work, the pitch contour algorithm was changed from the original approximated measure, which consisted of the euclidian norm of the difference between the magnitudes of the Short-Time Fourier Transform spectrum evaluated at two successive sound frames (Mean STFT Flux, Tzanetakis & Cook, 1999), to a new genuine measure of pitch contour (as described in Table 2).

[vi] It is important to note that by averaging across individual responses we focus on the common features of an “average” individual, i.e., the common trends in the responses of all individuals to the same stimuli.

[vii] Although this question was not part of our hypothesis, as part of our experimental procedure we asked participants how much did they enjoy listening to each piece. In order to gain some preliminary understanding on this question, we conducted a post hoc analysis to test weather higher ratings of enjoyment were related to higher valence ratings. Although a significant correlation was not found between the mean values valence and the level of enjoyment, valence tended to be higher for pieces that participants liked more ($r=0.620$, $p=0.075$). A more detailed analysis would be required to detect the type of relationships between pieces enjoyment and reported valence, in particular due to the variations of valence in different sections.

[viii] Each simulation is repeated 15 times in order to test the consistency of the results for each simulation.