2011-08-01

Musical Emotions: Predicting Second-by-Second Subjective Feelings of Emotion From Low-Level Psychoacoustic Features and Physiological Measurements

Coutinho, E

10.1037/a0024700
EMOTION

Running head: Musical emotions: psycho-physiological investigations.

Musical emotions: predicting second-by-second subjective feelings of emotion from psycho-physiological measurements

Eduardo Coutinho[1] and Angelo Cangelosi[2]

[1]University of Sheffield, Sheffield, United Kingdom

[2]University of Plymouth, Plymouth, United Kingdom

Words: 10693

Tables: 6

Figures: 5

Correspondence address:

Eduardo Coutinho

University of Sheffield, Department of Music

Jessop Building, 34 Leavygreave Road, Sheffield S3 7RD, United Kingdom

Phone: +44 1752 232580; Fax: +44 1752 232579

E-mail: ec@eadward.org

Date: February 12, 2010

Abstract

In this article we present a new methodology for replicating and predicting subjective feelings of emotion in response to music. We argue that music evokes emotion by creating dynamic temporal patterns to which our evolved socio-emotional brain is particularly sensitive, and we will show that the ways composers organize the acoustic building blocks of music, induce similar psycho-physiological responses in listeners. These claims will be supported by novel methodological investigations based on a combination of computational models and empirical psycho-physiological studies. We present evidence that the music psychoacoustic structure can account for a large proportion of the emotion reported by human listeners, by showing that a significant part of the listeners' affective response can be predicted from a set of six low level features of music: loudness, pitch level, pitch contour, tempo, texture and sharpness. We will also analyze how 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. The work presented here provides a new methodology to the field of music and emotion research based on combinations of computational and experimental work, which aid the analysis of emotional responses to music, while offering a platform for the abstract representation of those complex relationships. Future developments may conduct to fundamental advances in different areas of research since they may provide coherent descriptions of the emotional effects of specific music stimuli, which can aid specific areas, such as, psychology and music therapy.

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

Musical emotions: predicting second-by-second subjective feelings of emotion from psycho-physiological measurements

For thousands of years, the affective concomitants of the musical experience, and especially its conspicuous capacity to elicit powerful emotions, have challenged philosophical thought. Yet, many of its aspects remain a mystery and, perhaps unsurprisingly, unveiling the underlying mechanisms supporting musical experience continues to be one of the great challenges of scientific research. To date, a complex (and confusing) mesh of ideas and arguments still dominates contemporary literature.

Even though in the last few decades important advances in several disciplines have provided fundamental insights and experimental evidence that allowed for an enhanced understanding of different aspects related to the experience of emotion with music (see Juslin & Sloboda, 2010, for an up-to-date compilation of perspectives and approaches to studies on music and emotion), the way music interacts with the human emotional systems, and especially the nature of the emotions it induces, also referred to as "musical emotions" (MEs), has been at the very centre of a long-standing controversial discussion: can music induce emotions; are MEs just like other ("real") emotions? While some claim that music cannot induce any emotions at all since it does not appear to have any goal implications (a idiosyncrasy attributed to emotions) (e.g., Konečni, 2003), others suggest that music induces a particular set of affective states specific to music through distinct mechanisms than the one associated with other emotions (e.g. Scherer & Zentner, 2001). Moreover, other researchers support the hypothesis that MEs are just like other emotions (e.g. Juslin & Västfjäll, 2008). This paper follows this last position, although it additionally proposes that indeed music can convey a much wider range of affective qualities than the ones comprised by discrete lists of emotions (e.g, basic emotions). The paper also

sustains that MEs are routed in the same mechanisms that support these and other emotional reactions. As it is out of the scope of this paper to discuss in detail such ideas, we refer the reader to a comprehensive discussion on these and other topics related to MEs by Juslin and Västfjäll (2008). These authors compile evidence showing that music can evoke the various subcomponents related to emotional reactions (cognitive appraisal, subjective feelings, physiological arousal, expression, action tendency and regulation), thus forcefully suggesting that music evoke "real" emotions (or at least can elicit reactions in the same mechanisms associated with them).

For the remaining of this article we will focus on the subjective feeling and physiological arousal components of MEs, and their relation to music composed structure, i.e., the ways composers organize the acoustic building blocks of music in "emotionally-meaningful" ways. We will suggest that music can induce similar psycho-physiological responses in listeners by engaging with our brain systems in very particular ways. As others (e.g., Dissanayake, 2008; Clynes, 1977), we argue that music evokes emotion by creating dynamic temporal patterns to which our evolved socio-emotional brain is particularly sensitive, and we will show that a great part of the emotional responses of a group of listeners can be predicted from the perceived spatio-temporal characteristics of the acoustic signal organized through music. We will also analyze how peripheral feedback in music (Dibben, 2004) can account for the predicted emotional responses, i.e., the role of physiological arousal in determining the intensity and valence of MEs. All these claim will be supported by novel methodological investigations based on a combination of computational models and empirical psycho-physiological studies.

Emotional expression in music: the role of the composed structure

We have discussed elsewhere (Coutinho & Cangelosi, 2009; Coutinho, 2009) that modern research has emphasized the individual and culture-dependent aspects of the musical experience, even though a considerable corpus of psycho-musicological 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. Naturally this leads to a focused investigation of the factors in musical structure which contribute to the perceived emotional expression. Such observations already have a long history (at least back to our ancient Greek ancestors Socrates, Plato, and Aristotle), but they gained particular attention after 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). The isolation of the perceptible factors in music which may be responsible for the many observed effects has been a core interest amongst music psychologists until our days.

The basic perceptual attributes involved in music perception 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 (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.

There is now strong evidence that certain music dimensions and qualities communicate similar affective experiences to listeners. Gabrielsson and Lindström's (2010) review[i] of more than one hundred studies indicates that the most unambiguous effects regard the link between tempo, loudness and timbre (particularly the first two) with arousal, such that an increase in these features leads to a higher activation, whilst the opposite leads to a lower activation. These authors also note that the results regarding pitch tend to be less clear than those on loudness or tempo (e.g., both high and low pitch levels may be associated with high and low activation). Additionally, the associations found seem to be more consistent for the arousal dimension than for the valence one. It is important to highlight that the clearest relationships relate to fundamental features of sound perception, rather than more complex musical variables (e.g., key or mode), which seem to be more context dependent. For instance, the so often reported assumed link between major/minor mode with happy/sad judgments, which does not emerge until the age of six to eight years, may be modulated by tempo and pitch level.

Insights into the universal aspects of music expression of emotion

Another important finding to consider is that the perception of emotion in music appears to be marginally affected by factors such as age, gender or musical training (e.g. Robazza, Macaluso, & D'Urso, 1994; Bigand, Vieillard, Madurell, Marozeau & Dacquet, 2005; see Gabrielsson & Lindström, 2010 for further references). The fact that musical training is not necessary for listeners to experience emotion in music suggests that a general mechanism that processes emotional stimuli is involved. This idea is supported by the finding that the ability to

recognize discrete emotions is correlated with measures of "Emotional Intelligence" (Resnicow, Salovey & Repp, 2004). Peretz, Gagnon, and Bouchard (1998). Even more compelling evidence shows that the immediate emotional judgments determined by musical structure (mode and tempo on the reported study) can resist brain damage affecting the perceptual analysis of the music input, i.e., separate pathways may be involving in processing emotional and cognitive information. Such finding was later supported by the discovery that music may recruit subcortical emotional circuits (Blood & Zatorre, 2001; Blood, Zatorre, Bermudez, & Evans, 1999), which are also associated with the generation of human affective experiences (e.g., Damasio, 2000; Panksepp, 1998), and can operate outside an individual's awareness.

In this line of reasoning, it is important to provide evidence from cross-cultural research, since the absence of cultural (learned) associations between emotion and music supports the argument that listeners rely on psychoacoustic features. For instance, Balkwill and Thompson (1999) reported that Western listeners, who had no familiarity with North Indian music, and who listened to Hindustani music (North Indian classical music style), were able to identify emotions of joy, sadness, and peace as also identified by listeners experienced in Hindustani music. Most importantly, the authors have shown that the associations between psychoacoustic properties of music and emotion are compatible with results from intracultural studies. Later, Balkwill, Thompson and Matsunaga (2004) have also shown that Japanese listeners are sensitive to anger, joy and peace in Western, Japanese and Hindustani music, and that, for the music of all three cultures, judgments of emotions were associated with loudness, tempo, timbre as well as melodic complexity. More recently, Fritz, Jentschk, Gosselin, Sammler, Peretz, Turner, Friederici & Koelsch (2009) conducted a comparative study with participants from a native African population (Mafa) and Western participants (both groups being naive to the music of the other

respective culture). Results show that the Mafas recognize happy, sad, and scared/fearful from Western music excerpts. The authors suggest that the expression of three basic emotions (happy, sad, scared/fearful) can be recognized universally in Western music and that consonance and permanent sensory dissonance are also universal determinants of the perceived pleasantness of music.

Overview of the Present Study

We believe that the music psychoacoustic structure can account for a large proportion of the emotion reported by human listeners. Unquestionably, one cannot ignore the fact that most listeners appreciate music through a diverse range of cortico-cognitive processes, which rely upon the creation of mental and psychological schemas derived from the exposition to the music in a given culture (e.g. Meyer, 1956). However, cumulative evidence suggests that the same music stimulus induces similar affective experiences in listeners, somehow independently of acculturation, context, individual features or personal preferences.

In our previous work, we have shown evidence supporting such claim in two computer modeling experiments: one considers classical music (Coutinho & Cangelosi, 2009) and another considering multiple genres (baroque, classical vocal, dance music, death metal, film music, pop and rock; see Coutinho, *in press*). We proposed that the fundamental information about listeners' affective responses to music could be conveyed from nonlinear spatio-temporal patterns among the psychoacoustic features of sound organized through music. In order to test this hypothesis, we created a computational model (spatio-temporal connectionist network) sensitive to the temporal structure of psychoacoustic features that could predict the subjective feelings of emotion of human subjects while listening to music. Analyses of the computational model performance have shown that a significant part of the 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.

In this article we extend that work further. Firstly, we will describe an empirical experiment ("Experimental Study"), in which participants are asked to report their subjective feelings while listening to full pieces music. Simultaneously, their heart rate and skin conductance response are also recorded. The data collected is then used in a second study ("Computational Study") in which we use recurrent neural network models to mimic and predict participants' subjective feelings of emotion from the spatiotemporal properties of a set of psychoacoustic music features of sound, and assess the relevance of physiological activation for the subjective feeling responses. The dynamics of emotional responses to music are then investigated as computational representations of perceptual processes (psychoacoustic features) and self-perception of physiological activation (peripheral feedback).

Experimental Study

In this study the continuous response methodology was used to obtain listeners' affective experience with music on the basis of experimenter selected music. This methodology involves participants giving real-time responses to the music stimuli using a computer-interfaced device, and it has been frequently used in music psychology studies (e.g., Grewe, Nagel, Kopiez, & Altenmüller, 2007a, 2007b; Schubert, 2004; Krumhansl, 1997). The music pieces were chosen to induce differentiated subjective feelings of emotion in the listeners. Emotion was measured continuously in time by tracking two dimensions simultaneously - arousal and valence – following Russell's (1980) cognitive dimensions of affect. A major difference from our previous experiments regards the type of report asked to participants. The perception of emotional expression in music must be distinguished from one's emotional experience with it, an aspect

which is not always clearly defined, and that should be addressed since it may conduct to different results (see Kallinen & Ravaja, 2006; Evans & Schubert, 2008; Gabrielsson, 2002). As we are interested is relating physiological arousal with the motional response to music, participants were asked to report the emotion "felt" while listening to the music (rather than the emotion "thought" to be expressed by the music). We monitored participants' autonomic responses through their heart rate (measured in *bpm*) and skin conductance responses (measured in μS).

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 (classical) 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 music pieces, chosen by two professional musicians (one composer and one performer, other than the authors), attempting to illustrate the widest range of emotional responses possible distributed throughout a two-dimensional

emotional space (2DES) formed by arousal and valence (equivalent to Russell's (1980) model of affective space, which combines arousal and valence qualities to describe affective experiences). The pieces were chosen so as to be from the same musical genre, classical music (a style familiar to participants), and to be diverse within the style chosen in terms of instrumentation and texture.

The music pieces used are shown in Table 1, and the emotions they are expected to elicit in the listeners are described in the following paragraphs. The expected emotion produced by each piece is indicated by the labels which represent four main area of resultant from diving 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 --

Albinoni's "Adagio" (Piece 1) is a piece for strings and organ[ii], in the key of G minor, which is solemn in mood, with occasional outbursts of melancholy (and tragedy). This piece is expected to belong to quadrant three (low arousal, negative valence).

Grieg's "Peer Gynt Suite No. 1" fourth movement (Piece 2) begins slowly and quietly evolving through low registers, with careful and quiet movements. Then, the theme is slightly modified, the tempo gradually speeds up and the music becomes increasingly louder. This piece is expected to elicit responses within the first quadrant of the 2DES (positive arousal and positive valence).

Piece 3 is a prelude in G-major from the 1st book of Bach's "Well-Tempered Clavier". This short piece evolves at a fast tempo, slowing down towards the end. This piece should also elicit responses in the first quadrant of the 2DES, with higher arousal during the second part.

Beethoven's “Romance No. 2” (Piece 4) is a music piece notated as an “adagio cantabile” and called romance for its light, sweet tone. This piece is expected to induce low to high levels of arousal and positive valence (quadrants one and four).

Chopin's “Nocturne no 2” (Piece 5) is a piece with a romantic character, and with an expressive and dream-like melody. This piece is expected to elicit low arousal and positive valence (quadrant four).

The second movement of Mozart's “Divertimento”' (Piece 6) has some dance-like rhythms and simple harmonies. The “happy” character of this piece is expected to elicit in listeners emotional states of positive valence and moderate to high arousal (quadrant one).

Piece 7 (“Jeux de Vagues” or “Frolics of waves”) suggests a lively motion (a metaphor for the waves movements and games), conveying sensations of both bizarre and a dreamy atmospheres (the mysteriousness of the sea). It is a piece of variety and “color” expected to elicit a variety of sensations in listeners of both positive and negative valence, and low to high arousal (quadrants 1, 2 and 4; only low arousal and negative valence is not expected).

“Liebesträume No. 3” (Piece 8) is the last of three solo piano works published by Liszt in 1850, composed to describe mature love. This piece is expected to elicit responses of low arousal (quadrants three and four).

The “Ciaccona” (“Chaconne”, Piece 9) is the concluding movement of Bach's “Partita no. 2” that lasts some 13 to 15 minutes. The excerpt used here (from a transcription for piano by Ferruccio Busoni and performed by Mikhail Pletnev) are expected to elicit responses in all four quadrants.

Procedure

Each participant sat comfortably in a chair inside a quiet room. The goal of the experiment was explained through written instructions to explain the quantification of emotion and the self-report framework to be used during the listening task. The physiological measures were obtained using a WaveRider biofeedback system (MindPeak, USA). 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 the heart rate and skin conductance responses. Participants reported their emotional state by using the EMuJoy software (Nagel, Kopiez, Grewe, & Altenmüller, 2007), which consist 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 the 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 (for a discussion of emotion felt versus perceived see Kallinen & Ravaja, 2006 and Evans & Schubert, 2008). 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 small break of 15 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 experimented started.

Data Processing

Description of the psychoacoustic measures

Our hypothesis to this experiment is that low level music structural features have causal relationships with the listeners' reports of emotion. To extract such information from the music pieces we analyzed the perceptual experience using the same psychoacoustic variables used in our previous work (Coutinho & Cangelosi, 2009)[iii], which consist of a set of six features: loudness, tempo, pitch level (power spectrum centroid) and 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 raging from -1 to 1, with 0 as neutral. Then, 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 music piece.

Physiological responses. The physiological variables had to be processed to rule out the effects of individual differences on physiological levels. The first method was applied to both variables and consisted of dividing the individual heart rate and skin conductance response 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 comparing between subjects without further calculations. These are the values for heart rate (HR) and skin conductance response (SCR) that we report in this article.

Results

Figure 1 shows these 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 expected quadrants 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 (except piece 8) elicited responses in the predicted quadrants (see description in page 10). 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 (see for instance pieces 7, 8 and 9, which 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. A possible explanation of this 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 not a likely possibility given that, at least 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[iv].

Another possible cause of this apparent compression may be a positive effect of music on mood perhaps derived from the pleasantness of listening to the pieces, which is consistent with the fact that participants reported emotions felt and with the fact that most ratings belong to quadrant 1 (pleasant emotions). As a matter of fact, some of the participants (mostly expert musicians) reported difficulties using the left side of the valence scale justifying that with the experience of positive sensations with all pieces. To check for a possible influence of musical training on the average ratings of valence, we analyzed the mean values for each piece diving the participants into two extreme groups: those with five or more years of musical training/practice, and those with less than one. We observed that the average reported valence for seven of the nine pieces (except pieces 2 and 9) was higher among musicians group[v]. Future experiments should look in more detail to possible factors influencing the self-report of emotion, not only regarding musical training effects, but also other aspects that may influence participants reports of emotion (e.g. personality traits and emotional intelligence).

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 different segments by focusing on criteria related to its form and perceived affective value. 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 segments mean data in order to observing how the emotion dimensions reported 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 ($r = .41$, $p < .05$).

Overall, these results are coherent 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 also relates to the valence ratings as well as with arousal ones. Such a result is also congruent with previous studies (and suggests also complex relationships with emotion), as for example those which report its associations 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).

The test of the relationships between physiological features and emotion has yielded only one significant correlation, implying that increased HR relates with reports of higher arousal ($r = .46$, $p < .05$). This result is coherent Krumhansl's (1997), who has shown that increased heart rate levels related to fear and happy excerpts in comparison 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 also verified that the changes in heart rate has clear associations with sound features, and we found that the hear rate had a propensity to be higher during louder segments ($r = .51$, $p <$

.01). This liaison is coherent with the fact higher loudness relates with higher subjective arousal, and suggests that the heart rate may to some extent mediate this interaction.

No significant correlations were found relating the skin conductance response with either the emotion dimensions.

Music segments: classification analysis

Before introducing the computational study, we conducted another examination on 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 recurred to a Linear Discriminant Analysis (LDA) (Mclachlan, 1992), a classical method of classification using categorical target variables (features that somehow relate to or describe objects). The categories 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 (one to four, rotating anti-clockwise, starting at positive arousal and positive valence). 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 features 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 first analysis shows that the mean levels of the sound features allow for a classification of the segments in 85% of the cases, with a cross validation of rate of 70%. The second test yields a success rate of 89% with a cross validation of 78%. These results indicate that the inclusion of the physiological variables conducted to an improvement in the emotional classification of the music segments (in this analysis, by increasing the separability of the groups – the mean distance between group centroids increased from 2.4 to 3.5). This effect is nevertheless small (8% increase in the cross-validation results), and sound features hold clearly the strongest discriminatory power.

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 response. 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.

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 conquered great popularity, especially as patterns 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 neurons activations consists 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 interests of this article we will focus on the later.

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 connections 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 will vary 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)[vi]. 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.

Simulations 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 consists is the fact 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 5 shows the mean *rmse* across all 15 trials for each of the two simulations condition. In addition, we also indicate the calculated the correlation between model outputs and experimental data, using the Pearson product-moment correlation using coefficient (*r*).

-- Insert Table 5 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 SCR 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 remaining 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 of 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. It is especially remarkable the good performance for the pieces used to test the model predictions to the new set of unknown stimuli ("test set"). 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 an unheard piece to 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 between model and experimental data 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 the each segment (defined earlier; see Table 4), 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 times.

-- 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 correlation analysis results are shown in Table 6.

-- Insert Table 6 here --

As it 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. The only noticeable difference relates to the correlation between valence and pitch contour: the model predictions seem to be more dependent on this variable than the experimental data. Regarding

the correlations between physiological variables and emotional dimensions we also found similar relationships between the model and experimental data: only the heart rate had a significant correlation with arousal ($r_{HR,A} = .46$ and $r_{HR,A'} = .38$; $p < .05$).

Discussion and Conclusions

We believe that one of the main factors that leads to divisive and confusing findings regarding the relationships between music structure and emotional response is related to the fact of placing the main focus of attention on high level features of music and on generalizing the stationary characteristics of the low level ones. Regarding the former, as we have mentioned earlier in this article, a review of more than one hundred studies has not revealed systematic associations between emotion and features such as key or mode. As for the later set of features (e.g., tempo, loudness, timbre, pitch), their relationships with emotional qualities seem to be more complex than it is often assumed. In this article we focused on the low level music features to predict emotional responses to music.

We have started by suggesting that the common use of averaging methods to compare music structural features with emotional responses can obscure important information regarding its dynamics, especially because these can be intense and momentary (e.g. Dowling, 1986). As we have shown in this article, the average arousal and valence over full pieces masked important variations on the emotional character within the pieces. Our analysis of music segments has shown clear-cut associations between sound features and emotional dimensions that would be visible when looking at the full pieces: arousal systematically increased for segment with higher loudness, faster tempi, higher pitch levels and sharper timbres. Valence correlated positively with tempo and pitch level ($r = .41$, $p < .05$).

The fact that acoustic factors are often analyzed in terms of extreme levels (e.g., high/low, slow/fast) may conduct to a misleading assumption that their effects may be generalized to intermediate values. This is because extreme levels lead to the assumption that sound features do not interact, which is not the case, and especially that they show simple relationships with emotional states, which the fact that we still know little about those associations also suggests to be incorrect. Moreover, the transient effects are barely investigated, despite the fact that they are a prominent component of music, and are overwhelmingly important for expressive purposes (Gabrielsson & Lindström, 2010). We need 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 (certainly they are not only emotional ones, since, for instance, music can also induce mood and motivational states), emerge from the dynamic qualities of the music stimulus. This idea is consistent with a number of studies that show temporal variations in affective responses (e.g., Goldstein, 1980; Nielsen, 1987; Krumhansl, 1997; Schubert, 2004, Korhonen, 2004, Grewe et al., 2007a, 2007b).

We suggest that the complexity of experimental data on music and emotion studies requires adequate methods of analysis, which support the extraction of relevant information from experimental data. It seems that linear models do not suffice to predict emotional responses to music. Rather, non-linear models are needed that assess global sound characteristics as well as relationships between sound features to account for continuous changes in experienced emotions. In that direction, we have considered spatio-temporal connectionist models as a possible platform for the analysis of the interactions between sound features and the dynamics of

emotional ratings, since it supports the investigation of both temporal dimensions (the dynamics of musical sequences) and spatial components (the parallel contribution of various psychoacoustic factors).

In this article, consistently with our previous modeling work, we identify a group of six variables – loudness, tempo, pitch level, pitch contour, texture and sharpness - which represent low level psychoacoustic dimensions of music, and are fundamental for the predictions of emotional responses. We could not only train a model to reproduce experimental data, but also to predict the emotional responses from human listeners. The amount of data predicted and the fact that we have already applied our model to three sets of data (apart from the experimental data presented here we also successfully modeled other data sets; see Coutinho & Cangelosi, 2009 and Coutinho, in press) are notable, especially due to the fact that the model extracts relationships between sound features and emotional responses coherent with most empirical studies. Overall, the model responses to novel data validate the model and support the hypothesis that sound features are good predictors of emotional experiences. Our model brings the prediction of emotional responses to an appropriately high level, including predictions of experienced emotional valence.

By testing the inclusion of physiological cues, we tested the peripheral feedback hypothesis, and the influence of visceral input on the self-report of emotions experienced while listening to music. Although the improvement was rather small compared with the supremacy of sound features, we have shown that the model could perform better when adding the extra heart rate and skin conductance level inputs. These results support previous work on peripheral feedback in music (Dibben, 2004), and reveal that physiological cues may be an important path to explore in future studies. We suggest that the low increase in the model performance may be

due to the fact that physiological cues are themselves affected by the music, and so much of their information is redundant for the model. Even if it may be this the case for the model, it is important not to consider them in such way for other studies, since understanding their dynamics may be fundamental to convey important information regarding the impact that music has in listeners at different levels, especially considering peripheral routes that may exert important interferences with high level cognitive processing.

The work presented here provides a new methodology to the field of music and emotion research based on combinations of computational and experimental work, which aid the analysis of emotional responses to music, while offering a platform for the abstract representation of those complex relationships. Future developments may conduct to fundamental advances in different areas of research since they may provide coherent descriptions of the emotional effects of specific music stimuli, which can aid specific areas, such as, psychology and music therapy.

Due to the fact that the sound features in use by the model constitute a basic set of psychoacoustic features not exclusive to music, but rather general to the auditory domain, we are investigating the possibility that the relationships extracted by the mode may also serve other channels conveying emotional information through sound, as is the case of speech prosody. This research will determine the acoustic similarity between affective speech prosody and music shed light on the extent to which they share mechanisms involved in evoking emotional response. The identification of shared characteristics of emotion expression in music and speech prosody may contribute to evolutionary perspectives on music and language.

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.

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.

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

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

Coutinho, E. (in press). Modeling psycho-physiological measurements of emotional responses to multiple music genres. *Manuscript submitted for publication*.

Coutinho, E. (2009). *Computational and Psycho-Physiological Investigations of Musical Emotions*. Unpublished doctoral dissertation, University of Plymouth, UK.

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

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.

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. (Ed.), *Handbook of Music and Emotion* (pp. 367-400). Oxford: 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.

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.

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

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

Juslin, P.N. & Timmers, R. (2010). Expression and communication of emotion in music performance. In Juslin, P. & Sloboda, J. (Ed.), *Handbook of Music and Emotion* (pp. 453-490). Oxford: 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.

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 dissertation, 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., Bradley, M., & Cuthbert, B. (2005). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.

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

Meyer, L. (1956). *Emotion and Meaning in Music*. Chicago: 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.

Nielsen, F. (1987). Musical tension and related concepts. In Sebeok, T. A. & Umiker-Sebeok, J. (Eds.), *The semiotic web '86: An international year-book* (pp. 491-513). Berlin: Mouton de Gruyter.

Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions*. New York: Oxford University Press.

Parncutt, R. (1989). *Harmony: a Psychoacoustical Approach*. Berlin: 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.

Resnicow, J., Salovey, P., & Repp, B. (2004). Is Recognition of Emotion in Music Performance an Aspect of Emotional Intelligence?. *Music Perception*, 22 (1), 145-158.

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: 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 Juslin, P.N. & Sloboda, J. (Eds.), *Music and Emotion: Theory and Research* (pp. 361-392). Oxford: Oxford University Press.

Schubert, E. (2004). Modeling Perceived Emotion With Continuous Musical Features. *Music Perception*, 21, 561-585.

Tzanetakis, G., & Cook, P. (2000). Marsyas: a framework for audio analysis. *Organised Sound*, 4 (03), 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: Springer-Verlag.

Table 1

Music pieces used for the experimental study. The pieces were numbered consecutively, so as to serve as aliases for reference in this article. For each piece we indicate the composer and title, their duration, and also the 2DES quadrant in which we expect that they will elicit emotional responses in listeners (which is on the basis of their selection).

Piece ID	Piece Details (Composer and Title)	Duration	2DES Quadrant
1	T. Albinoni – Adagio (G minor)	200s	Q_3
2	E. Grieg - Peer Gynt Suite No. 1 (Op. 46): IV. "In the Hall of the Mountain King"	135s	Q_1
3	J. S. Bach - Prelude and Fugue No. 15 (BWV 860): I. "Prelude" (G major)	43s	Q_1
4	L. V. Beethoven - Romance No. 2 (Op. 50, F major)	123s	Q_1, Q_4
5	F. Chopin - Nocturne No. 2 (Op. 9, E flat major)	157s	Q_4
6	W. A. Mozart – Divertimento (K. 137): "Allegro di molto" (B flat major)	155s	Q_1
7	C. Debussy - La Mer: II. "Jeux de vagues"	184s	Q_1, Q_2, Q_4
8	F. Liszt - Liebestraum No.3 (S. 541, A flat)	183s	Q_3, Q_4
9	J. S. Bach - Partita No. 2 (BWV 1004): "Chaconne" (D minor)	240s	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

Pieces 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 they refer to (e.g. piece 1 - segment b alias is 1b).

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 on experimental data: psychoacoustic and physiological features were compared with the arousal and valence dimensions. The values under comparison 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*	-
SCR	-	-
HR	.46**	-

Table 5

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 ran for each simulation and the human participants responses. 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, SCR	.069	.063	.885	.714

Table 6

*Correlation analysis on the model predictions: psychoacoustic and physiological features were compared with the arousal and valence dimensions. The values under comparison 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*	-
SCR	-	-
HR	.38**	-

Figures captions

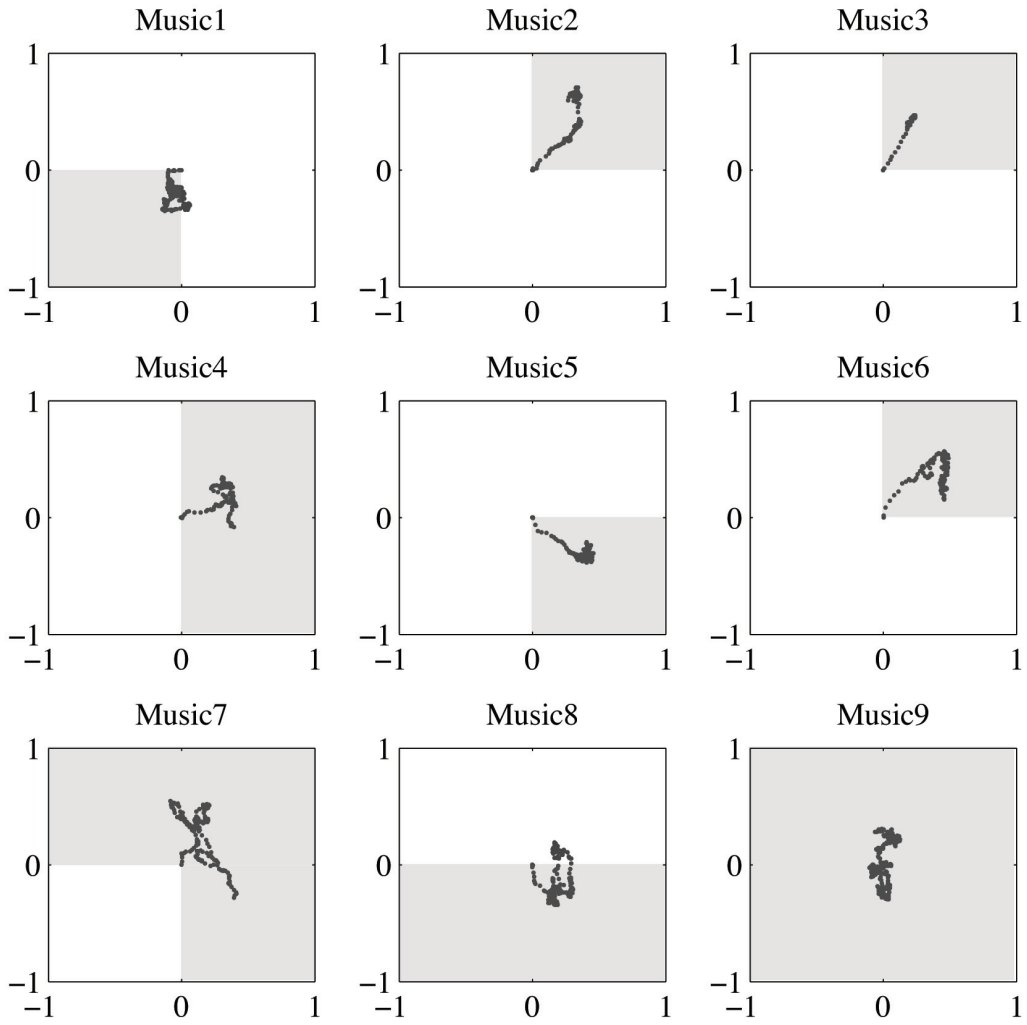
Figure 1. The figure shows the second by second values of Arousal and Valence, averaged across participants, for each piece used in the experiment. The grey rectangles indicate the areas of the 2DES, which correspond to the core affective states expected to be elicited in the listeners (see Table 1).

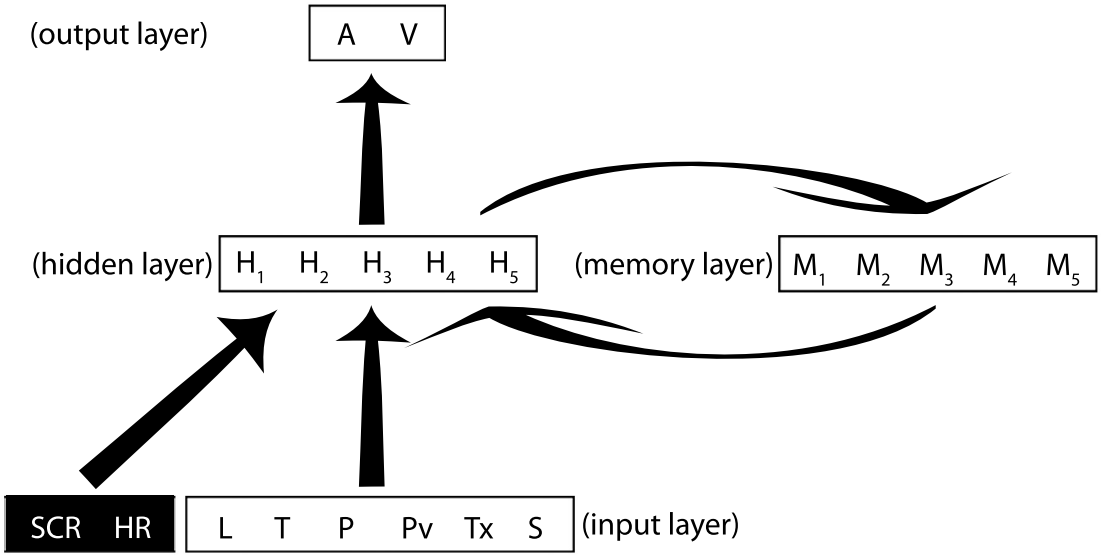
Figure 2. Model architecture of the neural network used in the simulation experiments. Input units: sound features (T, Tx, L, P, S and C) and physiological variables (SCR and HR); Hidden units - H_1 to H_5; Memory (context) units - M_1 to M_5 ; Output units: arousal (A) and valence (V).

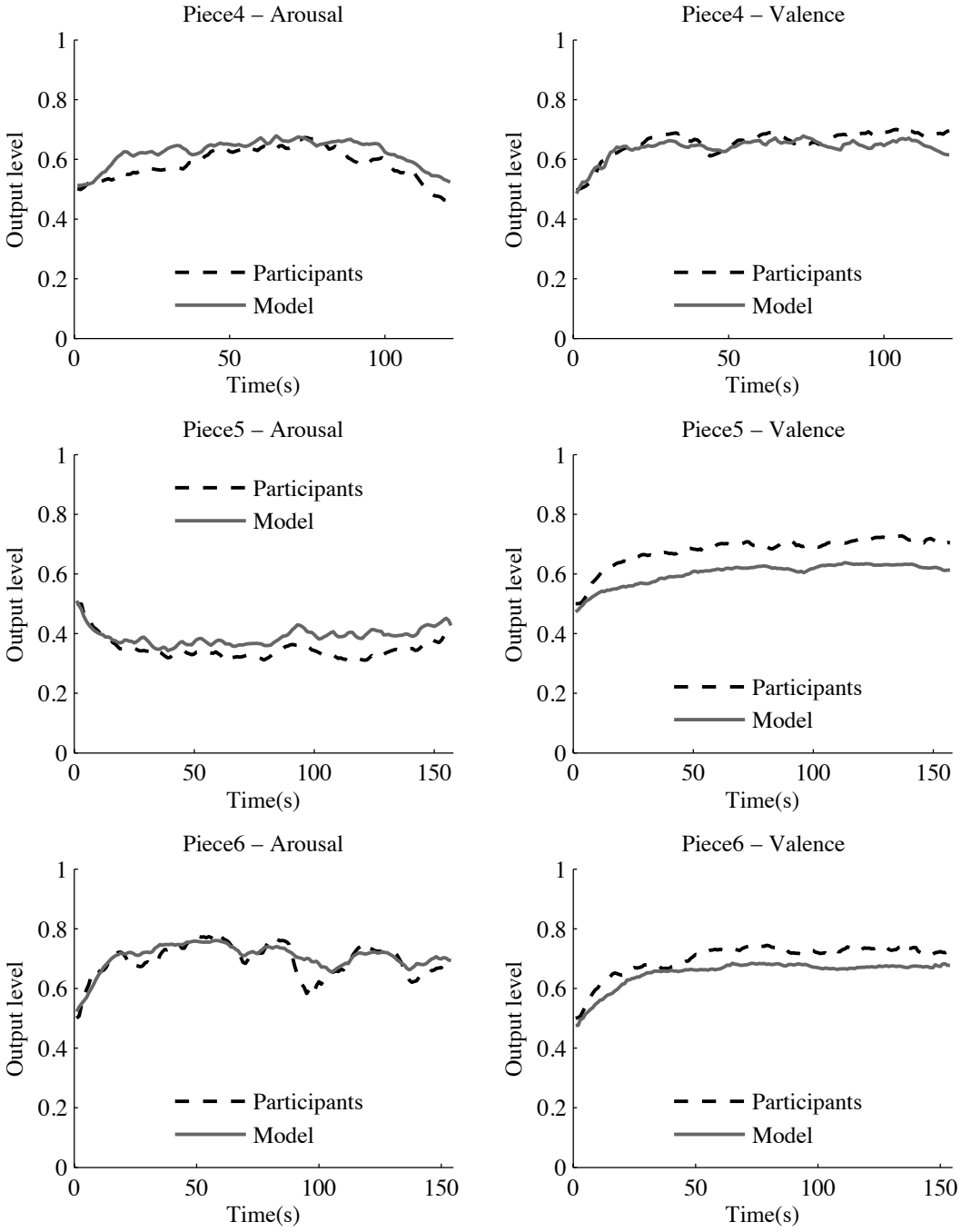
Figure 3. Comparison between the model arousal and valence outputs and experimental data for three samples pieces from the training data set (from top to bottom): Piece 4 (Beethoven - Romance No. 2), Piece 5 (Chopin - Nocturne No. 2) and Piece 6 (Debussy - La Mer, "Jeux de vagues"). The arousal and valence 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.

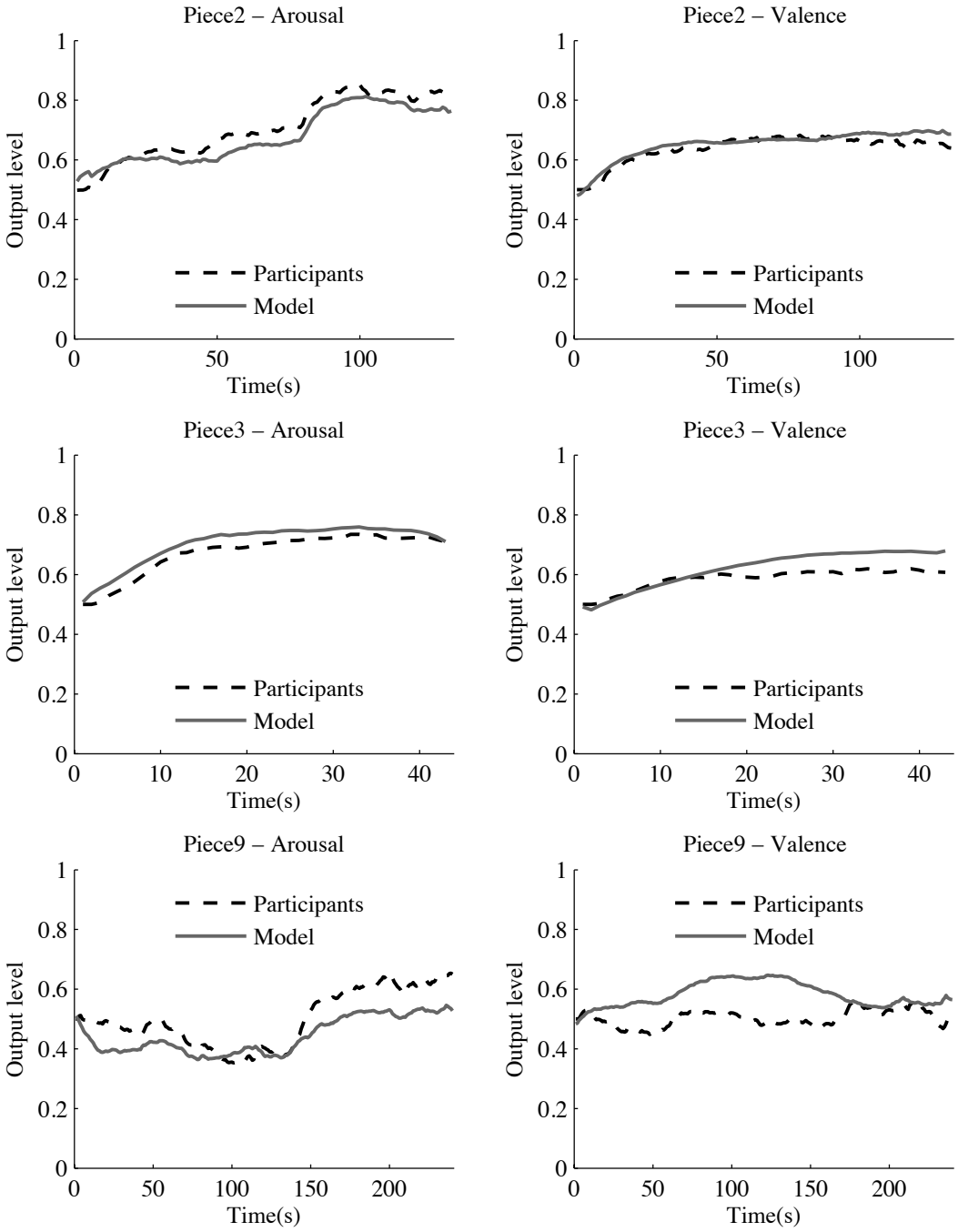
Figure 4. Comparison between the model arousal and valence outputs and experimental data for three samples pieces from the test data set: Piece 2 (Grieg - Peer Gynt Suite), Piece 3 (Bach - Prelude No. 15), and Piece 9 (Bach - Partita No. 2, "Chaconne"). The arousal and valence values shown correspond 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.

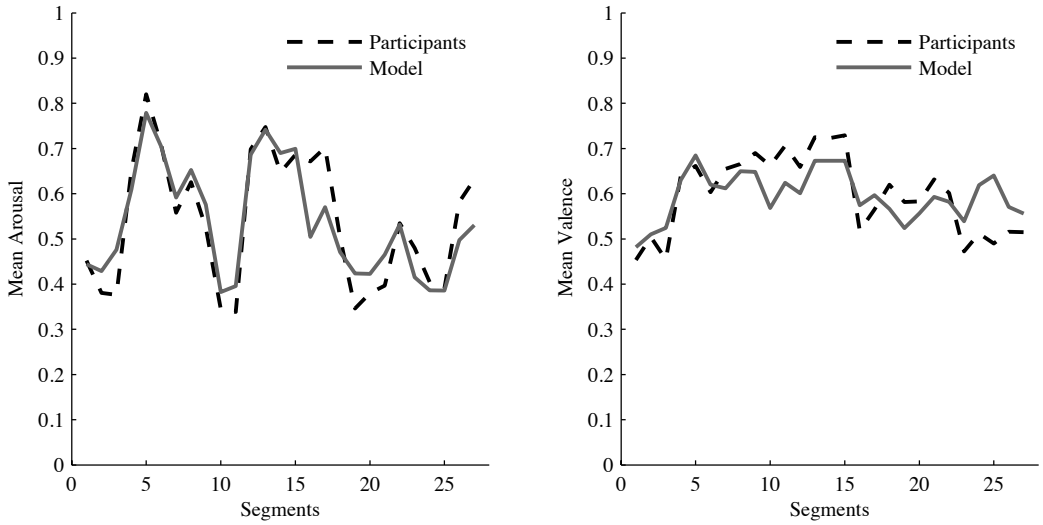
Figure 5. Comparison between experimental data and model predictions: average arousal (left) and valence (right) for each music segment as indicated in Table 3.


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


Participants
Model
Mean Arousal
Segments
Participants
Model
Mean Valence
Segments

Author Notes

The authors would like to acknowledge the financial support from the Portuguese Foundation for Science and Technology (FCT). We are also very grateful to Dr. Nicola Dibben for her pertinent insights and suggestions.

Notes

[i] Gabrielsson and Lindström's (2010) review focuses on the properties of the composed structure, although the relationships between emotional expression and music also relate to performance features (see Juslin & Timmers, 2010). By using real music pieces the performance attributes are inevitably included, and they relate to the specific version used. In this article, we won't make specific comparisons between different performances of the same pieces, and so our focus is also on the composed structure.

[ii] This piece is attributed to Albinoni but composed by the also Italian Remo Giazotto, who came across a manuscript fragment which he later presumed had been composed by Albinoni. The piece is constructed as a single-movement work around the fragmentary theme.

[iii] 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).

[iv] It is important to notice 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.

[v] The average responses of both groups are highly correlated, although more strongly for arousal ($r = .97$, $p < .001$) than for and valence ($r = .85$, $p < .004$).

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